Training in Peacekeeping Operations Using Virtual Environments

Dr. R. Bowen Loftin, Dr. Mark W. Scerbo, Dr. Rick McKenzie, Ms. Jean M. Catanzaro, Mr. Nathan R. Bailey, Mr. Mark A. Phillips, and Dr. Gaye Perry[1]
Virginia Modeling, Analysis, and Simulation Center
Old Dominion University
Suffolk, VA 23435 USA

Summary

The present paper describes two studies aimed at evaluating Virtual Environment (VE) technology for training individuals to perform military checkpoint duty. Participants stood guard at a fictitious base in which simulated drivers in vehicles approached seeking entrance. Participants inspected each vehicle, interacted with the drivers, verified their identification, and made a decision to allow the driver to enter the base, detain the vehicle, or asked the driver to turn around and leave. The first experiment was conducted in a CAVE environment with stereoscopic visual and auditory displays, participant tracking, and voice recognition. The second experiment provided the same training on a desktop system. The results of both studies showed that participants learned quite effectively with either interface, but that overall levels of performance were better with the fully immersive VE. These findings suggest that VE technology holds promise for activities that are more like experience-based training and which place a greater emphasis on social interaction skills.

Introduction

Military training has been traditionally aimed at preparing soldiers to apply doctrine and to react instinctively to accomplish objectives. This type of training is necessary for soldiers to win wars and to minimize casualties and collateral damage. Unfortunately, these are the same soldiers who become ambassadors, peacekeepers, and police in a disrupted state where the complexities of the environment are so great that instinctive or skill-based behavior is simply not enough to cope with unexpected and complex situations. Too often, soldiers are faced with difficult and politically-sensitive decisions for which they have received no training at all.

During the past 20 years, the United States has engaged in two wars but has been a participant in nearly thirty major peacekeeping operations. Further, since most military operations today are intensely scrutinized by the news media, it has become evident that the actions of even the most junior members of a military unit may profoundly impact world opinion and affect the most senior levels of leadership.

This was made abundantly clear in the recent war in Iraq. On March 29, 2003, a suicide bomber at a checkpoint near Najaf driving a taxicab feigned engine trouble. When soldiers approached to inspect the situation, the driver blew up the vehicle. Four soldiers were killed. Within 48 hours, another vehicle near Najaf failed to heed warnings to stop as it approached a checkpoint. After several unsuccessful attempts to try and slow the vehicle, the soldiers fired into the vehicle, killing seven women and children. Each of these incidents made headlines and required military commanders to scrutinize their rules of engagement. Thus, proper training of military personnel, at all levels has never been more crucial.

A recent review of military applications of virtual environments (VEs) indicates that most efforts have focused on training teams, leadership skills, mission rehearsal, and navigational skills (Knerr, Breaux, Goldberg, & Thurman, 2002; Pew & Mavor, 1998). Although the potential of VE technology for addressing the interpersonal skills needed in many military activities has certainly been acknowledged, few systems have actually been developed to meet this need. Accordingly, the present paper describes the application of VE technology to a novel task, the military checkpoint.

[1] University of Houston

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **00 APR 2004**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Training in Peacekeeping Operations Using Virtual Environments**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Virginia Modeling, Analysis, and Simulation Center Old Dominion University Suffolk, VA 23435 USA**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001667, NATO RTO-MP-HFM-101 Advanced Technologies for Military Training (Technologies avancées pour lentraînement militaire)., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**12**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

The primary objective was to recreate the kind of experiences a military guard would encounter while standing watch and to determine whether a VE would prove to be an effective learning medium. Toward this end, computer-controlled virtual humans and live human participants took part in a peacekeeping task whereby various skill-based activities were trained and evaluated in a VE. The task used was a checkpoint operation in a typical third world urban area. Simulated drivers in vehicles would approach a checkpoint seeking entrance to a fictitious military base. Trainees would inspect each vehicle, interact with the drivers, and ask for proper identification. The driver would produce an ID card and the trainee had to verify that it was valid. The scenario would end when the trainee made a decision to allow the driver to enter the base, to pull over the vehicle, or ask the driver to turn around and leave.

The training addressed six specific objectives (see below). In the first study, the training took place in an immersive environment using CAVE (CAVE Automatic Virtual Environment) technology. The system incorporated speech recognition software with a focused natural language interface. Trainee movements within the environment were monitored by an Ascension Flock of Birds magnetic tracking system. This tracking information was provided to the virtual human agents in the environment. The technology permitted a high level of interaction between trainees and the human models. These virtual agents answered questions, knew where the trainees were in the environment, and replied while looking the trainees “in the eye.”

In a typical scenario, a trainee would approach the car and ask the virtual driver for identification. The trainee’s virtual partner provided cover for the trainee during the identity check. The driver produced an ID card and the trainee verified that it was appropriate.

Two groups of individuals participated. The first group participated in one session, received feedback on their performance, and then returned for a second session. It was expected that these individuals would perform better (i.e., make fewer errors) in the second session if the VE was an effective training medium. A second group of individuals participated in only a single session. These participants were trained with the same scenarios that the first group performed in their second session. This second group served as a control for the specific scenarios performed by the first group on their second session. Thus, if the participants in the first session truly benefited from their training, one would expect their performance on the second session to be superior to that of the second group who only performed a single session. Further, performance levels for Group 2 should be similar to those of Group 1 on their initial session.

The second study was a replication of the first experiment with one important difference. The scenarios were presented on a desktop VE system. A new interface was created to allow the trainees to navigate and inspect the vehicles. This study was intended to provide a comparison between the immersive CAVE and desktop VE.

The Training System

Virtual Human Agents

Virtual human agents were created with Jack Tool Kit, a 3D modeling environment with support for high degrees of freedom human models. These models are typically used to evaluate ergonomic factors pertaining to the modeled environment. The human models within Jack were selected for this project because of the range of dynamic motion available. Jack includes utilities for locomotion, head and eye movement, arm and leg movements, and movement of all joints. The extent of motion of the human models is always within the physical constraints of selectable human body types. As a result, one is assured of gestures and positions that are within the realm of possibility, given the particular human in a particular environment.

Behaviors in Jack are supported through layers of interfaces with decreasing complexity. At the lowest level, rotations and translations of 68 joints in the human figure are supported. Above this layer are a number of primitives that control movement of individual human body parts such as Move Arm, Move Head, Bend Torso, Rotate Pelvis, etc. These primitives are combined to create an executable behavior in the Jack agent instantiated for the target application. A network of these executable behaviors provides the activities and reactions that the agent will exhibit during part or possibly throughout an entire scenario. The network consists of basic transition nodes as well as nodes that can execute in parallel. Thus, the behavioral network is called a Parallel Transition Network (PatNet; Badler et al., 2000). Another layer of generalized capability called Parameterized Action Representation (PAR) is available in Jack that supports both natural language commands and automatic behavioral animation (Badler et al., 2000).

Training Scenarios

The present set of experiments examined the application of VE technology to a military checkpoint task. The primary objective was to reproduce the kind of experiences a military guard would encounter while standing watch. Thus, the task, setting, and virtual characters were created to match typical checkpoint conditions as closely as possible.

The checkpoint locale was recreated from the U.S. Marine training town in Quantico, VA. A pre-existing Quantico model was updated by remodeling existing structures for improved real-time performance and by applying texture map created from photographs obtained at the site. Scene graph construction and rendering was done with VrTool. There are two scene graphs, one in VrTool and the other in Jack. VrTool's scene graph is what is actually seen and rendered on the screen, whereas Jack's is used internally for dynamic character animation calculations.

The task required the creation of many distinct training scenarios. In actuality, the process of manning a checkpoint can be a highly repetitive, mundane activity. Rarely does anything out of the ordinary occur. Accordingly, the training scenarios were designed from this perspective. Specifically, a general or neutral scenario was created which begins when a vehicle approaches and stops at the checkpoint. The trainee inspects the vehicle and asks the driver, a virtual human intelligent agent, for identification. The driver produces an ID card. The trainee verifies that it is valid and, if so, allows the driver to pass. The trainee's partner (another virtual agent) provides cover for the trainee during the interaction.

In order for the training experience to more closely reproduce the true conditions of this activity, each neutral interaction had to be unique. Thus, a pool of neutral scenarios were generated that varied in vehicle type, vehicle color, driver's sex, skin color, hair color, and shirt color. Although the characteristics of the neutral scenarios varied, they all unfolded in the same manner. In this regard, every attempt was made to create a training environment that would reproduce the *experience* of standing watch at a checkpoint.

The participants' ability to follow protocol and exercise judgment was examined by including a variety of critical scenarios. The critical scenarios appeared at random intervals throughout the training session and unfolded without any cues to distinguish them from the neutral scenarios. Specifically, these scenarios addressed the following training objectives:

a) the ability to handle matters of situational urgency according to procedure;
b) the ability to resist social pressures that conflict with procedure;
c) the ability to recall and identify vehicles, people, and license plates from a predefined target list;
d) the ability to perceive inappropriate objects/contraband or the absence of required information;
e) the ability to maintain situation awareness.

Thus, for example, in one scenario an ambulance arrives without proper authorization via radio alert and the driver advises the trainee that he does not have time to go through the normal identification verification routine because he has an injured passenger. The trainee is responsible for the security of the base and must follow proper procedure and perform an identification check on both the driver and passenger even if confronted with an urgent situation. In another type of critical scenario, the trainee was presented with specific information about a vehicle that he/she needed to remember and was instructed to watch out for during his/her shift. Other scenarios required the trainees to detect a missing base sticker or identify the presence of contraband items (see Figure 1). The trainees were also evaluated on their ability to maintain situational awareness by attending to two events simultaneously. Specifically, in one scenario the driver of a second vehicle awaiting entry to the base exhibits suspicious behavior. The trainee must interact with the driver in the first vehicle while monitoring the activity in a second vehicle. In this particular scenario, if the participant failed to ask for backup from their virtual partner, the driver in the second vehicle would pull out a gun and fire. Although the scenarios are mainly skill-based, they are still representative of the kinds of judgment a checkpoint guard must make.

Experiment 1: Fully Immersive CAVE Environment

Virtual Environment Implementation

The VE interface used in the first study was the CAVE (CAVE Automatic Virtual Environment).

Figure 1. Example of a vehicle approaching the checkpoint. Note the missing license plate.

The system configuration for the first experiment is illustrated in Figure 2. There are three main computing systems connected through a 100-mbps network switch as described below:

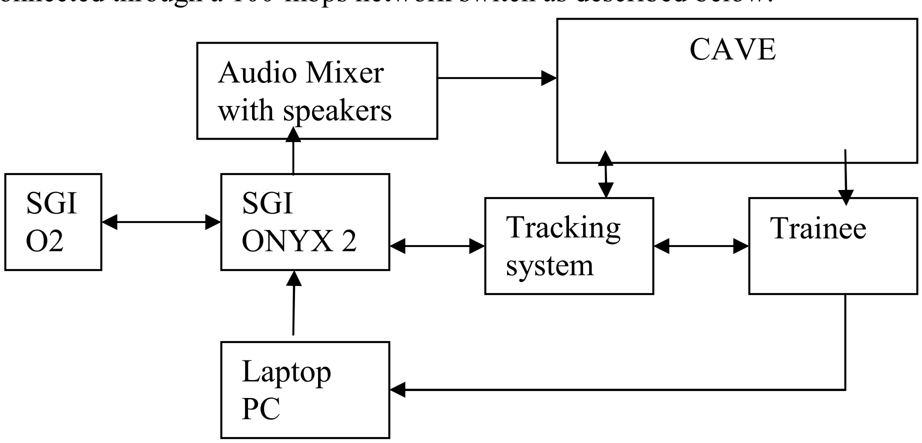

Figure 2. Experiment 1 hardware configuration.

- An SGI ONYX 2 computer was used to display the application in the CAVE, provide the sound playback, and read the information from the tracking devices. This computer used VrTool, TrackD, Jack, Python, Open Inventor, and IRIX 6.5.
- An SGI O2 computer was used as the experiment main console. From that machine one could launch the application and have override controls during the simulation. This computer used IRIX 6.5, Motif, and buttonfly.
- A PC computer was also used for the voice recognition software and to communicate the information to the SGI ONYX 2 through a network socket. This computer used Windows 2000, IBM ViaVoice, and VrSpeech.

Images were presented on two 10x10 ft walls of the CAVE with a resolution of 1280x1024. The images were viewed stereoscopically with LCD CrystalEyes stereo shutter glasses. Positional tracking was provided through Ascension Technology's Flock of Birds software, a six-degree-of-freedom (6DOF) tracker able to track one to four sensors simultaneously. A single head sensor was attached to the CrystalEyes LCD shutter glasses. The participant wore a wireless headset microphone to communicate the voice commands to the PC running IBM's Via Voice Recognition software.

Audio Elements

The IBM ViaVoice Speech Recognition required two components—a grammar and a dictionary. The most difficult challenge concerned creating the grammar. The SRCL (Speech Recognition Control Language) used was a particular type of the BNF (Backus-Naur Form) generic grammar representation (IBM Corporation 1997). It supports substitutions and repetition and can generate very complicated sentences while at the same time addressing a wide selection of the possible commands. The dictionary provides the software with pronunciations for each word to be recognized. In this study, it was designed for an East Coast USA accent. For each word, it was necessary to say the word to record it. The software would then convert it into a corresponding 'baseform' representation according to its lexeme (spelling).

Communication from the laptop program to the main program running on the SGI was established using sockets. It connected to VrSpeech (a component of VrTool) that is designed specifically to receive ASCII strings via socket communication.

Audio files were created for the virtual humans and for sound effects. For the speech files, male and female voices were recorded for a variety of phrases such as: "Here's my ID, I don't have a pass, I don't understand," etc. For the neutral scenarios, a group of four male voices and four female voices were recorded, using an identical set of scripts for each individual. Because only eight unique voices were recorded for the neutral scenarios, there was some repetition of the individual voices. For each of the critical scenarios, a unique script was written and recorded. For the critical scenarios, 12 male and seven female voices were used. Again, in some cases, the same voice was used in more than one critical scenario. However, when repetition did occur, it took place in critical scenarios that were presented in different experimental sessions.

Each voice was recorded in mono at a 22.1 kHz sampling rate. Background and other supplemental audio sounds including gunfire, airplane flybys, wind, and the approach of a car on a gravel surface, were created using a combination of existing sound samples and environmental sound recordings. A noise reduction algorithm was used to eliminate unwanted noise (including hiss, clicks and pops), for each of the audio files. The files were converted to Audio Interchange File Format Version C (.AIFFC) for final presentation in the CAVE environment

.All sounds were displayed in the CAVE environment through a standard four-channel soundboard, though only two of those channels were used for the auditory presentations in this experiment. With the experimental participant facing forward in the CAVE, the left and right speakers were placed at approximately 225 and 315 degrees, respectively. The speakers were mounted on speaker stands at an elevation of approximately five feet. Throughout the majority of the experimental session, the speech files and background samples did not exceed 85dB.

Experimental Method

Participants. Thirty-two undergraduate students from Old Dominion University with normal or corrected-to-normal vision participated in the study. They were offered either (a) four hours of extra credit or (b) $30 as compensation for their time. This population was chosen because they were representative of the type of individual who would likely be assigned to guard duty in the military. Participants who were predisposed to simulator sickness as indicated by the Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993) were excluded from participating. In addition, individuals who reported previous experience with checkpoint duty were excluded from the analyses and four were replaced due to system malfunctions.

General procedure. Participants were all run individually. They were asked to complete a background survey addressing demographics and military experience and then review a three-page training manual outlining procedures for their shift as a checkpoint guard. They then watched a 7.5-minute video that provided information about their checkpoint location, existing threat conditions, proper radio call procedures, and the need to maintain vigilance and alertness. Participants were told their job was to assess the vehicle and all of its occupants and render a decision as to whether they could enter the base. They were also told that they were part of a team and that they had a virtual teammate who would provide cover during their interactions with the drivers.

They were then taken to the CAVE and fitted with their equipment: an inert pistol and holster as well as a walkie-talkie to communicate with the base. Next, the participants were given a log sheet depicting a time log of events that occurred on “the previous shift” and “Be On the Lookout” (BOL) information. The BOLs on the log sheet described events the participant was required to remember and look for throughout the session. Additional BOL events were presented aurally during the checkpoint task. Next, the participants were given a chance to familiarize themselves with a typical scenario and the equipment used to interact with the virtual people in the scenario. They were given ample time to repeat this process until they were comfortable with the task.

A video camera was used to record each participant’s performance during the scenarios. The camera was positioned to record a fixed image of the participant within the CAVE. A pair of LCD CrystalEyes shutter glasses was taped over the camera lens to record a single image from the stereoscopic display.

Experimental design. Training was assessed in two ways. Participants were randomly assigned to one of two groups. Those in the first group performed a 45-minute shift, received feedback on that shift, and then performed a second 45-minute shift. Performance was compared between the two shifts. Participants in the second group performed only a single session. The performance of these participants was compared to the performance obtained from the second session of the participants in Group 1.

Procedure: Group 1. The first experimental session contained 23 neutral scenarios and 12 critical scenarios, and each interaction took 1 to 2 minutes, on average, to execute. Upon completion of the first session, participants were given an after-action review (AAR) in which they received corrective feedback regarding the nature of their errors and the proper resolution of those errors. The videotape of the participant’s performance was replayed if necessary. The participants were also given a second questionnaire to assess postexperimental levels of stress and symptoms of simulator sickness. These participants returned 48 hours later and performed a second 45-minute shift. The second experimental session contained 22 neutral scenarios and a different set of 11 critical scenarios that were conceptually similar to those from the first session and addressed the same training objectives.

Procedure: Group 2. Participants in Group 2 performed only one session. Specifically, they performed the same session that Group 1 performed on their second shift. Group 2 was included because it could be argued that the unique scenarios contained in session 2 were easier than those in session 1. If that were true, one would expect the performance of Group 2 to be similar to that of Group1 on session 2. On the other hand, if the participants in Group 1 truly benefited from their training, one would expect their performance on their second session to be superior to that of Group 2 who only performed a single session. Further, the level of performance of Group 2 should be similar to that of Group 1 on their *initial* session. Other than the set of scenarios, the experimental procedures for Group 2 replicated those used in the first session with Group 1.

Results

Performance was assessed by the total number of errors made by each participant on each scenario type. An alpha level of .05 was used for all statistical comparisons. Tukey post hoc tests were used to analyze differences among the means.

Group 1. Performance data were analyzed in two ways. The first analysis addressed performance differences between session 1 and session 2 for only the Group 1 participants. The number of errors for critical and neutral scenarios in both sessions was compared using a 2 x 2 within-subjects ANOVA. A significant effect for scenario type indicated that participants committed significantly more errors on critical scenarios ($M = 4.21$, $SD = 5.17$) than on neutral scenarios ($M = .80$, $SD = 1.81$), $F\ (1, 15) = 34.31$, $p < .01$. In addition, a significant session effect indicated that more errors were made in the first session ($M = 2.94$, $SD = 4.14$) than in the second session ($M = .92$, $SD = 2.67$). $F\ (1, 15) = 12.93$, $p < .05$. Further, there was a significant interaction between session and scenario type, $F(1,15) = 12.14$, $p < .05$. The nature of that interaction is shown in Table 1. As can be seen in the table, the mean number of errors dropped considerably from session 1 to session 2 for both types of scenarios, but the decline was slightly more pronounced for the critical scenarios.

Table 1.
Mean Total Errors for Critical and Neutral Scenarios in Each Session.

	Session1	Session 2
Critical Scenarios	6.36 (6.0)	1.87 (4.51)
Neutral Scenarios	1.15 (2.53)	.44 (1.1)

Standard deviations appear in parentheses.

Group 2. The second analysis compared performance between the participants in Group 2 and Group 1 in their second session. The analysis followed the same format as that described above. Thus, in the first analysis the mean number of errors for critical and neutral scenarios was examined for participants in Groups 1 and 2. This analysis used a 2 x 2 mixed-factor ANOVA with group analyzed as a between-subjects factor and scenario type analyzed as a within-subjects factor.

As hypothesized, Group 1 participants made significantly fewer errors (M = .92, SD = 2.67) in their second session than participants in Group 2 (M = 2.31, SD = 2.62), $F\ (1, 30) = 9.58$, $p < .01$. A main effect for scenario type was also found, $F\ (1,30) = 52.44$, $p < .001$. Participants committed more errors in critical scenarios (M = 3.52, SD = .3.96) than neutral scenarios (M = .66, SD = 1.54). Again, there was a significant interaction between condition and scenario type, $F\ (1, 30) = 11.16$, $p < .005$. The interaction is shown in

Table 2. As can be seen in the table, the participants in Group 2 made more errors on both types of scenarios, but their poorest performance was on the critical scenarios.

Table 2.
Mean Number of Errors for Critical and Neutral Scenarios in Each Group.

	Group 1	Group 2
Critical Scenarios	1.87 (4.51)	5.17 (3.41)
Neutral Scenarios	.44 (1.1)	.88 (1.98)

Standard deviations appear in parentheses.

The better performance of participants in Group 1 during session 2 over those of Group 2 indicates that they benefited from their training experience. If the scenarios used in the second session were "easier" than those used in the first session, one would have expected both groups to perform at the same level. That did not happen. This argument is also supported by comparing the performance of participants in Group 1, session 1 with those of Group 2. An independent t-test used as a manipulation check indicated there was no significant difference in errors between Group 1, session 1 and Group 2, $t(30) = -.34$, $p > .05$. Thus, the scenarios used in each session can be considered equivalent.

Experiment 2: Desktop Interactive Training

Experiment 2 was intended to provide a comparison between the immersive CAVE and VE desktop interfaces. In an effort to examine the benefits of an immersive environment for this task, all other parameters were held constant to the greatest extent possible. As a result, there were few differences in the software, scenarios, and protocols. Specific differences between the two interfaces are described below.

Desktop Virtual Environment Implementation

There were two main differences in the hardware configurations between Experiments 1 and 2. First, the SGI ONYX 2 Image Generator (IG) was replaced with an SGI Octane desktop computer. Also, the images were not displayed stereoscopically and instead were presented on an 18-inch Sony flat panel display. Therefore, no shutter glasses were used. Second, there was no positional tracking with the desktop system. Instead, a new interface had to be developed to allow the participant to navigate within the scene and inspect the vehicles (see Khan, 2002).

The Graphical User Interface (GUI) was placed over the rendered scenarios as shown in the Figure 2. The GUI had two drop-down menu options in the toolbar: menu (play, pause, speech toggle, or quit) and zoom (zoom in, zoom out, or zoom reset). There were also push buttons that provided features such as *walk to driver*, *walk to neutral*, *look at driver*, and *change view*. The first two allowed the avatar to walk to a location that is an offset from the driver and to walk back to a pre-determined neutral location. *Look at driver* bends the avatar's torso so that the driver comes into his line of sight. Change view simply toggled between the original camera view (overlooking the checkpoint scene) and the avatar's view. In addition, the GUI had buttons (large directional triangles) to pan the camera left, right, up, and down, and to reset it to a default position. Using these buttons, a trainee could inspect the front and back seats of a vehicle.

A second GUI (see Figure 3) was required for the experimenter's computer to execute driver and passenger responses and to force commands that the voice recognition software failed to recognize. This coordination between trainee and experimenter GUIs and the scenarios was achieved using a client-server relationship between the two machines implemented by communication sockets. The experimenter's GUI had all the capabilities of the trainee's GUI so that appropriate adjustments, prompting, and demonstrating could be accomplished.

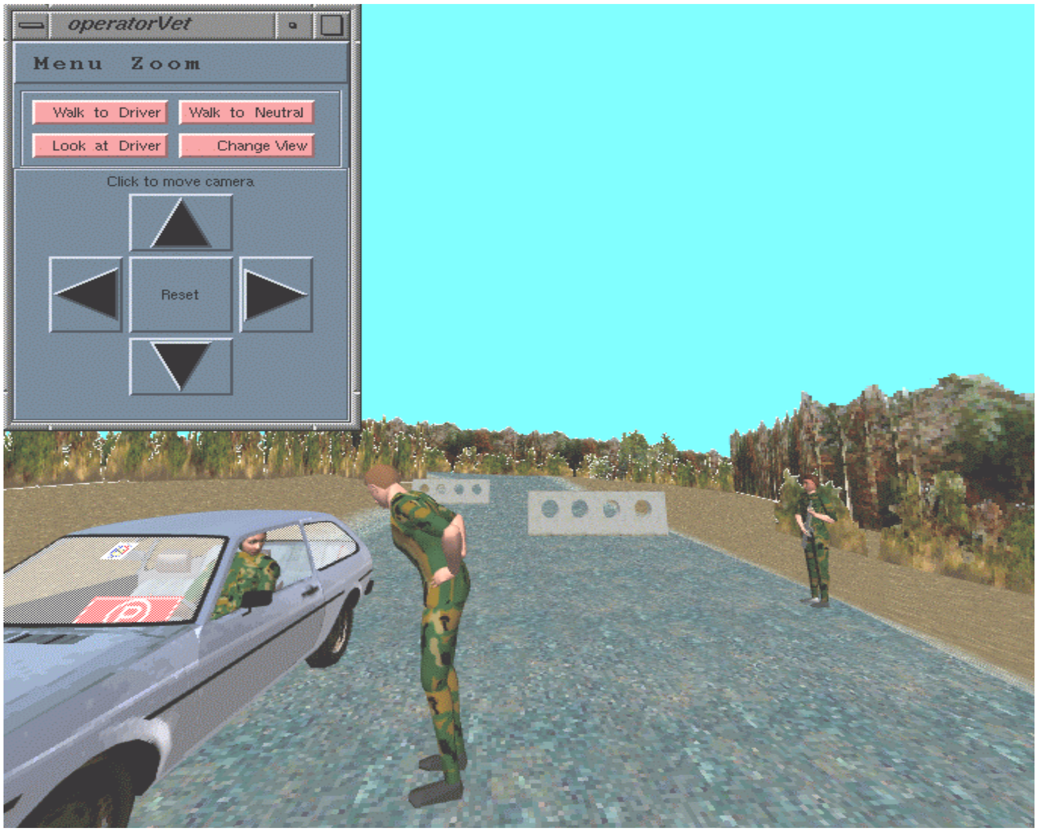

Figure 2. Example of the GUI enabling user navigation in the desktop system.

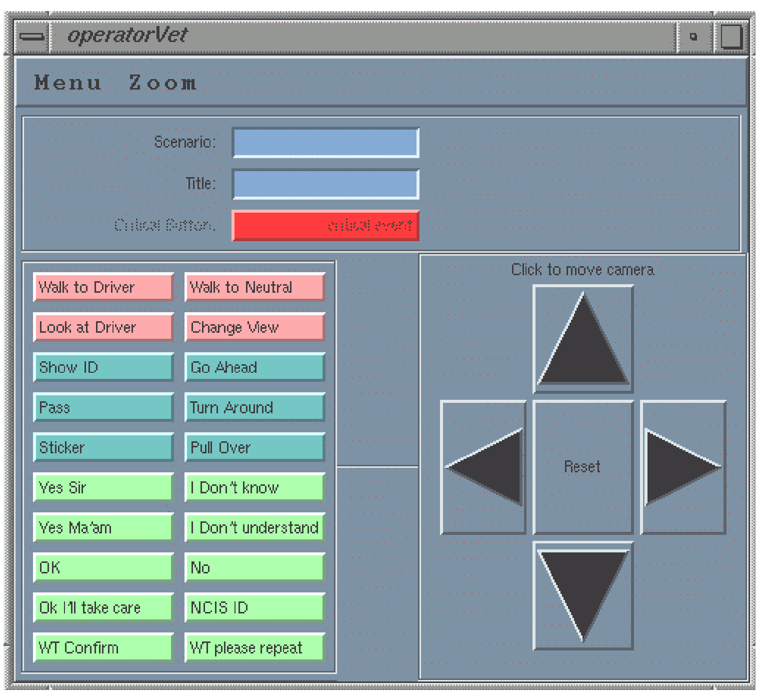

Figure 3. Example of the experimenter's GUI used to intervene and maintain flow of action.

Experimental Method

Fourteen undergraduate students from Old Dominion University with normal or corrected-to-normal vision participated in the study. The task and experimental design were identical to those used in Experiment 1 with a few exceptions. First, participants in the desktop experiments were not administered a simulator sickness questionnaire. Second, in order to decrease the preparation period, the participants were not shown the briefing video. Instead, the critical information from that video was added to their prebriefing package. Third, the sessions were not videotaped. Instead, the sessions were run in adjoining sound attenuated chambers. The participants sat at a desk in one room and the experimenter sat a control panel in the next room. A window between the two rooms allowed the experimenter to "look over the shoulder" of the participant and keep a log of their performance. This log was then used for the AAR for those who participated in two sessions.

Results

The analyses for the desktop experiments were handled differently because there were fewer total participants. Performance was assessed by the total number of errors made by each participant on only the critical scenarios. This approach was taken because so few errors were made on the neutral scenarios. A one-tailed t-test indicated that participants in Group 1 made significantly more errors in their first session ($M = 7.43$, $SD = 3.26$) as compared to their second session ($M = 3$, $SD = 2.71$), $t(6) = 2.8$, $p < .025$.
The data from participants in Group 2 were compared to those of Group 1 from their second session. As hypothesized, the participants in Group 1 made fewer errors ($M = 3$, $SD = 2.71$) than those in Group 2 ($M = 8.14$, $SD = 5.24$). A one-tailed t-test indicated that this difference was significant, $t(6) = 2.31$, $p < .05$.

The major findings from the immersive and desktop studies are reproduced in Table 3. The table shows the mean number of errors for only critical scenarios. Due to the large difference in the number of participants in each study, assumptions for statistical analyses are unlikely to be met and therefore no statistical comparison was performed. However, an informal comparison between the two studies shows that the overall pattern of results was consistent. Those individuals who participated in two sessions showed marked improvement in their second session and performed better than another group of individuals who only participated in a single session. More important, however, the overall performance levels for all participants in all conditions were better with the immersive environment than with the desktop system.

Table 3.
Mean Number of Errors for Critical Scenarios in the Immersive and Desktop Studies.

	Group 1	Group 2	Group 1	Group 2
Session 1	6.36 (6.0)	5.47 (3.41)	7.43 (3.26)	8.14 (5.24)
Session 2	1.87 (4.51)		3.0 (2.71)	

Standard deviations appear in parentheses.

Discussion

The primary goal of the present set of studies was to evaluate virtual environment technology as a training tool for military checkpoint duty. It was expected that if participants learned from the experience and feedback they obtained in their first session, they would commit fewer errors in a subsequent session. The results from both sets of experiments support this idea. Participants, on average, made about 60% fewer errors on their second session. These findings clearly show that individuals from an undergraduate college population, with little or no military experience, were capable of learning the fundamentals of performing checkpoint duty in an experiential context. Moreover, these findings are not likely the result of the specific scenarios chosen for study in each session. In both experiments, the Group 2 participants who performed only one session made over twice as many errors on the identical scenarios that Group 1 performed in their second session. Moreover, the performance of the Group 2 participants was also similar to that of the Group 1 participants on their first session.

It could be argued that some of the improvement observed for Group 1 across sessions might be attributable to increased familiarity with the task and procedures over the course of their initial session. If that were true, one might expect these individuals to also show an improvement in performance within their initial session. An analysis of performance on scenarios from the first and second halves of the initial session for these participants, however, revealed no differences. Thus, the performance improvements of the Group 1 participants in their second session were most likely due to knowledge acquired through training and the feedback provided during their AAR.

The pattern of results obtained in the immersive VE and VE desktop studies were consistent. Those individuals who received two sessions performed significantly better on their second session and better than another group who performed only a single session. However, a comparison between studies revealed one important difference. The overall level of errors was higher with the desktop system than with the immersive VE system. Moreover, this difference was observed with both groups in all sessions. Thus, although participants were able to learn effectively with both platforms, the overall levels of performance were better in the immersive VE.

The objective data indicate that the participants responded well to the VE. Most participants were initially unfamiliar with virtual environment technology, yet they acclimated quickly to the environment, became accustomed to the methods of interaction, and interacted with virtual objects rather naturally. On a more subjective level, there was evidence that suggested the participants were "immersed" in the task. Some individuals were observed using hand gestures to motion cars to pull up to the gate and others reached out to try and hold the ID card presented by the driver.

Despite these encouraging results, there were several technological problems with the system that affected how the participants interacted. For instance, the voice recognition/natural language interface was a source of many problems. The voice recognition software used in the present study could not recognize various voice tones, inflections, and accents equally among all participants. Consequently, participants often had to repeat commands. Further, it was apparent that participants, especially when well immersed in the environment and task, added extra words and conversational components to their interactions. Unfortunately, the additional utterances were sometimes misinterpreted by the speech recognition software, which in turn, executed unintended commands. Although a set of responses was preprogrammed to address several categories of unintended user dialogue, the response set was very limited and was insufficient for many utterances. Other participants indicated that they were aware they were dealing with a computer and used less natural speech during their interactions. Occasionally, some actions were not attempted because participants believed the system might not respond appropriately.

The limited fidelity of the vehicle models and virtual humans also had an impact on performance. For example, in the back area of one of the vehicles (the jeep model) there was a box-shaped wheel well that was often misconstrued as a suspicious package. This led some participants to question drivers in what were intended to be neutral scenarios. Also, the Jack agent used in the study was originally developed as an anthropometric/ergonomic model and was not designed for the subtleties of human expression or behavior needed to address a wider range of training objectives important for checkpoint duty. For instance, the poor quality of facial expressions and behaviors generated some ambiguity as what constituted "suspicious" behavior in the Jack agents. During the AAR sessions, participants had to be instructed to adjust their criteria to match the lower fidelity of the agents.

Conclusion

The results from the present studies indicate that individuals can benefit from training in a VE that places greater emphasis on social interaction skills. Individuals with little or no military training were able to learn some of the fundamentals for performing checkpoint duty in an experiential context. These findings should encourage those in the development community to continue to improve and refine the technology required for this class of VEs. Additional work on modeling body gestures, facial expressions, and voice recognition in real-time simulations is needed to develop training for more complex social interactions.

The results from the present study also showed that overall levels of performance were higher with the immersive VE as compared top the VE desktop system. Although this difference was found across groups and conditions, the magnitude of the difference was not dramatic. Thus, the ability to port a similar training experience to a less expensive PC platform without major performance differences underscores the potential for providing greater access to this type of VE training in a much more cost effective medium.

References

Badler N.I., Bindiganavale, R., Allbeck, J., Schuler, W., Zhao, L., & Palmer, M. (2000). Parameterized action representation for virtual human agents. In J. Cassell (Ed.), *Embodied conversational agents* (pp. 256-284). Boston: MIT Press.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*, 203-220.

Khan, T. (2002). Approaching high fidelity human tracking and representation in interactive virtual environments. Masters Thesis. Electrical and Computer Engineering, Old Dominion University.

Knerr, B. W., Breaux, R., Goldberg, S. L., & Thurman, R. A. (2002). National Defense. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation and applications* (pp. 857-872). Mahwah, NJ: Erlbaum.

Loftin, R. B., McKenzie, F., Scerbo, M. W., Badler, N., Phillips, M. A., Garcia, H., Lusso, A., & Seevinck, J. *Virtual Environments for Training: Final Report*. Prepared by Old Dominion University for University of Houston and Office of Naval Research (ONR) under Grant N00014-95-1-1044. December 30, 2002.

Pew, R. W. & Mavor, A. S. (Eds.) (1998). *Modeling human and organizational behavior: Applications to military simulations.* Washington D.C.: National Academy Press.