Re-viewing reality: human factors of synthetic training environments

ALEX W. STEDMON
HUSAT Research Institute, Loughborough University, Loughborough, LE11 1RG, UK

ROBERT J. STONE
Virtual Presence Ltd, Chester House, 79 Dane Road, Sale M33 7BP, UK.
email: r.stone@vrsolns.co.uk

(Received 18 December 2000, and accepted in revised form 31 May 2001)

Computer-based training (CBT) has become an important training tool and is used effectively in providing part-task activities. In the military domain virtual environments (VEs) have long been exploited, mainly through virtual reality (VR), to create realistic working environments. More recently, augmented reality (AR) and advanced embedded training (AET) concepts have also emerged and the development of “AR-AET” and “VR-CBT” concepts promise to become essential tools within military training. Whilst the advantages of both AR and VR are attractive, the challenges for delivering such applications are, generally, technology led. Equally as important, however, is the incorporation of human factors design and implementation techniques and this has been recognized by the development and publication of International Standard ISO 13407, Human-Centred Design Processes for Interactive Systems. Examples described in this paper serve to review Human Factors issues associated with the use of both AR and VR training systems. Whilst there are common issues between AR and VR applications in considering the potential of synthetic training environments, it is also necessary to address particular human-centred design issues within each application domain.

KEYWORDS: virtual reality; augmented reality; virtual environments; computer based training; human factors of synthetic training.

1. Introduction

From programmed text delivery to flight simulation, computer-based training (CBT) and computer-aided instruction (CAI) have become important “technology-based training” tools and are used effectively in supporting part-task activities, incorporating systems that endow basic computer aided design (CAD) or virtual models with realistic behaviours (Stone, 2001a). Defence establishments around the world have long exploited virtual environments (VEs), in large-scale simulators designed for operations planning, war gaming and command-control-communications and intelligence (C^3I). Indeed, many routine aspects of tri-service training now exploit VEs, in various activities such as
helicopter training, parachuting experience, explosive ordnance disposal, naval helicopter deck landing, submarine and surface ship blind piloting, officer of the watch (OOW) training, etc. (Stone, 2001a). VEs are “computer-generated models for participants to experience and interact with intuitively in real-time”, and can be generated via a host of technologies (see Wilson, 1997) of which virtual reality (VR) is one of the most common (Nichols, Haldane & Wilson, 2000). More recently, however, augmented reality (AR) and advanced embedded training (AET) concepts have also emerged in the military domain (Young, Stedmon & Cook, 1999). AET systems offer the potential to train operators in their real working environments rather than being trained elsewhere and, coupled with AR (the synthesis of computer images and the real world) for the provision of on-line feedback, offer dynamic and flexible alternatives to conventional training facilities (Zachary, Ryder, Hicinbothom & Bracken, 1997).

This paper sets out to present and review the current state of synthetic training applications that either supplement the real world (as in AR) or substitute the real world (as in VR). Results of research in the academic sphere and of successful applications in the industrial arena are provided. As such, these two perspectives highlight two themes that are central to the VE applications described within this paper:

- the need to address specific Human Factors issues prior to the application of advanced technologies (from the academic treatment of AR training);
- the importance of a human-centred approach in understanding how advanced technologies might be exploited in the training domain (from the industrial application of VR training).

The academic perspective of this paper focuses on traditional experimental studies whilst the industrial perspective draws more from applied fieldwork. By taking this approach it is apparent that these two perspectives yield different knowledge that is specific to the AR and VR training domains. The challenge of such an approach is to draw from both and to provide a review of the issues that both face in the development and use of advanced training technologies for the future.

2. Computer-based training in synthetic environments

Developments in augmented reality advanced embedded training (AR-AET) and virtual reality computer-based training (VR-CBT) show much promise. They are becoming essential tools within petrochemical, automotive, heavy engineering and military training, by helping to familiarize personnel with the spatial and behavioural aspects of various operations and procedures, especially those of a safety critical nature.

Whilst AR and VR are similar concepts (often using very similar hardware and software) they are inherently different in the way they command their environments. Depending on the user requirements of the application under investigation, VR applications can employ immersive headsets, desktop (Windows-based) workstations or special-purpose image projection systems, such as those described by Stone (2001a). AR technology is generally applied via a transparent or variable translucent head-mounted display (HMD) providing visual augmentation to the real-world environment (Stedmon, Hill, Kalawsky & Cook, 1999a). A typical AR HMD is illustrated in Figure 1. Although
FIGURE 1. A typical AR see-through HMD.

this is the most common manifestation of AR, reality may be augmented in a number of ways and through various sensory mechanisms such as tactile and even olfactory systems (Croft & Craig, 1998).

There are particular advantages implicit in the synthesis of real and virtual information. In the case of AR, orientation cues are still available to the user from the real-world scene and users are, therefore, unlikely to experience the feelings of vertigo and sickness that can be induced by traditional VR systems (Caudell, 1994). However, while AR eliminates the need to address some Human Factors issues that characterise VR applications, AR configurations produce new challenges; these include better information presentation (Feiner, MacIntyre & Seligmann, 1993) and registering the synthetic and real environments into a seamless and believable “mixed-reality” environment.

Whereas AR combines aspects of a synthetic environment with a real environment, VR replaces the real environment with a synthetic environment altogether. As such, the fundamental problems inherent in AR are overcome by modelling the whole environment which the user experiences. Although this procedure takes up more resources, in terms of programming all aspects of the environment, it does allow for complete control of the virtual world. For training this has a number of potential benefits:

- VR may be used where manipulation of real-life variables would risk damage to people, equipment or the environment (e.g. in hazardous environments);
- VR can provide an alternative where manipulation of real-life variables has an unacceptably high associated resource cost (e.g. logistics, finance, personnel or national security);
- VR allows users to experience views of micro or macroscopic entities in a variety of dimensions [e.g. using VR to explore the atomic surface of materials or human tissue at a nanoscopic level (Stone, 2001b), in a way that would not be physically or ethically possible in the real world].
VR may be used to enhance or degrade or otherwise alter some aspect of reality. As such, VR might be used to impose visual restrictions on the user (e.g. smoke effects in a burning compartment or reduced visibility due to a visual defect) or VR might be used to highlight components in the real world which could otherwise be missed (e.g. fire extinguishers or escape routes through buildings); VR can be used where real locations are impossible or difficult for users to physically occupy (e.g. hazardous environments, simulating outer space, under the sea, etc.); VR allows rapid prototyping and configuration of an environment where perceptual input might need to be changed frequently (e.g. reconfiguring interface details or instrument layouts); VR can be used to aid initial design and configuration of training environments even if VR is not ultimately used for actual training (e.g. using VR as a data collection and feasibility assessment tool for procurement exercises).

Whilst the advantages of both AR and VR are attractive, the adoption of VE-based training techniques has not been straight-forward. The push for classroom VR trainers, designed to replace ageing conventional techniques, such as “chalk-and-talk”, overhead projection, simple video and even 2D-CBT, brings with it new challenges. These challenges are, in general, technology led with issues such as the following:

- The delivery of high performance and visual fidelity on low-cost personal computer workstations;
- Open systems architectures (assuring the longevity and reusability of training applications);
- Standardizing techniques for 3D computer modelling;
- Developing protocols for the integration of behavioural simulations with multi-display rendering.

Equally as important, however, is the incorporation of “best practice” Human Factors design and implementation techniques. Only by doing this can the exploitation of AR and VR as evaluation environments be truly realized.

3. Human factors in synthetic training environments

Whenever new technologies introduce a new level of functionality into a training paradigm, whether it is a desktop CBT application or a full-scale simulation facility, the same typical questions apply (Stone, 2001a):

- will it improve the effectiveness with which knowledge is delivered or assimilated?
- will it reduce the reliance on scarce operational systems or costly hardware-based training material?
- does it offer anything over and above conventional training methods?
- can previous investments in technology be protected, or must new bespoke systems be procured?
- will students and trainers actually use the technology?
- will there be a positive transfer of training (or knowledge) from the computerized setting to the real operational environment?
These sorts of questions highlight the need to understand not only the wider aspects of using advanced technologies for training but also the more focused need to understand the Human Factors of how such technologies might be used and the demands they place on their users. The exploitation of VEs has not escaped this examination and, indeed, there are those who firmly believe that VR, in particular, has suffered as a result of early promises of an ultimate system. Fortunately, the application of VR in a number of industrial applications has stimulated a “revival” of interest in both immersive and desktop applications (Stone, 2001a) and also in the developing potential of AR.

In parallel with this “technology push” there has been a “market pull”, with potential CBT users demanding lower technology costs, more efficient utilization of students and training systems and hard evidence of the cost-benefits and manpower performance improvements the technology might offer.

4. Human-centred design processes for interactive systems

To answer the typical questions levelled at any new training technology application and to achieve validity and reliability in development and evaluation programmes, it is crucial to invite and incorporate the expertise of the Ergonomics and Applied Psychology community. An understanding of the Human Factors issues is essential for the appropriate design of experimental programmes (Meister, 1985, 1986; AIAA, 1993) and also for guidance in the structured analysis of real-world tasks and task elements to be included in a training schedule.

Various task analysis techniques (e.g. Kirwan & Ainsworth, 1993; Militello & Hutton, 1998) allow Human Factors researchers to describe the interactions between users and their working environment at a level of detail that is appropriate to a pre-defined end goal. Typically, such techniques are used for evaluating existing system and task requirements, but can also be employed and developed for the definition of future system and task requirements (Stone, 2001a). The type of analysis employed depends on a number of variables such as the particular Human Factors specialist involved, whether the task exists in reality, the goal of the analysis, and any constraints imposed by the analysis environment.

The task analysis should form an early and central component of any project that takes a human-centred perspective. Indeed, recognition of this has recently been formalized by the publication of International Standard ISO 13407, Human-Centred Design Processes for Interactive Systems. ISO 13407 (Earthly, Sherwood Jones & Bevan 2001) specifies four general principles of human-centred design:

- ensure active involvement of users and a clear understanding of user and task requirements (including context of use and how users might work with any related future system);
- allocate functions between users and technology (recognizing that today’s technology, rather than de-skilling users, can actually extend their capabilities into new applications and skill domains);
- ensure iteration of design solutions (by involving users in as many stages of the design and implementation cycle as is practical);
• ensure that the design is the result of a multidisciplinary input (emphasizing the importance of user feedback, but also stressing the need for input from such disciplines as marketing, ergonomics, software engineering, technical authors, etc.).

Whilst task analyses can provide an abstraction of the real-world task elements into their corresponding AR and VR environments, it is only when they are implemented within a coherent framework, such as ISO 13407, that human-centred design issues are fully appreciated. Without such a framework, there is a risk of specifying or designing a VE system that fails to support the users’ understanding of the target application (Stone, 2001c).

The examples described in the remainder of this paper serve to review Human Factors issues associated with the use of both AR and VR training systems. Whilst there are common issues in AR and VR applications, when considering them collectively as synthetic training environments, the examples also show that there are particular issues within each separate domain. The examples have been chosen to reflect the diverse range of Human Factors topics typically involved in advanced technologies and to highlight the potential benefits of both AR and VR in the training domain.

5. AR-AET concepts and research

The provision of cost-effective training for the operators of future systems is a particular concern of both the UK and US Navy and, more generally, for other military organizations (O’Shea, Cook & Young 1999). Within the Royal Navy, most Command Team Training (CTT) is currently undertaken within high-fidelity shore-based training centres. Whilst such training is extremely effective, it is also a high resource drain on manpower, time, logistics and finance.

There is a growing emphasis on training at sea, and the US Navy is developing ship-board, CTT strategies such as AET (Zachary et al., 1997). These AET systems, where the training system is embedded within the operational equipment, represent the forefront of training concepts and technology.

AET systems have several potential advantages over more conventional CTT (O’Shea et al., 1999), including:

• reduced instructional staff costs (since fewer supervisors are necessary);
• reduced equipment procurement and through-life costs (since the maintenance of independent high-fidelity training simulators is eliminated);
• the provision of “just in time” training;
• increased transfer of training (as training takes place within the operational domain);
• increased availability of training facilities at both the individual and team training levels.

Various methods can be used to present on-line feedback, including the following:

• a human instructor providing verbal feedback;
• auditory feedback through headphones;
• visual feedback presented on a secondary display;
• visual feedback incorporated on a primary task display;
• visual feedback presented in the line-of-sight on a personal display.
Whilst some of the above options are conventional training techniques, their practicality is called into question in the Naval CTT domain. Trainees may be monitoring radio communications channels through headphones (sometimes with different communication channels to each ear) and, as such, it may be impractical to have supervisors standing over them.

Implicit in the philosophy of AET is the notion that trainees learn from the system they are being trained on. As a consequence, human supervision and intervention is not anticipated to the same degree as in other training platforms. Furthermore, as trainees have to monitor their communication channels and are also expected to keep abreast of other communications within an Operations Room, providing auditory feedback is not necessarily the best solution.

Some form of visual feedback could be a candidate method but even this has potential problems. Providing training feedback on a display positioned to the side of a primary task display is one option, but there is a risk that feedback messages might not always be noticed since the operator’s focus is on their primary task display. Providing feedback on the primary task display itself would overcome this problem, but could prove expensive and impact on primary screen “real estate” available for primary task fidelity. Furthermore, problems arise in situations where individuals in a team might share the use of a display with other team members. Since each individual team member may require personalized feedback in the training environment, the presentation of this information on shared operational screens could interfere with the performance of other team members.

AR-AET has the potential to enhance a conventional, real-world, training environment with additional, synthetic information. Such a facility places the trainees in their actual working environment, allowing them to carry out required tasks and ensure that feedback is delivered in the trainee’s line-of-sight. In this way, training feedback can be delivered on-line and in real time to enhance the training process (O’Shea et al., 1999). Vreuls and Obermayer (1985) advocate that measurement information for feedback and performance diagnosis must be provided in real, or near-real, time if it is to be useful for training.

A potential advantage of AR over more conventional methods of providing additional information is that it can create a training environment without major modifications to the operating interface, or without having to create expensive simulation facilities. Of course, this potential can only be realized if the possible problems mentioned above are resolved appropriately. Another potential advantage of AR is that the learner experiences training in the real working environment, so transfer to the operational situation may well be enhanced. The training support provided via AR may also be reduced subtly and gradually until the trainees find themselves in the full working situation.

5.1. COGNITIVE ISSUES IN AR

Much of the development within AR has been technology and applications driven and, as a consequence, Human Factors issues of AR have not received much research attention (Stedmon et al., 1999a). Technological limitations of synthesizing real and virtual images in the same visual field present particular problems, such as image registration, which raises concerns of where an operator is looking in order to present an
AR image in an appropriate location against the real world. Any error in the registration, such as time lags, perspective, size or collimation (even conflicting texture or colour cues), may cause user performance difficulties, eye strain, fatigue and disorientation (Azuma, 1993).

However, there are other limitations associated with the use of AR. Of particular interest are the excessive human information processing demands that AR may impose. Overlaying one synthetic image upon another may exacerbate any processing difficulties for the user as the fusion of two synthetic images might create an unacceptable cognitive load (brought about by fusion uncertainty and perceptual rivalry). If the perceptual load placed upon the user is too great, it is unlikely that AR would prove to be an effective training aid.

Research into the Human Factors issues surrounding the use of AR systems appears very limited, and no set guidelines exist for any application of AR technology (Stedmon et al., 1999a). In order to address some of the cognitive ergonomics issues of using AR, a number of experimental studies have been conducted. These have attempted to identify effects on human cognitive performance when an operator was required to process information presented via AR overlaid on a primary display, and in the processing of information on the primary display itself.

The following experimental trials were conducted (see Table 1). Summary accounts of the experimental rationales and results are given below. Full accounts of the experiments and further discussion of the findings are given in Hill and Kalawsky (2000), Kalawsky, Hill, Stedmon, Cook and Young (2000) and Stedmon et al. (1999a, b).

5.1.1. Information comprehension and retention. These experiments investigated whether an AR headset was a viable information provider in terms of comprehension and retention of text and graphics formats, the interpretation of simple prose and the learning of icons through AR-assisted explanations.

Experiments 1.1 and 1.2—recall of unrelated nouns (text and graphics)
- Rationale—it is known that the short term memory (STM) capacity for chunks of uni-dimensional stimuli is approximately 7 ± 2 (Miller, 1956), and that displaying blocks of more than seven unrelated stimuli should produce a degradation in STM recall performance. Strings of stimuli, containing 4, 7, 10 or 13 items, were displayed as single words (Experiment 1.1) or graphics (Experiment 1.2).
- Results—as expected, as the number of words or graphics in a string of stimuli increased (and hence STM load increased), the number of stimuli correctly recalled decreased significantly. The results from both experiments showed no significant difference between the AR headset and primary display for STM recall.

Experiment 1.3—comprehension and retention of simple prose
- Rationale—the smooth and efficient operation of human–machine systems often depends on the processing of written language such as reading instructions, comprehending labels, exchanging information with other team members, etc. (Wickens, 1992). To examine simple comprehension and retention effects, two short prose passages were presented to participants on either the AR or primary display, and questions relating to the passages were presented after each trial.
• Results—the data showed no significant difference between the AR headset and primary display for comprehension and retention of simple prose.

Experiment 1.4—comprehension and retention of abstract symbols
• Rationale—Standing (1973) reported an advantage for icons or symbols over words in the amount that can be remembered and for rapid processing of information from computer displays. This experiment assessed the comprehension and retention of pictorial information when descriptive information was provided in order to establish whether such information could be learned, understood and retained with the aid of an AR display.
• Results—a significant main effect was observed for retention, illustrating that immediate post-test scores were significantly higher than either the pre-test scores or those one week later. These results indicated that the learnt meanings had not been committed to long term memory (LTM). There were no significant differences between the primary display combination and the AR headset or a side monitor combination.

5.1.2. Perceptual interference caused by AR. These experiments combined information presented on a primary display format (VDU) and an AR headset using target identification tasks to assess interference or clutter effects. Display clutter is the interference of information on a display when too much, or different types of, information are presented at the same time (saturating the display and/or overwhelming the operator’s cognitive abilities); or because information is presented or updated at such a rate that it overlays other information, causing obscuration effects (Stedmon et al., 1999b).

Experiment 2.1—interference: increasing clutter effects
• Rationale—this experiment determined the extent to which clutter (non-targets) interfered with a visual search task and was conducted using the display formats in isolation.
• Results—a significant main effect was observed for the level of clutter, which confirmed that target location times were greater when there were more non-targets or clutter. There was no significant difference in target location times between the AR headset and the primary display.

Experiment 2.2—interference: cumulative and compounding clutter effects
• Rationale—this experiment was designed to examine how clutter displayed on the AR headset affected task performance on the primary display when both displays were used at the same time.
• Results—there were no significant differences in target location times between the AR headset and the primary display. Therefore, when the two displays were combined and it was not known on which display the target was to be located, there was no difference between the two displays.

Experiment 2.3—interference: AR secondary task effects
• Rationale—the aim of this experiment was to determine how a secondary AR task might affect primary task performance and what the implications would be for divided attention.
<table>
<thead>
<tr>
<th>Series/ exp no.</th>
<th>Description</th>
<th>Main effects under evaluation</th>
<th>Measures taken</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>STM recall—text</td>
<td>• STM load</td>
<td>• STM recall, correctly recalled items</td>
<td>• No difference between displays</td>
</tr>
<tr>
<td>1.2</td>
<td>STM recall—graphics</td>
<td>• STM load</td>
<td>• STM recall, correctly recalled items</td>
<td>• No difference between displays</td>
</tr>
<tr>
<td>1.3</td>
<td>Comprehension and retention—prose</td>
<td>• Prose comprehension</td>
<td>• Comprehension measures</td>
<td>• No difference between displays</td>
</tr>
<tr>
<td>1.4</td>
<td>Comprehension and retention—icons</td>
<td>• Learning</td>
<td>• Comprehension measures</td>
<td>• No difference between displays</td>
</tr>
<tr>
<td>2.1</td>
<td>Interference—increasing clutter effects</td>
<td>• Clutter effects on visual search task</td>
<td>• Response times • Correct responses</td>
<td>• Clutter impaired task performance</td>
</tr>
<tr>
<td>2.2</td>
<td>Interference—cumulative and compounding clutter effects</td>
<td>• Clutter effects on visual search task</td>
<td>• Response times • Correct responses</td>
<td>• Combined displays impaired task performance</td>
</tr>
<tr>
<td>2.3</td>
<td>Interference—AR secondary task effects</td>
<td>• Divided attention</td>
<td>• Compensatory tracking performance</td>
<td>• AR task impaired dynamic task performance</td>
</tr>
<tr>
<td>Series/ exp no.</td>
<td>Description</td>
<td>Main effects under evaluation</td>
<td>Measures taken</td>
<td>Outcome</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------</td>
<td>------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>3.1</td>
<td>Response times to alerts</td>
<td>• Divided attention</td>
<td>• Response times</td>
<td>• AR headset effective for presentation of alarms</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Re-accommodation issues</td>
<td>• Subjective workload measures</td>
<td>• No re-accommodation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Task performance</td>
<td></td>
<td>• Increased workload effect condition</td>
</tr>
<tr>
<td>3.2</td>
<td>Simultaneous events</td>
<td>• Divided attention</td>
<td>• Response times</td>
<td>• AR headset effective for task cueing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Workload</td>
<td>• Subjective workload measures</td>
<td>• Increased workload effect for no alarms condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Task performance</td>
<td></td>
<td>and side-VDU</td>
</tr>
<tr>
<td>4.1</td>
<td>Alarms in multiple display environments</td>
<td>• AR cueing</td>
<td>• Critical moment analysis</td>
<td>• AR headset effective for presentation of alarms</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Workload</td>
<td>• Subjective workload measures</td>
<td>• Increased workload effect for no alarms condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Task performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Head-tracked dedicated feedback</td>
<td>• AR head-tracked feedback</td>
<td>• Critical moment analysis</td>
<td>• Head-tracked feedback most effective for task performance effect for</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Workload</td>
<td>• Subjective workload measures</td>
<td>no alarms condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Task performance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.1.3. Stimulus detection across displays. In typical Naval operational environments, operators deal with their primary tasks while often communicating with other team members and/or monitoring tasks and communications on other systems. AR was explored to determine how such a facility might be used to provide cues to operators, perhaps advising them of a high priority situation occurring on another display, such as that of a colleague, or another display outside their primary task view.

Experiment 3.1—response times to alerts

- Rationale—this experiment assessed the time taken to respond to an alarm presented on either the AR headset or primary task display, in isolation, concentrating on focused attention and information processing limitations. Alarms were presented at either the same focal length as the primary task display or 4 ft behind it. Conducting the experiment in this way addressed any visual re-accommodation effects.
- Results—no significant difference was observed for the focal length of the AR headset, illustrating that re-accommodation did not appear to be a significant performance factor. No significant effect was observed between the AR headset and the primary display format and, as such, alarms presented on either display format were equally salient.

Experiment 3.2—simultaneous events

- Rationale—this experiment sought to determine how AR alarms interfered with a primary display task. To achieve this, participants carried out a continuous pursuit task on the primary display and attended to alarms presented on the AR headset.
- Results—a significant main effect was observed for alarm type and the data demonstrated that complex alarms took longer to respond to than simple alarms. A significant main effect was also observed for the simultaneous tasks, indicating that there was a decrement in primary task performance when an alarm was attended to. No significant effect was observed between the AR headset and the primary display format and, as such, alarms presented on either display format were equally salient.

5.1.4. Multiple display environments. In a team-working, multiple-display, environment it is likely that an operator’s attention could be diverted away from their primary task display. If an operator is dealing with a difficult situation on the primary display, other displays might be ignored to the extent that an alarm is missed. Displaying an AR warning in the line-of-sight could allow the operator to monitor the status of other displays without the need to break away from the primary task.
Experiment 4.1—responses to alarms

- Rationale—participants were required to maintain performance on a continuous tracking task, whilst ensuring that critical situations on peripheral displays, either side of the primary task, did not go undetected. Critical information, presented via AR in the line-of-sight was compared to no alarms and information presented on a side-monitor.

- Results—in the “no alarms” condition, participants constantly had to check peripheral displays, whereas in the AR and side monitor conditions they could concentrate on the primary task, secure in the knowledge that they would be informed of a critical situation if/when it arose. In conditions where alarms were presented, less “critical time” was spent on peripheral task displays. There was no significant difference in performance scores between the AR headset and the side monitor and, as such, it would appear that neither display format was superior. However, the subjectively reported workload of the line-of-sight cueing information on the AR headset was significantly lower than that of either the side monitor or the no alarms condition. As the warnings on the AR headset were presented in the line-of-sight, the operator had no need to look away from the primary task in order to attend to the alarm information.

5.1.5. Advanced issues of AR. This experiment investigated the use of AR to cue an operator to ongoing activities in a multi-console multi-task environment. AR was employed to provide timely and dedicated feedback to operators, in their line-of-sight, depending on where they are looking and what task they are carrying out at the time.

Experiment 5.1—head-tracked dedicated feedback

- Rationale—in this experiment comparisons were made between head-tracked vs. non-head-tracked and cueing vs. explanatory information across multiple displays. In the head-tracked condition feedback was dedicated to where operators were looking and the task they were conducting at the time. In the non-head-tracked condition, without knowing where the operator was looking, the feedback was standardized over the displays and tasks so participants received all the information pertaining to all of the displays.

- Results—in the “no feedback” condition, participants constantly had to check peripheral displays, whereas in both AR conditions (head-tracked and non-head-tracked), they could concentrate more on each of the separate tasks with the knowledge that they would be informed of a critical situation if/when it arose. This point was supported by a significant main effect observed for task consistency. This illustrated that, in the condition with head-tracked AR feedback, performance across multiple tasks and displays was enhanced, whilst in the other condition the task consistency suffered.

5.1.6. Overview. As a whole, these experiments demonstrated that an AR display provides information as efficiently as a standard computer monitor. Although it was found that dealing with a secondary task on an AR display interfered with a primary task, the proposed use of AR is to aid primary task performance. As such, the AR task would not be a competing task and, with careful implementation of ISO 13407 methods, the AR
component of the training task would be complementary to and contextually bound by the primary task.

Consequently, when the AR display was used for this purpose (in the multiple display experiment), it was indeed found to enhance performance. In future it seems likely that, with increased system complexity and “optimized manning”, fewer operators will be required to carry out multiple tasks with multiple displays. AR appears to offer potential for managing user attention and enhancing task performance.

5.2. ON-LINE TRAINING FEEDBACK IN NAVAL CTT
In conjunction with the examination of the cognitive ergonomics of AR, further work was carried out into the potential of on-line training feedback in a complex, decision-making task based upon real-world Naval operator roles (O’Shea et al., 1999; Young et al., 1999; Cook, O’Shea, Young & Stedmon, 1999). Of particular concern were any comparative benefits of different types of individual on-line feedback with off-line feedback (i.e. post-exercise) for improving training performance and retention; and the relative benefits of providing the feedback directly in the trainee’s line-of-sight or via a dedicated side display. Participants received either real-time feedback in the line-of-sight (AR); real-time feedback via a side-VDU; no additional feedback (Control); or the same feedback after completion of the scenario (post-exercise).

To overcome some of the technological limitations with AR technology in its current state, an experimental facility was developed at the Defence Evaluation and Research Agency (DERA) that simulated an AR training application. The Tactical Projection Graphical System (TPGS) incorporated a large screen display with two task interfaces, a simulation of a synthetic task abstracted from a number of Naval operator roles, and a feedback editor which produced feedback overlays for the large screen display. The configuration is shown in Figure 2.

Head and eye movements were tracked during task performance in order that the relevant feedback could be provided in the line-of-sight. In this way, the relative benefits of providing timely and relevant feedback could be assessed without the limitations of current image registration and collimation or the visor resolution problems.

![Figure 2. The DERA TPGS facility.](image-url)
In this study, participants generally found on-line AR training feedback more useful than feedback presented in the other conditions. Participants in the AR condition thought they read all or most of the feedback messages, whilst two-thirds of participants in the side monitor condition thought they read only a few of the messages. Feedback delivery in the line-of-sight seemed to result in more messages being read than when the feedback was delivered via a separate display (side-VDU).

Questionnaire data were also collected as a measure of participants’ knowledge of the training rules. Although participants in the AR condition performed best, the differences observed were not significant. An indication of the extent to which participants were applying the rules was provided by the measure of training process. This measure was derived from evidence of specific instances when training rules were broken, in those cases where a rule should clearly have been applied. During the experimental scenarios, participants in the AR condition were found to have performed best, followed by participants in the side-VDU condition. Participants in the post-exercise condition performed worst in the experimental scenarios, but best in the retention trial. However, it must be noted that none of the differences observed was significant. Nevertheless, the fact that performance was at least equally good using the AR compared with conventional methods is a success in and of itself.

With regard to a final distance from rendezvous point, participants in the AR condition performed best in both the experimental and retention phases of this study, with participants in the post-exercise condition performing worst in the experimental phase. The task of maintaining a rendezvous position may have been perceived as a secondary task, since participants were subjected to a number of other activities that may have focused their attention. The finding that participants in the AR condition performed better on this task is important, since it indicates that the feedback reminded them of this second objective, and maintained their focus on both tasks. There was a trend for better performance in the AR condition in the application of the training rules, for the number of messages read and rated as useful to the task.

This study demonstrated the provision of automated real-time trainee feedback in a complex military decision-making task, and evaluated two media of delivery (AR and side-VDU). It is suggested that the provision of such feedback can enhance existing military training programmes (Young et al., 1999), and that this is likely to become more feasible as the technology develops.

The investigation into the cognitive ergonomics of AR utilized a more constrained task and provided more significant results, particularly for the utility of AR for cueing a trainee to attend to other critical events. In general, AR was not found to impose any additional cognitive load on the human operator, and was found to enhance task performance in a multi-task and multi-console environment whilst also lowering reported workload levels.

Findings from this research provide a basis for future AR-AET developments and also have implications for the use of AR in other military applications. These may include the use of AR overlays on shared large-screen displays, to provide individual operators with information that could support their dedicated tactical decision-making roles, or the development of wearable computing applications with integrated AR headsets (Stedmon et al., 1999c).

RE-VIEWING REALITY 689
6. VR-CBT concepts and applications

The previous section described an extensive series of laboratory research studies, stimulated by military—especially Naval—wishes to move training gradually nearer to being “on-the-job” as it were and more integrated with the actual working situation (e.g. at sea). This section illustrates, with an Air Force and two Naval examples, the current stage of implementation of VE technologies in military training. As will become evident, advanced technology training still mainly uses VR and is mainly used as part of advanced training courses rather than integrated with the operational working environment.

6.1. THE TORNADO F3 AVIONICS TRAINING FACILITY

The UK Royal Air Force’s Tornado, Air Defence Variant (ADV), is the UK’s principal air defence aircraft, pending the introduction of the Eurofighter (Typhoon).

As with many other applications in the military sector, gaining access to appropriate hardware for maintenance training, be it a complete aircraft or even individual functional components, presents resource and logistics problems. As such there is a growing need for more flexible training delivery mechanisms. A new training device, the Avionics Training Facility (ATF), was specified. The ATF initiative arose as a result of limited access to airframe hardware and a requirement to reduce training times and costs. Prior to ATF, students at the Tornado Maintenance School (TMS) at RAF Marham were trained using a variety of systems, such as a comprehensive selection of instrumented mock-ups relating to the Tornado GR4 strike aircraft.

6.1.1. Example of previous facilities. To put the ATF facility into context, some of the recent findings (see later) can be contrasted with those of another training unit in place at Marham, namely that focusing on the Tornado GR4 strike aircraft. At the time of writing, this facility had been in existence for over 2-years and provides students with a 13-week course. However, as the facility is based on physical rigs with removable line replaceable units (LRUs) and quite limited bench avionics training devices, only two students plus one instructor can be present on a rig at a time. A typical example is the Command Stability Augmentation System (CSAS). CSAS sub-system training requires LRUs to be removed from the cockpit mock-up and inserted into a nearby bench trainer. Faults can be simulated by means of a complex set of wiring combinations, reminiscent of an old telephone exchange, and as a result of these practices and the system design, each student experiences a 3-week downtime in their training schedule. The CSAS and the bench trainer are shown in Figure 3.

![Figure 3. The CSAS bench trainer.](image)
6.1.2. The Avionics Training Facility (ATF). The main requirements laid down by the RAF for the ATF system included the following:

- ten networked Windows NT workstations (two instructor stations and eight student stations) supporting student collaboration in real time, the “injection” by the instructor of LRU faults and a capability for the instructor to “snoop” on individual workstations;
- multi-screen display (depicting external aircraft, pilot and navigator stations, if required);
- real-time 3D navigation and interaction by simple mouse-and-function key interface;
- integration of underlying ATF simulation with the virtual Tornado, via a core real-time event manager;
- provision for future interface upgrades (head-mounted display, stereo projection, special immersive displays, etc.);
- a software guarantee to allow for ATF support, modification and upgrades over a minimum 10-year period (based on the somewhat fluid nature of VR toolkits between 1998 and 2001).

One typical ATF workstation is shown in Figure 4; each screen displays different working views of the aircraft, avionics bays, LRUs and/or virtual test equipment. Alternatively, the entire aircraft can be displayed as a “panorama” across the three screens. The workstations allow a maximum of eight students to be trained and supervised by two instructors in basic and advanced Tornado F3 avionics maintenance routines (Moltenbrey, 2000; Shepherd, 2000).

In addition to the virtual aircraft shell itself, around which students are free to move, all moving surfaces are present and fully interactive (removable and hinged panels, flight control surfaces, toggle switches, safety covers, rotary switches, push buttons, pedals, throttles and joysticks). Furthermore, over 450 LRUs feature in the simulation, located in equipment bays around the aircraft and as control and display units within the
cockpit. The LRUs can be tested using virtual test equipment and subsequently refitted or replaced.

Control inputs are, wherever possible and ergonomically acceptable, restricted to a conventional mouse (2-button and wheel) and single function key, delivering discrete, momentary and continuous input functions, supporting the following:

- aircraft walk-around;
- panel opening;
- withdrawal, inspection and test of LRUs;
- cockpit, LRU and test set display/control operation.

Every control input made by the student results in a realistic and accurate change of state within the virtual Tornado, from the movement of external flight surfaces down to the illumination of individual LRU test indicators.

6.1.3. The potential of the ATF facility. The ATF facility has, at the time of writing, been in existence nearly 2 years. In that period, the course time has been reduced from 13 weeks (GR4 course) and 11 weeks (Avionics Ground Training Rig course) to 9 weeks, in each instance, with no downtime. The TMS Marham trainers believe that the course could be shortened even further, but are reluctant to do so, choosing instead to increase course content and promote retention through “consolidation breaks” and extra-mural self-paced refresh trials.

Initial feedback from the TMS Marham trainers suggests that, in contrast to previous courses, ATF students “grasp the concept” (i.e. gain enhanced spatial and procedural knowledge) after only two-thirds of the time taken by previous non-ATF students. This finding is in line with some of the CBT performance claims made by Fletcher (1996) and, in a related study, by Boeing investigating the use of desktop and immersive VR as potential training delivery mechanisms for the Joint Strike Fighter (Barnett, Helbing, Hancock, Heininger & Perrin, 2000) (see Stone, 2001d, for more details).

Whilst this is an obvious advantage for the use of such a training facility, there could well be effects that impact on the trainees such as increased stress through tighter training schedules (brought on by the quicker training lifecycles). This might be addressed, if schedules allow, by follow-up training on the real-world rig in those aspects for which the VR Tornado is not so well suited at present.

The modified Tornado F2 rig, as used by the GR4 course students, is still an important training tool. It is employed to replace some of the obvious shortcomings in the VR Tornado such as Health and Safety risks associated with lifting some of the heavier LRUs. The real aircraft can also be used in other Health and Safety training procedures, such as highlighting very hot areas once the aircraft has landed, or components capable of delivering electrical shock. Whilst these aspects could be included in the VR Tornado through various highlighting techniques and even the use of on-line feedback, it is still important (until issues in the transfer of training are identified) to relate such Health and Safety issues back to the real world. Whilst VR may be used for the training of hazardous procedures it should not necessarily be seen as a total substitute for training in the real world.
In terms of financial cost and resource loading, the ATF facility has amounted to £1.5 million and can train a minimum of eight students with two instructors. This is a fraction of the cost of non-VR-based facilities, such as the GR4 facility at UK£14 million (US$35 million); the latter is a physical rig with limited bench avionics training devices and can only accommodate two students plus one instructor at a time.

6.2. SUBMARINE QUALIFICATION TRAINING

Based on similar experiences in delivering the ATF system for the RAF, a recent project has been undertaken for the UK Royal Navy’s Flag Officer Submarines (FOSM). Although the original project was based on the early training and vessel familiarization needs for submariners destined to serve on the UK’s fleet of Trafalgar Class boats, the results of the project are now being implemented for future platforms such as Astute and are under consideration for the future NATO Submarine Rescue System.

6.2.1. Virtual Environment for SUBmarine (VESUB). VESUB, a US project, was a precursor to the UK SMQ project (see Section 6.2.2). VESUB is used primarily as a submarine handling training technology demonstrator. The immersive characteristics of VR were used to create virtual environment views as if the user were located on the fin of a typical US submarine. Over 40 Naval personnel (involved with submarines) were used in the trials of this device. Improvements between training and test sessions ranged from 33 to 57% (see Stone, 2001d, for details).

6.2.2. The Submarine Qualification (SMQ) Training Facility. Following a period of close liaison with the potential customer (as defined in the Human-Centred Design Activities section within ISO 13407) and shore/boat-based users of a Submarine Qualification (SMQ) training system, an approach was recommended for developing and delivering a hybrid VR-multimedia training system.

Videos, 3D engineering data, digital images and panoramas are all structured within a 3D virtual geometric “submarine” shell (itself based on converted CAD data from the submarine developers). Interaction with, and the display of, hybrid data sources is carried out via a “minimalistic” user interface (for both students and instructors, as with the Tornado ATF system), the design of which will also adhere to the guidelines laid down in ISO 13407 and related usability standards and best practices.

The SMQ lesson style is anticipated to take the form of an initial series of virtual “tours” of the 3D submarine model, to demonstrate the user interface components and the functionality of the software. A series of exercises will be developed which will require students to locate and (where necessary) actuate shipboard components. Students can “walk” through the virtual submarine on predefined paths where they are free to turn, look around and interact with “active” scene items, as appropriately identified on-screen, such as lockers, fire equipment, consoles, ladders, large-bore piping and unique components. The student is then able to initiate any animated or interactive features to demonstrate operation and, in the case where the operation of safety equipment can originate from elsewhere in the submarine (e.g. operations compartment), warning displays are also presented. A representation of this is shown in Figure 5.
In contrast to the VESUB project, the Royal Navy requirements relate to the initial provision of a PC-based trainer which will enable students to become familiar with the layout of the target class of boat, including decks, compartments, equipment, main service routes (e.g. high-pressure air), safety equipment and so on. Furthermore, the system is required to preserve the RN’s investment over a long period, by demonstrating features that support:

- upgrading (to account for boat-to-boat differences, refit planning, special system upgrades, equipment additions, enhancing the overall fidelity of the model as technology and resources permit);
- database links (with existing and planned training material).

Reusability of the PC-based trainer will be expected to provide, in whole or in part, for:

- new submarine classes;
- diving training (e.g. hull inspection);
- in-service training (extension of the model for operational and contingency planning);
- navigation (including channel and “blind” pilotage, as with VESUB);
- officer of the Watch/Deck training (as with VESUB);
- dockside/deck procedures;
- special incident rehearsal (e.g. helicopter casualty evacuation).

7. Future directions and conclusions

7.1. Future directions

From the examples discussed in this paper it is apparent that, without the necessary human-centred knowledge and expertise, any technology-based training approach may well not be fully utilized or exploited. Implicit in the way that technology can be used to support training is the way that trainees themselves will use that technology. This is
based on their individual capabilities and limitations (as highlighted in the application of AR for training) and the impact that human-centred approaches have on the way that technology is then integrated or exploited (as highlighted in the application of VR for training).

AR has demonstrated potential for supporting trainee attention management in a multi-task situation, particularly when the trainee is focused upon conducting a primary task. The findings also demonstrated that AR technology did not impose any additional cognitive load on the human operator. In addition, the display of supplementary cueing information via this medium was found to aid task performance, in a multi-task and multi-console environment, by directing the operator’s attention to a secondary task when appropriate. Using findings such as these, it is important to develop a better understanding of which applications would benefit most from AR and to identify further issues that underpin human cognitive performance in an AR environment. Following a human-centred approach during training needs analysis is essential for determining where AR might prove beneficial in the training cycle.

To date, the research has concentrated on single operator use of AR, and no work has been carried out to assess how AR technology might impact on social aspects of team performance and communication. Indeed, with peripheral vision possibly obscured, subtle cues and feedback may be lost, disrupting more traditional communication patterns. So research on this type of problem is one of the next steps towards the implementation of AR laboratory results into practical situations.

In relation to CBT, Fletcher (1996) listed a series of military studies that demonstrated the potential of CBT when compared to standard lectures, text-based materials, laboratory work or hands-on experience with real equipment. Furthermore, the use of “interactive multimedia instruction” (Fletcher, 1996) resulted in increased performance figures and a reduction in time required to achieve course objectives, when compared to conventional delivery techniques. This supports some of the early findings of the Tornado ATF VR simulator and the planning for the Submarine Qualification Training facility. The growing evidence indicates that VR CBT technologies are capable of demonstrating true cost-benefits, in terms of improved human skills and financial returns from improved training efficiency.

As mentioned earlier, whenever technology introduces a new level of functionality into the training domain there are particular Human Factors questions to be answered (see Section 3). These questions concern the way in which technology could and should be used to support the training process and enhance the training experience. Through the combination of an academic and industrial perspective to the understanding of human-centred principles in the training domain, it has been shown that utilizing VE-based training technologies:

- can potentially improve the effectiveness with which knowledge is delivered and assimilated;
- can reduce the reliance on scarce operational systems or costly hardware-based training material;
- may offer enhanced training methods over and above conventional training methods;
- can utilize previous investments in technology with the minimum of bespoke systems procurement.
Other questions are related more to the uptake of advanced technologies in training and their implementation for training in non-military situations. Such questions include: whether students and trainers will actually use such systems; whether they will use them in the way they are intended; and whether there will be a positive transfer of training (or knowledge) from the computerized setting to the real operational environment. The answers to these and similar questions await the necessary research.

Understanding the capabilities and limitations of the individual user and the role of the individual within their organization is essential to the future development of VR as a stable form of computer-based training. For example, it is all too easy to fall into striving for visual excellence at the expense of usability and content, not to mention losing sight of the wider market needs of the user organisation. As Stone (1997) argues, “the drive for visual impact appears to have over-shadowed the crucial issue of concentrating on the underpinning human factors issues surrounding the need for sophisticated 3D display formats ...”.

With the aid of a human-centred approach, such as ISO 13407, some of these issues can potentially be dealt with in a systematic fashion by adhering to the guidelines. However, as Stewart (1998) states, while “such a standard is an attempt to solve the problem of developing ergonomics standards quickly enough in a fast changing technical environment, … many people find them difficult to understand”. This highlights a problem with interpreting standards and applying them across different domains; but, as Stewart concludes, standards “often represent important constraints and may give some guidance for what has worked in the past”. As a result, gaps in understanding may only be closed when more training applications are considered, or those that are being developed are understood in more detail.

7.2. CONCLUSIONS

Understanding the capabilities and limitations of individuals and how these translate into performance issues within the training domain is essential to the future development of the VE as a form of CBT.

Future lean-manned systems are likely to require operators to carry out multiple tasks in parallel, and therefore further research must be conducted to identify the task characteristics for which this type of support is likely to be of most benefit. There are other research issues associated with training feedback, for example the most appropriate type of feedback for different levels of trainee expertise, and when the feedback should be faded to encourage learning (Cook et al., 1999).

AR and VR are, first and foremost, a suite of technologies which provide the Ergonomics and Human Factors community with a “toolkit” for optimizing the design of the human–system interface for numerous applications. Ergonomics and Applied Psychology has a significant contribution to make to the development of VEs into this Millennium, not just as a means of alerting users to negative issues, such as the potential side effects of “immersion”, but in the development of methods to measure and report the positive effects of applying this exciting human-centred technology throughout industry. ISO 13407 offers a potential framework for integrating the “best practice” Human Factors methods in addressing issues such as these and identifying the best applications of AR-AET and VR-CBT.
As concluded by one of the authors in a recent encyclopaedia (Stone, 2001a), Ergonomics is often defined as the study of the relationship between the individual and their working environment. It should make no difference whatsoever if that working environment is real, virtual or a synthesis of the two.

References


MILLER, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. The Psychological Review, 63, 81–97.


