Maximizing Return on Training Investment in Mixed Reality Systems

Christina Padron, Ada Mishler, Cali Fidopiastis, Kay Stanney
Design Interactive, Inc.
Orlando, FL
christina.padron, ada.mishler, cali.fidopiastis, kay@designinteractive.net

Gino Fragomeni
Research, Development, and Engineering Command (RDECOM) Simulation and Training Technology Center (STTC)
Orlando, FL
gino.f.fragomeni.civ@mail.mil

ABSTRACT

Many simulated team training systems developed across the Department of Defense, e.g., the Army's Close Combat Tactical Trainer (CCTT), focus on high fidelity systems that require high initial expense and maintenance costs, require trainees to travel to a specific training site, and cannot be easily updated or reconfigured to reflect changes to the operational equipment/environment. The Army's future Synthetic Training Environment may leverage advances in virtual and augmented reality technology to provide mixed reality training that balances physical components and virtual assets, which would decrease the cost of networked training environments and increase reconfigurability and mobility. However, indiscriminate use of virtual technologies could remove sensory cues critical to task performance, thereby decreasing training value. Further, without an understanding of tasks and users, inappropriate virtual or augmented reality headsets could result in negative training and user sickness. A human-centric sensory task analysis can be effectively used to identify and optimize system fidelity- virtualizing what can be, while maintaining physical components required for training value. The purpose of this article is to introduce a human factors approach that capitalizes on sensory task analysis to maximize training effectiveness while minimizing cost, leading to maximal return on training investment. The goal is to provide practitioners with guidance for the effective use of innovative mixed reality technologies in training systems. In this study, using a sensory task analysis, design guidelines were derived for a mixed reality tactical trainer for the M1 Abrams, identifying constituent cues that needed to remain veridical and those that could be virtualized. These guidelines specified (a) when VR/AR would best be implemented within a complex training simulator based on training objectives, tasks, and environment, (b), VR/AR headset parameters to consider for optimized training value, and (c) software design elements unique to VR/AR, such as how to design to minimize user sickness.

ABOUT THE AUTHORS

Christina K. Padron is a Senior Research Associate II in the Extended Reality (XR) Division for Design Interactive, Inc., and has worked on the design and development of virtual assessment and training tools for over 10 years. She holds a Masters in Industrial and Manufacturing Engineering from Pennsylvania State University and a BS in Industrial Engineering from Purdue University.

Ada D. Mishler is a Research Associate I in the XR Division for Design Interactive, Inc., with experience in the design and development of virtual training programs. She holds a PhD in Human Factors and Cognitive Psychology from the University of Central Florida and a BS in Psychology from Colorado State University.

Cali M. Fidopiastis is a Chief Scientist for Design Interactive, Inc., who specializes in creating virtual environments for training. She holds a PhD in Modeling and Simulation from the University of Central Florida and a Masters in Sociology with focus on Experimental Psychology, as well as Bachelor degrees in Biology and Psychology from the University of California, Irvine.

Kay M. Stanney is President and Founder of Design Interactive (DI), Inc., a woman-owned, small business focused on human-systems integration. She received a BS in Industrial Engineering from SUNY Buffalo and Masters and PhD in Industrial Engineering, with a focus on Human Factors Engineering, from Purdue University.

Mr. Gino Fragomeni has over 34 years of military experience and is currently a Sergeant Major with the 3rd Battalion 20th Special Forces Group. He is also a Science and Technology Manager at Research, Development, and Engineering Command (RDECOM) Simulation and Training Technology Center (STTC) and is conducting research and development in Soldier training simulations. Mr. Fragomeni holds an M.S. from the University of Central Florida (UCF), and has specialized training in Systems Engineering.

Maximizing Return on Training Investment in Mixed Reality Systems

Christina Padron, Ada Mishler,
Cali Fidopiastis, Kay Stanney
Design Interactive, Inc.
Orlando, FL
christina.padron, ada.mishler, cali.fidopiastis,
kay@designinteractive.net

Gino Fragomeni
Research, Development, and Engineering Command (RDECOM) Simulation and Training Technology Center (STTC)
Orlando, FL
gino.f.fragomeni.civ@mail.mil

INTRODUCTION

The United States Army faces significant training challenges, now and in the years ahead. Some of these challenges involve ever evolving conflict and deployment pressures on training requirements. An increased diversity of operational environments, the complexity of possible domestic and global missions, and competing requirements for training resources are all key factors that have a potential impact on future Army training. The Army requires methods for providing Soldiers and leaders with effective training and opportunities to practice tasks effectively and efficiently. Consequently, Army trainers are increasingly turning to emerging technologies to exploit lower cost, technology based solutions to rapidly fill critical training gaps and increase the impact and effectiveness of training.

Technologies such as Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR; Extended Reality, or XR, for the sake of brevity), combined with increasing reliance on Artificial Intelligence (AI), create a myriad of opportunities to radically transform the way the Army both trains and operates. For example, the Army's Synthetic Training Environment (STE), a coherent single immersive training environment that is capable of delivering relevant training to Soldiers in a timely manner, seeks to leverage these and other capabilities to provide next-generation training that reduces costs and enables new ways for Soldiers to train collectively, and over distributed space and time. A vision for future usage of the STE in the warfighter training cycle is to create a simulation based training system that can be used to exercise traditional instruction and get a pass/fail grade from a distributed immersive simulation based on demonstrated performance. The theory is that training solutions can leverage XR technologies with the richness and fidelity needed to properly exercise critical thinking skills and allow Soldiers to apply their classroom training in a collective training environment. To date, Army systems such as the Close Combat Tactical Trainer (CCTT) and Aviation Combined Arms Tactical Trainer (AVCATT) for virtual collective vehicle simulation have tried to achieve this goal via a hardware centric approach. Specifically, they rely heavily on hardware for detailed, physical replication of the real world environment. New advances in XR, as well as touchscreens, motion tracking and gaming technologies may allow a software centric approach to virtual vehicle simulation, while providing a sufficient level of fidelity for collective training. The Army is conducting research in support of STE, such as the effort described in this paper, with regard to the efficacy of such software centric approaches, as there are still many unanswered questions regarding the training effectiveness of immersive technologies. Much of the common wisdom concerning the effects of XR training is anecdotal, lacking an empirical foundation to validate the way forward. The majority of articles and reports written and published on these types of environments are based on limited data of potential prototypes; there is a paucity of doctrinally correct use cases. Further, the majority of research that does exist has been focused on enhancing the understanding of how to incorporate these technologies into the design of training environments, with less emphasis placed on how to best use these technologies to facilitate acquisition of specific knowledge and skills.

This paper describes a human factors approach to XR training system design, which is intended to maximize training effectiveness (i.e., reach a high level of competency), while minimizing cost. Unlike most current approaches to return on training investment (ROTI), the described approach is heavily focused on early investment in training outcomes as an important means of optimizing returns. The application of this approach to the design of a MR training system for collective training in the M1 Abrams tank is described. Finally, the approach is summarized into guidelines to support the optimal design of future XR training systems.

FRONT END ANALYSIS

Historically, training system design has been technology centered instead of user centered, which can result in fanciful technologies that do not lead to optimal operational performance (Clark, 1994; Padron, Champney, Sinagra, & Hart, 2016). XR training system design that is user centered and maximizes ROTI does not necessarily employ all of the latest technological capabilities nor the highest fidelity virtual imagery; in fact, overly interesting and overly realistic virtual environments can hinder training by distracting trainees from their goals (Stone, 2008), and virtualization of some system components could eliminate cues critical to task performance (e.g., Champney, Carroll, & Surpris, 2014). Instead, user centered training system design requires an up-front analysis that gives designers an understanding of 1) users, 2) their tasks, 3) their training objectives, and 4) the context/environment of use.

An understanding of users can be accomplished by involving them continuously throughout the development lifecycle (Brehl, Meth, Maedche, & Werder, 2015). Methods such as observation, interviews, surveys, and focus groups can be used to understand critical aspects of users, such as preferences, needs, and culture (Brehl, Meth, Maedche, & Werder, 2015; Hix & Gabbard, 2002). Users' tasks can be understood via task analysis, which decomposes tasks into their constituent subtasks and required actions. Task analysis generally involves users by questioning them about their workflows, but can also be accomplished via review of existing technical documentation (Hix & Gabbard, 2002). In the context of simulation based (including XR) training systems, sensory task analysis can add value. In sensory task analysis, tasks and subtasks are linked to critical sensory cues, i.e., the sensory information that is needed to support performance of a subtask or task (Champney, Carroll, Milham, & Hale, 2008). This is particularly relevant to XR training systems, given that low-cost visual or auditory virtualization of a physical system component (e.g., representing a switch on a machine) removes haptic cues that may or may not be critical to task performance in the operational environment. One function of sensory task analysis, therefore, is to prevent inappropriate virtualization that might cut up-front costs but increase long-term risk, as an inability to provide the necessary cues at an appropriate level of realism can lead to negative training (Champney, Stanney, Milham, Carroll, & Cohn, 2017).

Training objectives can be understood via a training needs analysis, during which instructors and other subject matter experts (SMEs) are interviewed and existing documentation reviewed to determine the knowledge and performance outcomes that are needed from a training system. In addition, gaps between current training system outcomes and desired outcomes, and relative importance of each outcome to real-world performance can be identified (Padron, Champney, & Carroll, 2016; Reed & Vakola, 2006). As part of the training needs analysis, relevant aspects of the training context/environment should be considered, both for selection of appropriate hardware and software components (Stone, 2008), and for selection of appropriate psychological fidelity (Champney et al., 2014). For example, a training environment in which trainees are highly mobile might require ruggedized hardware, whereas a training environment that is highly stressful, such as a battlefield, might require auditory cues to mimic battlefield noise. Here, as above, early and continuous end user involvement will assist with understanding; practices such as training observation and SME interviews can help to determine what the training environment and operational environment are like and how context will affect training system requirements. The user needs, task, sensory task, and training needs analyses conducted in a front end analysis result in system requirements that must be considered in the design of XR training solutions. The next two sections outline how the resulting system requirements may be met within an XR solution.

SENSORY CUE SPECIFICATION

Haptic and Auditory Cues

XR training technology must be able to support critical sensory cues identified in the front end analyses for each sensory system. Low-cost full virtualization of operational system interfaces is desirable as a means of meeting budgetary constraints, but it must be balanced against the need for sensory cues that cannot be effectively supported by fully virtual solutions. For example, the ability to provide haptic feedback increases up-front cost compared to a purely visual/auditory XR solution, but may be important for training effectiveness if haptic cues are a critical component of the tasks being trained (Stone, 2008). This may lead to the need for virtualized haptic feedback to be included in the system design, such as force feedback gloves to simulate the haptic feedback that would occur in the natural environment (haptic simulation) or haptic metaphors (e.g., using a haptic cue to provide an alert that would be visually presented in the real world; Muller et al., 2006). However, non-virtualized haptic cues may also be provided by implementing some components of the environment via hardware, such as joysticks or steering wheels. Hardware solutions may include relatively low-cost commercial off-the-shelf products or more costly but more realistic custom hardware designed to closely mimic components of the operational environment (e.g., a remote weapon controller with the same hardware button configurations as the real controller; Stone, 2008). Designers must also consider whether or not auditory cues are necessary, and, if so, which type of auditory cues would be most effective to include in the training system design. Non-spatialized auditory cues could be presented via a headset with a simple microphone arrangement, whereas auditory cues that provide important information about spatial layout may require more complex 3D audio capabilities. A need for auditory clarity and realism vs. the acceptability of low-quality auditory cues may also be considered in auditory cue specification.

Visual Cues

XR headsets do not always represent visual information in a veridical manner. Thus, visual cue requirements need to be objectively considered in XR training technology. Specifically, XR training system designers have to be cautious in selecting headsets that can represent visual cues with the level of accuracy and clarity needed for the types of tasks being trained and the types of cues that are critical for those tasks. A number of display parameters in XR headsets affect the user's visual perception. For example, display resolution and field of view (FOV) will affect visual acuity (Fidopiastis, Rizzo, & Rolland, 2010), which will affect the size and detail of visual virtual cues that can be effectively presented in a display. Headsets vary widely in the visual acuity that they can achieve (Fidopiastis, Fuhrman, Meyer, & Rolland, 2005; Livingston, Gabbard, Swan, Sibley, & Barrow, 2013), and tasks that require highly detailed visual cues, such as air vehicle identification tasks, will require headsets that can provide high visual acuity; or they will require see-through (AR/MR) headsets that allow for presentation of detailed visual cues via non-virtual means. By contrast, a lower-acuity and potentially lower-cost headset may be appropriate for tasks in which visual cues can be represented by simplified virtual images without sacrificing any information critical to task performance, such as tasks that require detection, but not identification.

Field of View

Field of view is also important to consider in terms of the size of the display area needed. Training requiring the ability to simultaneously monitor a wide swathe of visual space may necessitate a headset with a wide FOV. However, because a wide FOV may come at a cost to visual acuity (Fidopiastis et al., 2010) and an increase in cybersickness (Rebenitsch & Owen, 2016), designers must consider the trade-offs between a broader viewable context and the ability to view fine scene detail without inducing cybersickness. Brightness and contrast levels determined by the display type can also affect the level of detail that can be seen in a display, and may be affected by display type, e.g., LCDs (Liquid Crystal Displays) vs. brighter and higher-contrast OLED (Organic Light Emitting Diode) displays (Fidopiastis et al., 2010). Training systems requiring higher detail will benefit from headset technology that improves brightness and contrast.

Depth Perception

Depth perception is affected by multiple XR display factors as well, including clutter in the environmental scene, which can make depth ordering difficult and cause scene distortion (Swan et al., 2006); color fidelity in the display, which may alter the effectiveness of depth cues (Livingston et al., 2013); and human depth perception cues such as retinal disparity, which can cause depth distortion if they conflict with visual information provided by a display device (Fidopiastis, 2006; Fidopiastis et al., 2010). If depth cues are important to the performance of a trained task, e.g., the trainee must be able to reach for a lever and grasp it, then it is critical to use XR displays that can provide accurate and easy to read depth information, as inaccurate depth information can cause long-lasting corresponding inaccuracies in motor learning (Welch, Carterette, & Friedman, 1978). Scene clutter and low color fidelity can also reduce visibility of virtual objects, thereby potentially compromising training effectiveness (Swan et al., 2006; Livingston et al., 2013). For example, tasks that require color judgment will necessitate high color fidelity in order to accurately represent task-critical color cues.

Augmented vs. Virtual Reality

Another important consideration for visual cues is whether to employ VR displays, AR displays, or a combination of the two. (For the purposes of this discussion, VR refers to a fully immersive virtual visual display, whereas AR refers to a visual display in which virtual content is overlaid on a user's view of the real world.) If it is important for users to see their hands and track them in 3D space- for example, if they need to use their hands to interact with complex control systems- then either AR that allows them to view their actual hands, or VR that closely tracks hand position and translates it to virtual hands, is necessary. Additionally, if an operational system employs hardware components that users typically interact with partially by sight, then AR may be necessary to allow users to see through virtual components to the hardware outside the headset.

Along with sensory cue specification, system requirements will be influenced by the context of use. The next section outlines training context considerations in the design and selection of XR training technology.

CONTEXT OF USE CONSIDERATIONS

Context of use will affect the type of XR hardware that is appropriate for training. For example, training in which participants are seated and make minimal body movements may be accomplished by headsets that are tethered to a computer. By contrast, training that involves more than minimal body movement requires a tetherless solution, such as the Microsoft HoloLens (Microsoft, 2016). Additionally, usage in potentially rough conditions, such as outdoor use in mixed live-virtual-constructive (LVC) training exercises, may indicate a need for ruggedized headsets that will not drive up costs by requiring frequent repair or replacement.

The length of training sessions should also be taken into consideration. Hours-long training indicates a strong need for lightweight headsets, as even the weight of lightweight modern XR headsets increases the risk of neck pain and headache and feels uncomfortably heavy over a 3- to 4-hour performance period (Wille, Adolph, et al., 2014; Wille, Grauel, & Adolph, 2014). Headset users may also restrict their head movements (Wille, Adolph, et al., 2014), suggesting that the weight and comfort of an XR headset must be considered if naturalistic head movements are an important component of task performance in the operational environment. An additional consideration for long training sessions is visual fatigue. Headset factors that minimize perceptual problems and visual fatigue, such as accurate depth cues, will be even more important when training occurs in long stretches of time.

Finally, factors of XR signal intensity, such as audio volume range, should also be considered in light of the training context. In a noisy training environment, quiet auditory cues could be missed, whereas loud auditory cues could be startling and distracting in a quieter environment- especially during learning (Woodhead, 1964). Similarly, in a complex visual environment or outdoors, low display brightness can make visual cues difficult to see or interpret (Feiner, MacIntyre, Höllerer, & Webster, 1997), although overly bright objects may be inappropriate, as virtual objects that are too bright or too dim can be misperceived as depth cues and distort depth perception (Drascic & Milgram, 1996). Thus, XR training system requirements will be heavily influenced by the context of use. In addition, these requirements will be further influenced by the need to minimize cybersickness, as discussed in the following section.

CYBERSICKNESS CONSIDERATIONS

In designing XR training solutions, a critical consideration is how best to minimize cybersickness (i.e., sickness associated with immersion in an XR environment), which tends to increase with XR headsets as compared to desktop and large screen displays (Rebenitsch & Owen, 2016). Aside from causing unnecessary discomfort to the user of XR systems, cybersickness can have consequences for vision, motor performance, and disorientation that can last for hours or days after exposure (Stanney & Kennedy, 1998; Gower & Fowlkes, 1989), which increases the danger of accidents to personnel and increases the need for recovery time after XR training. Cybersickness may also affect ROTI by reducing the ability to learn in XR environments (Barrett, 2004).

Cybersickness and XR Headsets

Contrary to popular belief, minimizing cybersickness is not simply a matter of using newer and better hardware (Stanney & Kennedy, 1997). Instead, several display design factors affect the likelihood and severity of cybersickness. A recent review by Rebenitsch and Owen (2016) identified the following three parameters of XR headsets that contribute to cybersickness. First, blocking real-world peripheral vision beyond the edges of the XR display increases cybersickness when users traverse a virtual environment. Moss and Muth (2011) suggested that peripheral occlusion creates a visual-vestibular conflict by preventing users from gathering visual information from the real world; if users can view the real world in the periphery of the headset, this provides a rest frame to maintain spatial orientation while immersed in a virtual world, which has been demonstrated to reduce cybersickness symptomatology by as much as 400% (Bonato, Bubka, & Krueger, 2015). This indicates that training systems for tasks involving considerable virtual movement should utilize open-sided VR headsets or see-through AR headsets to allow users to gather real-world visual information relevant to their virtual position.

Second, stereoscopic displays (displays that present slightly disparate viewpoints to each eye to create the appearance of depth) increase oculomotor disturbances, possibly as a result of imperfect alignment of the display's interpupillary distance (IPD) with that of the user's, as well as increased disparity between vergence and accommodation compared to monoscopic screens (Wann, Rushton, & Mon-Williams, 1995). This suggests that designers may want to avoid stereoscopic displays as long as stereoscopic depth cues in virtual objects are not important for task performance, and may want to choose displays with adjustable IPD if they do need to provide stereoscopic depth cues. Third, a linear relationship exists between FOV and cybersickness, with a wider FOV substantially increasing cybersickness (Lin, 2002). Thus, headsets with a narrow FOV, while limiting the amount of virtual content that can be viewed at once, may lead to better ROTI than a FOV that covers more of the real-world visual field.

Based upon the specified sensory cues, training context, and need to minimize cybersickness, XR training system designers can select hardware solutions that best fit trainees' needs. Some considerations for software design that are unique to XR are discussed in the next section.

DESIGN OF XR TRAINING SOFTWARE

Stanney, Mollaghasemi, Reeves, Breaux, and Graeber (2003) outlined usability issues in the design of XR solutions. These include system interface considerations such as interaction design (i.e., object manipulation, wayfinding, and navigation) and design of multimodal system output (i.e., visual auditory, haptic, etc). They also include XR user interface design considerations such as engagement (i.e., presence and immersion) and side effects (i.e., comfort, cybersickness, and aftereffects, which were discussed above).

Object Manipulation

In terms of object manipulation, a primary consideration is whether or not realistic manipulation of virtual objects is necessary for task performance (Stanney et al., 2003). Simplified object selection and manipulation may come with advantages such as making objects easy to move, but more complex and realistic interactions can enhance immersion in the virtual environment and might be necessary for training skills in which the precise form of object manipulation is important. Object manipulation may need to include the ability to view multiple perspectives of an object, and use inputs that are body centric, rather than using inputs such as a joystick that do not move with reference to the user's body- with the exception of manipulations that involve devices such as joysticks in the real world. Additionally, direct manipulation may need to be used for selection of spatial objects (e.g., a lever in a particular location that trainees will need to reach for) in a display. It can be also difficult for users to understand which objects in an XR environment are active (i.e., which objects can be interacted with). If it is not clear, then active objects may need to be distinguished from passive objects, for example, by providing a distinctive visual attribute, such as a spotlight.

Wayfinding and Navigation

Wayfinding and navigation are XR interaction usability considerations that can impede training effectiveness if not appropriately addressed in the training system design. Wayfinding involves the ability to maintain knowledge of one's location and orientation while navigating (i.e., moving about) a virtual space. If trainees need to focus their interaction time on trying to figure out the spatial layout of a virtual training environment, this will detract from their ability to focus on training objectives. It is thus essential to provide means (e.g., compass, you are here maps, god's eye views, etc.) to easily orient and move about in immersive three-dimensional spaces. Oftentimes multimodal cues can be used to help maintain orientation. For example, use of effective 3D spatialized content when positional or directional information is important to training tasks can be effective. However, such spatialized auditory content needs to be cautiously designed, as individual anatomical differences can affect the localization of sounds by different users. The use of generalized head-related transfer functions may not be appropriate as an auditory localization method, whereas the addition of visual information may improve sound localization (Majdak, Goupell, & Laback, 2010). Stereoscopic visual content must be cautiously designed, however, as distortions in spatial layout can hinder completion of tasks that require spatial manipulation of objects or a precise understanding of the spatial environmental layout. If virtual haptic feedback is employed, then kinesthetic and tactile feedback may need to be paired so that congruent information is provided to both senses. However, the use of hardware to provide real-world haptic cues avoids the issue of integrating different types of virtual haptic feedback, and also simplifies the issue of how to provide a sufficiently fine-grained level of haptic detail.

Engagement and Communication

In terms of engagement, Stanney et al. (2003) argue for maintaining high presence, or a sense of being in an XR environment, and immersion, or the amount of sensory information occurring in the environment. Factors of presence relevant to training systems include multisensory output, predictable and realistic system responses to user actions, and non-distracting presentation of visual information. Relevant factors of immersion include the use of intuitive and unobtrusive controls to allow users to concentrate on sensory information from the XR environment and avoiding distraction by activities occurring in the real world.

In an era of complex team and multi-team systems, the designer of an XR training solution also needs to consider how to integrate teammate communication and cooperation into the training system, especially if teammates are trained in separate locations or if a VR headset is being used that occludes the view of a real teammate. Users should be able to see the training system respond in realistic and predictable ways to their team members' actions as well as their own, and the form of communication (e.g., real-time verbal communication, asynchronous textual communications) should fit the timing constraints of the task being trained as well as the types of information processing that need to be performed (Maruping & Agarwal, 2004). Avatars can be used to represent teammates who are in the same location or would be in the same location in the operational environment. This can reduce training costs by both reducing the need for all team members to travel to the same location, and potentially allowing avatars to stand in for real teammates who are not available for training (Roche, Kurt, & Ahrens, 2010). When avatars are used, evidence suggests that they should both be visually similar to the user, and have high visual similarity to other avatars on the team. The combination of self-similarity and team-similarity increases task performance, strategic communication, teammate motivation, and social attraction between teammates (van der Land, Schouten, Feldberg, Huysman, & van den Hooff, 2015). The next section discusses an effort to design an MR tactical training system for collective training in the M1 Abrams tank based on a user centered front end analysis focused on maximizing training effectiveness and ROTI.

USE CASE: MIXED REALITY TACTICAL TRAINER

Front End Analysis

The goal of the Abrams training effort was to design and develop an XR based system to train collective tasks in the Abrams, while providing similar or better training effectiveness and reducing development and maintenance costs as compared to the high fidelity training systems that are currently used (i.e., the M1 Abrams module of the Close Combat Tactical Trainer [CCTT]). The first step in the development effort was to conduct a front end analysis in order to understand Abrams users, tasks, training objectives, and context of use. Initial interviews with SMEs allowed for prioritization of collective training objectives that are most commonly trained using high fidelity simulation, tying those to training objectives for individual crew members that are required to support prioritized collective tasks, and selecting individual training objectives that occur most frequently across Abrams collective tasks. Next, Army training manuals for the selected individual training objectives were referenced to break down training tasks into subtasks in order to help specify sensory cues necessary for successful task completion. To further define XR system requirements, an additional six (6) hours of time with three (3) SMEs at the Clarke Simulation Center at Fort Benning, GA, which holds the CCTT, consisting of Abrams, Bradley, and Reconfigurable Vehicle Simulator (RVS) manned modules. The SMEs had an average of 17.3 years of experience, an average of over four (4) deployments, and each had experience holding all four (4) of the roles within an M1 Abrams crew (i.e., loader, driver, gunner, tank commander). One had also been a Platoon Sergeant, commanding a Platoon of four (4) tanks. During this time, pre-prepared documentation based on information in training manuals guided interviews and cognitive walkthroughs using the high fidelity Abrams manned module as a baseline. Data from SME interviews allowed for specification of detailed task breakdowns and associated visual, auditory, and haptic cues that were necessary to support each of those tasks. With this analysis complete, the next step was to decide how best to apply XR technologies in the design of the Abrams training solution, including what type of headset to utilize (e.g. fully immersive VR, optical see-through AR) and whether all of the crew positions should use the same headset or if task and cue requirements for the different crew members would be best supported with different headsets.

System Requirements and Technology Selection

In the case of Abrams training, the most important interaction (in terms of sensory cues) between crew members is auditory verbal communication. Because this is done essentially through a microphone and headphones, it can be easily simulated within a networked XR environment. Therefore, crew positions could be treated as individual training stations because the networked XR environment provides the perception of co-location. This allowed for the headset for each crew position to be considered individually. Designing the crew member stations as individual stations also provided additional benefits that the current high fidelity manned modules are not able to support, such as being easily reconfigurable and mobile. SME interviews revealed that the loader position is not a priority in the CCTT high fidelity training system, thus the decision was made to focus the XR training solution on the other crew member positions (i.e., driver, gunner, and tank commander).

Driver

The driver's main task during most collective training is to drive the tank according to commands from the tank commander, including moving forward and in reverse, pivoting, slowing, and stopping. The majority of the driver's visual focus is on either vision blocks that provide windows to the world outside, or the Driver's Vision Enhancer (DVE), a screen that can replace middle vision blocks that use a camera to display the outside world. From an auditory perspective, the driver focuses mostly on communications within the crew, but also listens for feedback from the tank, such as engine sounds. In terms of haptic cues, the driver's hands are mostly placed on handgrips to steer the tank, while a transmission shift is used to switch control between drive, reverse, neutral, and pivot. The driver's feet are also used to slow the tank via a brake pedal. The driver remains seated in a reclined seat throughout training, and rotates his/her head as needed.

Fully immersive VR technology was chosen for the driver station for the following reasons. Hardware that replicates the feel and functionality of the seat, handgrips, transmission shift control, and brake pedal will provide the haptic cues necessary to support performance of the driver's tasks. Associated visual and auditory cues can be represented virtually within the VR headset. Further, the vision blocks and DVE in the driver station can be supported by a narrow FOV VR display, which reduces the likelihood of cybersickness.

Gunner

The Abrams gunner's main tasks during collective training include identification and prosecution of targets using a main gun and coaxial gun. The majority of the gunner's visual focus is on sights, including the Gunner's Primary Sight (GPS), meaning that many of the settings needed to successfully complete the gunner's tasks (e.g. magnification) are done with muscle memory and do not rely on visually locating a button or switch. From an auditory perspective, the main focus is on team communication, but the gunner also listens for other cues, such as the gun being loaded and fired. The gunner remains seated straight-up throughout training.

A fully immersive VR technology was chosen for the gunner station. Hardware that replicates the feel and functionality of the seat, headrest for the GPS, and most commonly used buttons and switches provide the haptic cues necessary for the gunner. Virtual visual content will be provided when the gunner looks through the GPS (as well as the other sights) and when he/she lifts his/her head up from the sight, he/she will see virtual buttons and switches in the same position as the actual hardware. The key is that the gunner will still be able to manipulate buttons and switches without removing visual focus from the sight as he/she would naturally do in the real environment, which is expected to support training effectiveness.

Tank Commander

The Abrams tank commander sits straight up on a seat, but does not always remain stationary. The commander sometimes needs to stand up to look outside the hatch. The commander also views and interacts with two major display screens, as well as an extension of the GPS. Along with the tank commander's windows and displays, the commander's tasks include handling two complex gun control systems- the Commander's Control Handle Assembly (CCHA) and the control grip of the Common Remotely Operated Weapon Station. Because the commander is viewing and interacting with multiple displays as well as using multiple hand-operated controllers via touch, the commander's station requires both a high amount of visual and haptic information. Like the gunner, the tank commander's primary auditory focus is team communication (both within the tank as well as between tanks in his/her Platoon), but he/she must also listen for the gun being loaded and fired.

A see-through AR headset was chosen for the tank commander's station, along with hardware that replicates the feel and functionality of the seat and the two hand-operated gun controls. The AR headset was chosen for a few reasons. First, the tank commander can safely see to stand and move within commander station; second, the tank commander can move within the virtual environment while viewing real-world balance and location cues to minimize visual-vestibular conflict, reducing the risk of cybersickness; and third, the tank commander can view the two different hardware gun controls and thus use vision-controlled reaching to switch between them, even if he/she only does so with peripheral vision while monitoring a display. The tank commander's display screens are controlled mainly via simple pushbuttons located immediately next to the screens; these buttons will be implemented as virtual objects so the tank commander is able to view them while viewing the display. Thus, the tank commander will be able to monitor and control display screens while simultaneously operating hardware based gun controls via touch as would be done in the operational environment. As with the gunner and driver, auditory cues will be presented via the AR headset.

Design of XR Training Software

The design of the software for this particular use case was relatively straightforward in the sense that there was not much room for creative liberties to be taken. It is a constrained environment, with each crew member working in a small area throughout the training, so wayfinding and navigation on a micro level is not necessary. Though wayfinding and navigation on a macro level (directing where to drive the tank) could potentially support training goals, embedded training was out of scope for this effort. In terms of object manipulation, the driver and gunner are spending the majority of their time with their hands on the main controls (steering handgrips and gunner handles, respectively), while sporadically manipulating small buttons and switches. Similarly, the tank commander spends most of the time using two joysticks, and then also manipulates buttons and switches. The majority of the objects in this case were either actual hardware, or required simple manipulations (e.g. pressing a button or moving a switch). The software design challenge was in aligning the software elements to the hardware components such that there was perceptually no difference to the users in what they see and feel. This alignment also helped to induce a sense of presence, as did the accurate and timely reactions from both the software and hardware components. Finally, the system was designed for team performance, communication, and cohesiveness by representing teammates in the virtual world as static avatars wearing appropriate uniforms. The uniforms create both similarity to the trainee and similarity between teammates.

XR TRAINING SOLUTION GUIDELINES

Maximizing ROTI is far more than a consideration of acquisition and maintenance expenses. Up-front costs, maintenance and operation costs, training effectiveness- both in terms of competency levels and in terms of how quickly competency is reached- and training recovery periods, among other considerations, all affect ROTI. In terms of up-front and continuing expenses and training time, current hardware centric solutions such CCTT are not optimal, and XR solutions have the potential to greatly decrease these costs without sacrificing the level of realism needed for effective training. However, indiscriminate adoption of XR also has the potential to decrease competency through sensory or cognitive deficiencies in the training system; increase training time, recovery periods, and risk to personnel through cybersickness; and decrease team effectiveness through lack of communication and cohesion. Thus, the approach taken here requires increased investment in the early stages of training system development, as a means of avoiding the later monetary and performance risks associated with incorporating virtual content into a training system. Table 1, below, summarizes actionable guidelines based on the information and the use case presented in this paper. It also presents some challenges and tradeoffs that may be faced when following the guidelines, as well as the payoff from the ROTI perspective. These guidelines summarize a novel approach to the integration of new XR technology into training systems, which combines training system fidelity, side effects, interaction design, and cost into one generalized approach to maximizing ROTI.

Table 1. XR Training Solution Guidelines with Challenges and Benefits

Guidelines	Challenges and Tradeoffs	ROTI Benefit
Front End Analysis Conduct a front end analysis prior to designing an XR training solution	Identifying appropriate SMEs and reserving their time can be challenging; front end analysis requires time, delaying the beginning of the design and development effort	Increases likelihood of training effectiveness by targeting training needs and decreases costs and delays associated with rework when training needs are not met
Critical Sensory Cues Ensure that critical sensory cues are supported	Might require combination of software/hardware (or integration of special components, such as haptic devices), which can be a more difficult development effort.	Increases likelihood of training effectiveness by ensuring that training needs are met and decreases costs and delays associated with rework when training needs are not met
Context of Use Conduct a deep dive into the head worn display (HWD) parameters and compare them to training needs in the context of use when selecting a HWD	May be time consuming and seem tedious; may be difficult to obtain information on headset parameters; might turn out that a more expensive HWD is a better fit	Increases likelihood of training effectiveness by ensuring that content is displayed appropriately, decreasing chances of cybersickness and perceptual problems, and decreases costs and delays associated with rework when training needs are not met
Cybersickness Consider common causes of cybersickness and design to minimize them	May require a bit of creativity in the design, may require a trade-off between cybersickness and immersion considerations, or may require a different HWD than expected depending on the task	Unless sickness is a common aspect to the tasks being trained, it will be completely distracting and detrimental to the effectiveness of the training. It will waste the time of the trainees and instructors, and will increase costs due to having to redo the training later.
Usability Maximize usability via optimal interaction design, multimodal system output, and the induction of engagement via presence and immersion based on the tasks and context of use	May require a bit of creativity in the design, or require a different HWD than expected depending on the task	Usability increases likelihood of training effectiveness, decreases costs associated with extra training needed to learn to use the training system, and decreases time to train per trainee

ACKNOWLEDGEMENTS

This work is sponsored by the Office of the Secretary of Defense and managed by the Army Research Laboratory under contract W911NF-17-C-0020.

REFERENCES

Barrett, J. (2004). *Side effects of virtual environments: A review of the literature* (Report No. DSTO-TR-1419). Edinburgh, South Australia: Defence Science and Technology Organisation Information Sciences Laboratory.

Brehl, M., Meth, H., Maedche, A., & Werder, K. (2015). Exploring principles of user-centered agile software development: A literature review. *Information and Software Technology, 61,* 163-181.

Bonato, F., Bubka, A., & Krueger, W.W.O. (2015). A wearable device providing a visual fixation point for the alleviation of motion sickness symptoms. *Military Medicine*, 180, 1268-1272.

Champney, R.K., Carroll, M., Milham, L., & Hale, K. (2008). Sensory-Perceptual Objective Task (SPOT) taxonomy: A task analysis tool. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting, 58,* 2355-2359.

Champney, R.K., Carroll, M., & Surpris, G. (2014). A human experience approach to optimizing simulator fidelity. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting, 58,* 2355-2359.

Champney, R.K., Stanney, K.M., Milham, L., Carroll, M.B., & Cohn, J.V. (2017). An examination of virtual environment training fidelity on training effectiveness. *International Journal of Learning Technology, 12,* 42-65.

Clark, R.E. (1994). Media will never influence learning. *Educational Technology, Research, & Development, 42,* 57-66.

Drascic, D., & Milgram, P. (1996). Perceptual issues in augmented reality. In *Proc. SPIE Vol 2653: Stereoscopic Displays and Virtual Reality Systems III* (pp. 123-134). San Jose, CA.

Feiner, S., MacIntyre, B., Höllerer, T., & Webster, A. (1997). A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Personal Technologies, 1,* 208-217.

Fidopiastis, C.M. (2006). *User-centered virtual environment assessment and design for cognitive rehabilitation applications* (Doctoral dissertation). Retrieved from the University of Central Florida Showcase of Text, Archives, Research, & Scholarship (STARS) (Accession No. 911).

Fidopiastis, C.M., Fuhrman, C., Meyer, C., & Rolland, J.P. (2005). Methodology for iterative evaluation of prototype head-mounted displays in virtual environments: Visual metrics. *Presence: SI Immersive Projection Technology, 14,* 550-562.

Fidopiastis, C.M., Rizzo, A.A., & Rolland, J.P. (2010). User-centered virtual environment design for virtual rehabilitation. *Journal of NeuroEngineering and Rehabilitation, 7:*11.

Gower, D.W., & Fowlkes, J.E. *Simulator sickness in the UH-60 (Black Hawk) Flight Simulator* (Report No. AD-A214 434). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Hix, D., & Gabbard, J.L. (2002). Usability engineering of virtual environments. In K. Stanney (Ed.), *Handbook of virtual environments: Design, implementation and applications* (pp. 681-699). Mahwah, NJ: Lawrence Erlbaum Associates.

Kooi, F.L., Bijl, P., & Padmos, P. (2006). Stereo acuity and visual acuity in head mounted displays. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, 50,* 2693-2696.

Lin, JJ-W., Duh, H.B.L., Parker, D.E., Abi-Rached, H., & Furness, T.A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings of the 2002 Virtual Reality*. IEEE.

Livingston, M.A., Gabbard, J.L., Swan, E., II, Sibley, C.M., & Barrow, J.H. (2013). *Basic perception in head-worn augmented reality displays.* Report prepared for the Naval Research Laboratory, Washington, DC.

Majdak, P., Goupell, M.J., & Laback, B. (2010). 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. *Attention, Perception, & Psychophysics, 72,* 454-469.

Maruping, L.M., & Agarwal, R. (2004). Managing team interpersonal processes through technology: A task-technology fit perspective. *Journal of Applied Psychology, 89,* 975-990.

Microsoft. (2016). *Microsoft HoloLens* [Mixed reality headset]. Redmond, WA.

Moss, J.D., & Muth, E.R. (2011). Characteristics of head-mounted displays and their effects on simulator sickness. *Human Factors, 53,* 308-219.

Muller, P., Cohn, J., Schmorrow, D., Stripling, R., Stanney, K., Milham, L., Jones, D., … ,Fowlkes, J. (2006). The fidelity matrix: Mapping system fidelity to training outcome. *Proceedings of the 28th Annual Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC),* Orlando, FL.

Padron, C.K., Champney, R.K., & Carroll, M.B. (2016). Quantifying return on simulation-based training investment. *Proceedings of the 38th Annual Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC),* Orlando, FL.

Padron, C., Champney, R., Sinagra, M., & Hart, J. (2016). *RITE-MT (Return on Investment Tool for Effective Medical Training)* (Report No. AD1001288). Arlington, VA: Office of Naval Research.

Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality, 20,* 101-125.

Reed, J., & Vakola, M. (2006). What role can a training needs analysis play in organizational change? *Journal of Organizational Change Management, 19,* 393-407.

Roche, L., Kurt, C., & Ahrens, S. (2010). An avatar approach to distributed team training for mission operations. *Acta Astronautica, 67*, 1164-1169.

Stanney, K.M., & Kennedy, R.S. (1997). The psychometrics of cybersickness. *Communications of the ACM, 40*(8): 66-68.

Stanney, K.M., & Kennedy, R.S. (1998). Aftereffects from virtual environment exposure: How long do they last? *Procedure of the Human Factors and Ergonomics Society 42nd Annual Meeting, 42,* 1476-1480.

Stanney, K.M., Mollaghasemi, M., Reeves, L., Breaux, R., & Graeber, D.A. (2003). Usability engineering of virtual environments (VE): Identifying multiple criteria that drive effective VE system design. *International Journal of Human-Computer Studies, 58,* 447-481.

Stone, R. (2008). *Human factors guidelines for interactive 2D and games-based training systems design: Edition 1* (Report for the Human Factors Integration Defence Technology Centre). Somerset, England: Human Factors Integration Defence Technology Centre.

Swan, J.E., II, Livinston, M.A., Smallman, H.S., Brown, D., Baillot, Y., Gabbard, J.L., & Hix, D. (2006). A perceptual matching technique for depth judgments in optical, see-through displays. *Proceedings of IEEE Virtual Reality 2006* (pp 19-26). Retrieved from https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=11055

van der Land, S.F., Schouten, A.P., Feldberg, F., Huysman, M., & van den Hooff, B. (2015). Does avatar appearance matter? How team visual similarity and member-avatar similarity influence virtual team performance. *Human Communication Research, 41,* 128-153.

Wann, J. P., Rushton, S. K., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Vision Research*, 19, 2731–2736.

Welch, R.B., Carterette, E.C., & Friedman, M.P. (1978). *Perceptual modification: Adapting to altered sensory environments.* New York, NY: Academic Press.

Wille, M., Adolph, L., Grauel, B., Wischniewski, S., Theis, S., & Alexander, T. (2014). Prolonged work with head mounted displays. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program* (pp. 221-224). New York, NY: ACM.

Wille, M., Grauel, B., & Adolph, L. (2014). Strain caused by head mounted displays. In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, … A. Toffetti (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference* (pp. 267 – 276). Retrieved from https://hfes-europe.org.

Woodhead, M.M. (1964). The effect of bursts of noise on an arithmetic task. *The American Journal of Psychology, 77,* 627-633.