ELSEVIER

Applied Ergonomics 33 (2002) 559–570

APPLIED ERGONOMICS

www.elsevier.com/locate/apergo

Using virtual reality technology for aircraft visual inspection training: presence and comparison studies

Jeenal Vora[a], Santosh Nair[a], Anand K. Gramopadhye[a,*], Andrew T. Duchowski[b], Brian J. Melloy[a], Barbara Kanki[c]

[a] *Department of Industrial Engineering, College of Engineering and Science, Clemson University, 29634-0920 Clemson, SC, USA*
[b] *Department of Computer Science Engineering, Clemson University, Clemson, SC 29634, USA*
[c] *NASA Ames Research Center, USA*

Received 10 May 2001; received in revised form 2 May 2002; accepted 28 May 2002

Abstract

The aircraft maintenance industry is a complex system consisting of several interrelated human and machine components. Recognizing this, the Federal Aviation Administration (FAA) has pursued human factors related research. In the maintenance arena the research has focused on the aircraft inspection process and the aircraft inspector. Training has been identified as the primary intervention strategy to improve the quality and reliability of aircraft inspection. If training is to be successful, it is critical that we provide aircraft inspectors with appropriate training tools and environment. In response to this need, the paper outlines the development of a virtual reality (VR) system for aircraft inspection training.

VR has generated much excitement but little formal proof that it is useful. However, since VR interfaces are difficult and expensive to build, the computer graphics community needs to be able to predict which applications will benefit from VR. To address this important issue, this research measured the degree of immersion and presence felt by subjects in a virtual environment simulator. Specifically, it conducted two controlled studies using the VR system developed for visual inspection task of an aft-cargo bay at the VR Lab of Clemson University. Beyond assembling the visual inspection virtual environment, a significant goal of this project was to explore subjective presence as it affects task performance. The results of this study indicated that the system scored high on the issues related to the degree of presence felt by the subjects. As a next logical step, this study, then, compared VR to an existing PC-based aircraft inspection simulator. The results showed that the VR system was better and preferred over the PC-based training tool.

Keywords: Virtual reality; Eye tracking; Visual inspection; Training; Presence

1. Introduction

Aircraft inspection and maintenance, essential to safe, reliable air transportation, is a complex system with many interrelated human and machine components (Drury and Gramopadhye, 1990; Drury, 1991). One of the major factors contributing to this complexity is the aging fleet. Scheduled repairs to an older fleet account for only 30% of all maintenance compared to the 60–80% for a newer one. This difference can be attributed to the increase in the number of age-related defects (Drury, 1991). In such an environment the importance of inspection cannot be overemphasized.

In the aircraft industry, 90% of all inspection in aircraft maintenance is visual in nature and is conducted by human inspectors whose reliability is fundamental to an effective maintenance system. However, inspection tends to be less than 100% reliable because of human fallibility (Chin, 1988; Drury, 1992). This human element, however, cannot be entirely eliminated because of our superior decision-making ability (Thapa et al., 1996), our ability to adapt to unforeseen events and our ability to judge using the sense of touch. Since it is critical that these visual inspections be performed effectively, efficiently, and consistently over time, continuing emphasis has been placed on developing interventions to make inspection more reliable and

---

*Corresponding author. Tel.: +1-864-656-4716; fax: +1-864-656-0795.

E-mail address: agramop@ces.clemson.edu (A.K. Gramopadhye).

0003-6870/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0003-6870(02)00039-X

error tolerant, with training being the strategy of choice for improving the performance of aircraft maintenance inspectors (Gramopadhye et al., 1998). Patrick (1992) has identified training content, training methods and the trainee as the important components of the training program. Drury (1992) includes the training delivery system as another key component while for Gramopadhye et al. (1997a, b), training methods along with an appropriate delivery system comprise an effective training system.

However, training for improving the visual inspection skills is generally lacking at aircraft repair centers and maintenance facilities even though its impact on visual inspection skills has been well documented in both the aircraft and manufacturing industries (Moll, 1980; FAA, 1991, 1993). It has been shown to improve the performance of both novice and experienced inspectors (Wiener, 1975; Drury and Gramopadhye, 1990; Gramopadhye et al., 1995). In particular, the success of off-line training/retraining with feedback suggests that this method can play an important role in aircraft inspection training.

Existing training for inspectors in the aircraft maintenance environment tends to be mostly on-the-job (OJT) (Latorella et al., 1992). However, this method may not be the best one because feedback, so important to training, may be infrequent, unmethodical, and/or delayed. Moreover, in certain instances feedback is economically prohibitive or impractical because of the nature of the task. Since the benefits of feedback in training have been well documented (Wiener, 1975; Gramopadhye et al., 1997a, b), and for other reasons as well, alternatives to OJT are sought. One of the most viable approaches in the aircraft maintenance environment, given its many constraints and requirements, is computer-based training which offers several advantages over traditional training approaches: it is more efficient while at the same time facilitating standardization and supporting distance learning.

With computer technology becoming cheaper, the future will bring an increased application of advanced technology to training. Over the past decade, instructional technologists have offered numerous technology-based training devices with the promise of improved efficiency and effectiveness. These training devices are being applied to a variety of technical training applications, including computer-based simulation, interactive videodiscs, and other derivatives of computer-based applications in addition to compact disc read only memory (CD-ROM) and digital video interactive (DVI), two technologies that will provide the "multi-media" training systems of the future. Many of these training delivery systems such as computer-aided instruction, computer-based multi-media training and intelligent tutoring systems are already being used today, thus ushering in a revolution in training.

In the domain of visual inspection, the earliest efforts using computers for off-line inspection training were reported by Czaja and Drury (1981), who used keyboard characters to develop a computer simulation of a visual inspection task. In order to study inspection performance in a laboratory setting, other researchers have used similar simulations. Latorella et al. (1992) and Gramopadhye et al. (1994) have used low fidelity inspection simulators with computer-generated images to develop off-line inspection training programs for inspection tasks. Similarly, Drury and Chi (1995) studied human performance using a high fidelity computer simulation of a PCB inspection task. Another domain that has seen the application of advanced technology is that of inspection of X-rays for medical practice (e.g., Kundel et al., 1990). The use of computer-based simulators for aircraft inspection training has a short but rich history (Latorella et al., 1992; Gramopadhye et al., 1994, 1998; Blackmon and Gramopadhye, 1996; Nickles et al., 2001), the most advanced and recent example being the automated system of self instruction for specialized training (ASSIST), a training program developed using task analytic methodology and featuring a PC-based aircraft inspection simulator (Gramopadhye et al., 2000).

The results of a follow-up study conducted to evaluate the usefulness and transfer effects of ASSIST were encouraging as to the effectiveness of computer-based inspection training, specifically in improving performance. Performance of the training group improved significantly on the criterion inspection task, the inspection of an aft-cargo bin of L-1011 aircraft, after training. Of greatest interest was the increase in the percentage of defects detected and the reduction in the number of misses for the training group in comparison with the control group. Moreover, the training system also scored highly on various usability measures (FAA, 2000; Gramopadhye et al., 2000).

Despite the advantages, the simulator is limited by its PC-based technology. It lacks realism as it uses only two-dimensional (2D) sectional images of airframe structures and, therefore, does not provide a holistic view of the aft-cargo bin. More importantly, the inspectors are not immersed in the environment, and, hence, they do not get the same look and feel of conducting an actual inspection. To address these limitations, virtual reality (VR) technology has been proposed as a solution, and in response a high fidelity VR-based inspection simulator has been developed (Duchowski et al., 2000).

VR, described by several researchers (Kalawsky, 1993; Burdea and Coiffet, 1994; Durlach and Mavor, 1995; Heim, 1998), is most applicably defined as immersive, interactive, multi-sensory, viewer-centered, three-dimensional (3D) computer-generated environments and the combination of technologies required to build them (Cruz-Neira, 1993). As this definition suggests, creating a virtual environment (VE) requires immersing humans into a world completely generated by a computer. The human user becomes a real participant in the virtual world, interacting and manipulating virtual objects. Therefore, human performance is one of the most important considerations in defining the requirements for a VE. As the user's senses and body are involved in the task, it becomes essential, then, to focus on user-centric performance measures. While for human–computer interaction, abstract values such as ease of use, ease of learning, presence and user comfort become significant (Kalawsky, 1993), for virtual environments *presence*, the subjective experience of being in one place or environment even when one is physically situated in another (Singer and Witmer, 1996), becomes the most important measure.

This concept of experiencing "presence" as a normal awareness or attentional phenomenon is based on the interaction between external stimuli and immersion factors. It is also stated as the mental state in which a user feels physically present within the computer-mediated environment (Slater and Steed, 2000). The involvement tendencies depend on focusing one's attention and energy on a coherent set of VE stimuli while the immersion tendencies lead to the perception of being a part of the VE stimulus flow. According to Witmer and Singer (1998), both involvement and immersion are necessary conditions for experiencing presence. It is generally held that human performance in VEs is directly proportional to the degree of presence induced by the environment, which, in turn, is influenced by the individual's level of immersion in the VE (Witmer and Singer, 1998; Stanney et al., 1998). In essence, a transitive relationship appears to emerge where immersion affects presence, which in turn affects performance.

Fully immersed observers perceive that they are interacting directly or remotely with the environment. Thus, presence becomes a subjective sensation or mental manifestation that is not easily amenable to objective physiological definition and measurement, with its strength varying both as a function of individual differences, traits, and abilities and the characteristics of the VE. In general, the more control a person has over the task environment or in interacting with the virtual environment, the greater the experience of presence (Schloerb, 1995). The success of using VR as a tool for training and job aiding, therefore, is highly dependent on the degree of presence experienced by the users of the virtual reality environment. In the light of this situation, it is critical that we measure the degree of presence of the VR simulator to support training.

If the VR simulator is to be proposed as a solution for off-line training, it is essential that this environment accurately mimic the real world as perceived by the user/trainee. Only then can the effects of training be expected to transfer from the VR environment to the real world. In order to measure the degree of presence of the VR simulator, this study solicits the subjective opinion of humans on the applicability of the VR simulator in supporting training in the aircraft maintenance environment. This is done by using presence questionnaires (PQ) to ascertain the degree to which individuals experienced presence in a VE and the influence of possible contributing factors on the intensity of this experience (Witmer and Singer, 1998). This study then took the next logical step, comparing the VR simulator and ASSIST using subjective evaluation to see if the subjects preferred one training program to the other. This paper describes the VR simulator developed and the two controlled studies conducted.

2. Description of the VR system—hardware platform

The primary rendering engine used to develop the VR simulator is a dual rack, dual pipe, SGI Onyx2 Infinite Reality system with 8 raster managers and 8 MIPS 12000 processors, each with 4 Mb secondary cache. It is equipped with 8 Gb of main memory and 0.5 Gb of texture memory (Duchowski et al., 2000). Multi-modal hardware components include a binocular ISCAN eye tracker (Duchowski et al., 2001) mounted within a virtual research V8 (high-resolution) head mounted display (HMD). The V8 HMD offers 640 × 480 resolution per eye with separate left and right eye feeds order to produce a simulated stereo effect for the user. HMD position and orientation tracking is provided by an ascension 6 degree-of-freedom (6DOF) flock of birds (FOB), a DC electromagnetic system with a 10 ms latency. A 6DOF tracked, hand-held mouse provides a means to represent a virtual tool for the user in the environment. Fig. 1 below shows the laboratory settings of the VR environment.

2.1. Development of the VR environment

The development of the VR environment was based on a detailed task analytic methodology (FAA, 1991; Nickles et al., 2001). Data on aircraft inspection activity were collected through observation, interviewing, shadowing, and digital data capturing techniques. More detail on the task description and task analytic methodology can be found in Nickles et al. (2001).

2.2. Geometric environment modeling

The goal of the construction of the virtual environment was to match the appearance of the physical inspection environment, an aircraft cargo bay. This physical environment is a complex 3D cube-like volume with airframe components (e.g., fuselage ribs) exposed

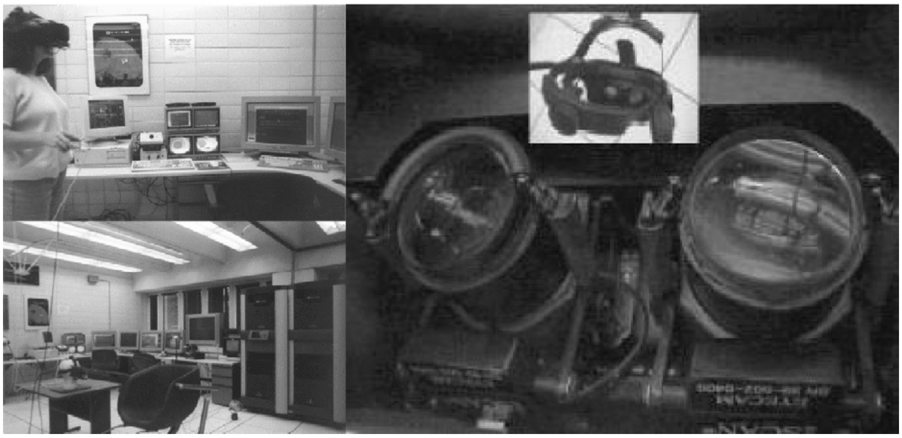
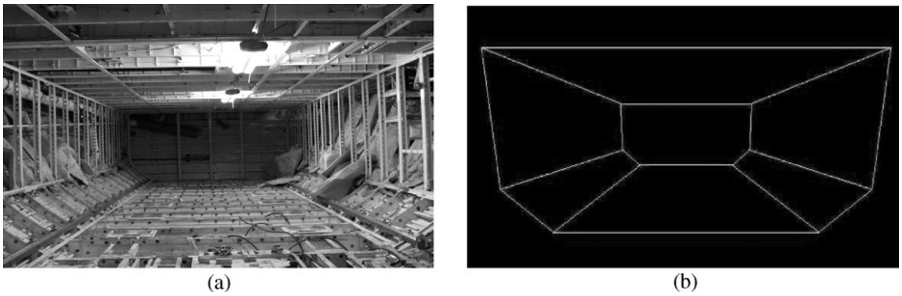


Fig. 1. Virtual Reality Eye Tracking (VRET) Lab at Clemson University.



(a) (b)

Fig. 2. (a) Physical environment of the Cargo Bay, (b) 3D box-like virtual environment.

for inspection. Fig. 2(a) shows the actual physical environment of the aft-cargo bay.

2.3. Model geometry

The geometry of the virtual inspection environment is contained in a plain text file, formatted according to a subset of the Alias|Wavefront Object files specification. This model geometry was patterned after a simple 3D enclosure (e.g., a cube), with dimensions matching the real inspection environment, that is, an aircraft's cargo bay, as shown in Fig. 2(b). The resulting model was built entirely out of planar polygons. There are two pragmatic reasons for this design choice. First, since the representation of the true complexity of the airframe structure is avoided, fast display rates are maintained (on the order of 25–30 fps), while tracking latencies are minimized (on the order of 10–30 ms for head and eye trackers). Second, planar polygons (quadrilaterals) are particularly suitable for texture mapping. To provide a realistic appearance of the environment, images of the physical environment were used for texture maps.

2.4. Lighting and flashlight modeling

The SGI Onyx2 host provides real-time graphics rendering performance, while simultaneously processing tracking information sent to the host via the RS-232 serial connection. To generate the environment, no specialized rendering algorithms are invoked beyond what is provided by the OpenGL graphics library application program interface (API). Standard (1st-order) direct illumination is used to light the environment. Additionally, an OpenGL spotlight is used to provide the user with a "virtual flashlight". The flashlight's position and orientation are obtained from the 6DOF electro-magnetically tracked "flying mouse" from Ascension. Fig. 3 represents the use of flashlight shown over a 2D polygon.

Because OpenGL relies on the Phong illumination model coupled with Gouraud shading to generate lighting effects, large polygons produce a coarse (blocky) flashlight beam (Foley et al., 1990). To correct this problem, the polygons were subdivided to smooth out the spotlight, producing a more aesthetically pleasing circular effect. The level of polygonal subdivision is user-adjustable.

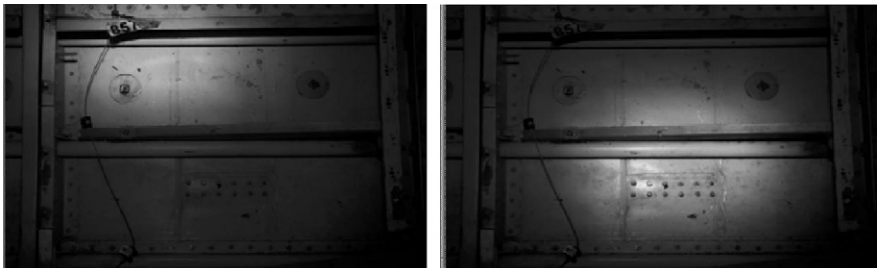


Fig. 3. Flashlight over a 2D-polygon.

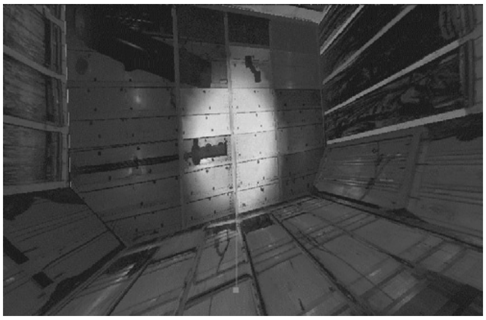


Fig. 4. Flashlight in virtual cargo bay environment.

2.5. *Realistic texture maps*

To texture map, the simple 3D box-like environment, images of the physical environment were obtained. With permission from a commercial airline, images of an aircraft's cargo bay were taken while the aircraft was withheld from operation for inspection and servicing. Care was taken to photograph the environment in sections by translating a hand-held camera; however, inaccuracies in image alignment were inevitable and were later resolved by careful labeling of photographs and digital post-processing using the GNU image manipulation program (Gimp) software (www.gimp.org). The resulting environment, with the user-held "virtual flashlight", is shown in Fig. 4.

2.6. *Scenarios developed*

Four virtual environment scenarios were developed, each representing different artificial airframe defects. A library of defects consisting of corrosion, cracks and broken conduits was created with several defects occurring at varying severity levels and locations. The representative defects are shown in Fig. 5 with an example of the corrosion defect and the broken conduit defect scenarios in Fig. 6. By manipulating the type, severity, location and defect mix, experimenters can now create airframe structures that are representative of real world situations to be used for running controlled studies.

3. Experiment 1—presence study

The objective of this first study was to measure the degree of presence of the VR simulator through subjective evaluation using a presence questionnaire (Witmer and Singer, 1998).

3.1. *Subjects*

Fourteen subjects between the ages of 20 and 30 were randomly selected from a population of graduate and undergraduate students at Clemson University. Subjects were screened for 20/20 vision, corrected if necessary, and were paid $25 for participating in the experiment.

3.2. *Stimulus material*

The task was a simulated visual inspection task of an aircraft cargo bay, similar to the one in an L1011 aircraft, implemented in a VR environment viewed through a head-mounted display.

3.3. *Experimental design*

The study used a within-subject design consisting of a single factor (VR training session). Each VR training session consisted of a calibration and a familiarization scenario for getting acquainted with the system followed by four inspection scenarios. The four scenarios with defects were designed such that three contained each defect (corrosion, cracks and damaged conduits) type and the fourth, a multi-defect scenario, had all the three types of defects. The locations of defects were kept similar to their real life occurrences. Each defect was defined as a target that the subjects were asked to identify, and each scenario contained 12 defects. The number was set to this value after a pilot test determined the amount of time needed by three subjects to search for all the defects. The experiment duration was limited to 30 min, the subjects being immersed in the VE for 25 min in two different sessions (5 min familiarization session and 20 min inspection scenarios session). The immersion duration was limited because of health and safety guidelines, which suggest that subjects should be

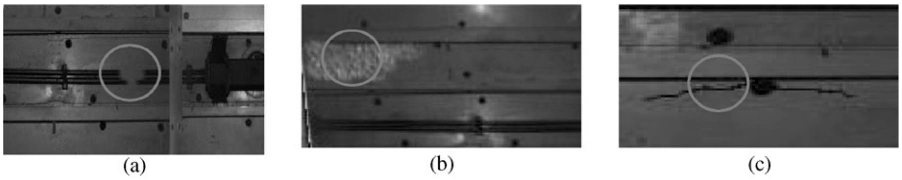


Fig. 5. Sample defects: (a) damaged conduit, (b) corrosion and (c) crack.

Fig. 6. (a) Corrosion scenario and (b) broken conduit scenario.

immersed in a VE for no longer than 10–20 min (Stanney et al., 1998).

3.4. Procedure

Subjects were initially greeted and asked to sign the informed consent form. Prior to commencement of the experiment, they were taken to the hangar site of Delta Airlines in Atlanta and shown the real aft-cargo bay of a wide-bodied aircraft. They were then taken to the VR Lab at Clemson University where the simulator was set up. Each subject was briefed on the experiment through a written description of the objectives of the study and the task. The researcher then answered all of the subjects' questions. Before being immersed in the VR environment, subjects filled out an immersive tendencies questionnaire (ITQ) (Witmer and Singer, 1994). Next, the researcher commenced the experiment with a short five-point eye tracker calibration routine. Subjects were asked to perform a walk through in the aft-cargo bay environment in which they were shown the entire search area that was to be inspected and were provided with a verbal description of all the defects. Subjects were then allowed to practice with the VR simulator and the 3D mouse as part of the software familiarization training. After the subjects became comfortable, they were randomly given the five scenarios with and without defects. The subjects had to walk through the environment, identify the defects present, if any, and use the 3D mouse to click on the defect to indicate their selection. When the subjects felt that there were not any more defects in the scenario, they notified the researcher, and the next scenario was presented. After completing all scenarios, subjects were asked to fill out a PQ (Witmer and Singer, 1994).

3.5. Data collection

Using the two questionnaires, the tendency of immersion and sense of presence of the subjects on the inspection task was evaluated as they used the VR environment, focusing on their perceptions of the system. The ITQ was used to identify and measure possible individual differences in the abilities or tendencies of subjects to immerse themselves in different environmental situations, while the PQ was used to address their subjective experience in a simulated environment, identifying and measuring to what degree aspects of the virtual environment engendered a sense of presence. Both the ITQ and the PQ use a 7-point Likert scale based on the semantic differential principle (Singer and Witmer, 1996). Each item is end-anchored by opposing descriptors, but unlike the semantic differential, the scale also includes an anchor at the midpoint. Items marked with an asterisk (*) (see below) have to be reverse scored in order to contribute to the subscale and overall totals. Based on the reliability analysis and the cluster analyses performed by Singer and Witmer (1996), the three clusters used to collect data for the virtual reality environment were identified as focus, involvement and games for the ITQ and involved/control, natural and interface quality for the PQ. The questions for each questionnaire are categorized below:

IMMERSIVE TENDENCIES QUESTIONNAIRE
Total number of questions = 16.
ITQ-focus: questions 1, 3, 7, 8, 11, 13, and 16.
ITQ-involvement/immersion: questions 2, 4, 5, 9, 10, 14, and 15.
ITQ-games: questions 6 and 12.
PRESENCE QUESTIONNAIRE
Total number of questions = 20.
PQ-involved/control: questions 1, 2, 4, 6, 8, 9, 10, 13, 14*, 15, and 16.
PQ-natural: questions 3, 5, and 7.
PQ-interface quality: questions 11, 12, 17*, 18*, 19, and 20.

3.6. Results

Mean scores obtained on the individual questions were combined into aggregate measures as suggested by Witmer and Singer (1996). Cronbach's coefficient alpha, shown in the first row of Table 1 (Cronbach, 1951) was calculated for the group of questions within the cluster to ensure that it was appropriate to group the scores from all the questions into a single aggregate measure.

The correlation analysis was conducted on the mean scores for the clusters of the pre- and the post-questionnaires. The results, reported in Table 1, revealed significant positive correlation between ITQ involvement and PQ involvement ($r^2 = 0.5446, p < 0.0441$), PQ involvement and PQ natural ($r^2 = 0.935, p < 0.0001$), PQ involvement and PQ interface quality ($r^2 = 0.7835, p < 0.0009$) and PQ natural with PQ interface quality ($r^2 = 0.7813,\ p = 0.001$). The mean scores for the individual responses to the 20 questions of the post-questionnaire are shown in Fig. 7.

The Wilcoxon Test was conducted to see if the mean scores on post-questionnaire were significantly different using a Likert scale of 1–7 with the anchor at the mid-point. The results of the test are tabulated in Table 2, with column 6 showing whether the mean response was significantly different from the anchor value (4).

3.7. Discussion

Analysis of the correlation revealed that the subjects who experienced a sense of involvement in real world experiences also felt involved in the VR experience. In addition, the subjects who experienced significantly greater level of involvement in the simulator felt that the experiences were as natural as the real world ones. The interface quality and the naturalness of the environment were also significantly correlated, indicating a significant level of interface quality in the simulator. However, it was interesting to note that the focus and games clusters from the pre-questionnaire did not correlate well with any of the clusters of the post-questionnaire, implying that it is not necessary to be familiar with video games to use the VR simulator. Also, the mental and physical state of the person and the tendency to avoid distractions while performing a particular task, that is, "being focused", did not affect the performance of the subjects on the assigned inspection task.

Analysis of the Wilcoxon Test revealed that the system scored significantly greater on 11 of 20 questions from the anchor points. The results bode well for the level of presence experienced by the subjects. The system scored significantly greater on the following presence-related issues. The most significant point was the subjects' response to Question 7, when they indicated that the experiences with the VR environment were consistent with real world ones. The subjects experienced a significantly greater level of involvement in the VR environment as indicated by the difference from the anchor points (Questions 4, 6, 8, 9, 10 and 13). Issues like the visual aspects of the environment, the sense of objects, the anticipation of the response of the system,

Table 1
Results of the Cronbach's and the correlation analyses

		Pre-questionnaire			Post-questionnaire		
		Focus	Involvement	Games	Involvement	Natural	Interface quality
Cronbach's Alpha		0.56	0.79	0.91	0.95	0.91	0.63
Pre-questionnaire	Focus		0.3971	0.4499	0.1111	0.0362	−0.0173
			(0.1598)	(0.1064)	(0.7053)	(0.9023)	(0.9533)
	Involvement			0.4367	*0.5446*	0.3855	0.0509
				(0.1184)	*(0.0441)*	(0.1735)	(0.8626)
	Games				0.5078	0.4911	0.2459
					(0.0638)	(0.0746)	(0.3967)
Post-questionnaire	Involvement					*0.935*	*0.7835*
						(<0.0001)	*(0.0009)*
	Natural						*0.7813*
	Interface quality						*(0.001)*

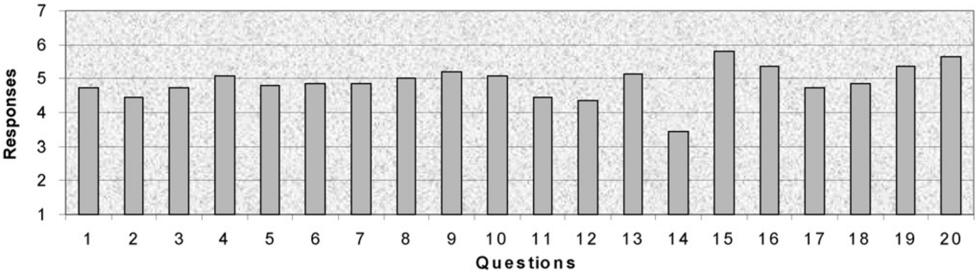


Fig. 7. Mean values—presence post-questionnaire.

surveying, and the experience in the VR environment, all contributed to the significant sense of involvement. The VR system scored significantly greater from the anchor points on the issues related to the subject's concentration on the assigned task and on the adjustment to the control devices, relating to a high level of realism, which is a testament to the high quality of the interface (Questions 15, 16, 19 and 20).

However, the system did not show significant differences on certain aspects of the interface suggesting and thereby needing improvement (Questions 11, 12, 17 and 18). On the other hand, the subjects indicated that the interactions were not significantly natural (Questions 3 and 5) nor were they able to control the events (Question 1). These results can be attributed to the fact that the subjects were not very familiar with nor felt comfortable using control devices like the 3D mouse, walking around the simulator and/or identifying defects correctly. In addition, the system was not very responsive to the actions initiated by the subjects (Questions 2 and 14), the primary reason being that the simulator was running on a Unix machine, a multi-user system that allows many processes run on it at the same time, causing a time lag during the experiment.

The VR system scored significantly greater than the anchor points on most aspects of presence and immersion and, hence, can potentially be used as a training tool for visual inspection tasks. Even the issues that need to be addressed can be improved easily, the primary one being the delay in system response experienced by the subjects during their task performance because the VR system runs on a Unix-based multi-user platform. This delay, caused by parallel users, can be avoided by setting priority levels to the works running on the Unix machine, by threading (Shivashankaraiah, 2000) or by providing a CPU dedicated to the system.

4. Experiment 2—comparison of ASSIST with VR simulator

The objective of this study was to compare the VR and the ASSIST prototypes using subjective evaluation on a 7-point Likert scale questionnaire. This study focused on analyzing the subjects' perception of the two systems with respect to each other and to the actual aft-cargo bay environment in terms of its applicability to support training.

4.1. Subjects

Fourteen subjects between the ages of 20 and 30 were randomly selected from a population of graduate and undergraduate students at Clemson University. Subjects were screened for 20/20 vision, corrected if necessary, and were paid $25 for participating in the experiment.

4.2. Equipment

ASSIST: a computer-based training program, was developed using Visual C++, Visual Basic and Microsoft Access. The development work was conducted on a Pentium 120 MHz platform with a 17-in high resolution monitor, 32 MB RAM, 2 MB video RAM, ATI Mach 32 VLB advanced graphics accelerator card, 810 MB hard drive, multi-speed CD drive, 210 MB Bernoulli drive and a Reveal multimedia kit. The training program used text, graphics, animation and audio. The input devices were a keyboard and a two-button mouse (Gramopadhye et al., 2000).

For the Virtual Environment the SGI Onyx2 systems were used as described in Section 2.

4.3. Stimulus material

This experiment, consisting of two parts, involved a visual search for three defects—corrosion, cracks and damaged conduits—in the aft-cargo bay of an aircraft. For the first part, the subjects conducted a simulated visual inspection of the airframe using a VR environment viewed through a head-mounted display. For the second, the same visual inspection task was implemented on the 2D environment ASSIST.

4.4. Experimental design

The study utilized a split-plot design with all subjects using both the VR and the ASSIST system and the

Table 2
Results of the Wilcoxon Test—Experiment 1

Question #	1	Anchor point 4	7	Mean (SD)	Wilcoxon Test
1. How much were you able to control events?	Not at all	Somewhat	Completely	4.71(1.07)	($p > 0.05$)
2. How responsive was the environment to actions that you initiated (or performed)?	Not at all	Moderately responsive	Completely	4.43(1.02)	($p > 0.05$)
3. How natural did your interactions with the environment seem?	Extremely artificial	Borderline	Completely natural	4.71(1.64)	($p > 0.05$)
4. How much did the visual aspects of the environment involve you?	Not at all	Somewhat	Completely	5.07(1.49)	($p<0.05$)
5. How natural was the mechanism that controlled movement through the environment?	Extremely artificial	Borderline	Completely natural	4.79(1.53)	($p > 0.05$)
6. How compelling was your sense of objects moving through space?	Not at all	Moderately compelling	Very compelling	4.86(0.95)	($p<0.05$)
7. How much did your experiences in the virtual environment seem consistent with your real world experiences?	Not consistent	Moderately consistent	Very consistent	4.86(0.95)	($p<0.05$)
8. Were you able to anticipate what would happen next in response to the actions that you performed?	Not at all	Somewhat	Completely	5(1.04)	($p<0.05$)
9. How completely were you able to actively survey or search the environment using vision?	Not at all	Somewhat	Completely	5.21(1.25)	($p<0.05$)
10. How compelling was your sense of moving around inside the virtual environment?	Not compelling	Moderately compelling	Very compelling	5.07(1.44)	($p<0.05$)
11. How closely were you able to examine objects?	Not at all	Pretty closely	Very closely	4.43(1.4)	($p > 0.05$)
12. How well could you examine objects from multiple viewpoints?	Not at all	Somewhat	Extensively	4.36(1.08)	($p > 0.05$)
13. How involved were you in the virtual environment experience?	Not involved	Mildly involved	Engrossed	5.14(1.1)	($p<0.05$)
14. How much delay did you experience between your actions and expected outcomes?	Long delays	Moderate delays	No delays	3.43(1.22)	($p > 0.05$)
15. How quickly did you adjust to the virtual environment experience?	Not at all	Slowly	<1 min	5.79(0.89)	($p<0.05$)
16. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?	Not proficient	Reasonably proficient	Very proficient	5.36(0.93)	($p<0.05$)
17. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?	Prevented	Interfered somewhat	Not at all	4.71(1.54)	($p > 0.05$)
18. How much did the control devices interfere with the performance of assigned tasks or with other activities?	Interfered greatly	Interfered somewhat	Not at all	4.86(1.56)	($p > 0.05$)
19. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?	Not at all	Somewhat	Completely	5.36(1.5)	($p<0.05$)
20. How easily did you adjust to the control devices used to interact with the virtual environment?	Difficult	Moderate	Easily	5.64(1.5)	($p<0.05$)

treatment factor being the type of system. The order in which the subjects performed the task was counterbalanced to cancel out any order effects. For consistency in environments, only one multi-defect scenario was used for the VR system.

4.5. Procedure

Subjects were initially greeted and asked to sign the informed consent form. Prior to the commencement of the experiment, they were taken to the hangar site of Delta Airlines in Atlanta where they were shown the actual aft-cargo bin of a wide-bodied aircraft. They were then taken to the VR Lab at Clemson University where the simulator was set up. Following this step, each subject was briefed on the experiment through a written description of the objective of the study and the upcoming task. After the researcher answered the subjects' questions, the experiment began with a short five-point eye tracker calibration routine. Subjects were asked to walk through the aft-cargo bay environment where they were shown the entire search area to be

Table 3
Results of the Wilcoxon Test—Experiment 2

Questions	ASSIST	VR	Wilcoxon results
	Means (SD)		
1. The experiences with the environment seem consistent with the real world experiences	4.21 (1.58)	5.29 (1.14)	$p = 0.034$
2. It was easy to comprehend the defects in the environment	3.14 (0.95)	5.21 (1.48)	$p = 0.0018$
3. It was easy to understand the location of the defects in the environment	3.36 (1.39)	5.71 (1.27)	$p = 0.0013$
4. The environment was very responsive to the actions that I initiated (or performed)	*5.86 (1.51)*	*4.14 (1.29)*	$p = 0.0114$
5. The visual aspects of the environment involved me very much	3.36 (1.5)	5.29 (1.07)	$p = 0.0037$
6. I was completely involved in the environment experience	3.0 (1.24)	5.14 (1.29)	$p = 0.002$
7. There was no delay experienced between my actions and the response from the system	*5.93 (1.64)*	*3.86 (1.29)*	$p = 0.002$
8. It was easy to understand the navigational aspects of the environment	5.29 (1.14)	5.64 (1.28)	$p = 0.9427$
9. The software is very applicable for the visual inspection task of an aft-cargo bay for training and job aiding	4.0 (1.41)	5.5 (1.16)	$p = 0.0107$
10. I would personally prefer the environment for training of visual inspection task	3.36 (1.45)	5.21 (1.67)	$p = 0.0086$

inspected and were provided with both a graphical and verbal description of all three defects. Subjects were allowed to practice for some time with the virtual reality simulator and the 3D mouse as part of the software familiarization. When the subjects felt comfortable, they were given the multi-defect scenario for inspection. The subjects had to walk through the environment, identify the defects present, if any, and click on them using the 3D mouse to indicate their selections. After completing this task, the subjects practiced with the ASSIST software before conducting a similar defect detection experiment in the ASSIST environment. After completing both tasks, the subjects were asked to fill out a comparison questionnaire.

4.6. Data collection

Data were collected using a subjective evaluation questionnaire with a 7-point Likert scale, comparing the subjects' perception of the two systems to each other and to the actual aft-cargo bay in terms for applicability as training support. Performance data (search times and number of defects detected) were collected for each individual subject while performing criterion task in VR as well as ASSIST system.

4.7. Results

The mean scores for both the environments determined from the questionnaire are compared in Fig. 9. The Wilcoxon Test was performed on all comparative questions to determine if the difference in means was significant. The results of the test are tabulated in Table 3. Table 4 presents the performance data of subjects for the VR and ASSIST comparison analysis.

Table 4
Performance results on the VR with ASSIST comparison analysis

	Mean (SD)	
	Search times (s)	Percentage defects detected
ASSIST	425.01 (204.32)	33.57 (13.36)
VR	275.70 (73.640)	52.38 (13.64)
F-value	6.61*	13.57**

*$p<0.05$, **$p<0.001$.

4.8. Discussion

The results revealed that the VR system was preferred to the ASSIST on 7 of the 10 questions while the ASSIST system was preferred on 2 of them. The remaining question was not significantly different, favoring neither system.

As these results indicated, the VR system was viable and preferred to ASSIST as an aircraft inspection-training tool. Although both systems compared favorably to the real world environment, the VR system scored significantly higher than the ASSIST (see Question 1). There was no significant difference noticed for Question 8, indicating that both the systems were easy to navigate. The issues related to the description of the defects, showed significant differences between the VR and the ASSIST systems, with the VR system scoring higher (see Questions 2 and 3). The ASSIST system scored higher on the issues related to the response time, probably because ASSIST runs on a PC platform where all CPU resources are dedicated to the software, whereas the VR system runs on a Unix-based multi-user platform. Parallel users on the Unix platform had a significant impact on the speed, with which the VR simulation ran and responded. This was reflected in subjects' responses to Questions 4 and 7, where the subjects preferred ASSIST to VR. The VR system scored significantly high on issues related to the involvement and immersion experienced by the subjects (Questions 5 and 6). As a result, the subjects felt that the VR system was more applicable as a training tool and job aid for the visual inspection task (Question 9), and when asked for their personal preferences, most preferred to use the VR system (Question 10).

Analysis of performance data revealed a significantly greater number of defects ($p < 0.001$) and significantly lesser visual search time ($p < 0.05$) in the VR environment in comparison with the ASSIST environment. These results coupled with subjects' perception of the two systems potentially indicate the effectiveness of VR environment over ASSIST environment in improving both speed and accuracy of visual search.

Although there were significant differences as revealed by the subjects between the two systems, the ASSIST system did reasonably well on the issue of comparison with the real world and navigational aspects. This result indicates that though the VR system is a viable preference for aircraft inspection training, ASSIST can still be used as a low-cost training tool. In addition, the VR system needs to be more realistic. Since the VR system was developed using 2D images, it lacks the feeling of being in a complete 3D environment, an environment that can be achieved by developing the depth perception of the simulator more fully.

Moreover, a study that looked into the transfer effects of training using ASSIST for aircraft inspection performance was encouraging, resulting in improved performance and faster search times (Gramopadhye et al., 2000). Based on this, it can be hypothesized that a more realistic environment like the VR simulator, when used as a training tool for inspection might show greater improvements in the performance of inspectors.

5. Conclusions

The VR system scored well most aspects of presence indicating that it suitably mimics the real world environment. Moreover, it was favored over the existing PC-based ASSIST system. Thus the VR system has the potential for use as an off-line training tool for aircraft visual inspection tasks. The use of the VR-based inspection environment will facilitate in conducting controlled studies off-line and in understanding human performance in aircraft inspection. Results obtained from these studies will yield interventions, which can be used to improve aircraft inspection performance and ultimately, aviation safety.

Acknowledgements

This research was funded by grants to Dr. Gramopadhye, Dr. Duchowski and Dr. Melloy from the NASA AMES Research Center (Program Manager: Barbara Kanki) and from the Office of Aviation Medicine, Aviation Maintenance and Inspection: Human Factors Research Program, Federal Aviation Administration (Grant 1-20-1906-51-4087,NASA Ames grant task NCC 2-1114, and NSF CAREER award 9984278).

References

Blackmon, R., Gramopadhye, A.K., 1996. Using the aircraft inspector's training system to improve the quality of aircraft inspection. In: Proceedings of the Fifth Industrial Engineering Research Conference, Minneapolis, MN. Institute of Industrial Engineers, pp. 447–451.

Burdea, G., Coiffet, P., 1994. Virtual Reality Technology. Wiley, New York.

Chin, R., 1988. Automated visual inspection: 1981 to 1987. Comput. Vision Graphics Image Process 41, 346–381.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of the test. Psychometrika 16 (3), 297–334.

Cruz-Neira, C., 1993. Virtual Reality Overview. ACM SIGGRAPH '93 Course Notes: Applied Virtual Reality. ACM SIGGRAPH '93 Conference, Anaheim, CA.

Czaja, S.J., Drury, C.G., 1981. Training programs for inspection. Human Factors 23 (4), 473–484.

Drury, C.G., 1991. The maintenance technician in inspection. In: Human Factors in Aviation Maintenance—Phase One Progress Report, DOT/FAA/AM-91/16. Washington, DC: Office of Aviation Medicine, pp. 45–103.

Drury, C.G., 1992. Inspection performance. In: Salvendy, G. (Ed.), Handbook of Industrial Engineering. Wiley, New York, pp. 2282–2314.

Drury, C.G., Gramopadhye, A.K., 1990. Training for visual inspection. In: Proceedings of the Third Federal Aviation Administration Meeting on Human Factors in Aircraft Maintenance and Inspection: Training Issues. Atlantic City, NJ.

Drury, C.G., Chi, C.F., 1995. A test of economic models of stopping policy in visual search. IIE Trans. 27, 382–393.

Duchowski, A.T., Shivashankaraiah, V., Rawls, T., Gramopadhye, A.K., Melloy, B., Kanki, B., 2000. Binocular eye tracking in virtual reality for inspection training. In: Proceedings of the Eye Tracking Research & Applications Symposium, Palm Beach Gardens, FL, November 6–8, 89–96.

Duchowski, A.T., Medlin, E., Gramopadhye, A.K., Melloy, B., Nair, S., 2001. Binocular eye tracking in VR for visual inspection training. In: Virtual Reality Software & Technology (VRST), Banff, AB, Canada. ACM, New York.

Durlach, N.I., Mavor, A.S., 1995. Virtual Reality Scientific and Technological Challenges. National Academy Press, Washington, DC.

Federal Aviation Administration FAA, 1991. Human factors in aviation maintenance. Phase One Progress Report, DOT/FAA/AM-91/16. Office of Aviation Medicine, Washington, DC.

Federal Aviation Administration (FAA), 1993. Human factors in aviation maintenance. Phase Two Progress Report, DOT/FAA/AM-93/5. Office of Aviation Medicine, Washington, DC.

Federal Aviation Administration (FAA), 2000. Using ASSIST for aircraft inspection training: performance and usability analysis. Technical Report submitted to the Biomedical and Behavioral Sciences Division, Washington, DC 20591. Human Factors in Aviation Maintenance Research Program, Phase X Progress Report.

Foley, J.D., van Dam, A., Feiner, S.K., Hughes, J.F., 1990. Computer Graphics Principles and Practice. Addison-Wesley, Reading, MA.

Gramopadhye, A., Drury, C., Sharit, J., 1994. Training for decision-making in aircraft inspection. Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, Seattle, WA, pp. 1267–1272.

Gramopadhye, A.K., Ivaturi, S., Blackmon, R.B., Kraus, D.C., 1995. Teams and teamwork: implications for team training within the aircraft inspection and maintenance environment. FAA-1995 Technical Report. Office of Aviation Medicine, FAA.

Gramopadhye, A.K., Drury, C.G., Sharit, J., 1997a. Feedback strategies for visual search in airframe structural inspection. Int. J. Ind. Ergonomics 19 (5), 333–344.

Gramopadhye, A.K., Drury, C.G., Prabhu, P.V., 1997b. Training for visual inspection. Human Factors Ergonomics Manuf. 7 (3), 171–196.

Gramopadhye, A.K., Bhagawat, S., Kimbler, D., Greenstein, J., 1998. The use of advanced technology for visual inspection training. Appl. Ergonomics 29 (5), 361–375.

Gramopadhye, A.K., Melloy, B.J., Chen, S., Bingham, J., 2000. Use of computer based training for aircraft inspectors: findings and recommendations. In: Proceedings of the HFES/IEA Annual Meeting, San Diego, CA, August '00.

Heim, M., 1998. Virtual Realism. Oxford University Press, Oxford.

Kalawsky, R.S., 1993. The Science of Virtual Reality and Virtual Environments. Addison-Wesley, New York.

Kundel, H., Nodine, C., Krupinski, E., 1990. Computer displayed eye position as a visual aid to pulmonary nodule interpretation. Invest. Radiol. 25, 890–896.

Latorella, K.A., Gramopadhye, A.K., Prabhu, P.V., Drury, C.G., Smith, M.A., Shanahan, D.E., 1992. Computer-simulated aircraft inspection tasks for off-line experimentation. In: Proceedings of the Human Factors Society 36th Annual Meeting. The Human Factors Society, Santa Monica, CA, pp. 92–96.

Moll, R.A., 1980. Product liability: a look at the law. Eng. Educ. 66, 326–331.

Nickles, G., Marshall, J., Gramopadhye, A.K., Melloy, B., 2001. ASSIST: training program for inspectors in the aircraft maintenance industry. International Encyclopedia for Ergonomics and Human Factors, Vol. 2. Taylor & Francis, UK, pp. 1178–1180. ISBN: 0748408487 (Invited Chapter).

Patrick, J., 1992. Training Research and Practice. Academic Press, New York.

Schloerb, D.W., 1995. A quantitative measure of telepresence. Presence 4 (1), 64–80.

Shivashankaraiah, V., 2000. Multi-threading a binocular eye tracking virtual environment for aircraft inspection training. Masters of Science, Final Report, Department of Computer Science, Clemson University, August '00.

Singer, M.J., Witmer, B.G., 1996. Presence measures for virtual environments: background & development. Draft United States Army Research Institute for the Behavioral and Social Sciences.

Slater, M., Steed, A., 2000. A Virtual Presence Counter, Presence: Teleoperators and Virtual Environments, Vol. 9 Issue 5, MIT Press, 413–434.

Stanney, K.M., Mourant, R.R., Kennedy, R.S., 1998. Human factor issues in virtual environments: a review of the literature. Presence 7 (4), 327–351.

Thapa, V.B., Gramopadhye, A.K., Melloy, B., Grimes, L., 1996. Evaluation of different training strategies to improve decision making performance in inspection. Int. J. Human Factors Manuf. 6 (3), 243–261.

Wiener, E., 1975. Individual and group differences in inspection. In: Drury, C., Fox, J. (Eds.), Human Reliability in Quality Control. Taylor & Francis, London.

Witmer, B., Singer, M., 1994. Measuring immersion in virtual environments (Tech. Rep. 1014) Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Witmer, B.G., Singer, M.J., 1996. Presence Measures for Virtual Environments: Background & Development. Draft United States Army Research Institute for the Behavioral and Social Sciences.

Witmer, B.G., Singer, M.J., 1998. Measuring presence in virtual environments: a presence questionnaire. Presence 7 (3), 225–240.