IMPROVING PERFORMANCE MEASUREMENT WITH VIRTUAL ENVIRONMENT TECHNOLOGY: VIRTUAL METRICS

Elizabeth Sheldon, Ph.D.; Sonalysts, Inc; Orlando, Florida
LT Joseph Cohn, Ph.D.; Naval Research Lab; Washington, DC
Scott Casey, Janice Giebenrath, John Burns, Ph.D.; Sonalysts, Inc.; Orlando, Florida
Robert Breaux, Ph.D; Naval Air Systems Command Training Systems Division; Orlando, Florida

Abstract

Virtual Environment (VE) training systems are fast becoming the tool of choice for supporting a wide array of military training needs. As the technology advances, this trend will become even stronger, with these systems ultimately competing with more traditionally accepted methods of training for precedence in assorted curricula. One area in which this faith in VE systems is warranted is performance measurement. In particular, the data collection capabilities afforded by VEs readily lend themselves to the implementation of new, advanced measures of performance that were not possible to develop in the operational environment. The critical challenge that developers now face is identifying and validating these Virtual Metrics and determining how best to include them in the ever-expanding array of evaluation tools available to the operational training community.

The current paper explores this process, using the research, development, and evaluation phases of the Conning Officer Virtual Environment (COVE) shiphandling training system as a case study. The process to be described progressed in three stages, each of which resulted in the testing and evaluation of a novel piece of technology, which ultimately led to Virtual Metrics, Alongside Box and Command Efficiency, for evaluating shiphandling performance. In the first stages, the data collection capabilities of the COVE testbed enabled the development of an objective quantification of basic UNREP performance, which is traditionally subjective in nature, using "track histories". In the final stages, the evaluation team used these basic measures to develop the more sophisticated Alongside Box and Command Efficiency measures, which serve collective variables for quantifying shiphandling performance. Finally, these Virtual Metrics were implemented to quantify the transfer of skills learned in the COVE system to the real world UNREP task.

Biographies

Dr. Elizabeth Sheldon is a Modeling and Simulation Engineer at Sonalysts, Inc. Dr. Sheldon earned a Ph.D. in Industrial Engineering and Management Systems at the University of Central Florida. Prior experience includes research positions at Institute for Simulation and Training and Naval Air Warfare Center Training Systems Division.

LT Joseph Cohn, Ph.D. is an Aerospace Experimental Psychologist, currently serving as Lead, Training Requirements and Evaluation, Virtual Reality Technologies, Naval Research Laboratory. He holds a Ph.D. in Neuroscience from Brandeis University and completed a Post-Doc at Florida Atlantic University's Center for Complex Systems.

Mr. Scott Casey focuses on R&D of PC-based simulation systems for training. His research has involved the investigation into the application of fuzzy logic and Data Driven Knowledge Engineering to understanding task performance.

Ms. Janice Giebenrath is a Research Analyst at Sonalysts, Inc. conducting research involving the application of virtual environment technology to the development of training programs.

Dr. John Burns is a senior scientist at Sonalysts, Inc. He received his Ph.D. in psychology from the University of Massachusetts in Amherst and has 9 years of experience in team training.

Dr. Robert Breaux is a Research Psychologist at the Naval Air Systems Command Training Systems Division, and he is currently working Network Centric Warfare and Human Performance Measurement programs.

IMPROVING PERFORMANCE MEASUREMENT WITH VIRTUAL ENVIRONMENT TECHNOLOGY: VIRTUAL METRICS

Elizabeth Sheldon, Ph.D.; Sonalysts, Inc; Orlando, Florida
LT Joseph Cohn, Ph.D.; Naval Research Lab; Washington, DC
Scott Casey, Janice Giebenrath, John Burns, Ph.D.; Sonalysts, Inc.; Orlando, Florida
Robert Breaux, Ph.D.; Naval Air Systems Command Training Systems Division; Orlando, Florida

1. INTRODUCTION

Virtual environments (VE) are becoming a popular platform for developing training systems, especially in the military training community. Factors contributing to this popularity include the ability to manipulate and control the training environment, the opportunity to train on dangerous tasks without the risks and costs of live training, and the reduced wear and tear on real vehicle platforms (Cohn, Burns, Helmick & Meyers, 2000). A less vaunted benefit of using VE systems is their ability to support the collection of quantitative data on all aspects of the user's execution of a task, which can lead to a better understanding of how to improve task performance. It is this latter application that serves as the focus of this paper.

One of the greatest difficulties of conducting real world performance assessment lies in the fact that the 'real world' laboratory contains many confounds. In the real world, it is simply not possible to safely and effectively evaluate the impact of high-g turning maneuvers on a pilot's motor control, while in the laboratory environment, it is relatively simple to measure these effects while spinning a subject in a rotating environment (Cohn, Dizio & Lackner, 2000). For this reason, researchers typically focus on abstracting essential components from a given real world task to a specific in vitro laboratory application. The hope is that the information gleaned from this minimalist approach will provide generalizable guidance for training.

Clearly, training research must advance beyond this basic level of abstraction to a more all encompassing, albeit less tidy one, if it is to truly provide much needed insights into developing techniques for enhancing performance. In the past, the solution to this challenge was to evaluate real world training through subjective measures, such as scores determined by an instructor evaluating a student's skills. Yet, by their very nature, these subjective measures fail to capture the essence of the training event (Martin, Sheldon, Kass, Mead, Jones, & Breaux, 1998). Nevertheless, focusing on these 'quick and dirty' measures is often times the only viable means for obtaining any indication of performance. In parallel, the use of advanced technologies, such as VEs, to provide training in areas like shiphandling (McGee, 1997) and helicopter training (Sullivan, Darken, & McLean, 1998) has only increased the pressure to develop performance measures to support instruction (e.g., Campbell, Buff, & Bolton, 2002).

While the introduction of VE systems to the training domain may represent an additional demand on already-overburdened training specialists, it may also provide the conduit through which this pressure may be relieved. VE systems, by virtue of being computer-based, are capable of collecting data on effectively any event and/or object within the environment. This ability gives VE–based training systems the potential to serve as a proving ground for a new type of performance metric, the Virtual Metric.

Our approach for describing how VE can be exploited to support this new and exciting capability is to focus on the development of a single VE training system (Conning Officer Virtual Environment), paying particular attention to identifying metric applications that occurred in phase with the development of the training system. This paper describes the development of Virtual Metrics, using a virtual simulation of a real-world task. The approach taken is to first define a basic set of Metrics in the VE, then validate and further refine these to collective variable-like Virtual Metrics, and finally evaluate these in a real world study.

2. BACKGROUND

2.1. The Need for a New Class of Performance Metric

The military is continually seeking new ways of enhancing performance across a range of tasks, with the tightly coupled requirement to quantify this improvement. Yet often, it is at this stage, the performance assessment stage, that current training approaches fail. There is good reason for this. Much of military training is completed in the operational environment through on the job training with senior crewmembers often providing supporting instruction. Additionally, the primary objective of operational equipment (e.g., radars, navigation equipment, etc.) is to support the execution of military missions, not to support the collection of performance data. Hence, using these systems for the collection of non-mission essential data, which may provide useful performance measures, is not only poorly supported, but often times impractical at best and dangerous at worst.

Training performance is typically evaluated using checklists, Likert scales, or frequency counts (e.g., CINCLANTFLT/CINCPACFLT Instruction 3501.1.A; COMCARAIRWING TWO INSTRUCTION 3740.1D). Aside from this type of evaluation being highly subjective and therefore susceptible to bias, it does not capture important information regarding those attributes of a student's performance that would contribute to an understanding of how to better support and deliver training. For example, a rating on a Likert scale provides a qualitative indication, such as "how well" a student performed a task. Thus, if a student is rated poorly on a Likert scale evaluation of shiphandling, it can only be concluded that something was done incorrectly because information relating to the underlying causes of this poor performance is not captured.

Clearly, there is a need for establishing meaningful performance metrics that convey useful feedback to the trainee and that also provide insight into how training can be better supported. Importantly, these metrics should be context specific, in that generic measures such as "communications are poor" are not as meaningful as specific comments within a given training context, such as "orders to the helm were not issued using standard commands". In order to provide such critiques, however, basic performance variables must be developed.

A typical approach for deriving these variables is to base them on metrics developed using simple, lab-based abstractions of complex, real world tasks and to make the assumption that these *in vitro* measures will capture the essence of the *in vivo* task. An alternative solution is to take advantage of VE technology to create a virtual laboratory that mimics both the complexities of a domain specific task and the interaction with subject matter experts (SMEs) to aid in the development and testing of Virtual Metrics. These Virtual Metrics can then be transitioned and evaluated in the real world arena.

2.2. Virtual Environment Technology

The applicability of VE technology to the training domain has long been recognized, while the role of VE as a virtual laboratory for better understanding human performance has only recently been considered (Cohn & Patrey, 2001). In a sense, this latter role for VE is the inverse of the first and should not require a significant paradigm shift. It does, however, require a dedicated program to identify those real world tasks that are ideally suited for serving as testbeds for developing both training guidelines and Virtual Metrics to validate these guidelines.

2.3. VETT and COVE

Virtual Environment Training Technology (VETT) was an Office of Naval Research (ONR) sponsored project, which was implemented at the Naval Air Warfare Center Training Systems Division (NAWCTSD) in the mid to late 1990's, that focused on research to support the development of an advanced, virtual environment (VE) training system. The VETT project was comprised of five research areas: VETT Enabling Research for the Human Operator (ERHO), Haptic Interface Design and Evaluation, VETT Side Effects Research and Development, Testbed Development and Utilization, and Training Effectiveness Research. The Training Effectiveness (TE) research area provided the background for developing virtual measures of performance. As the Surface Warfare Officer School (SWOS) was the supporting customer for the VETT project, shiphandling was chosen as the domain area, with the Underway Replenishment (UNREP) maneuver selected as the initial task based on an analysis of SWOS training needs (Bradley & Watters, 1998). Ultimately, the VETT project transitioned into a technology demonstration (ONR 6.3.) project that was called Conning Officer Virtual Environment (COVE) and has recently entered the acquisition stage.

For the purposes of this paper, the VE research and development conducted by the VETT and COVE researchers will serve as an example of how VE technology can be used as a tool to develop advanced performance measurement capabilities that can be applied in simulated and live training environments.

3. VETT TESTBED

3.1. Shiphandling and Underway Replenishment

Shiphandling is a complex task that requires fundamental sea-based knowledge, such as ship dynamics, navigation and nautical rules of the road, and seaman's eye, which refers to the ability to dynamically synthesize perceptual input from the environment with an understanding of ship and sea dynamics to control the ship (Crenshaw, 1975; Hays, Castillo, Seaman, & Bradley, 1997). Shiphandling training typically involves a combination of classroom education, non-

VE based simulation training (e.g., mock-up simulators), and on the job training while at sea.

One critically important, yet complex, shiphandling maneuver is the Underway Replenishment (UNREP). UNREP is a difficult shiphandling maneuver that requires a ship driver (the Conning Officer) to issue commands such that his/her ship (known as the receiving ship) matches course and speed alongside a supply ship that is moving at a constant course and speed (see Figure 3.1.1). The evolution is difficult because not only must the receiving ship maintain proper 'station' with the supply ship, but it must do so in spite of the action of the venturri effect, which creates a low-pressure area between the two ships, that, if not properly countered, can draw them together as they maintain the close separation distance necessary to facilitate the transfer of fuel and supplies. To maintain station, which is a pre-determined relative position between the two ships at which point supplies can be transferred, the conning officer must rely on visual judgment of relative distance and motion, a skill (seaman's eye) that is difficult to teach in the classroom. Since UNREP is primarily a perceptual task, UNREP training is well suited to the VE training domain (McGee, 1997).

Figure 3.1.1: Phases of Underway Replenishment, *Left* Approach: Receiving ship starts out 1,000 yds behind supply ship and closes distance; *Center* Slide-in: Receiving ship makes fine adjustments to match supply ship heading, speed and to arrive at correct separation and alignment; *Right* Alongside: Receiving ship maintains 'station' while supplies are transferred.

3.2. Virtual Environment UNREP

The VETT research effort was developed to demonstrate the power and flexibility of VE systems as training tools. A VE UNREP testbed was developed at the start of the project to enable this research. The virtual world was an ocean populated with ships. As the VETT researchers were focusing on the task of the conning officer, the viewpoint from which the scene was presented was on the port bridgewing of a ship, which was modeled as a guided missile cruiser.

The user viewed the environment through a head tracked helmet mounted display (HMD) and issued verbal commands, similar to a conning officer's interaction in the real world. Initially, the researchers manually executed the user's commands (e.g., selecting rudder increments from a pull down menu to steer the ship on the ordered course). As this was a demanding task, the ship often moved in directions other than the ordered course, making it difficult for the SMEs to perform an UNREP, and likewise, for the researchers to observe expert performance of the task. Consequently, the testbed was soon outfitted with an autopilot system, which controlled the heading and speed of the receiving ship (Martin et. al., 1998). Additionally, a prerecorded speech system was developed to supply the bridge and engine room teams' standard replies to the conning officer's commands.

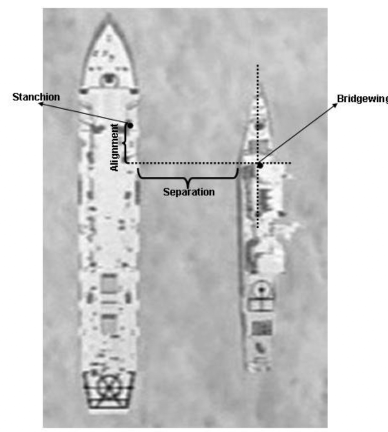

Figure 3.2.1: Coordinate system for UNREP. *Supply ship* is on the left, *receiving ship* is on the right. In the figure, the receiving ship is approaching, but not yet at, the station position. Perfect alignment would be zero, while perfect separation would be whatever the predefined value is (typically between 100-150 ft).

The basic UNREP scenario required the user to drive up alongside the supply ship at a pre-specified position (station), maintain the station position for two-minutes, and then drive his ship forward and away from the supply ship. The receiving ship was positioned 1000 yards directly behind the supply ship at the start of the scenario. The pre-specified position was designated along two axes, centered on the receiving ships' bridgewing, and measured in relation to a pre-designated point on the supply ship. One axis was the separation distance between the two ships, while the other was the alignment between the two ships (see Figure 3.2.1).

The supply ship maintained a constant speed of 15 knots throughout the entire scenario, so the user had to increase his speed during the beginning phase, called the approach, and then decrease his speed to 15 knots as he came alongside the supply ship in order to maintain station for two-minutes. The supply ship maintained a constant course throughout the evolution.

4. Primary Virtual Metrics

4.1. Knowledge Elicitation

Perhaps the most critical piece of the development of any training simulation is the extraction of the relevant pieces of information critical for designing a system that supports training. Specifically, information relating to how the task to be trained is accomplished, what cues from the real world environment support task accomplishment, and what the training goals are must be defined as clearly as possible before further action is undertaken. During this stage, information supporting the initial development of performance metrics was also captured.

4.1.1. Task Analysis. Typically, the first task in the knowledge elicitation (KE) process is a task analysis. There are many types of task analyses, ranging from the cognitive task analyses (CTA; Seamster, Redding & Kaempf, 1997), job analyses (Landy & Vasey, 1991) and behavioral analyses (Jonassen, Hannum & Tessmer, 1999). Which one is optimal depends to a large extent on the type of task being analyzed. For the UNREP task, a CTA approach was selected.

The cognitive task analysis (CTA) for the conning officer utilized traditional CTA methods, including, a literature and document review, SME unstructured and structured interviews, and conceptual graphing. However, shiphandling skills such as seaman's eye, which is the primary skill a conning officer uses to perform shiphandling maneuvers (Hayes, Castillo, Seamon & Bradley, 1997; Norris, 1998), is considered a perceptual-motor task that is difficult to verbalize (Crenshaw, 1975; Hays, et. al., 1997; Martin et. al., 1998). Consequently, the level of information elicited through this approach was not detailed enough to provide clear indications of how an actual UNREP could be evaluated (Martin et. al., 1998). The knowledge obtained with these methods did, however, provide a basic understanding of seaman's eye.

4.1.2. VE-Enabled Knowledge Elicitation. In an effort to obtain more detailed information supporting the development of metrics, a second stage of the CTA used a KE approach termed VE-Enabled KE (Martin et. al., 1998). Specifically, the UNREP simulation was used as a tool for identifying the steps a conning officer takes to conduct an UNREP and potential objective measures of performance for UNREP. This avoided the costs and safety issues associated with field observation of this task. Since the testbed, as originally designed, did not have an automated data collection system, the first VE-Enabled KE trials were recorded on videotape for a review by the researcher and the SMEs, which was similar to a retrospective verbal protocol (Sheldon, 1999). As these metrics became better defined, later versions of the testbed included increasingly sophisticated routines for data collection.

The first VE-Enabled KE trials were conducted with four shiphandling SMEs (two Surface Warfare Officers [SWOs] and two Merchant Mariners). The procedure required each SME to complete a single Virtual UNREP as it would have been done in the operational environment. Afterwards, the researchers conducted a debrief interview as well as a retrospective verbal protocol by reviewing the videotape of the UNREP with the SME.

The results from these trials provided a framework for determining basic performance variables and for defining distinct stages of an UNREP. The specific variables identified were the receiving ship's position (a composite of separation and alignment measures), velocity, and heading. Collectively, these variables can be considered as ship kinematic variables. These ship kinematic variables contributed to an analysis that indicated that the UNREP was a two-stage, higher-order tracking task (Sheldon, 1999). The first stage, which consisted of the approach and slide-in, was described as a pursuit tracking task, in that the goal was to guide a reference point (receiving ship) along a moving track (the path the receiving ship takes from start to station; Adams, 1989). For example, during the approach and slide-in, the shiphandler's goal was to steer his ship on a conceptual course that would bring his ship alongside the supply ship. The second stage, the alongside phase, was viewed as a compensatory tracking task, which is the process of nulling the error between two reference points (receiving ship and station position; Adams, 1989). In the case of UNREP, the two reference points were the shiphandler's position on the receiving ship and the station position, in which the separation distance was 100 feet and the alignment was 0 feet (neither fore nor aft of the alignment point).

The overall UNREP task was considered as a higher order control task since the system input (user's command) and output (latent change in speed or heading of the receiving ship) were not linearly related (Sanders & McCormick, 1987). This non-linearity is injected by the interaction of the ship and the ocean upon which it is moving. For instance, ships have a slow, non-linear response rate due to the effects of inertia and acceleration caused by the extreme weight of the ship and its interaction with the fluid dynamics of water. SMEs referred to this phenomenon during interviews and reported that course and speed changes must be planned well in advance to account for the sluggish responsiveness of operational vessels. This means that expert shiphandlers internalize an understanding of ship dynamics (position, velocity, acceleration, heading) in order to plan their course and speed changes in advance.

4.2 Primary Performance Measure Validation

Before developing higher-order metrics, it was critical to demonstrate that these basic measures (ship dynamics) adequately captured shiphandling performance. Consequently, the validation effort focused on regressing and correlating these measures with performance. Prior to this effort, enhancements to the VE UNREP scenario were made.

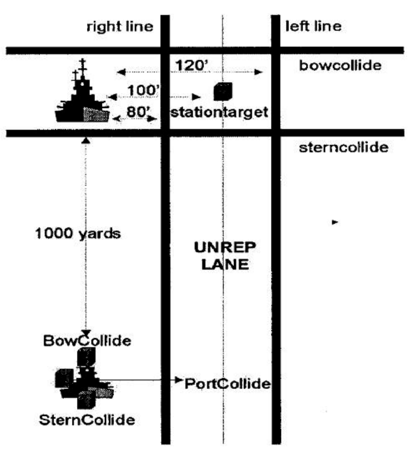

Figure 4.2.1.1. Initial UNREP Performance Metrics.

4.2.1. System Enhancements. The iterative process used by the VETT researchers enabled them to construct an environment that permitted the researchers to collect data in a controlled environment, while enabling observation of SME performance in a naturalistic setting, and providing a system for integrating "lessons learned" into successive design cycles. This upfront interaction with the SMEs allowed the researchers to update the VE scenario with task essential environmental cues, such as the inclusion of a gun turret on the bow of the ship that was reported by SMEs as reference in determining the relative position of the supply ship with respect to ownship.

Additionally, the interactions with the SMEs indicated that UNREP performance was related to the shiphandler's ability to control the ship on the separation and alignment axes described previously. Initial UNREP performance metrics (Figure 4.2.1.1.), specifically an "UNREP Lane" of acceptable separation and stationtarget (ideal separation and alignment alongside position), were proposed. This lead to the development of an automated data collection capability to enable quantification of these measures, while providing the flexibility to explore Virtual Metrics not yet proposed. This system implemented bounding box and collision detection features to determine when the user had obtained key positions in the scenario. Data were collected at one-second variables and included: position (e.g., heading, speed, and bearing), the relative distances between the ships, and the verbal communications made by the participant.

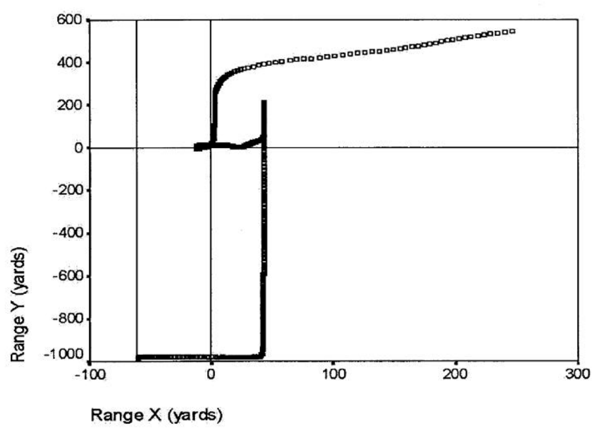

Figure 4.2.2.1. Individual UNREP Track History.

4.2.2. Pilot VE-Enabled UNREP Performance Study. The primary goal of this effort was to explore, in a pilot study, the validity of the virtually derived metrics as well as the experimental protocol. Eleven SWOs of varying experience levels participated in this study. The participants each completed one VE UNREP, prior to which they received a pre-brief of the UNREP scenario described previously. Scatterplots of the separation and alignment axes were used to plot two-dimensional track histories (Figure 4.2.2.1.) that provided a visual representation of the participant's performance. The tracks were also graphed as a composite (Figure 4.2.2.2.) for the purpose of a visual analysis.

The inspection of the composite track histories indicated that there was a great deal of variability in UNREP performance. As the participants only performed one UNREP trial, this variability was considered a function of the participants becoming acquainted with the VE, i.e. a learning or familiarization effect (Sheldon, 1999).

As a result of this finding, the experimental protocol for the formal study was revised to include two UNREPs per trainee, with the first UNREP considered as a practice run (Sheldon, 1999).

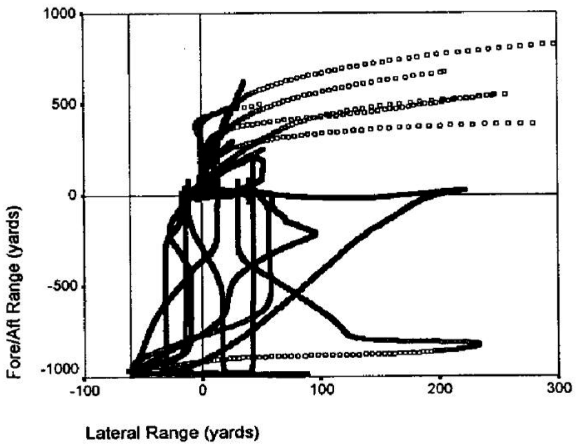

Figure 4.2.2.2. Composite UNREP Track History.

4.2.3. VE-Enabled UNREP Performance Studies. The first formal study was conducted at the SWOS in Newport, Rhode Island with the same system used in the pilot study. The participants were eighteen SWOs consisting of students and SWOS instructors. The participants followed the revised experimental protocol. Track histories were generated for each of the participant's two UNREP performances. These histories were then randomized and distributed to six experienced conning officers, who rated each of the track histories on a scale from 0 to 100 ("performance scores"), with 100 being a perfect UNREP. The inter-rater reliability of the officers' rating with Cronbach's alpha was reported as a=.96 (Martin, Davidson, & Jones, 1998).

The mean of the performance scores was used in an analysis that explored the relationship between UNREP performance and the receiving ship's dynamics during the UNREP. Specifically, measures of position, velocity, and acceleration were regressed against the mean performance scores. The resultant stepwise regression model (Martin, Davidson, & Jones, 1998) reliably predicted the UNREP performance ratings ($p<.001$).

A second experiment was conducted at SWOS to provide additional data and permit more extensive analyses. This time, twenty-six SWOs participated in the experiment. The procedures used in the first experiment were used in this second experiment with a few minor changes as follows. As with the first experiment, the participants completed two VE UNREPs. However, the first trial was considered as a warm-up trial, and consequently, only the second trial was used in the analysis (Patrey, Sheldon, Breaux, & Mead, 1998). As with the prior experiment, performance scores were obtained by having SMEs rate the track histories. Due to technical difficulties with the VE data collection system, the data for eight of the second trial UNREPs were unusable, which reduced the sample size for the analysis to eighteen.

Based on continued interactions with SMEs between the first and second experiments, the researchers considered the heading of the receiving ship in addition to its position, velocity, and acceleration in their analysis. The cumulative effects of these characteristics at a specific point in time (Patrey et. al., 1998) were described as thirteen "moments", which were defined by the receiving ship's range from the supply ship. Specifically, eleven of the moments were defined in terms of 100-foot increments, starting from a position of 1000 yards astern of the supply ship. The other two moments were respectively defined at a position of 84 feet and 50 feet astern of the station position, and these points signified the start (moment 84) and finish (moment 50) of the receiving ship's bow crossing the supply ship's stern. The four moment characteristics (heading, position, acceleration, and velocity) were regressed against the mean performance scores for each of the thirteen moments. The model for moment 84 (Patrey et. al., 1998) reliably predicted the performance ratings ($p<.001$), and each of the four ship dynamics characteristics were significant in this model.

These two experiments provided evidence of a relationship between quantifiable metrics of the receiving ship's dynamics and subjective assessment of UNREP performance. Specifically, heading, position, velocity, and acceleration were considered as quantifiable predictors of UNREP performance.

5. Virtual Metric Refinement

The research conducted with the VETT UNREP testbed was transitioned to a technology demonstration (e.g., prototype) program (Conning Officer Virtual Environment [COVE]). As part of the transition, further refinements to these metrics involved moving away from basing them on subjective scoring for performance and, instead, developing more objective, instructive metrics. Thus, the conclusions made by the VETT researchers regarding the link between UNREP performance and quantitative ship dynamics were refined and consolidated into two metrics. These metrics were Alongside Box, which provides an objective collective variable type performance measure, and Command Efficiency, which is based on this objective measure, but adds a level of feedback.

5.1. Alongside Box

Early efforts at integrating the basic Virtual Metrics into a single, collective variable included constructing an acceptable "UNREP Lane" and "stationtarget". Together, these notions were combined to form the first Virtual Metric, called the Alongside Box. This metric is defined by an acceptable range of lateral separation and vertical alignment for the alongside phase (Figure 5.1.1).

In a study designed to guide the selection of the boundary ranges for this box, Burns and Patrey (2000) based UNREP alongside performance on the lateral separation and vertical alignment maintained during the alongside phase. These measures were related to the UNREP task instructions to maintain 120 feet of horizontal separation and to keep the receiving ship in line with the supply ship for the alongside phase.

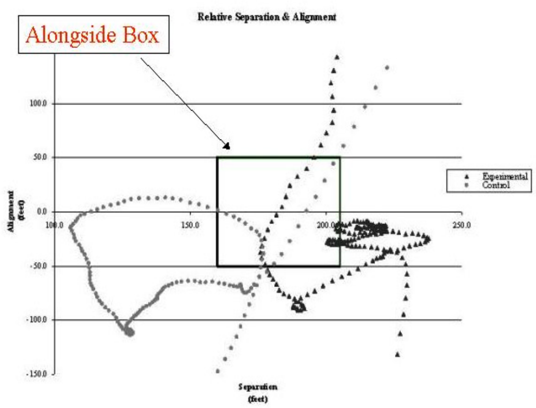

Figure 5.1.1. Alongside Box Virtual Metric.

An assumption was made, based on interviews and prior VETT UNREP performance data, that suggested the participants would deviate around the ideal station-keeping position (Burns & Patrey, 2000). To account for this, they used root mean square (RMS) calculations of the difference from the station position (120 feet) and the exact point of the receiving ship at a given time throughout the scenario. Burns and Patrey (2000) reported that RMS was used as it provided a measure of the magnitude of "error" (e.g., the difference from the optimal distance of 120 feet lateral separation) irrespective of whether or not the cruiser was too close or too far away from the supply ship.

In a second study, a different group of participants performed UNREP maneuvers in the VETT testbed, following procedures similar to those used in the previous performance studies. For this study, objective measures were defined according to explicit instructions that were provided to the participant that constrained the manner in which the UNREP was to be performed, which included the time for task completion. The rationale for this approach was that constraining the participants to perform the UNREP with the use of a single strategy would reduce the variability in the performance data, which was a result of the diverse UNREP performance strategies employed by participants in the previous experiments (Sheldon et. al., 2000). These objective instructions also served as a starting point for an objective evaluation of the participants' performance.

Of relevance to the development of the Alongside Box was the explicit instruction on which the objective performance measures were based that directed the participants (Sheldon et. al., 2000) to maintain an alongside separation of +/- 20 feet of the station position while simultaneously maintaining a vertical alignment position of +/- 20 feet of the station position. The alongside separation and alignment criterion defined a "box" of alongside phase acceptable performance, hence the name "Alongside Box".

5.2. Command Efficiency

The Alongside Box provides an independent method for assessing performance that, while based on subjective measures, provides a quantitative measure of performance. However, the UNREP maneuver is primarily perceptual in nature. Consequently, while ship kinematics and the more formal Alongside Box measure represent one set of measures by which performance can be judged, a second, potentially deeper measure concerns the relationship between the commands given and the point within the UNREP that supplies are transferred as well as the overall number and types of commands issued (Cohn et. al., 2000). Essentially, this degree of information provides a means for probing how well the trainees were able to develop a mental model of ship dynamics. Thus, the Command Efficiency variable builds on the refinement of the Alongside Box by introducing means for relating ship kinematics with timing and number of commands.

5.2.1. Command Extrapolation.

The process for evaluating Command Efficiency involves correlating the commands given against the ship's current and projected position (based on advance and transfer tables). To do this, "snapshots" of the ship's position (refers to the participant's ship) were taken between commands given. The extrapolated position of the ship (Figure 5.2.1.) was calculated based on the command executed. If the extrapolated ship position brought it within the Alongside Box, then the command was rated as "good" (efficient and appropriate). Otherwise, the command was considered a bad one (not efficient and inappropriate). If an additional command was required to bring the ship into the Alongside Box, it was considered an extraneous command, which was considered inefficient and lowered the Command Efficiency rating (Sonalysts, 2002).

Use of this method allowed the effectiveness of individual commands to be measured throughout an evolution, rather than simply focusing on the final outcome of a series of commands. As a performance feedback tool, this measure can provide students insight regarding the ineffective use of specific commands at various points of an evolution.

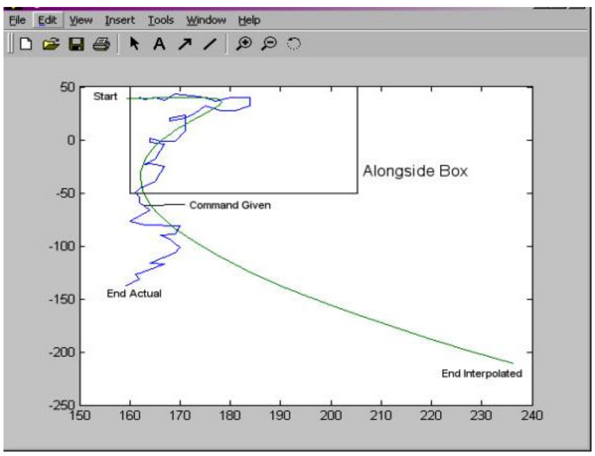

Figure 5.2.1. Command Efficiency.

6. VIRTUAL METRIC EVALUATION

An important aspect of any training system is the ability to support training transfer to the real world domain. In order to test the utility of these two Virtual Metrics and to refine the methodologies for use in a full scale training transfer study, a pilot study (Sonalysts, 2002) was conducted in which these virtually developed metrics were applied to real world shiphandling data.

6.1. Methodology

The study utilized a between subjects design with one independent variable (COVE training) with two levels (training, no training) by three dependent variables (separation, alignment, and Command Efficiency). The control group was the "no training group" because the participants in this group received no training on the COVE system prior to performing a real world "Leapfrog" shiphandling task, as did the training group participants. Both groups performed the Leapfrog task, which is comparable to UNREP, in which two ships take turns conducting simulated UNREPs, without transferring supplies. The participants in the "training group" received three, thirty-minute trials (total of ninety-minutes) of COVE training prior to their performance of the Leapfrog task. Data was only recorded during performance of the real world trials.

During the training transfer study, position data for the two ships was collected with global positioning system (GPS) hand-held devices that transmitted the GPS data to a laptop (one laptop per GPS) via a data cable. A software program enabled the position data to be recorded. The experimenters manually recorded the participants' commands. Data were reduced to provide measures of separation and alignment (Alongside Box) and Command Efficiency.

6.2. Results

The study was conducted for the primary purpose of testing the methodologies and refining the process. The rationale for providing the data in this paper is to demonstrate the feasibility of the analyses planned for the full experiment. Therefore, the separation and alignment performance data (Alongside Box) are only reported for one member of the "training group" and one member of the "no training group". For the Command Efficiency evaluation, data is presented for two participants from the "no training group" because the command data from the "training group" were incomplete (poor audio recording of the officers giving the command due to high wind).

6.2.1. Alongside Box. The Alongside Box was defined as + 25 feet/- 20 feet of the station position (separation) by +/- 50 feet of the station position (alignment). The "training group" participant was in the Alongside Box for 17.4% of the time spent in the alongside phase (6 minutes and 20 seconds), and the "no training group" participant was in the Alongside Box for 10.9% of the time spent in the alongside phase (3 minutes and 40 seconds).

6.2.2. Command Efficiency. The Command Efficiency scores for the two participants from the "no training" group were 1.36 (out of 10) and 1.29 (out of 10) (for a detailed description of the Command Efficiency computations, refer to Sonalysts, 2002). The participant who received a Command Efficiency score of 1.36 went outside of the Alongside Box one time, but was only inside of the box 17% of the time (total alongside time was 402 seconds). This participant issued eight commands: one was determined to be Good, none were Inefficient, and seven were Bad. The participant who scored 1.29 also went outside of the

Alongside Box one time, but was only inside of the box 10% of the time. This participant issued seven commands: one was determined to be Good, one was Inefficient, and five were Bad.

6.3. Implications

As stated previously, the primary purpose for conducting this pilot study was to verify the implementation of the procedures and methodologies planned for the formal training transfer study. As such, data analysis was not a focus of the pilot study and would not have been possible due to the small number of participants. Instead, the limited data that were collected were used to perform the analysis procedures to ensure that these measures would be appropriate for the training transfer study. This "walkthrough" of the data collection and analysis procedures demonstrated that these procedures produced the intended results.

7. CONCLUSION

In the first stage, which refers to the KE and performance studies, the data collection capabilities of the VETT UNREP testbed enabled the development of an objective quantification of basic UNREP performance based on the ships dynamics, specifically the concept of the Alongside Box, which was assisted by the development of the track history tool. The second stage (COVE transition) involved the application of these basic measures to the construction of a more sophisticated Command Efficiency measure, which serves as a collective variable for quantifying shiphandling performance. Finally, a real world training transfer study was undertaken to verify that the data collection procedures and performance evaluation methodologies were feasible for real world use.

An important aspect of this approach was the iterative process in which the researchers updated their methodologies and tools (e.g., the testbed) to reflect their continuous interactions with SMEs. For instance, the realistic fidelity that is easily achieved with VE technologies engages the SMEs in the research process and enables the researcher to identify the aspects of the real world environment that are critical to the training process to promote positive transfer to the operational environment.

Another equally important part of this approach was the involvement of an integrated, multi-disciplinary team (e.g., psychologists, engineers, computer scientists) in the research process. The frequent interactions of this integrated team enabled the lessons learned from each research task to be readily transitioned to the virtual data collection tool (VETT UNREP Testbed).

8. FUTURE WORK

The training transfer pilot study provided evidence, which verified that the Virtual Metrics were usable in the operational setting. The qualitative nature of these Virtual Metrics provides an objective means of diagnosing UNREP performance in addition to evaluating the quality of performance. As such, these types of quantitative measures can be incorporated in an on-the-job training environment to enhance training effectiveness.

The next step is the validation of the Virtual Metrics to obtain evidence that Alongside Box and Command Efficiency do indeed provide an accurate assessment of UNREP performance. Completion of the formal training transfer experiment would yield data required to support the validation of the UNREP Virtual Metrics. One approach to the validation is to compare the results (e.g., Alongside-Box and Command Efficiency) to the measures currently used in the real world.

It is reasonable that the Virtual Metric development process described in this paper can be adapted to a variety of perceptual tasks. For example, a very similar approach can be applied to the development of Virtual Metrics for the LCAC team in support of the VELCAC mission rehearsal system being developed as part of the Office of Naval Research's (ONR's) Virtual Environment and Technologies (VIRTE) program.

On a broader note, the work described here provides substantive evidence for the claim that VE technology is more than a souped up delivery medium. In the current effort, VE technology was used to present performance data in a novel graphical format, which in turn enabled more comprehensive knowledge elicitation from SMEs. VE technology also enabled the development of the Alongside Box and Command Efficiency quantitative metrics.

VE technology is a powerful and flexible medium for training system delivery. In the hands of creative developers, VE technology can also be a powerful medium for understanding performance and for iterative training system development.

9. REFERENCES

Adams, J.A. (1989). *Human Factors Engineering.* New York: Macmillan.

Bradley, S. & Watters, R. (1998). Examination of Surface Warfare Officers School (SWOS) (Technical Report 98-014). Orlando, FL: NAWCTSD.

Breaux, R., Martin, M., & Jones, S. (1998). Cognitive precursors to virtual reality applications. *Proceeding of SimTecT '98*, 21 – 26.

Burns, J. & Patrey, J. (2000). *A comparison of two interfaces for a VE-based training system.* Unpublished special report. Orlando, FL: NAWCTSD.

Campbell, G., Buff, W., & Bolton, A. (2002). Traditional mathematical modeling applied to human prioritization judgments: Quicksand for the unwary modeler. In, *Proceedings to the 11th Computer Generated Forces and Behavior Representation Conference*.

CINCLANTFLT /CINCPACFLT Instruction 3501.1.A Tactical Training Manual

Cohn, J.V., DiZio, P., & Lackner, J.R. (2000). Reaching during virtual rotation: Context specific compensations for expected coriolis forces. *Journal of Neurophysiology*, 83(6), 3230-3240.

Cohn, J.V., Helmick, J., Meyers, C. & Burns, J. (2000). Training-transfer guidelines for virtual environments (VE). *22nd Interservice/Industry Training, Simulation, and Education Conference Proceedings.*

Cohn, J.V. & Patrey, J. (2001). Virtual Environments as a multi-modal "real world" laboratory for training. *45th Annual Human Factors and Ergonomics Society Conference Proceedings*, MN.

COMCARAIRWING TWO INSTRUCTION 3740.1D (Training Qualification and Designation of Carrier Air Wing Two Integrated Warfare Mission Commander)

Crenshaw, R.S., Jr., CAPT, U.S. Navy (Retired). (1975). *Naval Shiphandling*. Annapolis, MD: Naval Institute Press.

Hays, R.T, Castillo, E.R., Seamon, A.G., & Bradley, S.K (1997, April). A virtual environment for submarine ship handling: perceptual and hardware trade-offs. In M.J. Chinni (Ed), *Proceedings of the 1997 Simulation MultiConference: Military Government, and Aerospace Simulations. Simulation Series Vol. 29, No.4*. San Diego, CA: The society for Computer Simulation International.

Jonassen, D.H., Tessmer, M., & Hannum, W.H. (1999). *Task analysis methods for instructional design*. Mahwah, NJ: Erlbaum.

Landy, F.J. & Vasey, J. (1991). Job analysis: The composition of SME samples. *Personnel Psychology,* 44(1), p 27-50.

Magee, L. (1997). Virtual reality simulator (VRS) for training shiphandling skills. Seidel, R (Ed.) & Chatelier, P. (Ed), *Virtual Reality: Training's Future?* (pp. 19 – 29). New York: Plenum Press.

Martin, M.K., Davidson, S., & Jones, S. (1998). *Virtual environments in training system design.* Unpublished manuscript. Orlando, FL: NAWCTSD.

Martin, M.K., Sheldon, E., Kass, S., Mead, A., Jones, S., & Breaux, R. (1998). Using a virtual environment to elicit shiphandling knowledge. *20th Interservice/Industry Training, Simulation, and Education Conference Proceedings*, 678-685.

Norris, S.D. (1998). *A task analysis of underway replenishment for virtual environment ship-handling simulator scenario development*. Unpublished Master's Thesis. Monterey, CA: Naval Post Graduate School.

Patrey, J., Sheldon, E., Breaux, R. & Mead, A. (1998). *Quantifying performance of a dynamic shiphandling perceptual-action task from spatial moments*. Unpublished manuscript. Orlando, FL: NAWCTSD.

Sanders, M.S., & McCormick, E.J., (1987). *Human Factors in Engineering and Design.* New York: McGraw-Hill.

Seamster, T.L., Redding, R.E. & Kaempf, G.L. (1997). *Applied cognitive task analysis in aviation*. VT: Ashgate Publishing Co.

Sheldon, E. (1999). *Virtual Environment Training Technology (VETT): Development of a virtual reality underway replenishment (UNREP) training simulator.* Unpublished special report. Orlando, FL: NAWCTSD.

Sheldon, E., Patrey, J., & Breaux, R. (2000, May). Applying virtual environment technology to the development of shiphandling navigation displays. *Proceedings of the International Marine Simulation Conference*.

Sonalysts, Inc. (2002). COVE: Real-world training transfer pilot study (Contract N61339-99-D-0012, DO # 0014, CDRL A009). Orlando, FL: Sonalysts, Inc.

Sullivan, J., Darken, R., & MacLean, T. (1998). Terrain navigation training for helicopter pilots using a virtual environment. *Third Annual Symposium on Situational Awareness in the Tactical Air Environment* (June 2-3, Piney Point, MD).