Technical Report 1227

Formative Evaluation of a Massively Multi-Player Persistent (MMP) Environment for Asymmetric Warfare Exercises

Michael J. Singer, Ph. D.
U. S. Army Research Institute for the Behavioral & Social Sciences

Rodney Long and Jeffrey Stahl
U. S. Army Research, Development, and Engineering Command

Laura Kusumoto
Forterra Systems, Inc

April 2008

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

BARBARA A. BLACK, Ph.D.	**MICHELLE SAMS, Ph.D.**
Research Program Manager	**Director**

Technical review by

Christian Jerome, U.S. Army Research Institute
Nicole Coeyman, U.S. Army Research, Development, and Engineering Command, Simulation and Training Technology Center

REPORT DOCUMENTATION PAGE

1. REPORT DATE: April 2008	2. REPORT TYPE: FINAL	3. DATES COVERED: February 2002 through September 2006
4. TITLE AND SUBTITLE: Formative Evaluation of a Massively Multi-Player Persistent (MMP) Environment for Asymmetric Warfare Exercises		5a. CONTRACTOR OR GRANT NUMBER: 5b. PROGRAM ELEMENT NUMBER: 622785
6. AUTHOR(S): Michael J. Singer (U.S. Army Research Institute for the Behavioral & Social Sciences); Rodney Long, &Jeffrey Stahl (Research Development, and Engineering Command); and Laura Kusumoto (Forterra Systems, Inc.)		5c. PROJECT NUMBER: A790 5d. TASK NUMBER: 233 5e. WORK UNIT NUMBER: H01
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: DAPE-ARI-IF, Simulator Systems Research Unit 12350 Research Parkway Orlando, FL 32826-3276		8. PERFORMING ORGANIZATION REPORT NUMBER:
9. SPONSORING / MONITORING ABENCY NAME(S) AND ADDRESS(ES): U.S. Army Research Institute for the Behavioral and Social Sciences 2511 Jefferson Davis Hwy Arlington, VA 22202-3926		10. MONITOR ACRONYM: ARI 11. MONITOR REPORT NUMBER: Technical Report 1227

12. DISTRIBUTION / AVAILABILITY STATEMENT:
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES:
Subject Metter POC: Michael J. SInger

14. ABSTRACT *(Maximum 200 words)*:

The U.S. Army RDECOM-STTC conducted research on distributed, multi-player simulations for training dismounted Soldier tasks. They requested U.S. Army Research Institute support formative evaluations by supporting briefings, demonstrations, and collecting usability information during exercises. Two usability exercises addressed a standard checkpoint scenario, a third evaluation was conducted during an Army Post Emergency Operations exercise, and a final evaluation supported a pre-deployment Battalion Staff exercise. During evaluation and development, changes were made to the voice system, Semi-Automated Forces (OneSAF) were added, and simulated radio networks were implemented. Added functionality enabled Soldiers to conduct standard Army tasks and trainers to insert threats and conduct AARs. The Soldiers' indicated that working with a simulated three dimensional environment dramatically changed their approach to the mission. Conducting rehearsals within a low-fidelity simulation was considered an enhancement to the training value of field exercises. The gathered opinion-based information indicates that the system can prepare troops for more expensive live drills and actual deployment; integrating basic Warrior skills with reinforced situational awareness, decision making, and asymmetric warfare skills. The program is continuing with new development by RDECOM-STTC, and ARI is using the AW-VTT to research challenges in the use of distributed, game-based simulations for training Soldiers.

15. SUBJECT TERMS:

formative evaluations, usability, distributed training, simulation

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	71	Ellen Kinzer Technical Publication Specialist. 703-602-8047

Technical Report 1227

Formative Evaluation of A Massively Multi-Player Persistent (MMP) Environment for Asymmetric Warfare Exercises

Michael J. Singer, Ph.D.,
U. S. Army Research Institute

Rodney Long and Jeffrey Stahl
U. S. Army Research, Development, and Engineering Command
Simulation and Training Technology Center

Laura Kusumoto
Forterra Systems, Inc

ARI-Orlando Research Unit
Stephen L. Goldberg, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

April 2008

Army Project Number	**Personnel Performance**
622785A790	**and Training Technology**

ACKNOWLEDGMENT

This research could not have been conducted without the cooperation and assistance of many people at the Sgt. Smith Simulation and Training Technology Center, Research, Development, and Engineering Command; and Forterra Systems, Inc. Special thanks go to Larry Ziock, Brian Comer, and Jim Grosse.

FORMATIVE EVALUATION OF A MASSIVELY MULTI-PLAYER PERSISTENT (MMP) ENVIRONMENT FOR ASYMMETRIC WARFARE EXERCISES

EXECUTIVE SUMMARY

Research Requirement:

The development of new approaches to simulation-based training requires multiple, iterative cycles of implementation, review and test, and revised development. Key in this process is acquisition of user input to insure that the development cycle is focused on the highest priority requirements and is usable by the target audience. The early and iterative application of these evaluations is used to guide the development path for usable systems. Once the U. S. Army Research, Development, and Engineering Command, Simulation and Training Technology Center (RDECOM-STTC) had decided to conduct research on the development of a commercial system into a distributed, multi-player simulation addressing the training and rehearsal of general dismounted Soldier tasks, the need for formative evaluations was evident. RDECOM-STTC arranged for long-term technical support from the U. S. Army Research Institute for the Behavioral and Social Sciences (ARI) from FY04 through FY06.

Procedure:

The formative evaluations were conducted by providing briefings, demonstrations, and exemplar exercises to selected prototypical users, and collecting information about the usability and applicability of the planned, demonstrated, or implemented systems. One limiting issue in acquiring prototypical users was the limited time of available organizations as a result of the current operating environment and resultant operational tempo. Several sessions were organized over the multi-year development cycle that enabled users to be informed about the system, have existing system features demonstrated, and personally experience the system. The initial two usability exercises addressed a standard checkpoint operations scenario, which required simulation of items of standard equipment used in common operations. Another usability evaluation opportunity was conducted during the support of an Army Post Emergency Operations exercise, and a final evaluation was a pre-deployment exercise for a Battalion Staff.

The early evaluations required a day of involvement by the available users. The usability sessions started with the collection of individual data about the users, including their background in training, video-game use, and simulations. The sessions then cycled several times through the presentation of information, practice at using the developed portions of the system, surveys addressing the features and use of the system, and general discussion about the usability of the system. The initial episode presented user training on the system interface as an introduction to the system. The final end-of-day episode typically

required the user team to develop and conduct a short training exercise against opposing forces (portrayed by the civilian developers and government employees).

The Force Protection Exercise was conducted during an Army Post Emergency Operations exercise at the invitation of an Army installation. The exercise involved repeated scenarios centered on Army Post gate operations, with civilian intrusion and attempted penetration. The trainees and role players provided standard biographical information and were given system training prior to conducting the exercise. The evaluation information was collected following the exercises, and addressed the usability and applicability of the system in exercises of that nature.

The final evaluation addressed an information gathering and planning exercise by a Battalion staff, using a pre-deployment scenario. This application was considered to be outside the targeted application of the developing system, but was addressed based upon the urgent need to continue evaluation of the developing system. The users provided standard biographical information and received system training prior to conducting the exercise. The exercise was conducted as a series of vignettes over the course of one day in the simulation environment, plus short interventions in the simulation over the following two days. The evaluation information was collected following the multi-day exercise, and addressed the usability and applicability of the system in exercises of that nature. The exercise was also followed by an after action discussion with the Battalion Commander regarding the effectiveness of the simulation, and information that he had gained about staff capabilities.

Findings:

Each of the evaluations provided information on the ease of use and quality of control provided by the Graphical User Interface (GUI). The ratings, rankings, and short answers provided during the evaluations were used to produce lists of simulated equipment, weapons, vehicles, and functionality needed for properly rehearsing or training exemplar tasks. While the physical implementation of the user interface was locked to standard computer and game interface equipment, the evaluation information was used to make considerable changes to information presentation and execution schemes, including menus for user interaction and capabilities for tailoring the user/trainee interface. As a result of both noted problems and suggestions by users, dramatic improvements were made to the voice system, the simulation was linked to OneSAF for automated entities, and simulated radio networks with easy controls were implemented. Weapons, vehicles, environmental objects, explosive devices, and control functionality were added to the system to enable Soldiers to better conduct standard Army tasks, as well as enabling trainers to insert and control threats that can be encountered in asymmetric operations. Controls and functionality were also developed to record sessions and replay the recordings using time tags in support of After Action Reviews.

Utilization and Dissemination of Findings:

Large amounts of usability information were gathered and used during the multi-year cycles of evaluation and development. Much of the usability and applicability information was immediately provided to the project team, and used by the developers to prioritize and guide the iterative development process. The RDECOM-STTC program has been continued and expanded to new applications that are currently funded by other Department of Defense (DOD) organizations (for example, the Joint Improvised Explosive Devices Defeat Organization). The system has also been demonstrated to multiple organizations throughout the DOD, including the Training and Doctrine Command, and an Army Science Board presentation to the Secretary of the Army. The commercial developers have added features for first responder and medical training with funding from the U. S. Army Telemedicine and Advanced Technology Research Center (TATRC), and they are pursuing multiple applications for first responder training at the municipal, state, and federal levels.

As a result of the increasing development and potential use of game-based simulation systems, the ARI Orlando Research Unit has initiated a research program that will use the OnLine Interactive Virtual Environment (OLIVE^{tm}), which is the basis of the Asymmetric Warfare-Virtual Training Technology (AW-VTT) implementation, to address the continuing research challenges of identifying and quantifying the effects and uses of game-based simulation-system characteristics and features on learning, skill acquisition, retention, and transfer of U.S. Army Soldier tasks.

EVALUATION OF MASSIVELY MULTI-PLAYER PERSISTENT (MMP) ENVIRONMENTS FOR ASYMMETRIC WARFARE EXERCISES

CONTENTS

Page

APPENDICES

LIST OF TABLES

LIST OF FIGURES

Introduction

In 2003, the Chief Technology Officer for the U.S. Army Program Executive Office for Simulation, Training, and Instrumentation (PEOSTRI) proposed that the Army needed a high-level training capability for asymmetric missions. Training programs at that time were focused on conventional warfare and mainly limited to units that were co-located, and steps were being initiated at the National Training Centers and Joint Readiness Training Centers to change training to better address asymmetric warfare and cultural awareness (Wunderle, 2006). Where training capability existed, the scenarios had limited interactivity and failed to address a variety of cross-cultural communication issues that troops encounter in the real world.

The U.S. Army's Research, Development and Engineering Command, Simulation and Training Technology Center (RDECOM-STTC) in Orlando, FL, has been conducting an Army Technology Objective (ATO) using massively multi-player (MMP) gaming technology to address these issues. The objective of the ATO was to develop a large-scale, persistent, distributed simulation environment to train Soldiers. The research focused on evaluating the use of MMP Simulations for U.S. Army training for operations in asymmetric warfare environments. Weapons of Mass Destruction, terrorists' actions, crowd and hostage situations, peacekeeping, psychological operations, and civil affairs are possible interactions faced by the users. One Semi-Automated Forces (OneSAF) Objective System computer generated entities are used augment the large numbers of real people who are needed to populate the scenarios. The various Armed Forces are able to engage in such simulation environments anytime, anywhere, using standard personal computers (PCs) connected via the Internet.

The paper details research in the formative evaluation of internet-based training using Soldier participants and MMP gaming technologies. Initial test results with the 101st Airborne Division of Fort Campbell, Kentucky and the Illinois Army National Guard are presented to indicate the potential such technology has to meet new asymmetric training needs and optimize use of Soldiers' time while preparing for live training events and actual deployment. Subsequent results from experiments with alternative applications, including Force Protection and a Battalion Staff Exercise, also are presented. The report also addresses tools needed to build the training environments and required After Action Review (AAR) capabilities.

The warfare that our Armed Forces are facing in Iraq and Afghanistan today is drastically different from the conventional warfare trained for during the "Cold War" era. Under conventional warfare, the opposing forces represent politically recognized states, and execute their action based on the presumption that the front lines of each side are well defined, and that military doctrine of each side is fairly well-known and only evolves slowly over the course of many years. In conventional warfare we can see and identify who our enemies are, we are familiar with their locations, and we know the uniforms they wear. We know and understand their strategic interest, their weapons' capabilities, and their tactics. If a country attacks our forces in a conventional fashion, then our Armed Forces are likely to respond in a prescribed way. The doctrine of conventional warfare is informed by centuries of experience: the previous wars and tactics of our potential adversaries have been analyzed time and again.

By contrast, asymmetric warfare, which encompasses terrorism inflicted on civilians, is a different sort of threat. For the first time since the Vietnam War, American forces are facing a substantial enemy whose actions are not governed by the principles of conventional warfare. Terrorist leaders are by nature unpredictable, and their followers adapt their tactics to keep up the element of surprise. Our forces are left with many unknowns and have to adapt their tactics on-the-fly. They are left only to imagine what tactics the enemies will use, what unconventional weapons might be employed, and even what clothes they will wear. The Armed Forces may not know where the next incident will take place or when it will happen. It is this unpredictability combined with limited understanding that is a critical challenge to our military training today. Even though the military teaches the history of asymmetric warfare, and continually adapts doctrine for operations, by the very nature of asymmetric threats the doctrine must be flexibly applied.

To help prepare our forces for ongoing operations in the Current Operating Environment (COE), primarily in Iraq and Afghanistan, the U.S. Army trains its Soldiers in the basic Warrior Skills and other skills they may require in battle. The U.S. Army captures lessons learned from recent action and eventually updates training for the Soldiers preparing to deploy. However, the current skills training lacks emphasis on certain factors that could strengthen our Soldiers' advantage over the enemy, including specific training for the unpredictable behavior of the enemy and regional and cultural training that would help units to know both the enemy and the civilian population who can help them locate the enemy. Furthermore, the process of transferring lessons learned from active units into current training for units preparing to deploy is seemingly slower than the rate at which the enemy is adapting its tactics. The inference is that any training system that enables our forces to shorten the learning curve can be cost-effective in terms of lives and materiel.

Training for Asymmetric Warfare

Today, the U.S. Army trains Soldiers on conventional warfare through schoolhouse courses and through live training events at its Combat Training Centers (CTCs) and Military Operations in Urban Terrain (MOUT) sites, but these courses take considerable time and effort to alter in response to the COE (e.g., Basic Officer Leaders Course, 2005). Recently the U.S. Army has begun to utilize Arabic speaking role players to enhance the training at its CTCs. The schoolhouse teaches doctrine at the various levels of Soldier roles. Current doctrine is based on recent wars and tactics, and only recently has begun to address asymmetric threats. Units that are about to deploy may receive training for the general kinds of threats they will be exposed to during unit exercises at training centers (Cahlink, 2004).

What the U.S. Army does not have is a virtual simulation environment for Dismounted Infantry (DI) that allows Soldiers to train tactics, techniques, and procedures (TTPs), perform mission planning and rehearsal operations, and practice decision-making tasks against current enemy tactics. Virtual training for DI has lagged behind that of vehicles (e.g., F/A-18 flight simulators) because of the complexity of the multiple team tasks and the levels of interaction between the avatars (graphical representations of users interacting physically in virtual space).

The U.S. Army trains for asymmetric warfare by communicating knowledge gained through experience and providing experience-based training. The sources for experiential knowledge range from reviewing documents from the Center for Army Lessons Learned (CALL), consulting U.S. Army-sponsored websites (e.g. companycommander.com and platoonleader.com, now incorporated as the Battle Command Knowledge System within Army Knowledge Online), talking with other units already deployed, and exercises in learning how to deal one-on-one with civilians in the countries of deployment. Training for asymmetric warfare must be realistic, and for maximum training effectiveness, it should be a first-person experience. While some knowledge can be transmitted through text, lecture, and slide presentations, the best learning is embodied in challenges that are personally experienced, and the most optimal method for experience-based learning is often simulation (Hays & Singer, 1988). Soldiers must learn how to prevent or adapt to creative, unpredictable asymmetric threats. Training must challenge leaders to think and to take appropriate measures to provide security and force protection. Units must strive to understand, defend against, and operate in a new, unconventional warfare environment.

Currently, the U.S. Army has CALL, located at Fort Leavenworth, Kansas, whose mission is to disseminate those documented lessons. Many of the asymmetric attacks that occur are captured and documented through written reports that are placed on the U.S. Army's websites. These reports take quite a bit of time to generate and get into the system for use throughout the U.S. Army. U.S. Army units can read these reports to garner information about the types of things that deployed units are facing. Even longer is the process for getting these lessons learned into the training simulations used by the U.S. Army. Currently, the most effective incorporation of these lessons is at the National Training Centers and Joint Readiness Training Centers, during field exercises that train units for these asymmetric threats.

Simulation-Based Training

First-person, experience-based training is being conducted at the National Training Center (NTC) at Fort Irwin, California and the Joint Readiness Training Center (JRTC) at Fort Polk, Louisiana, where skilled opposition forces (OpFors) and Arabic-speaking role players challenge leaders to protect their own forces while continuing to prosecute their mission. These exercises simulate the changing environment that deployed troops will face, with the goal of improving their successful response and increasing survivability. The major problems with this approach are the huge costs, the time involved in conducting this training, and the extremely limited frequency with which this training can be conducted. Soldiers and equipment have to be onsite and face a well-trained cadre that continually changes their tactics in order to represent the ever-changing threats that will be faced in the COE. Even these training exercise efforts suffer from the same lag time faced by CALL and the web-log based information provided by companycommander.com and platoonleader.com.

To address these issues, the RDECOM-STTC program began researching the ability to leverage and adapt a commercial MMP game to build a simulation that could bridge these gaps between distribution of lessons learned and simulation-based training based on the lessons that have been learned. The research effort is known as the Asymmetric Warfare – Virtual Training Technology (AW-VTT). The focus of the effort was to provide an easy-to-use, internet-based simulation framework that leaders can use to introduce and review new TTPs for responding to the asymmetric threat. The simulation could become a training multiplier when a Soldier in Iraq or Afghanistan participates in a virtual training exercise being conducted by a unit preparing to deploy to that location and provides real-time subject matter expertise on current enemy tactics. This can provide a powerful tool that augments and supports Situational Training Exercises (STX), without replacing the required "boots on the ground" training. ARI collaborated with RDECOM-STTC in these efforts by conducting the formative evaluations, focusing on the needed fidelity and instructional tools needed for training in the COE.

MMP Environments

There are a few MMP games currently on the commercial market (examples include Everquest, Battlefield 1942, and World of Warcraft). MMP games have some common characteristics that distinguish them from traditional games. The most obvious is that they are supposed to accommodate large numbers of users at a single time. An MMP can have hundreds of thousands of simultaneous users, although usually only a few dozen players may be in a particular part of the environment at a time. Typically the systems separate larger numbers of players wishing to operate in a particular environment into parallel worlds or instantiations of that game environment. Another feature of these game environments is that they run on a relatively standard PC (although more memory and better graphics cards are desirable) and they tolerate internet connections with limited bandwidth. There are typically no unique hardware requirements to run the software. Many other commercial games (e.g. America's Army) operate in peer-to-peer configurations, in which each computer must replicate everything done in the environment by all other participating computers.

The U.S. Army chose the Forterra Systems, Inc. MMP platform, the same technology used to run the commercial There (there.com) MMP, to determine if this technology can be used by the U.S. Army to provide a training capability that does not currently exist. The U.S. Army has been trying for years, with limited success, to develop a distributed simulation and training capability equivalent to that of the MMP game sector. The commercial game sector has successfully developed the architecture required to address the technology issues, but the challenge is how to best adapt the technology to meet the U.S. Army’s needs.

The AW-VTT environment is referred to as a “virtual world,” and it is in fact a model of the Earth. The architecture of AW-VTT is scalable, which enables the U.S. Army to develop terrain for as many parts of this world as are needed, and creation of multiple copies of the terrain as needed to accommodate large numbers of trainees wanting to train on the same piece of terrain at the same time. Currently, five areas on this globe have been modeled: a square km of urban terrain representative of Baghdad, a valley in Afghanistan, a MOUT-style camp set in the Philippines, a suburban California city (for emergency response training) and the New York City harbor. Because of the technology's scalable cluster architecture, there is no limit to the terrain that can be added.

The AW-VTT world is persistent – meaning that when you return to it, it does not automatically reset itself or restart from the point in time where you left off before. When an individual logs back into the environment, the state of the world reflects the changes other users have made while that individual was not there. In essence, time goes on with or without any particular individual's presence. Optionally, you can reset the state of a scenario, or let it run indefinitely. These options provide considerable training flexibility.

The environment is available 24 hours a day, seven days a week. With an "on-demand" 24/7 available persistent environment, exercises are not limited to short intervals but rather can last for days, months -- even indefinitely. Significantly, this also means that the Soldiers can log into the environment from any time zone, any location, and train together.

Once the virtual terrain has been modeled for a particular area such as South Korea, users could run an exercise there while a second exercise is being conducted in a model of Mosul, Iraq. These exercises could be run in conjunction, sharing resources, modeling communication between the two, playing all echelons in the chain of command to include higher headquarters. If it were necessary, units could simulate flying from one location to the other.

Technology Capabilities

In order to develop an MMP simulation capable of training Soldiers against asymmetric warfare threats, there are numerous features that the technology must integrate seamlessly together to create an environment that is scalable, viewable from the air and from the ground, and immersive enough in its representation of people and terrain to be effective for experience-based training. These capabilities include realistic virtual representation of participants (representing Soldiers, coalition forces, terrorists, OpFor and non-combatants), real-time communication between participants, vehicles, aircraft and weapons, health management, computer-generated entities to populate the environment, environmental effects, and an AAR system. The technology must also be implemented in a scalable way so as to allow unlimited participants and simulated entities, as well as a diversity of terrain to enable training to deploy in any part of the real world.

Avatars

A key feature of the AW-VTT is the ability to represent avatars in the simulation as unique individuals. In the real world, people identify each other through distinctions in voice, appearance, demeanor, height, weight and many more attributes. The AW-VTT allows users to create avatars with many distinctive features. Figure 1 shows the same avatar with medium and dark complexions, respectively.

Figure 1. Different complexions can be used on the same avatar to produce different appearances.

Users in the AW-VTT are able to modify their avatars in many ways to look like the type of character that they are representing in the environment. Users can modify facial features (eyes, ears, chin, hair), body features (weight, muscularity), flesh tone (light, fair, dark) as well as the overall shape and even the outfit the avatar is wearing. All these changes can be made in real-time during the simulation, if desired. The male and female avatars vary in appearance, with the sex determined at login, and the assigned sex limiting the appearance, clothing and equipment that is available.

There are other methods used in commercial games to identify individuals. The most commonly used method that also has been tried in AW-VTT is a simple floating name above the avatar. Other possible approaches could include name tapes and rank on the uniforms of Soldiers. Regardless of the approach, identification of others is a crucial capability. Figure 2 shows a group of avatar role-players assembling in an Afghan-like environmental scene, preparing for a medical relief scenario that required a variety of appearances.

Another way that individuals can be represented is through user profiles. This could lead to constructive simulation that includes individual Soldier actual capabilities. In the AW-VTT, Soldiers enter into the world as themselves. AW-VTT includes their visual representation as mentioned above, but potential additions could include their actual physical capabilities (i.e., Physical Training scores) and privileges such as a high mobility multipurpose wheeled vehicle (HMMWV) license. By being able to utilize the Soldier's actual physical capability the simulation could model such limitations as how fast a user could run through a particular type of environment, how quickly the Soldier became fatigued or even how much weight could be carried. By incorporating user profiles, the simulation could limit which users can operate vehicles such as tanks, aircraft or weapon systems.

Figure 2. Many different clothing choices can be used to identify individuals and roles.

Communications

The AW-VTT enables real-time communications between users, including both verbal communications and non-verbal gestures. It currently enables face-to-face communications between avatars using voice over internet protocol (VOIP), and also simulates radio communications using the same technology, directing the voice streams to users on the selected radio channels in the virtual world. A limitation with the current voice technology in the "local" mode is that it does not portray highly realistic auditory information. The technology does provide stereo audio signals that provide some directionality, and the system models attenuation over distances. However, the problem of distinguishing close audio sources (from nearby avatars) can be difficult in the virtual world. For cases where the users are not equipped with headsets and microphones for voice input, AW-VTT also allows participants to communicate via typed text, through a common text chat interface.

The technology also enables non-verbal communication through automated and user-controlled gestures and emotional representations. Automated gestures are triggered when an avatar speaks, causing lips, arms and hands to move. An automated gaze system causes avatars to look at other avatars who are speaking to them. Users can also meaningfully communicate specific gestures using the menus provided or by typing in the names of the gestures they wish to enact. Figure 3 shows a Soldier avatar saluting and a civilian modern middle eastern appearing avatar performing a gesture of greeting.

Figure 3. Gestures like saluting and the Salaam greeting can be selected by the user.

Future capabilities will look at incorporating an autonomous emotion state dependant upon the users' actions and dialog. The system can currently monitor what is typed and look for words such as "yes", "no" and when these words are identified the avatar displays the appropriate gesture. Future enhancements could include visual representation of the avatar's emotional state based on the context of the dialog.

Health Management

The AW-VTT provides for wounding and killing avatars, and modifying their behavior depending on emotional state. The current wounding capability in the AW-VTT is divided into 4 different hit zones: head, torso, arms, and legs. The extent of the wounding is based on the type of round used, the zone that is hit and the number of hits received. Currently the wounding outcome is not based on actual statistical wounding data, but instead on a configurable data table. This table can be modified to resemble actual real world data. Just as avatars can represent being injured, they can reflect being in an emotional state: avatars can be angry or terrified, in which case their animations and facial expressions differ from avatars that are in the normal state. Currently the emotional animations and expressions are controlled through menu selections or chat window commands on the user interface.

Computer Generated Forces

What distinguishes the AW-VTT from other commercial game MMP environments are the alterations for use in military training, including the fact that it can be networked with other, existing military simulations. Part of the reason to use MMP technology is that it can be populated with Soldiers receiving training as well as role players to fill out the scenario. Since numerous role players may not always be available to populate the environment, some sort of computer generated forces (CGF, also referred to as Semi-Automated Forces or SAF) need to be linked with the MMP. It was not the intent of this research effort to actually build these CGF behaviors in the MMP environment, but rather to use the behaviors available in the U.S. Army's premier constructive application, known as OneSAF, by linking to AW-VTT. OneSAF populates the MMP environment with non-critical personnel such as hotel clerks, cab drivers, and people in the marketplace. A simulation gateway conforming to the Distributed Interactive Simulation (DIS) standard currently enables AW-VTT to interoperate with OneSAF, and the technology has thus far been tested with hundreds of OneSAF controlled avatars.

Environmental Effects

The AW-VTT has great flexibility that enables users to build any specific environment needed for training. The entire world can be modeled, and the system is distributed, so troops from all over the world could access specific locations any time they require training or mission planning. With the flexibility of the AW-VTT, Soldiers can access the same location in the simulated world at the same time, or access multiple copies of the same terrain simultaneously. At the time of evaluation, simple lighting and visual effects are available (fog, etc.) and constructed artifacts can be developed (e.g. buildings, roads, walls, etc.). In addition, terrain formations, vegetation, and the appearance of water features can be structured in the environment.

Other environmental factors or functions are being considered, such as the environmental effects of heat on a virtual Soldier's performance. The AW-VTT also models the effects of the Soldier's actions on the environment to include dynamic terrain (bullet holes, building and minimal terrain deformation, etc.). Dynamic terrain and multiple copies or versions of terrain areas will be an issue in the future because the AW-VTT world is designed to be persistent: what happens in the world stays in the world. To enable repeatable training, functionality is being developed to reset certain parts or copies of the virtual world back to some "initial" state.

After Action Review

The key to producing an informative AAR is not just in collecting the data but in knowing how to display the right information at the right time to the trainees. This issue is the same whether the training is being conducted in a live environment or in a virtual world. In the case of MMPs the collection of data is not a problem: every keystroke, step taken, word spoken, or weapon fired is recorded. Although every aspect is recorded, the storage space required for an exercise is typically less than that of a digital video recording of the exercise from a single point of view. However, a storage space limitation could arise if the exercise is lengthy (e.g., days long) or there is an extremely large number of participants and entities that must be recorded for a single exercise.

For AW-VTT, the AAR design is evolving increasing capability for the trainer to configure the data being collected. This ability to customize the collection of data would allow a tailored AAR based on the task being performed and the skills being trained. Initial capabilities consisted primarily of recording the exercise and replaying with the use of video-like controls (play, pause, fast-forward, reverse). The system has evolved the capability of tagging the time and location of events which enables "jumping" to the pertinent event for AAR. Other considerations include when and where to conduct the AAR. The flexibility of the environment puts no constraint on the length of time to run an exercise. As mentioned earlier, the environment should be available 24 hours a day, seven days a week. In theory, exercises could run for days, although that would not be considered as a typical use case. With long duration training exercises, it becomes critical to identify when to start and stop recording, when to deliver AARs, and what to deliver for user feedback.

The U.S. Army has procedures and guidance on how AARs are to be conducted. These procedures are based on the premise that all participants are co-located. Since the AW-VTT environment provides a distributed training network, it is very likely that all the trainees, role players, and observer/controllers will not be located at the same facility. Rather, it is envisioned that Soldiers will participate from around the world in these training events. One of the strengths of the environment is that it can function as a virtual conference room, all users would be able to attend the AAR from their remote locations. Additional AAR trainer controls and functions are being developed for use in training research, and those features that promote improved training will be incorporated, along with guidelines for application and use.

Scalable Architecture

AW-VTT is based upon the OLIVE (On-Line Interactive Virtual Environment) architecture that underlies the consumer service “There” (http://www.there.com). The There consumer service has operated continuously since 2003 and as of Summer, 2007 supports hundreds of thousands of users, with peak loads of multiple thousands of simultaneous participants. OLIVE is designed for scalability both in the size and density of the virtual world that is represented and the number of simultaneous players in the world.

By their nature, large-scale virtual worlds require multiple computers to operate. One client computer is needed for each user and servers are required to host simulated objects. Some virtual environments use the client machines to do all or part of the simulation work thus minimizing or eliminating the need for servers. While this peer-to-peer configuration appears advantageous from a hardware requirement standpoint, this approach has some serious performance and scalability limitations. One of the major problems is that all information about changes to the environment from moment to moment have to be transmitted over the network. This grows to a large load on the system, and limits the number of participants in a particular environment. Instead, OLIVE centralizes the simulation into a cluster of co-located servers that represent the entire virtual environment and all of the entities within it. OLIVE then requires its client software to co-simulate that part of the world that the client can "see" or has interest in, with changes to the environment communicated through a central server.

Within the server cluster, OLIVE splits the world into geographic sectors whose size and shape vary depending on the virtual environment. The granularity of a sector may be as small as 50 meters or extend out for thousands of kilometers and, in fact, cover the entire world. OLIVE sectors are not static and may be adjusted at runtime to maintain a consistent load balance amongst the servers. This is important for simulations where large numbers of objects can converge on a single location whose associated server might otherwise become overloaded. This type of architecture both enables the kind of scalability required for training and the dynamic load balancing needed to maintain performance in highly kinetic situations such as combat operations, where individuals, ground vehicles and air vehicles are all in motion in unpredictable ways, often in urban environments.

On the other end of the scale, OLIVE also supports small installations. If only a few dozen participants (or entities) are to be supported and the terrain is limited in scale, then OLIVE can perform well on a single machine. For even smaller numbers of participants, the OLIVE client and server can be installed on a single machine.

Formative Evaluation of MMP Technology

Typically, for a training technology to be evaluated the training curricula must be identified and sample applications built. Until recently there were no working examples of multi-person simulation environments for Dismounted Soldier training purposes, making summative evaluation impossible. Some virtual environment systems for dismounted Infantry have been developed and examined, for example the collaborative effort to evaluate new technologies for Dismounted Soldier Simulation (Knerr, et al., 2003). That work, as with this effort, did not actually have a "curricula" established as a target, but used U.S. Army Field Manuals (FM) and U.S. Army Training and Evaluation Program (ARTEP) as the basis for skilled activities on which Soldiers continually work to improve. That work also took the standard approach of asking for estimations from the users about their skill improvement and learning resulting from participation in exercises. To date, this project has included four formative evaluations of the AW-VTT prototype. These evaluations are described in the next few sections. During these evaluations Soldiers and trainers rated the potential training effectiveness of the technology and provided detailed feedback and prioritization on features and improvements needed to fulfill that potential.

There has been at least one other formal evaluation conducted on another multi-player game technology, namely Full Spectrum Command (Beal and Christ, 2004). While this technology is not large-scale as is the case with AW-VTT, and it lacks the flexibility of scenarios inherent in the use of human role-players, it is the closest relative to the AW-VTT technology that has a documented evaluation. Unfortunately, that evaluation did not produce a usable comparison of training effectiveness against the control group as had been intended. However, it did include a formative evaluation with valuable insights into the training potential of the technology as well as evaluation procedures themselves. In their 2005 Technical Report (Bonk and Dennen, 2005), the Advanced Distributed Learning Initiative acknowledges the dearth of published research on Massively Multiplayer Online Games ((MMOGs), which are based on MMP technology) for military training and education, and lays out a framework for future research in this area.

Major Issues

Simulations have been used for training and rehearsal for many years, and the evaluation issues in development have not really changed (Hays & Singer, 1989; National Research Council, 1985). The simulation technology does keep changing, as the development of the AW-VTT clearly shows. Simulations based on game engines are now being developed for individual cognitive training (Christ & Beal, 2005), and evaluated based upon initial user's subjective opinions of training effectiveness. Establishing the resulting effectiveness through a summative evaluation is still the desired standard, although it cannot always be met (Boldovici, Bessemer, & Bolton, 2002). Using the formative evaluation process is more common during development because larger-scale comparative evaluations are inefficient when functionality is not complete and the time/expense available is limited (Nielsen, 1993). With the system examined in this paper, encompassing large scale, virtual simulation-based training for groups of individuals engaged in complex interactions, the formative evaluation is even more problematic than usual. The key aspect in the formative evaluation process is the acquisition of data and information from prospective users, which requires that they validly represent the user population in order to accept the evaluation information (Nielsen, 1993).

Users for Evaluations

One of the major issues in conducting any kind of evaluation is the representation of the target population evidenced by the test group (Campbell & Stanley, 1963; Boldovici, Bessemer, & Bolton, 2002). In formative evaluations, the intent is focused on gaining sufficient information to guide the development process; and therefore issues of statistical power do not enter in as they do in summative evaluations. Still, the group(s) providing the information should have some validity in representing the target population. Our U.S. Army groups are obviously not randomly recruited from the U.S. Army population, but are opportunistically selected for trials by the U.S. Army. In order to establish some validity for our participant sample, our demographics are matched with available U.S. Army information (see Table 1, http//www.defenslink.mil/prhome, Undersecretary of Defense for Personnel and Readiness).

In addition, there is some additional information available about U.S. Army Demographics and relevant skills previously published in reports on the computer skills of

Soldiers (e.g. Fober, Bredthauer, & Dyer, 2001) and user data about a prior evaluation of immersive and distributed simulation (Knerr & Lampton, 2005) which can be used to establish a baseline information about the recruited population. The information and a few basic comparisons about the early user evaluations of the AW-VTT will be presented below.

Table 1.
Relevant U.S. Army Demographics, 1998

Average Age	27.5 yrs.
Average Time in Service	7.33 yrs.
E3 – E7	81.6% of Enlisted
O1 – O3	58% of Officers

In developing U.S. Army training devices, it is typical to recruit subject matter experts who are validated through the presentation of experience in using actual equipment during operations. In evaluating the development of a distributed online simulation for dismounted Soldiers, the emphasis has to be on an easily presented and common set of tasks, common in this context referring to the frequency or critical nature of the tasks for the desired user population. Equally important, the evaluation questions should be asked of Soldiers in order to demonstrate that they adequately represent the potential user population.

While calls have been made for more organized and continuously updated information to be gathered during U.S. Army training evaluations (e.g. Boldivici, et al., 2002), there is still no clear program in place to do so (at least known to these authors). The two-fold result is that any research and development effort without a direct U.S. Army proponent, must be “sold” to users, and those users may not be the most appropriate evaluators for the development effort. The “sold” participation in a formative evaluation to prospective analysts and evaluators in this effort has been primarily focused upon Leadership vision of gaining training benefit for immediate training needs, and the concept of contributing to the U.S. Army development process. As a result there is limited information available about the evaluators that can truly be used in order to determine whether someone participating in a user evaluation actually represents the proposed population of users for a system.

Initial User Evaluations 2004

In order to demonstrate and evaluate the AW-VTT prototype, a suitable scenario was needed to illustrate the technology’s potential. Candidate scenarios that would be supported by the prototype technology and extant features included military checkpoint operations, convoy operations, cordon-and- search, building assault, intelligence operations, and force protection. RDECOM-STTC selected checkpoint operations in urban environments (drawn from ARTEP 7-5-MTP) as the test example. The scenario was constructed for Soldiers’ use, as the basis for evaluating the technology’s suitability for a variety of applications for introductory training, skills training, situational awareness exercises, and mission rehearsal.

Setting up and operating checkpoints are common missions, especially in asymmetric operations for dismounted infantry. These missions draw upon the basic Warrior Skills, require knowledge of the U.S. Army's Field Manual 25-4 and in addition, require implementation of a unit's specific TTPs that incorporate previous lessons learned. Soldiers manning checkpoints are required to follow rules of engagement, often making split-second decisions on which rules apply based on their situational awareness. They are required to observe and interact with the local population, and although they often are communicating through translators and civil affairs officers, any language skills and cultural knowledge they possess will enhance their ability to accurately read a situation and take appropriate action.

To support the training for checkpoint operations, Forterra modeled a geo-typical representation of one square kilometer of urban terrain, notionally in Baghdad, Iraq (referred to as Virtual Baghdad). To illustrate the ability of the technology to allow trainers to configure the environment to suit specific scenarios, the developers created a "checkpoint set" of moveable concertina wire, barriers, signs, traffic cones, and even white lines for the pavement. To populate each exercise, they developed avatar templates with skin and body types as well as appropriate clothing for Iraqi civilian men and women, Iraqi police, and U.S. Army Soldiers. Some of the template capabilities (skin, body shape, facial features, and clothing) were adapted from the commercial software used as the basis for the AW-VTT system. From these templates, an unlimited supply of avatars could be created to support the exercises. Culturally appropriate gestures were modeled and developed for the Iraqi and American avatars, and Forterra modeled and developed military and civilian weapons (M-16s, M9 pistols, and AK-47s) for use by role-players. In addition, Forterra adapted vehicle representations from the commercial game system as ground vehicles and helicopters.

First User Evaluation, June 2004

The first group of evaluators recruited by RDECOM-STTC managers was an Army National Guard (ARNG) unit. Their leadership saw the evaluation exercise as an opportunity to gain information and practice during their transition from a Field Artillery Battalion to a Military Police Battalion. The user evaluation briefings and test sessions were held at a conference center in Moline, IL, over three days, with the same presentations and exercises being held on each day with different groups of evaluators.

Method

Participants. The ARNG group consisted of 27 males, ranging in age from 21-52 (average age = 31) and ranging in rank from E3-E7 (22 enlisted participants) and O1-O3 (5 Officer participants). Their Military Occupational Specialties were Field Artillery (18), Supply/Services (4), Chemical (2), Infantry (1), MP (1), and Medical (1). Significantly, the ARNG group had not been deployed to Operation Iraqi Freedom (OIF) at the time of their participation in the evaluation.

Materials. The AW-VTT system (implementing the Virtual Baghdad Checkpoint scenario described above) was used for demonstrations and participant exercises. The AW-VTT client computers were relatively standard Pentium 4 (3GHz) systems with 1Gbyte

RAM and GeForce 6600 graphics cards (gaming systems). The system was connected over the internet from a commercial hotel conference room to the Forterra servers in California. Several additional role-players were based at the Forterra offices and participated in the exercises from that location.

The questionnaires were developed using Microsoft Access^{tm}, and presented on the same computers used for the demonstrations and exercises. The biographical information questionnaire consisted of a series of questions that encompassed experience, rank, training expertise, and computer familiarity. The computer familiarity issues were adapted from Fober, et al., 2000 (see Appendix A for a text version). The sessions were also video-taped for later review of discussion comments.

Procedures. The formative evaluations were structured as one-day, seven-hour sessions, with approximately one-third of the participants attending on each day. The sessions started with an overview of the AW-VTT system, the completion of a biographical questionnaire (Appendix A), and approximately one hour of training on how to use the technology, including hands-on practice in a specifically developed introductory training area in the virtual world (graphically placed on the outskirts of Virtual Baghdad). The next several hours were used in a cycle of presentations by contract and government personnel covering the AW-VTT system features and tools (both existing and under development). The presentations were followed by questionnaires addressing the system aspects and features covered during the preceding presentation, and structured discussions about the existing and planned individual tools, features, and capabilities of the system.

Following this complete review of the system, the Soldiers were given a mission to accomplish in geo-typical Virtual Baghdad requiring checkpoint operations. A leader was selected (by the group) or appointed (by the Battalion training officer), received the operational order (OPORD) for the mission, and then directed the Soldiers according to the OPORD. The exercise was conducted with Soldier participants acting as blue forces and with live role-players (some remotely logged in from California) playing red forces and civilian bystander roles. RDECOM-STTC personnel on-site filled the observer/controller (O/C), OPFOR, and civilian roles. After the exercise, Soldiers conducted an AAR (without the aid of a replay feature, as that was not implemented at that time). At the completion of the entire exercise, final questionnaires were administered and brief concluding discussions covered the exercise and system capabilities.

Results

The results presented here follow the structure of the presentations and discussions. They are based on 7-point scale and anchored responses to questionnaire items. The scales have all been adjusted so that 1 indicates very low ratings while 7 is the highest value available. The ARNG provided good initial ratings of the system's potential for augmenting normal unit training following the introduction and practice session on the system (see Table 2). They also provided generally favorable ratings for movement, menus, and general system interactions following the initial demonstrations and training sessions (shown in Table 2). The general evaluations of these functions were all above the middle anchor of "usable/good enough" for training with the exception of the "teleporting" function used for entering buildings (3.8). It should be noted that the “teleport” function is an artifact of the commercial version of the software and only intended to be used in setting up or positioning Soldiers for training in any final version of a military training system.

Table 2.
Ratings from First User Evaluation

Training Aid in Normal Ops.	6.11 (7 = MOUT site)
Movement Controls	5.19
Aiming Weapons	4.67
Visual Inspections	4.81

Following a more detailed briefing and demonstrations of the system tools and environmental features, ratings were obtained on the importance of terrain, weather, and lighting aspects in the environment. The ARNG rated these issues as highly important, with ratings from 5.4 to 6.5 on the 7 point scales (7 being "critical" for training). When questioned about possible training improvements based on easy trainer manipulation of these features, the ARNG assigned ratings averaging in the 5.6 to 6.4 (with 7 anchored as "twice as effective as current training"). During general discussions, the ARNG identified the need for personalized selection of individual equipment and individual control over avatar configuration as being particularly important.

When asked about visual factors in the simulation, the medical representative in the ARNG group had extensive comments about the need for simulated injury indications, arguing that everything in a mission or exercise changes when a Soldier is wounded, and that this aspect must be included in training. Other Soldiers pointed out that changes due to wounding may be the focus of an exercise.

Finally, the ARNG were asked whether the system could be effective for any U.S. Army training, and they responded with a resounding 88% positive. During the course of the open discussions, many needed pieces of equipment and functionality were identified and discussed. Key among these required items or needed functionality were explosives of all types, weapons typically available to the U.S. Army, and weapons typically found in use in Iraq. The Soldiers also wanted to see the physical effects from weapons on terrain, vehicles, environmental objects (houses, etc.) and, as noted above, avatars.

Discussion

One indication of the success of the system is that the ARNG evaluation cadre and leadership all agreed that even in the raw developmental state of the AW-VTT (circa June, 2004), they would be able to use the system to enhance their training preparation for the upcoming National Training Center rotation and their eventual deployment. Their request for installations for use in preliminary training were not supported due to lack of funds for maintaining and upgrading the system, as all funds were being channeled into development.

The rating, ranking, and discussion data presented above are, as is typical of formative evaluation data from subject matter experts, soft and subjective in nature. The data were also collected in the framework of system developers presenting both existing and projected features, tools, and capabilities. Nevertheless, the information was used to shift development sequences and prioritizations. Overall, the ratings, rankings, and comments provided support for the approach as a simulation supplement for training exercises, especially in the current era of changing threats and roles for Soldiers.

Second User Evaluation, November 2004

The second evaluation group was derived from active duty U.S. Army Soldiers. Significantly, 14 of the 15 Soldiers in the 101st Airborne Division had returned from Operation Iraqi Freedom within the previous year. The 101st as a whole was in the process of ramping up for their training rotation and return to Iraq within the next nine months. The major impetus for AW-VTT participation was increased information about the simulation from a more experienced group. The system had changed through the addition of several weapons and explosive devices, and a major change in the implementation of improved routines for VOIP.

Method

Participants. The Soldiers assigned from the 101st consisted of 15 males aged 25-40 (average age = 32), with ranks ranging from E6-E8. All of the Soldiers were on active duty, with thirteen Soldiers in the Infantry, and two with the military police. The Soldiers participated in the evaluations five at a time, on three consecutive days, as assigned by the division leadership.

Materials. The AW-VTT system (implementing the Virtual Baghdad Checkpoint scenario described above) was again used for demonstrations and participant exercises. The AW-VTT client computers were standard Pentium 4 (3GHz) systems with 1Gbyte RAM and GeForce 6600 graphics cards. The system was connected over the internet from rented convention center facilities to the Forterra servers in California. Several additional role-players were based at the Forterra offices and participated in the exercises from that location.

The same biographical questionnaire (see Appendix A) was administered, this time using an internet online program (Zoomerang™). The commercial system uses identification and password codes to maintain confidentiality, and information was deleted from the site after downloading for analysis. Again the questionnaires were accessible on the same computers that were used for the demonstrations and exercises. Other questionnaires (Appendix B contains Microsoft Word™ versions, as the questionnaire software enabled menus and button selections) focused on the usability of the interface, the features and fidelity of the environment, and the prototype tools and described planned functionality of the system. These questionnaires were similar to the questionnaires administered to the ARNG, but reflected changes made to the AW-VTT system, corrections and minor improvements in the structure of the questionnaire system, and additional features developed from discussions and suggestions made by the ARNG. The discussion sessions were also video-taped for later review of discussion comments.

Procedures. The procedures followed the same pattern and cycle as used with the ARNG (see above). The differences in procedure and content were based upon improvements to the system developed during the relatively short interval between evaluations. For example, additional weapons were added to the simulation, and several improvised explosive devices were implemented. In addition, the voice controls were improved, and several software errors were removed.

Results

The results presented here follow the structure used during the presentations and discussions, and are based on 7-point scaled and anchored responses to questionnaire items. The scales have all been adjusted so that 1 indicates very low ratings while 7 is the highest value available. The 101st Soldiers provided good initial ratings of the system's potential for augmenting normal unit training following the introduction and practice session on the system (see Table 3). They also provided generally favorable ratings for movement, menus, and general system interactions following the initial demonstrations and training sessions, as shown in Table 3. The general evaluations of these functions were all above the middle anchor of "usable/good enough" for training with the exception of the "teleporting" function used for entering buildings (3.33). Again, the "teleport" function is an artifact of the commercial version of the software that is used for direct travel between locations and as a mechanism for entering building spaces.

Table 3
Ratings from Second User Evaluation

Training Aid in Normal Ops.	5.13 (5 = Sand Table)
Movement Controls	5.73
Aiming Weapons	5.0
Visual Inspections	5.2

Following a more detailed briefing and demonstrations of the system tools and environmental features, ratings were obtained on the importance of terrain, weather, and lighting aspects in the environment. The 101st Soldiers rated these issues as highly important, with ratings from 5.4 to 6.5 on the 7 point scales (7 being "critical" for training). When questioned about possible training improvements based on easy trainer manipulation of these features, the assigned ratings averaged in the 5.6 to 6.4 (with 7 anchored as "twice as effective as current training").

The 101st Soldiers were given a more detailed series of questions (based on open-ended discussions with the ARNG) about AAR tools and typical exercise difficulties. In general, Soldiers from the 101st saw O/C control aspects (see Tables 4 & 5) as moderately important for implementation in exercises and AARs. Complementing these ratings were their responses to questions about typical exercise difficulties. The 101st Soldiers rated the difficulty of evaluating overlapping fires by individuals and teams as moderate (Table 4), and the evaluation of building clearing movement and squad formation shifts as easier.

When queried about the value of these types of information for AARs, the ratings provided were at the top end of the scales. The value of showing tracked events during replay was rated moderately high (see Table 5). The capability to measure and replay

Table 4
STX Features Desired

Issue Addressed (& Relevant Anchors)	AA Response
Stealth Point of View (5=Moderate, 7=Critical)	**4.79**
Time & Position Marks	**5.53**
Event Control	**5.60**
Soldier/Unit Tracking	**5.87**
Weapons Fire Data	**5.53**
Difficulty Tracking Squad Locations (1=Very easy)	**3.27**
Eval Formation Shifts (4=Requires O/C Monitor)	**3.4**
Eval Overlapping Fires	**4.67**
Eval Bldg Clearing Move	**3.67**

Note: The anchor values for the issues are provided in parentheses for clarity.

movement, rules of engagement (ROE) violations, and data graphics were all important for AARs. The Soldier/Trainers of the 101st also saw great value in the capability to observe an exercise replay while in the environment and immediately repeat the critical task following the review (6.2 rating, with 7 anchored at "Great Benefit"). This capability, while discussed, was not an easily implemented feature. Since this evaluation, a distributed replay capability has been implemented and is now available, see below.

Table 5
AAR Information Desired

Issue Addressed	AA Response
Wpns Fire Time/Place	**5.79**
OpFor Actions	**5.8**
Comm Events	**5.8**
Wall/Floor Transparent	**6.33**
Movement Tracking	**6.53**
ROE Violations	**5.6**
Visualization Graphics	**5.67**

Note: Information Value in AAR (7=Very High/Useful)

The 101st Soldiers were also given the opportunity to rank-order needed objects for development that the ARNG had identified in open discussions and open text entries. These rankings are presented in Table 6. The Soldiers had a hard time assigning non-repetitive rankings to requirements, often listing several items as "first." The items and capabilities are those that the Soldiers felt would be needed by themselves, the OpFor, or both during an effective Asymmetric Warfare exercise. It is obvious that the Soldiers wanted to be able to cause damage and see the results when running through their training exercises.

The Soldiers of the 101st also provided rankings that agreed with the ARNG general discussion on the need for selection of individual equipment and individual control over avatar configuration, rating these needs between 1st and 5th (but not seeing any need for lots of clothing selections, ranked 5th through 8th). The only other need that ranked higher was the need for overall Scenario Modification controls, ranked 1st -3rd by all Soldiers in a separate selection set.

Table 6

Object & Capabilities Rankings (101st Soldiers)

Object/Capability	% of Group	Ranking
Explosives	86	**1st / 2nd**
Weapons	71	**1st / 2nd**
RPGs	79	**1st /3rd**
Vehicle Damage	71	**1st /3rd**
Avatar Damage	71	**1st /3rd**
Terrain Damage	71	**1st /5th**
Helicopters	**71**	**1st /5th**

Note: Range of ranks is provided for the percentage of group.

When asked about visual factors in the simulation, the 101st Soldiers rated individual Avatar identification and accurate injury visualizations (both Avg.= 6.3) as near "Absolutely Critical" (the #7 anchor, see also the rankings provided in Table 4). This reflected the extensive comments that the medical representative in the ARNG group had for injury indications, arguing that everything in a mission exercise changes when a Soldier is wounded during a mission and that this aspect must be included in training. When discussing needed features for training in the environment, the Soldiers also were unanimous in requiring Night Vision Goggles, a visual effect that is difficult to generate, and has not been attempted as yet.

Finally, the 101st Soldiers were asked to evaluate the system's general adequacy for simulation and most appropriate level for training application (issues not considered by the ARNG). When asked about the adequacy of the system in supporting Introductory Training, 64% of the Soldiers thought it was appropriate for or could supplement that level of training. In addition, 86% thought the system was appropriate for Skills Training, and 93% thought the system was appropriate for or could supplement STX. An additional question, in the context of a discussion about using the system within the "Crawl – Walk – Run" sequence of training exercises, generated an evaluative rating average of 5.4 (with 4 anchored as "moderately good" and 7 as "great").

Initial Formative Evaluation Results

Overall, the ratings, rankings, and comments provide support for the approach as a simulation supplement for training exercises, especially in the current era of changing threats and roles for Soldiers. The rating, ranking, and discussion data presented above are, as is typical of formative evaluation data from subject matter experts, soft and subjective in nature. The data were also collected in the framework of system developers presenting both existing and projected features, tools, and capabilities. Nevertheless, the information can be, and has been, used to shift development sequences and prioritizations.

User/Evaluator Characteristics

Although these groups were not randomly recruited from the U.S. Army population, both the ARNG and 101st Soldiers general demographics (see Table 1, above) were not dissimilar from available U.S. Army information (Army Demographics, 1998). The age (averaging approximately 31 and 32) and experience (in terms of duty time, 101st = 11.53 years, ARNG = 8.23 years) of both groups of evaluation participants fell into the upper quartile when compared to the computer skills baseline for comparable ranks (Fober, et al., 2001). As another comparison, the 101st Soldiers age and experience are higher than the Soldiers involved in the evaluation of Soldier Visualization Stations (Knerr & Lampton, 2005), who averaged 22 years old and had been in the U.S. Army 30.6 months.

The level of computer use, expertise, and familiarity reported by the 101st Soldiers was comparable to the self-reported levels from specialists on the Battalion staff from the report by Fober, et al., 2001 (which broke the responses down by staff occupation). This represents a difference from the baseline information gathered by Fober, et al., in that significant differences were found in their results between staff and Infantry even when rank was held constant. The 101st Soldiers all owned computers and used them an average of 14.97 hours per week, an increase from the data presented by Fober, et al., with 83% ownership by the higher ranking enlisted Soldiers (time of use not reported). The 101st Soldiers also reported that thirteen of the fifteen used computers during duty hours as a part of their activities, compared to the 36% of Infantry reporting daily use on duty in the Fober, et al., report. In spite of this, the average self-rating of expertise for the 101st Soldiers was between "novice" and good with one program, which is below that reported for the staff specialists in the Fober, et al., report, and more consistent with the lower ranks of Infantry.

Although it was not specifically reported or discussed in Fober et al. (2001), our biographical survey explored the use of video games, and virtual reality entertainment. The ARNG group reported significantly greater levels of enjoyment from video game playing than the active unit (avg. rating of 4.26 vs. 2.60; Mann-Whitney $U=61.5$, $p<.001$). The ARNG group also reported significantly higher levels of PC expertise than the active unit (avg. rating of 2.41 vs. 1.6; Mann-Whitney $U=103.5$, $p<.006$). In this respect, the ARNG group seemed more like the Battalion staff than standard Infantry, even though the Infantry provided some apparent differences from the only available baseline (Fober, et al., 2001). In this effort, the biggest difference may be that the 101st Soldiers were highly experienced in ground operations in Urban Terrain from their tour in Iraq, while the National Guard were transitioning from a primarily Artillery role to a Military Police Battalion.

Discussion of Results

The 101st Soldiers agreed that the example exercise was something they had all done on a repetitive basis while deployed, and they also noted that the operation of a checkpoint/roadblock was not a simple exercise. Questions on specific uses and general applicability were developed from reviews of the open discussions held with the ARNG group, and the realization that the constructed scenario would continue to be the most useful example of system capabilities. When queried about their experience in conducting Checkpoint/Roadblock training, 12 of the 15 members of the 101st Soldiers (80%) reported six or more exercises, and none of the group had participated in less than three. When queried about participation in exercises, the same approximate level was reported (see the above comments about actual experiences in Iraq). The average hours involved in training during a duty week was reported as 32.8, probably reflecting the unit preparation for their scheduled return to Iraq. They rated themselves as having high levels of training expertise, averaging 6.5 on a 7 point (very experienced) scale. The group also reported averaging 36 hours on equipment or firearms trainers, and 24 hours in Mission or Operations simulations.

Minor indications of successful alterations for the system can be found in the general increase of ratings over the 6 months between the ARNG session and the 101st Soldier responses, with regard to the movement, menus, and weapons controls. The low ratings for the temporarily necessary "teleporting" function for entering buildings has increased the prioritization of developing the capability to directly enter buildings through representative doors. (Kicking down, blowing open, and other standard building entry techniques are even more difficult to implement.) The recent addition of easily modifiable explosives (Improvised Explosive Devices, IEDs) that allow refuse piles, bags, carcasses, and vehicles to blow up on command reflects the continued focus on Soldier-identified needs for Asymmetric threats in training. Other identified needs have been prioritized and are being addressed as funded capabilities become available.

Perhaps the most encouraging evaluation information comes from the Soldier's recognition of training benefit. The first indication of this comes from their evaluation of the benefit that comes from being able to replay a simulation exercise, and even repeat a critical task immediately following the AAR presentation. Secondary evidence comes from the high agreement with statements on the system's adequacy in supporting and supplementing STXs, and relatively good rating on usefulness in the crawl and walk phases of training.

These responses were used to establish feature development priorities (which were weighted by development cost, of course). In general, the responses generated by the 101st Soldiers were more detailed, and were perhaps of more use as a result. Although both groups were "green" in that they were members of the U.S. Army, it is not clear that either group should be construed as more representative than the other. In fact, while the groups do compare in a general way to the U.S. Army demographics (1998), and the computer capabilities found in Fober, et al. (2001), there is no clear way of knowing whether they actually represent U.S. Army trainers.

AW-VTT Improvements

In 2005, Forterra Systems released an upgrade to the technology that underlies the AW-VTT. During 2005-2006, RDECOM-STTC directed Forterra to move the AW-VTT to this upgraded platform and also to add significant new features of value to the U.S. Army. In the fall of 2005, the AW-VTT database was ported to run on a new generation of Forterra Systems' MMP technology, called OLIVE (for Online Virtual Environment). OLIVE is, by design, a platform for developing multi-player applications that can support anything from small scale, repeatable scenarios to massive, long-term persistent scenarios. Its architecture is also designed to import external databases and integrate external simulations. OLIVE applications can be packaged and adjusted to run on a range of cluster and network configurations, from a single-machine server to large, massively scalable clusters. AW-VTT on OLIVE has been tested on both single-machine and small cluster configurations. The capability seems to enable large scale, persistent terrain databases and objects to be created and maintained on the server side of the simulation network. The limitations still exist because of bandwidth limitations and client (end-user) computing capabilities. As might be expected in a software development spanning four years, processing and graphics capabilities have been upgraded several times with the requirements continually pushing to the commercially available high-end machines.

Voice Communications

Voice communication between users in the AW-VTT is based on the VOIP. The original implementation of this feature had a number of usability problems which were evident in earlier tests of the system. In particular, the need to calibrate the volume and dynamic range of each speaker's voice for each session was cumbersome for trainees. A new implementation has been engineered with an auto-calibration feature that adjusts the volume to a speaker's voice. Additional voice controls are provided, including user control over whether there are graphical indicators of who is speaking (shown over each avatar's head). Although some quality problems still remain, as is shown in the user test results, these changes have helped to reduce the time it takes to train users to operate the system, and increased the reliability of the voice system overall, during the period from 2004 through 2006. In addition, the simulation of radios in AW-VTT was redesigned to emulate a prototypical military radio communications system. The radio system can now be configured so that each user can operate up to three radios simultaneously, with multiple channels that can be operated in either simplex or duplex mode.

Reconfigurable Graphical User Interface

The original AW-VTT graphical user interface was based upon a menu-driven interface. While this interface excelled at enabling new users to learn the controls of a simple commercial virtual environment on their own, it was both difficult to modify and embodied many options that most users in the military will seldom or never need. The new implementation of AW-VTT, based on the OLIVE platform, provides a graphical user interface (GUI) framework that is more easily and rapidly configurable. Changes to the GUI are made by editing and re-loading an XML file that translates all user inputs (key strokes, mouse events, game controller events, etc.) into system commands. It is even possible to set up functions that swap between different GUI designs in real-time during a session.

After initial, informal testing of alternate GUIs based on this new capability, it was decided that the best use of this flexibility would be to tailor the user interface for each user's role, so that every participant in an exercise has all of the controls that are needed to perform their tasks, but no more. For example, an artillery specialist would not be given the controls for a full medic kit, and vice versa. Moreover, a trainee would see a limited subset of the controls that trainers and role-players would have access to. For example, while an instructor can use an overhead view, trainees are limited to the ground view; and while role-players can switch clothing and pitch-shift their voices, trainees cannot.

As can be seen in the user test results below, the first experiment with the new role-based user interfaces met with mixed results in terms of how ease-of-use was perceived. However, those responsible for providing system training to the users noticed a significant decrease in training time required to gain the ability to perform their basic tasks in the system as compared with the previous, menu-based, "one-size-fits-all" GUI. The flexibility of the new GUI framework in OLIVE suggests that user feedback can be accommodated by the technology easily and rapidly.

Computer Generated Forces

By utilizing the MAK Technologies VR-Link DIS/HLA gateway software, the AW-VTT environment has successfully been integrated with the latest version of ONESAF (as of late 2006). In May of 2006, RDECOM-STTC and Forterra Systems Inc. conducted a semi-automated forces (SAF) load test experiment distributed across the internet. ONESAF was used to populate a desert environment with over 100 entities in close virtual proximity to stress all aspects of the overall system – SAF, servers and clients. It is important to note the distinction between the entity metric described here and the user metric for virtual worlds described earlier. The challenge of generating 100 viewable entities while maintaining the required graphics and simulation performance is a different requirement than that of measuring the total number of simultaneous users in a virtual world (all of whom do not have to see each other at any given moment.) The scalable server architecture approach, described above, enables larger numbers of users and persistent objects.

Alternative Use Evaluations 2006

As the improvements to AW-VTT continued, two additional formative evaluations were planned and conducted in 2006. Both of these evaluations explored uses of the technology to support scenarios other than warfighting. Both of these experiments supported Military Police missions, one in a domestic setting, and one in an overseas setting.

Force Protection Exercise – February 2006

In February of 2006, the AW-VTT prototype was used to support a force protection and crisis management exercise at Ft. Riley, Kansas. The primary training audiences for the AW-VTT simulation were the Special Response Team (SRT), part of the military police company at Ft. Riley, and the Ft. Riley Operations Center (FROC) that is responsible for making tactical decisions. A suitable scenario was developed at their request to illustrate the technology's potential, support SRT training, and stimulate the activities of the FROC / Crisis Management Team that would integrate seamlessly with Live and Constructive exercises that were to be conducted at the same time. A vehicle gate at Ft. Riley was modeled for use as the basis for the training situations, and served as the basis for evaluations of the technology's suitability for skills training, situational awareness exercises, and mission rehearsal.

A series of vignettes were developed by RDECOM-STTC , derived from the weather emergency scenario focus of the Ft. Riley operations exercise (a tornado causing major damage in and around the Ft. Riley area). The vignettes included civilians wanting help during the emergency, protesting insufficient aid being delivered in time, escalating to violence, including sniper attacks and attempted gate entry – much like the events that occurred after Hurricane Katrina in New Orleans during 2005. The vignettes also included evidence of a terrorist group that was planning to attack the post, similar to the apparent terrorist plot against Ft. Dix that was exposed in 2007. A simulated reporter was also included to stress the response team and provide independent information to the FROC. Because the major focus was the activities at the FROC, limited time was provided for hot-wash AARs following vignettes in the AW-VTT system. The SRT group and role players therefore spent approximately six hours in the persistent simulation during the day.

Method

Participants. Formative evaluation questions were administered to two groups during this exercise, the SRT group and the supporting role players provided by Ft. Riley to support the exercise. The basic demographics of the two groups are presented in Table 7.

Table 7
Force Protection Exercise Soldier Participant Demographics

	SRT	**Role-players**
Group Size	5	**10**
Force Type	Regular U.S. Army & Civilian Police	**Regular U.S. Army**
Service Grades	E5-E6, GS-5	**E3-E4**
Age Range	**21 – 34**	**21 - 34**

Materials

The formative evaluations consisted of responding to a biographical questionnaire before the exercises began, and completing a questionnaire about the system following the exercises (on the next day). The materials used were substantially comparable to the questionnaires used during the prior evaluations.

Procedures

The use of the system was driven by the exercise flow from the Ft. Riley Operations Center exercise. Exercise events, when initiated, followed the prepared scenarios and escalated in complexity throughout the day. As with previous user tests, at the start of the day, the plan for the day was introduced and some minimal biographical information was collected. Unlike the user tests in 2004, this test included training U.S. Army personnel as role-players rather than employing only vendors and RDECOM-STTC to play roles. Both users and role-players were trained on the system for approximately one hour, although because they were given additional training on how to play the training vignettes, the role-players had a few more hours of exposure to the technology before the exercise began. No criterion test for user capabilities was administered.

Following the introduction and training, the exercise cycle followed a repetitive cycle for the vignettes. The leader would provide an operations order briefing based upon intent provided from the operations center (and vignette structure). The vignette or exercise would proceed to completion, and then the participants would conduct a verbal AAR. The AAR would enable rehearsal of what happened during the exercise, what should have happened, and how the participants could improve the procedures in a subsequent situation. Following the completion of the exercises, short final questionnaires were administered and user discussions were held. These sessions and data collection were conducted by RDECOM-STTC personnel.

Results

The results presented here follow the structure of the questions, and are based on 7-point scaled and anchored responses to questionnaire items. For brevity, results will be presented for each group, and in some instances were combined over multiple questions.

Several important issues need to be discussed at this point. First, the training group was not randomly selected with respect to the U.S. Army, the SRT have a special role in force protection at Ft. Riley. However, the role-players were enlisted personnel that were available and did not have special roles in the exercise, and therefore could be regarded as representative. Second, the numbers in the two groups are very unequal and very low. The subjective opinions are the only data available for influencing the trend of development and the only basis possible for suggesting controlled research that would provide statistically meaningful results from which conclusions could be inferred. Both groups (trainees and role-players) provided rating data about the system and the potential for training. The major issues were the voice functionality, control and movement, identification of avatars, and general judgments of training capability.

Both the five trainees and nine of the ten role players thought that the push to talk control in the voice system was superior to the automatic voice control (which broadcast every utterance of the user). This may have been influenced by the equipment environment, consisting of relatively cramped cubicles in a single room. There may also have been some desire by the participants to be able to talk "outside" the exercise. The groups disagreed on whether the voice quality was adequate and whether the voice system could support the training exercise. The trainees were slightly negative (three out of five) on the voice quality while the role players were satisfied with it (eight out of ten). The trainees thought that the voice system was only barely adequate or workable with problems (four out of five) while the role players rated the system workable to more than adequate (nine out of ten). It should be noted that the trainees relied more heavily on the system's simulated radios than on speaking to the avatars next to them – just as in an equivalent real life situation, the SRT would be communicating via their radios.

The control and user interface ratings by the two groups were in general agreement, with some interesting differences. It should be noted that users were provided with and trained to use both a game controller and a keyboard and mouse, and given a choice of which controls to use. The game controllers were not set up to handle all system controls for all roles, so most users who elected to use the game controller also needed to use the keyboard. The following results in part reflect choices the users made on whether or not to use the game controllers.

There was general agreement that learning to control movement was moderate to moderately easy, and that aiming and shooting was generally good. Also, the groups mostly agreed that it was easy to move after learning, and that detecting collisions with objects was both important and moderately easy. Overall, only one trainee disagreed with the description of the user interface as a good design, while the rest either agreed or strongly agreed (2 neither agreeing nor disagreeing).

Three more specific questions about the user interface revealed some disagreement about the system. The trainees found the interface moderate to hard to use, while six of the ten role players found the system quite to extremely easy to use. The same pattern was found in the ease-of-use evaluation question about the special function keys (gestures, object manipulation, etc.) with the trainees finding them moderate to hard and half of the role player group finding the function keys moderately to extremely easy to use. Finally, four out of five of the trainees preferred the hand held game controller, while six out of ten of the role players preferred the keyboard and mouse control setup. These basic preferences may have been reflected in the other disagreements to user interface questions. It is also of some interest that while only two of the five trainees reported detectable latency in the system, six of the ten role players detected latencies. A final point that may help explain the differences in the subjective assessments of the two groups is the level of involvement. The trainees were the focus of the efforts of the role players, who were also less involved in responding to the ongoing nature of the vignettes.

As previously noted, one of the main reasons for choosing the MMP technology is the ability to customize avatars so that they look different, primarily to support identification within the simulations. There was general agreement that the avatar appearance was adequate to very human looking (versus being cartoon-like) and that the animated gestures were fair to good. However, there was some disagreement between the groups about the ease of identifying the individual avatars. The trainees loaded the reasonable to not easy end of the scale (three of five), while the role players loaded the easily and very easily identified options (all ten). In spite of this minor disagreement, both trainees and role players generally agreed that the avatar appearance would be adequate or more than adequate for training.

In general, both groups found the reality of the simulation adequate and the training potential good. Participants were asked about the explosions and sound effects, and rated those aspects adequate to more than adequate. They were also asked generally about the adequacy of environment realism, and while three of the role players rated it mediocre, the remainder of both groups rated the environment fair to extremely good. The same pattern was found in response to a question about their overall impression of the system, with one mediocre rating and the remainder from moderately good to excellent. A final general question about how many military tasks there were in which the system could support training provided agreement again, with four of the five trainees and seven of the ten role players claiming that the system could support the basic tasks to most tasks.

Discussion

The rating data presented above are, as is typical of formative evaluation data, soft and subjective in nature. Other obvious problems in the representativeness and size of the groups were noted above. Nevertheless, the information can and is being used to shift development sequences and prioritizations, and can be used as indicators for research issues to be pursued in the future. Overall, the ratings provide support for the approach as a simulation supplement for training exercises, especially in the current era of changing threats and roles for Soldiers.

The general assessment of the voice system was inconclusive with these two groups, as they disagreed on quality and the adequacy for training. The basis of the disagreement may be that the trainees wanted more control over intermediate distance communications – being able to project verbalizations over intermediate distances by increasing their speaking volume, functionality that is not currently available. In addition, the radio functionality employed heavily by the SRT may have compounded the audio difficulties for them.

The control and user interface seemed to be adequate, with indications that the trainees found the system more difficult to use than the role players. The major and obvious difference between the two groups may be in their different roles and the amount of practice they had with the system prior to the commencement of the exercise. The trainees were being required to apply knowledge and skills within the context of an unfamiliar simulation and in the face of unexpected occurrences as a training exercise. The role players had to follow a general script when called upon, and were otherwise relatively free to explore the system. Both groups found the avatar appearance to be adequate, but differed on ease of identification. The argument presented above may also explain this difference. The trainees were acting as military police dealing with potential law breakers, and therefore may have had more reason to try to discriminate among the role player avatars for later identification of criminals.

Perhaps the most encouraging evaluation information comes from the Soldier's recognition of training benefit. Even though the evaluation question was generally worded, the trainees and role players that rated the system as highly capable of conducting general training. Additional weight to these subjective opinions probably should come from the fact that the groups had spent several hours in exercises and understood the system capabilities quite well.

Battalion Staff Exercise - July 2006

Although the program was initiated and funded as an engineering development Science and Technology Objective (now referred to as an Army Technology Objective), there was a requirement for evaluations of the training effectiveness of the developed system. As reviewed above, a general formative evaluation approach was followed during the course of the four year program. While multiple efforts were made to enlist active U.S. Army units in a culminating event that would use a summative and comparative approach, the coordination efforts all failed probably due to the ongoing pace of operational deployments.

The last large employment of the system that could be arranged was supporting a battalion staff exercise for a New York National Guard Battalion (BN). This group was in the midst of transforming from a Field Artillery BN to a Military Police BN. The exercise was a staff reconnaissance prior to deployment in a host country

The objective of this training exercise was to see if a virtual simulation could successfully drive a BN Staff exercise without Troops, essentially replacing paper Master Scenario Events List (MSELs) and interaction with constructive simulations such as the Brigade/Battalion Battlefield Simulation (BBS) or Janus. The hypothesis was that using a well-crafted scenario, a virtual simulation could exceed the capabilities of standard methods of conducting staff exercises by immersing a staff in a virtual environment and providing the information they need to execute their battle tasks in a more realistic and militarily relevant fashion. Unfortunately, no prior comparative data was available due to the ongoing conversion by the Battalion from Artillery to Military Police, and the unavailability of information about prior Staff Exercises by the Battalion. This exercise provided an opportunity for a win-win situation for the two organizations involved – the BN Staff and the AW-VTT program. The BN Staff received an opportunity to exploit a relevant training experience and the program has the opportunity to collect valuable data about the usefulness of the simulation along with technical user data about system performance. Unfortunately, there was no opportunity to collect hard, objective data that evaluated the staff performance after the exercise, nor was there any information for a comparison to standard paper-based exercises.

The scenario was developed by first collecting and analyzing ARTEP 7-8, FM 22-2.8 and FM 108. These references were used to address a multitude of battle tasks that a BN Staff would be required to execute as members of an Advance Party. Twenty short and informative integrated vignettes comprising a complete introductory scenario were developed that directly mapped to military police, Anti-Terrorism, Force Protection and Conventional Warfare Battle Tasks. These vignettes provided the same essential information that paper-based MSELs typically provide, using simulations of meetings, tours, and briefings.

It must be noted that a staff training exercise (STAFFEX) was never envisioned for this virtual simulation. However, when the team began thinking about how it could be used to drive a STAFFEX, it actually started to make a lot of sense. At the conclusion of Day 1 of the exercise, when the BN staff began conducting their military decision-making process (MDMP), it was very clear to everyone that the virtual simulation, by itself, had successfully provided the stimulus that the BN Staff needed to conduct their Battle Tasks.

There were only 3 days allocated to conduct this exercise. The BN Commander and executive officer decided early in the planning stages that the first day would be dedicated almost completely to utilizing the virtual simulation and collecting the information the Staff needed. Days 2 and 3 were to be used to conduct the staff MDMP, develop their Courses of Action (COAs) and finally their Operations Order for battalion deployment in-country.

Method

Participants. The BN staff consisted of twelve (12) Officers, Warrant Officers, and Non-Commissioned Officers with a median of fifteen years in the service (ranging from two to thirty-seven years). The staff positions (S1 through S4) were all Captains and Majors, with the lower ranks providing staff support. The median computer use for the group was 30 hours per week (ranging from 15-50 hrs/wk), with all participants claiming daily internet and email use. The group reported a fairly wide range of simulation experience, ranging from four with no experience to four reporting the “twenty-one to forty hours” category. They also varied in game-playing time and experience, with the median response being less than monthly, while the average was equivalent to playing one hour per week.

Materials. Minimal new materials were developed for the evaluation portion of the exercise. Similar task-based questionnaires were prepared, addressing the battalion staff tasks and subjective estimations of their improvement from the exercise were added (Appendix C). The same questionnaires as before (Appendices A & B) were used to collect the biographical information and usability estimations.

Procedures. As with the previous evaluation sessions, the biographical information was collected before the system training and actual exercises began. Short interviews were conducted with the Staff officers about their exercise position or role, and preparation for the exercise. These interviews were conducted during breaks in the training, during the exercise scenario, and prior to the staff preparation portion of the exercise. Following the entire exercise, a set of function and capabilities questions were completed by all participants, and an open discussion was held about the system capabilities. Some final comments about the system capabilities and the staff response were also acquired during the exercise hotwash (basically an on the spot review of lessons-learned) with the BN commander, executive officer, and S3 during the exercise review held on the morning after the three day exercise.

The AW-VTT Simulation Environment has the capability to operate over the Internet with participants logging in over various Wide Area Network (WAN) nodes. For this exercise, due to network security issues and the operational security of the scenario, it was decided that the exercise would be executed over a Local Area Network at the Training & Training Technology Battle Lab (T3BL) at Fort Dix.

The exercise was set up using 2 separate classrooms, one for the Exercise Control and one for the Training Audience. The MP BN Staff also had a separate planning room where they conducted their staff estimates, vulnerability assessments and briefings to the BN Commander.

Results

Preliminary Interviews. An interview was conducted before the exercise with the BN Commander about his estimation of his staff status for the coordination visit exercise. The overall conclusion was that the Staff was at “Walk” level of expertise and proficiency. The general level refers to the standard categories of capability in the U.S. Army. "Crawl": is used to indicate that personnel do not know all there is to know about a task, that they can't perform the task without guidance. "Walk" is used to indicate that the personnel understand the task, and can perform the task correctly given adequate time and situations with minimal stress. "Run" implies that the task can be fulfilled to standard in required time under battlefield conditions, in other words, done correctly at speed and under stress.

During the preliminary interview with the BN Commander, in particular he saw his new S1 (Personnel) as having little or no experience and needing to learn the basics of his role in the MDMP for Battalion courses of action in this area. His evaluation of the S2 (Intelligence) was that the individual was experienced, and the commander expected him to help the other staff in addressing and using information operations. The S2 had spent a tour in Iraq serving on a Battalion S2 staff, but was a lieutenant (and had not led the Intelligence staff before). The BN Commander also expressed some confidence in the expertise of the S3 (Operations and Planning), although he wanted to see the S3 delegate more to his staff. The newly assigned S4 was viewed as needing considerable focus on the duties and responsibilities of the position, having just moved from commanding a company into the staff position.

The interview with the executive officer (XO) covered the role he played in initiating the exercise, as well as personal and staff preparation for the exercise. Generally, he felt that the overall staff skills level was at the “crawl” level, including himself. This general opinion had been communicated to RDECOM-STTC staff, who had used it as the basis for organizing the information presentation and setting the pace in the Advance Party exercise. The executive officer is primarily responsible for leading the COA preparation and presenting the options to the commander during the briefing.

The interview with the S1 (personnel) was relatively short, as that officer had not had any opportunity for reviewing the duties and responsibilities for the new position. She was coming off two weeks of duty for the state during a weather disaster and stated that she was expressly relying on the XO to guide and mentor her during the course of the exercise. The implicitly self-rated level of expertise was pre-crawl and her major goal from the exercise was establishing a working level of understanding with her staff and the other BN staff positions concerning the duties and responsibilities of the position.

The S2 (Intelligence) officer was a Lieutenant that had been on a Brigade Intelligence staff during a tour of duty in Iraq (no total time was provided). The Lieutenant stated that he thought he had a good skill set and clear abilities for the job of Intelligence officer at the Battalion level. He maintained that although the ARTEPS for a Military Police Battalion were somewhat different than similar tasks for Infantry, that the Intelligence job didn't significantly change other than with the mission. He claimed that he had spent some time reviewing the Field Manuals for pre-deployment survey in order to better serve the BN Commander. He was also quite clear about not having any expectations of learning anything from the simulation exercise. He also made it clear that he considered himself an expert in computer use, as he used computers all day, every day in his civilian job as a stock analyst and trader.

The S3 is responsible for Operations and Planning, the major coordinating position on the staff. The Major had staffed the exercise with available staff personnel, following several planning efforts with the XO. He considered himself to be the lead on the 2 year BN transition from Field Artillery to Military Police. In that role he had studied the different FMs and ARTEPS, preparing and leading thirty to forty briefings for the training staff, NCOs and Officers in the BN. In preparation for the simulation, he had spent one recent day reviewing the Mission Training Plan for a Host Nation pre-deployment survey.

The S4 is responsible for Logistics, and has large amounts of information to acquire and process in preparation for deployment to another country. The S4 reported during the interview that he had varied and different experiences in both the Field Artillery and the Infantry. He had also served in Iraq, and had been involved in the preparation of intelligence analysis and briefings. He claimed prior experience with paper exercises and had a relatively poor opinion of the "canned" nature of the paper exercises.

A major portion of the Staff Exercise questionnaire (see Appendix C) focused on the perceived results of using the AW-VTT. In response to whether the AW-VTT was easier or more difficult to use than other approaches, the simulation was rated as “same” to “easier” than a paper-based Staff exercise, “easier” to “much less difficult” than a Field exercise, and “easier” than a training exercise without troops. They also judged the simulation as requiring equivalent or less preparation than a normal field training exercise or normal staff exercise.

Questions administered after the exercise addressed individual perceptions about their personal capabilities or change based on the simulation portion of the exercise. Overall, the group rated their ability to communicate with their staff or other teams as not changed to slightly better, and rated their ability to recognize hidden problems and prepare alternatives in the same range. They also rated their ability to gather information for the military decision making process, and to deal with the host nation representatives as slightly better (based on a single simulation exercise). No staff member regarded the exercise as impairing their skill, while more than half saw definite improvement in their performance. Furthermore, when rating the difficulty of evaluating environmental information or obtaining information in the simulation, the Soldiers rated the difficulty as moderate but saw extra and critical information that was not normally or easily obtained during staff exercises.

Many comments and insights were acquired from the interviews with the BN Commander and staff officers. The BN Commander commented that taking a significant amount of time in the virtual simulation was actually a good thing, in that it forced the staff to pay attention when and where critical information was provided. He added that just like in the real-world, if someone wasn’t paying attention when information becomes available, it will be missed. His opinion was that the course of the simulation forced his staff to pay attention, take notes and be able to think back to when and where critical information was provided. This requirement was also at odds with the normal staff exercise without troops, which centers on written information provided directly to the staff with minimal “clutter” or diversionary information.

More critically, the commander claimed a much better assessment of staff capabilities resulted from watching the simulation and processes. In addition, the executive officer was impressed with the way that the exercise drove the decision making process in comparison to paper exercises in which it is typically a struggle to focus on the required processes.

Post Exercise Evaluation Survey. Following the BN Staff Pre-Deployment Exercise, all participants completed the Staff Exercise survey (Appendix C). The questions grouped into several general categories; Exercise Effectiveness, Simulation Quality, and User Interface. Medians are the descriptive statistic used in the presentation, as the number of respondents was low (eleven) and the scales were ordinally anchored. The Median value seems to indicate the preponderance of responses better than the mean.

Several of the questions requested a comparison between the AW-VTT simulation and other types of exercises. When asked about ease of use, in comparison to "paper," "Field training," or a "Training Exercise without Troops" the responses were all at the positive end of the 5-point scale. The medians for these comparisons were 4, 5, and 4 respectively; with 5 anchored as "Much Less Difficult" than the comparisons.

A related set of questions asked for a comparison to Field Training Exercises in several areas. The scale anchors for this set of questions was 1 = "Much Better" and 2 = "Better", with 3 = "About Equal." The question stems with resulting Media scores were as follows: Diversity of Tasks Median = 3, Ability to Record Events Median = 1, Time required for Exercise Median = 1, and Ease of Change Median = 2.

When asked to compare the preparation level for the AW-VTT simulation with a normal FTX, the median was 4, with a normal paper staff exercise the median was 4, and with a computer course the median response was 3. The corresponding anchors for this question set were 3 = "About Equal" and 4 = "Less".

A set of questions were also asked about the information available in the AW-VTT simulation that was usable for this staff exercise. When queried about whether there was sufficient information provided, the response labeled "Somewhat Adequate" typified the average selection (Median = 4, Mean = 3.73). When asked about the difficulty of detecting important environment information the Median response was 4 (Moderate, on the only 7 point scale in this group of questions) and the amount of critical information was rated at a median of 3 (Moderate). The question of whether there was information available in the simulation that is not normally available was rated as a median of 3 (Sometimes). The question about the amount of information sought outside the simulation that could have been in the simulation was rated as a median of 3 (Moderate).

The staff was asked to address how well the simulation supported teamwork to accomplish the Staff exercise goals. The five point scale was anchored at 1 by "Prevented" and at 5 by "Enabled." 3 was anchored as "Neither" with 2 as "Hindered" and 4 as "Supported." The four question stems addressed were: Visual aspects, Gesture system capabilities, Communication aspects, and Movement system characteristics. The median scores assigned were all four, with no responses at the "Prevented" end of the scale. The responses thus indicated considerable positive agreement about the capability of the system to support teamwork, although some of the response options were neutral.

Another question set addressed individual self-assessment of capabilities change as a result of the simulation exercise. The question anchors ranged from 1 as "Much Worse," to 5 as "Much Better," with 3 anchoring "No Change." The question stems with resulting Median scores were: Communication with Team members = 3, Communicate with other Staff members = 4, Gather information necessary to support decisions = 4, Deal with Host Nation Representatives = 4, Recognize Hidden Problems = 4, and Prepare Alternatives for Command decisions = 4. The response set indicates that almost all participants deemed the exercise as aiding in improving their capabilities, although not to a great extent. The range of the responses was from the "No Change" (3) to "Much Better" (5) on the scale.

User Interface Evaluation. Questions were asked about the ease of use of the interface (see Appendix C) in relatively straightforward structure. The questions addressed the ease of learning, ease of use, and the appropriateness and usefulness of movement capabilities, environment and avatars. The questions used anchored opinion scales that addressed perceived difficulty of use or agreement with positive statements (e.g., Was the overall User Interface easy to understand and use, very difficult {1} to extremely easy {7}). The median response to the “understand and use” question was 5 (Moderately easy). The staff was asked to rate agreement with the statement "the user interface is a good design for this simulation" and responded with a median of "agree" (2 on the 5 point scale), with all responses at the positive end of the scale.

Questions and responses about the environment addressed buildings, physics, reality, and simulation latency. The realism of simulations and work-arounds to circumvent physics of door interactions, long distance movement, etc., were addressed with both specific and general questions. For example, the question of realism of buildings and objects was rated as Real Enough (median response of 5 on a 7 point scale), and the overall realism was rated as "Moderately good" (5 on a 7 point scale). The lack of functionality of doors (entry requiring either walking through an image of a closed door, or using a menu to appear inside a building environment) was given a median rating of "artificial in places" and the median response was "agree" with the statement that "building entry delay interrupted the simulation." The statement "using menus to enter vehicles or buildings" received a median response of "agree", indicating that the mechanism for dealing with these situations was acceptable. Other aspects dealt with movement from one scenario to another ("teleport" used as a means of transitioning between "distant" locations), which was not perceived as a problem as "waiting for teleport was irritating" rated a median response of "disagree" and "teleporting made it less real" received a median response of "neither agree or disagree." Ratings on the scale of the operations area, transportation speeds and movement in realistic time all received median ratings of "somewhat real" on a 5 point scale anchored from "artificial" (1) to "totally real" (5).

Movement and collisions are another important aspect of individual representation in a multiplayer environment. The movement control systems were rated "moderately easy" to learn, and ease of movement after learning to use the controls was rated as "quite easy." The importance of detecting collisions during movement in the environment was rated as "important" (a 2 on a 5 point scale) and the easy of detecting collisions was given a median rating of "Moderately easy" (a 5 on a 7 point scale).

Questions also addressed the visual aspects of the individual avatars in the environment. Questions about the "look" of the avatars was rated as "somewhat like humans" and although individual avatars in the environment received a median rating of "could barely be identified," overall the appearance of the avatars was rated as "will be more than adequate." The gestures were rated as "moderately good" with a median of 5 on a 7 point scale, and the realism of the gesture capabilities was given a median rating of "somewhat real." The avatar capabilities for Movement, Communication, Visual Inspection, and physical Inspection were also given median ratings of "somewhat real" (4 on a 5 point scale). In addition, the avatars were rated as "Hard" to recognize by physical features or at a distance in the environment, while voice made the identification of avatars "Moderately easy."

The voice system was rated as "workable, with problems" (the mid-point on a 5 point scale), although overall sounds in the environment were judged to be more satisfactory. No important sounds were judged to be missing, unexpected, or incorrect, and the explosions and special effects were rated as "adequate for minimal training." As noted above, the voice (sound) system was rated as the best way to identify avatars, indicating that a considerable amount of information was available through that system.

Overall, the statement "Overall, the system was easy to learn" was given a median rating of "agree", with all responses at the positive to neutral end of the scale. Similar agreement responses were given to a series of statements about the ease of use, system performance, ease of error correction and capability to focus on the mission information. Although errors in use were typically admitted (users disagreed with the statement "did not make many errors"), the system difficulties were not sufficient to create perceived interference with the exercise (raters disagreed with "difficulties in working with the simulation interfered with the exercise").

Post-Exercise Evaluation. On the morning following the three day exercise, a command hot-wash review with the developers, the simulation exercise staff, and the Battalion lead staff was held to discuss their perceptions of the training achieved. The biggest and initial question was whether the exercise objective had been met. This was definitively answered by the Colonel as affirmative, and supported by both the XO and the S3. During a relatively open discussion, the BN Commander surprisingly acknowledged that his assessment of the overall staff capabilities had to be downgraded, primarily due to recent staff losses and replacements coming in at lower levels than expected. In spite of this difference from expectations, he considered the exercise a success as all members of the staff had opportunities to learn processes, information gathering requirements, and the MDMP leading to course of action recommendations. In his opinion, one strength was that multiple staff members had to go through the information acquisition process together, giving them an opportunity to start integrating their respective processes – something that often didn't happen with paper exercises. The XO was especially pleased with the conduct and flow of the simulation, leading to reduced pressure on the leaders to drive the overall process. He regarded the biggest benefit from the simulation as the process being driven by the information in the simulation. In his opinion, paper exercises were much worse in that they required a continued focus on process and push from trainers to maintain the exercise and bring it to a training conclusion.

When specifically queried about comparisons to Janus the officers were unanimous in rating the simulation as superior, although for different reasons. The commander reasoned that the asymmetric nature of ongoing operations were more directly brought out in the simulation, while the executive officer and S3 both emphasized the flexibility of the simulation in responding to information presentation requirements. They did agree with a suggestion that more wide ranging and detailed visual information displays would have enhanced the exercise even further.

Several suggestions for improvements were offered by the staff. Primary among these were the standard support materials that individuals would have available in field exercises. One well-received example offered was the creation of standard topographical maps for understanding the different locations. Another was that they didn't have their typical note taking capabilities, as the keyboard interface almost precluded writing notes. Another suggestion was that the radio nets should have been more structured, enabling specific channels for teams and team coordination rather than having everyone on the same net.

Discussion of Alternative Exercises

It is obvious that realism can be sacrificed for increased training effectiveness, and this point has been made and discussed before (e.g. Hays & Singer, 1988). There are two points that were learned during the Battalion Staff Exercise. First, a Battalion would never send their entire staff as part of an Advanced Party. However, in this exercise, all 12 members of the staff participated in the exercise as a team. Even though this was unrealistic, according to the commander it provided a better training experience for all members to work together and get relevant experience conducting the MDMP. Secondly, the Advance Party conducted a ground reconnaissance of two separate cities. Although different methods were used to move the groups around the simulated environments, in the end it didn't seem to matter. The feedback received from the BN Staff was that they didn't care how they got from Point A to Point B (realistic or not); they were only interested in collecting the information they needed to conduct their staff estimates. Once again the lesson is that reasonable fidelity is only needed in connection with the stimuli surrounding or supporting the learning objectives.

Overall Conclusions

This U.S. Army project is leveraging a commercial MMP environment and evaluating its potential for training Soldiers for the asymmetric warfare missions that they are facing today in places like Iraq and Afghanistan. By representing participants as individuals and enabling the enemy role-players to model the unpredictable behaviors of terrorist groups, in a scalable, diverse environment, AW-VTT offers the U.S. Army training capabilities for the Global War on Terror that is not possessed today.

Soldier Evaluations

Minor indications of successful alterations can be found in the general increase of ratings over the 6 months between the ARNG session and the 101st interviews, with regard to the movement, menus, and weapons controls. The low ratings for the temporarily necessary "teleporting" function for entering buildings has increased the prioritization of developing the capability to directly enter buildings through representative doors. (Kicking down, blowing open, and other standard building entry techniques are even more difficult to implement.) The 2005 addition of easily modifiable explosives (Improvised Explosive Devices, IEDs) that allow refuse piles, bags, carcasses, and vehicles to blow up on command reflects the continued focus on Soldier-identified needs for Asymmetric threats in training. Other identified needs have been prioritized and are being addressed as funded capabilities become available. As noted above, the need for fidelity has to be driven by the stimuli surrounding and supporting the targeted learning objectives.

Perhaps the most encouraging evaluation information comes from the Soldier's recognition of training benefit. The first indication of this comes from their evaluation of the benefit that comes from being able to replay a simulation exercise, and even repeat a critical task immediately following the AAR presentation. Secondary evidence comes from the high agreement with statements on the system's adequacy in supporting and supplementing STX, and relatively good rating on usefulness in the crawl and walk phases of training. Additional evidence comes from the leaders of the Battalion Staff Exercise, in their re-evaluation of staff capabilities and estimations of increased understanding from the exercise.

Soldiers indicated that AW-VTT offers unique training capabilities by representing participants as individuals and enabling enemy role-players to model unpredictable behaviors in a scalable, diverse environment. The data were collected in the context of system developers demonstrating existing features and presenting projected features, tools, and capabilities. The system was evaluated by a small number of users (unavoidable given the current operating tempo of the U.S. Army), that seemed to match well in computer skills and training background, as well as demographics of the U.S. Army population. For the foreseeable future, limited Soldier input and a relatively rapid development pace will continue. The approach used here was to gather data which related the assigned evaluation sample to the relatively easily available information about the potential user population. As a result of the match in demographics, the developers feel confident in using the evaluative information to prioritize development.

Potential Usefulness

Clearly Soldiers require acceptable movement, shooting capabilities, and communication components as an acceptable foundation for even low-fidelity simulation-based training. It seems reasonable to accept ratings from user groups that are actively engaged in training for deployment. The Soldiers' general comments indicated that being able to look at an environment dramatically changed their approach to the (example) mission. This is an important point, given that typically Soldiers use map and sand table or parking lot exercises to review training mission orders. In addition, they thought that being able to conduct rehearsals within a low-fidelity simulation would enhance the training value of a wide variety of field exercises. This in spite of the discussion about problems in the actual use of the system in training – Soldiers feel the need to get their boots on the ground. A few Soldiers in the 101st group discussed their concerns about whether the use of a PC-based system might lead to fewer actual field exercises, which in their opinion would decrease the overall effectiveness of U.S. Army training. This also brings up the issue of comparison – there is no simulation that these Soldiers had available for comparison, and the evaluation questions were not comparative but based on how well the simulation seemed to support (or in some cases was described as eventually being able to support) individual Soldier training.

AW-VTT represents a new way to train within the U.S. Army. It could be used to supplement training at schools, at home-station, and military training centers. The opinion-based information gathered from questionnaires and interviews indicates that the system can prepare troops for more expensive live drills and actual deployment integrating basic Warrior skills with reinforced situational awareness, decision making, and asymmetric warfare skills acquired through experiential learning about the enemy and surrounding civilian population. Perhaps the most tantalizing benefit, however, is that by enabling deployed troops to log in and coach Soldiers who are preparing to deploy, the AW-VTT environment will enable the U.S. Army to rapidly develop, validate, communicate, and train new TTPs at a velocity that provides individual Soldiers and teams on the ground with an increasing advantage over insurgencies and terrorist organizations that are embedded in local populations.

References

(2005). Defense Population Representation 1998, Undersecretary of Defense for Personnel and Readiness, Retrieved October 4, 2005, from http://www.defenselink.mil/prhome/poprep98.

Basic Officer Leaders Course (2005). Retrieved June 10, 2005, from www.infantry.army.mil/BOLC.

Beal, S.A. & Christ, R.E. (2004). *Training Effectiveness Evaluation of the Full Spectrum Command Game*. Technical Report 1140. US Army Research Institute for the Behavioral and Social Sciences.

Beal, S.A. & Christ, R.E. (2005). Evaluation of a Game-Based Trainer for Infantry Platoon Leader Decision Making. . *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*. Santa Monica, CA: Human Factors & Ergonomics Society.

Boldovici, J.A., Bessemer, D.W., & Bolton, A.E. (2002). *The Elements of Training Evaluation*. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Bonk, C.J. & Dennen, V.P. (2005). *Massive Multiplayer Online Gaming:A Research Framework for Military Training and Education*, Technical Report 2005-1, Advanced Distributed Learning Initiative, Readiness and Training, Office of the Secretary of Defense for Personnel and Readiness.

Cahlink, G. (2004). Trial by Fire. *Government Executive.* Retrieved June 8, 2005, from http://www.govexec.com/features/0405-01/0405-01s4.htm#top

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally College Publishing Co.

Fober, G. W. Bredthauer, J. L., & Dyer, J. L. (2001). Computer background of Soldiers in FORSCOM units: FY00. (RR-1778, ADA399393) Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Hays, R. T., & Singer, M. J. (1988). *Simulation Fidelity in Training System Design: Bridging the Gap between Reality and Training*. New York: Springer-Verlag.

Knerr, B. W., Lampton, D. R., Thomas, M., Comer, B. D., Grosse, J. R., Centric, J. H., Blankenbeckler, P., Dlublac, M., Wampler, R. L., Siddon, D., Garfield, K. A., Martin, G. A., & Washburn, D. A. (2003). *Virtual Environments for Dismounted Soldier Simulation, Training, and Mission Rehearsal: Results of the FY 2002 Culminating Event.* Technical Report 1138 (ADA417360), Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knerr, B. W., & Lampton, D. R. (June, 2005). *An Assessment of the Virtual-Integrated MOUT Training System (V-IMTS).* (Technical Report 1163), Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Mayo, M., Singer, M. J., & Kusumoto, L. (December, 2005). Massively Multi-Player (MMP) Environments for Asymmetric Warfare. *Proceedings of the 27th Interservice/Industry Training Systems and Education Conference*. Arlington, VA: National Training Systems Association.

Nielsen, J., (1993). *Usability Engineering*. San Francisco, CA: Morgan Kaufmann.

Wunderle, William D. (2006). *Through the lens of cultural awareness: a primer for U.S. Armed Forces deploying to Arab and Middle Eastern countries.* Combat Studies Institute Press, Fort Leavenworth, KS.

Appendix A

Participant Information Questionnaire

Participant Information Questionnaire

Enter Session Number	
Enter Group Number	
Enter Participant Number	
PURPOSE: The U.S. Army Research Institute (ARI) Simulator Systems Research Unit is supporting the Research, Development and Engineering Command in conducting research on low fidelity, highly flexible, multi-player systems. Part of the information we need concerns biographical information, health-related information, and information on how you interact with computers. We appreciate your cooperation and your time. Your responses will remain anonymous in the processing of all data. DIRECTIONS: Please respond to as many of the following items as possible. Position the cursor over a response, then click the left mouse button to select it. If applicable, type in your answer. If you want to change a selection, you may need to make an entry in another question first. Click on the tab at the top of the page to move to the next set of questions. Please tell the experimenter when you are finished.	
1. Please type in your age.	textbox
2. Please select your rank/grade.	
3. Please type in your years and months of active duty in the Army, if any.	Textbox textbox
4. Please type in your current duty position.	
5. Please select your Army branch or profession.	
6. If officer, please select your source of commission.	
7. Please type in your Primary and Secondary MOS.	Textbox textbox
8. How many hours during the average week do you spend training others? (Include preparation and execution.)	textbox
9. How susceptible to motion or car sickness do you feel you are?	7 pt scale
10. Do you have a good sense of direction?	Yes/no
11. When did you use computers in your education? (Select all that apply)	Grade school, middle, high, college, technical, graduate
12. Type in the number of hours per week that you use a computer. Use a decimal format, e.g, 7.5, 8.0, etc.	textbox
13. Where do you currently use a computer? (Select all that apply)	Home/Barracks/BOQ Unit/Work Library/Learning Ctr/Training Facility Do not use
14. Do you own a personal computer?	Yes/no

15. How often do you: (Select how frequently you use each)	Mouse Computer games Icon-based programs Graphics/Drawing Features Programs w/ pull-down menus Email Internet
16. I enjoy playing video games (home or arcade):	5 pt scale Disagree – Unsure - agree
17. I am __________ at playing video games:	5 pt scale Bad – Average - good
18. Type in the number of hours per week that you play video games. Use a decimal format, e.g, 7.5, 8.0, etc.	
19. How many times in the last year have you experienced a virtual reality game or entertainment?	12 pt scale – 0 – 10+ times
20. Which of the following best describes your typing ability? (Select one)	Hunt & peck slowly Hunt & peck quickly Type slowly while not looking Type quickly while not looking
21. Which of the following best describes your expertise with computers? (Check one)	Novice Good with one program Good with several packages Program in one language & use several packages Expert – Bill Gates would hire me
22. If you are good with one or more software packages, please check those that apply	Word Processing Audio Media Spreadsheet Visual Media Database Internet Browser Slides Scheduling/Calendar Other
23. If you are good with one or more languages, please check those that apply.	Visual Basic HTML Java C++ Other
The remaining questions present icons, and the respondent is supposed to enter a label or description in the text box next to the icon.	

Spellcheck	Print
Savetodisk	Undo
Paste	Zoom
mousepoint	Cut
newfile	openfile
copy	drawarrow
recyclebin	centeralign
close	help
fill	group

Appendix B
Usability Questionnaire

Please enter the serial number (no letters) from a dollar bill in the blank at the right. This will be used to track your responses ANONYMOUSLY, and the same number should be used in all surveys. THANK YOU FOR YOUR COOPERATION!	__________

Please enter todays date.	________________________________

Please select the category that best describes your use of the system today.		
	○	Acting as a Trainer
	○	One of the Trainees
	○	Acted as a Role Player

What was your overall impression of the exercises conducted today?	[- Select One -]

Was there adequate time to prepare for the Exercises or Scenarios?					
	Very fast	Too fast	Fairly quick	Moderatly paced	Too slow
	○	○	○	○	○

Was the overall User Interface easy to understand and use?	[- Select One -]

The User Interface is a good design for this simulation.		
	○	Strongly agree
	○	Agree
	○	Neither agree nor disagree
	○	Disagree
	○	Strongly disagree

Was there sufficient time for "hot wash" after action reviews between the exercises?		
	None	○
	Not enough to recover	○
	Just enough	○
	More than enough	○
	Far too much	○

How much "boredom" was there between the exercises?		
	None	○
	Not enough to recover	○
	Just enough	○
	More than enough	○
	Far too much	○

Overall, which control system seemed to work the best for you?	○	Game Controller	○	Keyboard & Mouse

Were the special function keys easy to use?	[- Select One -]

Was the voice quality adequate?	[- Select One -]

Was the movement control system easy to learn to use?	[- Select One -]

Was it easy to move around in the environment after you learned to use the controls?	[- Select One -]

Is experiencing collisions in the simulation important in moving an avatar around?	[- Select One -]

Was it easy to detect collisions during movement (for example, hitting doorways during entry)?	[- Select One -]

Did aiming and shooting the simulated weapons work well enough to support the goals of the training exercises?	[- Select One -]

How good or appropriate were the animated gestures?	[- Select One -]

Were there any important gestures that were not implemented, or was there a gesture that needed improvement?

____
____
____

Please provide a short description of any controls that were NOT easily learned or did not work as expected.

____
____
____
____
____

Overall the avatars in the environment looked:	
very human-like	○
somewhat like humans	○
adequate in appearance	○
un-natural and artificial	○
like cartoons	○

Individual avatars in the environment:	
were not very easily identified	○
could barely be identified	○
could be identified	○
could easily be identified	○
were very easily identified	○

The appearance of the avatars in the environment:		
	will not support training	○
	will barely support training	○
	will be adequate for training	○
	will be more than adequate	○
	will enhance training effects	○

Was there any noticeable latency in the simulation	[- Select One -]

Did the explosions and special effects seem real enough for training?		
	Too Hollywood for training	○
	Barely adequate representations	○
	Adequate for minimal training	○
	More than adequate for training	○
	Good, will improve performance	○

Was the voice system adequate to support these training exercises?		
	Inadequate	○
	Barely adequate	○
	Workable, with problems	○
	Completely adequate	○
	More than adequate	○

Which worked better, the "hands-free" or the "push-to-talk" voice control?				
	○	Hands-free	○	Push-to-Talk

What voice or communication capability needs to be improved or added to this system for general Army training?		
	____________________ ____________________ ____________________	

How good was the environment realism?	[- Select One -]

Could this simulation support Army training as it works right now?		
	Incapable of Training	○
	Could support one Task	○
	Could support a few Tasks	○
	Could support Basic Tasks	○
	Could support Many Tasks	○
	Could support Most Tasks	○
	Could support all Tasks	○

What is the most important feature or capability needed by this system to better support Army training and rehearsal?		
	______________________________ ______________________________ ______________________________	

Appendix C
Staff Exercise Questionnaire

Please enter your name. We will only use this information to match your responses to the other questionnaires you have answered. All data afterward will be anonymous and only reported in aggregate form. THANK YOU FOR YOUR COOPERATION.		
	_________________________________	

Please enter todays date.	_________________________________

Please select the category that best describes your use of the system today.		
	○	Acting as a Trainer
	○	One of the Trainees
	○	Acted as a Role Player

Was the overall User Interface easy to understand and use?		
	○	Very Difficult
	○	Difficult
	○	Hard
	○	Moderate
	○	Moderatly easy
	○	Quite easy
	○	Extremely easy

Was the AW-VTT Simulation easier or more difficult than:						
		More Difficult	A Little Harder	About the Same	A Little Easier	Much Less Difficult
	a “paper” Staff Exercise.	○	○	○	○	○
	a Field Training Exercise	○	○	○	○	○
	Than a TEWT (Training Exercise Without Troops)	○	○	○	○	○

The User Interface is a good design for this simulation.		
	○	Strongly agree
	○	Agree
	○	Neither agree nor disagree
	○	Disagree
	○	Strongly disagree

Overall, which control system seemed to work the best for you?	○	Game Controller	○	Keyboard & Mouse

Were the special function keys easy to use?		
	O	Very Difficult
	O	Difficult
	O	Hard
	O	Moderate
	O	Moderatly easy
	O	Quite easy
	O	Extremely easy

Was the movement control system easy to learn to use?		
	O	Very Difficult
	O	Difficult
	O	Hard
	O	Moderate
	O	Moderately easy
	O	Quite easy
	O	Extremely easy

A a result of the AW-VTT Exercises, how do you think your Staff Capabilities changed?		
	O	Huge Decrease
	O	Moderate Interference
	O	Slight Hindrance
	O	No Improvement
	O	Slight Improvment
	O	Moderate Improvement
	O	Vast Improvement

Was it easy to move around in the environment after you learned to use the controls?		
	O	Very Difficult
	O	Difficult
	O	Hard
	O	Moderate
	O	Moderatly easy
	O	Quite easy
	O	Extremely easy

How realistic were the buildings/facilities?		
	O	Totally Artificial
	O	Generally Too Artificial
	O	Artificial in Places
	O	Minimal Level Needed
	O	Real enough
	O	Somewhat Real
	O	Totally Real

Is experiencing collisions in the simulation important in moving an avatar around?		
	O	Very important
	O	Important
	O	Unimportant
	O	Unnecessary
	O	Interfering

Is there a problem in not having functioning doors? Please rate the building entry capabilities of the system.		
	O	Totally Artificial
	O	Generally Too Artificial
	O	Artificial in Places
	O	Minimal Level Needed
	O	Real enough
	O	Somewhat Real
	O	Totally Real

How easy was it to recognize the avatars throughout the simulation ?						
		Difficult	Hard	Moderate	Moderately easy	Quite easy
	by physical features	O	O	O	O	O
	by voice	O	O	O	O	O
	at a distance	O	O	O	O	O

Was it easy to detect collisions during movement (for example, hitting doorways during entry)?		
	O	Very Difficult
	O	Difficult
	O	Hard
	O	Moderate
	O	Moderately easy
	O	Quite easy
	O	Extremely easy

Were the avatar's capabilities realistic?						
		Artificial	Somewhat Artificial	Minimal Level Needed	Somewhat Real	Totally Real
	Movement	○	○	○	○	○
	Communication	○	○	○	○	○
	Gesture	○	○	○	○	○
	Visual Inspection	○	○	○	○	○
	Physical Inspection	○	○	○	○	○

How good were the animated gestures?		
	○	Very Poor
	○	Unsatisfactory
	○	Mediocre
	○	Fair
	○	Moderately good
	○	Quite good
	○	Extremely good

How does the simulation compare to a field training exercise in the following areas?						
		Much Better	Better	Equal	Worse	Much Worse
	Diversity of tasks	○	○	○	○	○
	Ability to Record Events	○	○	○	○	○
	Time required for exercise	○	○	○	○	○
	Ease of Change in System	○	○	○	○	○

Were there any important gestures that were not implemented, or was there a gesture that needed improvement?		
	____ ____ ____	

Were there any important sounds:			
		No	Yes
	Missing?	○	○
	Missing when expected?	○	○
	Incorrect in characteristics?	○	○
	Unexpected when they occurred?	○	○

Please provide a short description of any controls that were NOT easily learned or did not work as you expected.
____
____
____
____
____

Some simulation features are difficult to implement. Please rate your agreement with the following statements.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Waiting for teleport was irritating.	○	○	○	○	○
The building entry delay interrupted the simulation	○	○	○	○	○
Using menus to enter vehicles or Buildings is OK.	○	○	○	○	○
Teleporting made it less real.	○	○	○	○	○

Overall the avatars in the environment looked:

very human-like	○
somewhat like humans	○
adequate in appearance	○
un-natural and artificial	○
like cartoons	○

Did the simulation require more or less preparation: ?

	A Lot More	More	About Equal	Less	A Lot Less
than a normal FTX	○	○	○	○	○
than a normal staff exercise	○	○	○	○	○
than a computer course	○	○	○	○	○

Individual avatars in the environment:

were not very easily identified	○
could barely be identified	○
could be identified	○
could easily be identified	○
were very easily identified	○

Overall the system was easy to learn.		
	○	Strongly agree
	○	Agree
	○	Neither agree nor disagree
	○	Disagree
	○	Strongly disagree

The appearance of the avatars in the environment:		
	will not support training	○
	will barely support training	○
	will be adequate for training	○
	will be more than adequate	○
	will enhance training effects	○

Was there any noticeable latency in the simulation?		
	○	System was very fast
	○	System was fast
	○	Adequately fast
	○	Moderately slow
	○	Too slow

Did the explosions and special effects seem real enough for training?		
	Too Hollywood for training	○
	Barely adequate representations	○
	Adequate for minimal training	○
	More than adequate for training	○
	Good, will improve performance	○

Was the voice system adequate to support these training exercises?		
	Inadequate	○
	Barely adequate	○
	Workable, with problems	○
	Completely adequate	○
	More than adequate	○

Which worked better, the "hands-free" or the "push-to-talk" voice control?				
	○	Hands-free	○	Push-to-Talk

What voice or communication capability needs to be improved or added to this system for general Army training?		
	____ ____ ____	

How good was the environment realism?		
	○	Very Poor
	○	Unsatisfactory
	○	Mediocre
	○	Fair
	○	Moderately good
	○	Quite good
	○	Extremely good

In general, could this simulation support Army training as it works right now?		
	Incapable of Training	○
	Could support one Task	○
	Could support a few Tasks	○
	Could support Basic Tasks	○
	Could support Many Tasks	○
	Could support Most Tasks	○
	Could support all Tasks	○

What is the most important feature or capability needed by this system to better support a wide range of Army training and rehearsal?		
	____________________ ____________________ ____________________	

Please indicate your level of agreement with the following statements.						
		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
	Once I got used to the simulation, I could easily focus on the necessary information for accomplishing my part of the mission.	○	○	○	○	○
	It was easy to remember how to do things in the simulation, once I had done them a couple of times.	○	○	○	○	○
	The system performed as I expected.	○	○	○	○	○
	It was easy to correct any errors made during operation of the simulation.	○	○	○	○	○
	I did not make many errors in using the simulation.	○	○	○	○	○
	The difficulties in working with the simulation interfered with the exercise.	○	○	○	○	○
	Using the simulation in support of the Staff Exercise was a good experience, I would like to use it in other exercises.	○	○	○	○	○

What other gestures in the system are needed for MP Staff exercises like this one?

____
____
____
____
____

How well was rank and authority reflected in the simulation?						
		Totally Inadequate	Somewhat Inadequate	Minimal Level Needed	Somewhat Adequate	Totally Adequate
	Were indications of rank clearly available.	○	○	○	○	○
	Were there indications of civilian status.	○	○	○	○	○
	Was it possible to exercise authority to accomplish goals?	○	○	○	○	○

What is the most important aspect of the visual displays to improve?		
	______________________________ ______________________________ ______________________________ ______________________________ ______________________________	

How real or artificial were the following major aspects of the simulation?						
		Artificial	Somewhat Artificial	Minimal Level Needed	Somewhat Real	Totally Real
	Was the area of operations realistically scaled?	○	○	○	○	○
	Was the transportation speed reasonable for training?	○	○	○	○	○
	Did you cross physical distances in realistic time?	○	○	○	○	○

What problems/opportunities does an FTX present that the simulation doesn't?		
	______________________________ ______________________________ ______________________________ ______________________________ ______________________________	

Rate the areas below in terms of supporting teamwork to accomplish exercise goals.						
		Prevented	Hindered	Neither	Supported	Enabled
	Visual aspects.	○	○	○	○	○
	Gesture system capabilities.	○	○	○	○	○
	Communication aspects.	○	○	○	○	○
	Movement system characteristics.	○	○	○	○	○

The AAR system made it easy to review and determine what happened in the simulation during the exercise.					
	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
	○	○	○	○	○

The AAR system made it easier to determine which areas to focus upon during future exercises.					
	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
	○	○	○	○	○

As a result of the simulation portions of the exercise, how do you think your capabilities have changed in the following areas?						
		Much Worse	Slightly Worse	No Change	Slightly Better	Much Better
	Communication with team members.	○	○	○	○	○
	Communicate with other Staff members.	○	○	○	○	○
	Gather the information necessary to support decisions.	○	○	○	○	○
	Deal with host nation representatives	○	○	○	○	○
	Recognize hidden problems.	○	○	○	○	○
	Prepare alternatives for Command.	○	○	○	○	○

What types of training or rehearsal tasks do you think this simulation system (not just this exercise) is BEST suited to support?		
	____________________ ____________________ ____________________ ____________________ ____________________	

What types of training or rehearsal tasks do you think this simulation system (not just this exercise) is LEAST suited to support?		
	____________________ ____________________ ____________________ ____________________ ____________________	