

## **Automated Performance Assessment of Teams in Virtual Environments**

Peter W. Foltz, Noelle LaVoie, Rob Oberbreckling, and Mark Rosenstein

In Dylan Schmorrow, Joseph Cohn and Denise Nicholson, eds., *The PSI Handbook of Virtual Environments for Training and Education: Developments for the Military and Beyond*. Praeger Security International, 2008.

## Introduction

Multiplayer virtual environments provide an excellent venue for distributed team training. They provide realistic, immersive, engaging situations which can elicit the complex behaviors that encompass teamwork skills. These environments permit trainers the opportunity to target particular skills in order to assess and improve a team's performance in situations that are difficult to create in live environments. In addition, because virtual environments provide fine-tuned control of the training situation and automate the collection of data, training teams in virtual environments can save effort and money when compared to live training. As the military and other large collaborative organizations incorporate greater network-centric methods, operations, tactics and technologies, virtual environments become an essential means to monitor, train and assess teams.

However there are numerous challenges to effectively identify, track, analyze, and report on teams in complex virtual environments. For example, many current methods of assessing team and group performance rely on both global outcome metrics and hand-crafted assessment techniques. These metrics often lack information rich enough to diagnose failures, detect critical incidents, or suggest improvements for the teams for use in their collaborative aids. It is also problematic for these techniques to produce assessments in the near real-time-frame that is necessary for effective training feedback because of the reliance on time-consuming hand-coding. Thus, while there has been an explosive increase in the availability of team information that can be obtained from a virtual environment, there needs to be a concomitant development in tools that can leverage the data to monitor, support and enhance team performance. In this chapter, we discuss the issues of evaluating teams in virtual environments, describe an automated communications-based analysis approach that we have found fruitful in tackling these issues, and finally detail the application and evaluation of this approach in predicting team performance in the context of three task domains.

### Team performance measurement in virtual environments

Complex team virtual environments provide an ideal venue for team training. Orsanu and Salas (1993) identify a number of critical characteristics for training teams, including having interdependent members with defined roles, using multiple information sources and sharing common goals. Because of the inherent automation in virtual environments, they afford better ability to measure performance of teams, both in recording what is being done by the team as well as what is communicated by team members. Nevertheless, while a virtual environment can produce a record of what team members have done and said, there are challenges in converting that information into measures of performance and difficulties in determining how those measures can be used to give feedback.

Team performance can be seen as a combination of taskwork and teamwork. Taskwork, which is the work a team does to accomplish its mission, is often more amenable to automated analysis from a virtual environment event log. For example, a system can provide information on whether a person moved from x to y at time t and whether an objective was completed. Teamwork, on the other hand, encompasses how the team members coordinate with each other. In order to measure teamwork within virtual environments, the critical aspects of teamwork must be identified along with how they can be measured, assessed and trained (e.g., Salas & Cannon-Bowers, 2001). These skills include leadership, monitoring, back-up behavior, coordination, and communication (e.g., Cannon-Bowers, Tannenbaum, Salas & Volpe, 1995; Curtis, Harper-Sciarinin, DiazGranados, Salas & Jentsch, 2008; Freeman, Diedrich, Haimson, Diller and Roberts, 2003; Hussain, Weil, Brunyé, Sidman, Alexander & Ferguson, 2008). Curtis et al., (2008) identifies three teamwork processes that have major impacts on teamwork and which appear to be strongly predictive of team performance: communication, coordination and team leadership. These processes are typically assessed by Subject Matter Experts (SMEs) watching and checking off behaviors associated with the processes. This protocol can be quite time-consuming and is often performed after the exercise is completed rather than in real-time, limiting the ability to incorporate teamwork performance measurement into virtual environments or provide timely feedback to teams. Thus, methods are required

to automatically measure teamwork in an accurate and responsive manner. This chapter focuses on the aspects of communication that can be used to predict performance and how analyses of communications can be automated to provide rapid measurement of teamwork.

### Communication as an indicator of performance

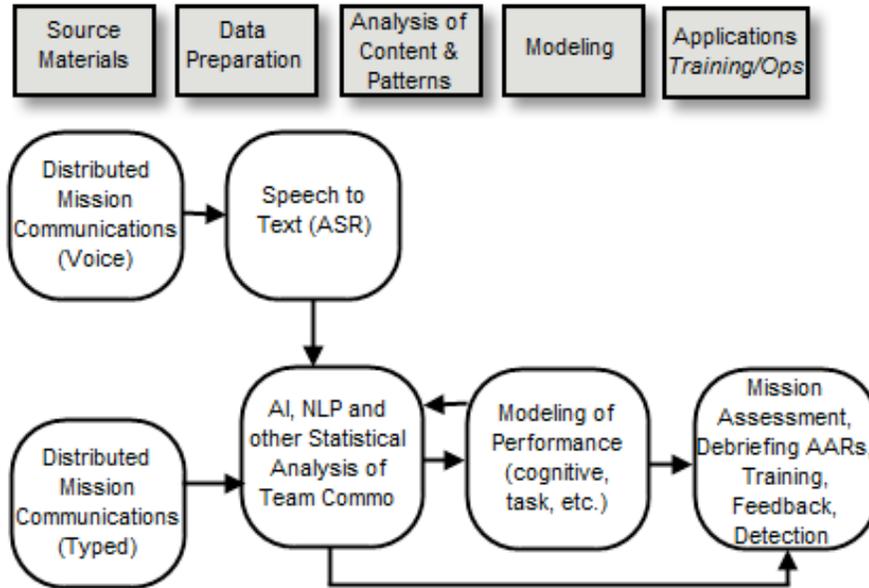
Networked teams in virtual environments provide a rich source of information about their performance through their verbal communication. The communication data contains information both about the actual structure of the network and the flow of meaning through the network over time. The structure and communication patterns of the network can provide indications of team member roles, paths of information flow and levels of connectedness within and across teams. The content of the information communicated provides detailed indications of the information team members know, what they tell others, whom they tell, and their current situation. Thus, communication data provides information about team cognitive states, knowledge, errors, information sharing, coordination, leadership, stress, workload, intent, and situational status. Indeed, within the distributed training community, trainers and Subject Matter Experts typically rely on listening to a team's communication in order to assess that team's performance. Nevertheless, to effectively exploit the communication data, technologies need to be available that can assess both the content and patterns of the verbal information flowing in the network and convert the analyses into results that are usable by teams, instructors and commanders.

In this chapter, we provide an overview of ongoing research and development of a set of tools for the automatic analysis of team verbal communication and discuss their application in measuring team performance in virtual environment training systems. The tools exploit team communication data and use language technologies to analyze the content of communication, thereby permitting characterization of the topics and quality of information being transmitted. To explore these ideas further, we describe how these tools were incorporated into three application environments and the results of their use.

### **Verbal Communication Analysis**

The overall goal of automated verbal communication analysis is to apply a set of computational modeling approaches to verbal communication in order to convert the networked communication into useful characterizations of performance. These characterizations include metrics of team performance, feedback to commanders, or alerts about critical incidents related to performance. This type of analysis has several prerequisites. The first is the availability of sources of verbal communication. Second, there must be performance measures which can be used to associate the communication to standards of actual team performance. These prerequisites can then be combined with computational approaches to perform the analysis. These computational approaches include computational linguistics methods to analyze communication, machine-learning techniques to associate communication to performance measures, and finally cognitive and task modeling techniques.

By applying the computational approaches to the communication, we have a complete communication analysis pipeline as represented in Figure 1. Proceeding through the tools in the pipeline, spoken and written communication are converted directly into performance metrics which can then be incorporated into visualization tools to provide commanders and soldiers with applications such as automatically augmented After Action Reviews (AARs) and briefings, near-real-time alerts of critical incidents, timely feedback to commanders of poorly performing teams, and graphic representations of the type and quality of information flowing within a team. We outline the approach to this communication analysis below.



**Figure 1.** The Communication Analysis pipeline

### Communication data

For analysis purposes, communication data includes most kinds of verbal communication among team members. Typed communication (e.g., chat, email, or instant messages) can be automatically formatted for input into the analysis tools. Audio communication includes the capture of many kinds of spoken data, including use of voice over IP systems, radios, and phones.

Because a majority of communication in virtual environments is typically spoken, two classes of information can be gleaned from the audio stream: content and audio features. Automatic speech recognition systems (ASR) convert speech to text for analysis of content, while audio analysis extracts characteristics such as stress or excitement levels from the audio. ASR systems often also provide measures such as rate-of-speech and ASR uncertainty. All this processed information can be input into the communication analysis system.

### Performance metrics

In order to provide feedback on team performance, the toolset learns to associate team performance metrics with the communication streams from those teams. Thus, the system typically requires one or more metrics of team performance. There are a wide range of issues in determining appropriate metrics for measuring team performance (e.g., Brannick, Salas, & Prince, 1997). For example, metrics need to be associated with key outcomes or processes related to the team's tasks; they should indicate and provide feedback on deficiencies for individuals and/or teams, and they need to be sufficiently reliable so that experts can agree on both the value of the metric and on how it should be scored for different teams (Paris, Salas & Cannon-Bowers, 2001).

Objective measures of performance can be used as metrics to indicate specific aspects of team performance. These measures can include threat eliminations, deviations from optimal solution paths, number of objectives completed, and measures derived from task specific artifacts such as SALUTE and ACE reports. One advantage of computer-based environments is that they are able to automatically track and log events and then generate such objective measures.

Subjective measures of performance can also be used as metrics. These can include Subject Matter Experts' ratings of such aspects as command and control, management of engagement, following

doctrine, communication quality, and situation understanding. Additionally, SME evaluations of AARs and identification of specific critical incidents, failures or errors can be used to measure performance. Care must be taken, as all metrics will have varying levels of reliability as well as validity. For new metrics, it is often advisable to obtain ratings from more than one SME in order to determine reliability.

### Computational modeling tools

Communication data is converted into a computational representation which includes measures of the content (what team members are talking about), quality (how well team members seem to know what they are talking about) and fluency (how well team members are talking about it). This process uses a combination of computational linguistics and machine learning techniques that analyze semantic, syntactic, relational, and statistical features of the communication streams.

While we will discuss a number of tools, the primary underlying technology used in this analysis is a method for mimicking human understanding of the meaning of natural language called Latent Semantic Analysis (LSA) (see Landauer, Foltz & Laham; 1998, for an overview of the technology). LSA is automatically trained on a body of text containing knowledge of a domain, for example a set of training manuals, and/or domain relevant verbal communication. After such training, LSA is able to measure the degree of similarity of meaning between two communication utterances in a way that closely mimics human judgments. This capability can be used to understand the verbal interactions in much the same way a Subject Matter Expert compares the performance of one team or individual to others. The technique has been widely used in other machine understanding applications, including commercial search engines, automated scoring of essay exams, and methods for modeling human language acquisition.

The results from the LSA analysis are combined with other computational language technologies including techniques to measure syntactic complexity, patterns of interaction and coherence among team members, audio features, and statistical features of individual and team language (see Jurafsky & Martin, 2000). The computational representation of the team language is then combined with machine-learning technology to predict the team performance metrics. In a sense, the overall method learns which features of team communication are associated with different metrics of team performance and then predicts scores for new sets of communication data employing those features.

### Performance prediction with the communication analysis toolkit

Tests of the toolkit's use for communication analysis have shown great promise. Tests are performed by training the system on one set of communication data and then testing its prediction performance on a new data set. This procedure tests that the models generalize to new communication. Over a range of communication types, the toolkit is able to provide accurate predictions of the overall team performance and individual team metrics. It makes reliable judgments of the type of statements each team member is making and it can predict team performance problems based on the patterns of communication among team members (Foltz, 2005; Gorman, Foltz, Kiekel, Martin & Cooke, 2003). In addition to the approaches described above, there have been other approaches used to analyze communication in teams that have shown great promise. These have included modeling communication flow patterns to predict team performance and cognitive states (see Gorman, Weil, Cooke, & Duran, 2007; Kiekel, Gorman & Cooke, 2004).

The communication analysis toolkit has been tested in many environments including an Unmanned Aerial Vehicle (UAV) synthetic task environment (see Gorman et al., 2003; Foltz, Martin, Abdelali, Rosenstein & Oberbreckling, 2006), in Air Force simulators of F-16 missions (Foltz, Laham & Derr, 2003; Foltz et al., 2006), and in Navy Tactical Decision-Making Under Stress (TADMUS) exercises (Foltz et al., 2006). The tools predicted both objective team performance scores and SME ratings of performance at very high levels of reliability (correlations ranged from  $r=0.5$  to  $r=0.9$  over 20 tasks). It should be noted that the agreement between the toolkit's predictions and SMEs is typically within the range of one SME to another. In addition, the tools are able to characterize the type of communication for individual utterances, (e.g., planning, stating facts, acknowledging) (Foltz et al., 2006).

### Issues and current limitations of this approach

While the next section describes successful applications of this approach, there are a number of issues and current limitations. For verbal communication, this approach requires automatic speech recognition, and that technology currently has a number of limitations. The state-of-the-art requires building acoustic models and speaker independent, but task specific models, currently requiring about 20 hours of speech to train the ASR system, which increases the startup time for a new task domain.

The second prerequisite of the approach is performance measures. If objective measures are available, than as soon as the ASR is available, teams can begin to execute the task, communication data and performance data can be collected, and then a performance model can be built. If expert ratings are preferred, than protocols for scoring communications need to be developed and then SMEs must score a set of missions to be used as a training set, though this limitation will impact any approach that uses experts.

Besides these startup costs, there is also the issue of the accuracy of the ASR. The communication analysis technologies have been tested with ASR input from a number of datasets of spoken communication (see Foltz, Laham & Derr; 2003). The results indicate that even with typical ASR systems degrading word recognition by 40%, the model prediction performance degraded less than 10%. Thus, the approach appears to be quite robust to typical ASR errors.

### **Applications of the Communication Analysis Toolkit in Virtual Environments**

A number of applications have been developed to test the performance and validate the use of the toolkit in virtual and live training situations. Below we describe three applications, one monitoring and assessing learning in online discussion environments, another providing real-time analyses and visualizations of multi-national Stability and Support Operation (SASO) simulation exercises, and the third providing automated team performance metrics and detection of critical incidents in both convoy operations in simulators and in live training environments. These applications cover the range of immersion in virtual environments. At one end are collaborative discussion environments, which permit use and evaluation of the planning, communication, and coordination aspects of teams, but do not providing the full immersive qualities of a simulator environment. At the other end, are virtual convoy environments and similar live training environments where the approach is tested as teams move from virtual to real-world training and operations.

### Knowledge Post

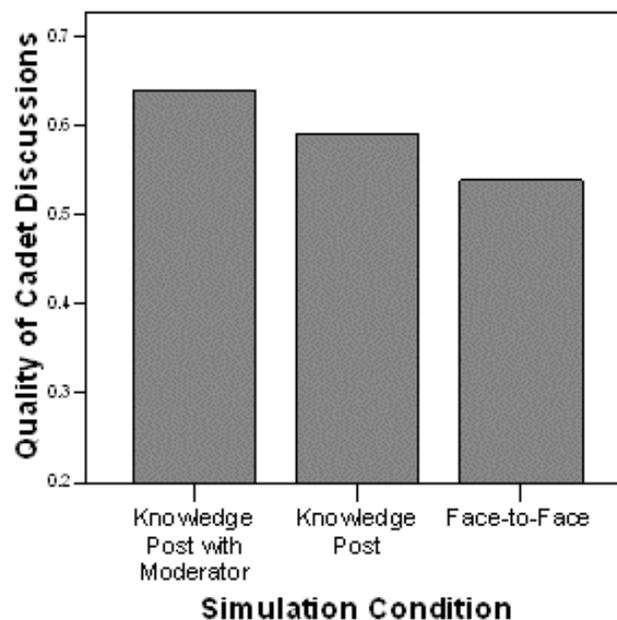
In large networked organizations, it is difficult to track performance in distributed exercises. Knowledge Post is designed for monitoring, moderating and assessing collaborative learning and planning. The tools within Knowledge Post have been tested in a series of studies at the U.S. Army War College and the U.S. Air Force Academy (LaVoie, Psootka, Lochbaum & Krupnick, 2004; LaVoie, Streeter, Lochbaum, Wroblewski, Boyce & Krupnick, in press; Lochbaum, Streeter & Psootka, 2002). The application consists of an off-the-shelf threaded discussion group that has been substantially augmented with Latent Semantic Analysis-based functionality to evaluate and support individual and team contributions in simulated planning operations.

Knowledge Post supports the abilities:

- To automatically notify the instructor when the discussion goes off track.
- To enhance the overall quality of the discussion and consequent learning level by having expert comments or library articles automatically interjected into the discussion at appropriate places.
- To locate material in the discussion or electronic library similar in meaning to a given posting.
- To automatically summarize contributions.
- To assess the quality of contributions made by individuals and groups.

The utility of each of the aforementioned functions were empirically evaluated with senior officers, either in research sessions or participating in distributed learning activities at the U. S. Army War

College, or with cadets at the U. S. Air Force Academy. Among the findings of the studies were the superiority of learning in a Knowledge Post environment over a face-to-face discussion with significantly improved quality of discussion and the usefulness to the participants of the Knowledge Post searching and summarizing features (Lochbaum, Streeter & Psotka, 2002). The research conducted with the Army War College established the usefulness and accuracy of a software agent that automatically alerts moderators when groups and individuals are floundering, by identifying on- and off- topic comments in a discussion (LaVoie et al., 2004). A human rater coded over one thousand comments as either on- or off-topic. A second rater coded a random 10% of these comments. The correlation between the two raters for this task was  $r(162) = .85, p < .001$ , while the correlation between the LSA-based model and one human rater was  $r(1605) = 0.72, p < .001$ , showing that the model was able to accurately determine when a group's discussion was off-topic. The work with the Air Force Academy demonstrated improved solution quality of a group of cadets as a result of exposure to automatically interjected relevant expert comments (LaVoie et al. in press). Cadets participated in a discussion of a challenging leadership scenario. The discussion was conducted in one of three ways: 1) face to face in a classroom with a live human moderator, 2) in Knowledge Post with an automated moderator that added relevant comments from experts, or 3) in Knowledge Post without the automated moderator. The quality of the discussions was evaluated by using LSA to determine the similarity of the cadets' discussion to that of senior military officers, and the highest quality discussions were found for groups which used Knowledge Post with the automated moderator to conduct their discussions (see Figure 2).



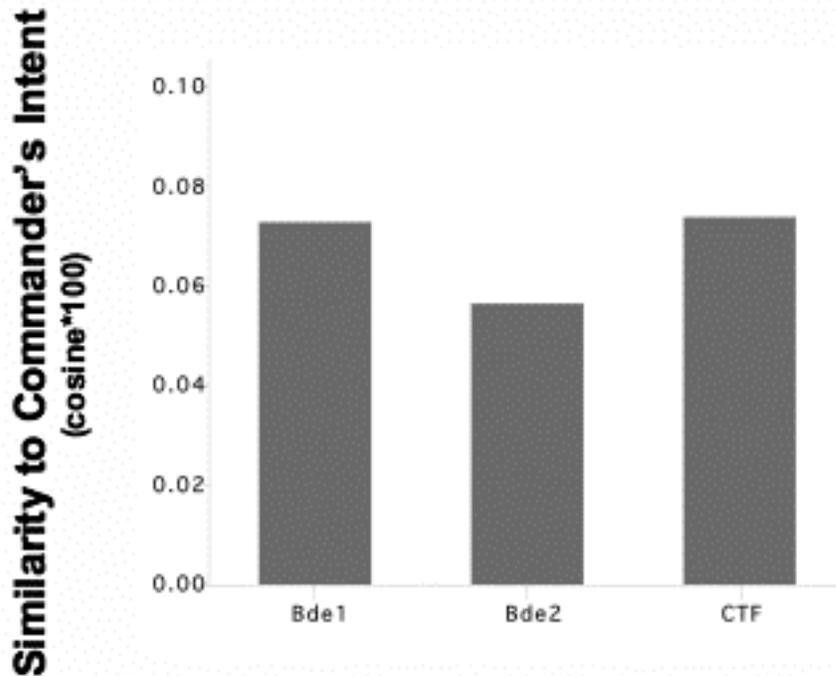
**Figure 2.** Quality of discussion comments in the three discussion conditions.

Although customized for distributed learning activities, the tools developed within Knowledge Post can be incorporated into other virtual environments for automated analysis and monitoring of teams performing planning-based discussions.

### TeamViz

The TeamViz application provides teams and evaluators ways of monitoring performance in large collaborative environments using a set of visualization tools and enhancements built on the Knowledge Post toolset. TeamViz ran live during a U.S.-Singapore simulation exercise designed to evaluate collaboration among joint, interagency, and multinational forces conducting combat and stability operations (Pierce, Sutton, Foltz, LaVoie, Scott-Nash & Lauper, 2006). The system automatically

analyzed the content and patterns of information flow of the networked communication. It also provided automated summarizations of the ongoing communications as well as network visualization tools to improve situation understanding of team members. Analyses showed that the technology could track the flow of commander's intent among the team members by comparing the commander's briefing to the content of communication of different parts of the team. For example, the commander stressed the importance of naval facility defense in his briefing to three groups: two brigades under his command and the Coalition Task Force command staff. Comparing the content of the communications in each group following this briefing shows that Brigade 1 followed the commander's intent more closely than did Brigade 2 (see Figure 3).



**Figure 3.** Communication analysis shows that Brigade 1 followed the commander's intent more closely than Brigade 2.

It was also possible to detect the effects of scenario information injects on performance within the coalition task force and brigades by comparing the communication within each group to the content of the scenario inject. Figure 4. shows the response to an inject about a chemical weapons attack. It is clear that the coalition task force responded more quickly to the inject, and with a greater degree of discussion, than did either brigade.

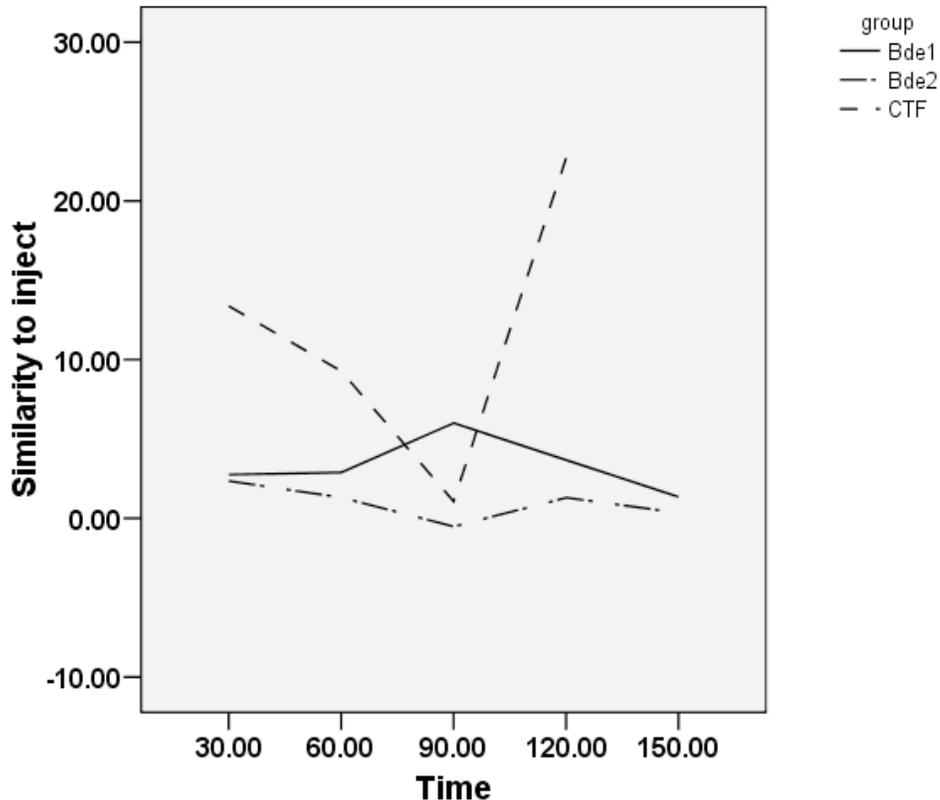


Figure 4. Communication analysis shows that the Coalition Task Force responded more quickly to a scenario inject than either brigade.

Singapore staff officers used TeamViz in real-time to monitor the communication streams and to inform their commanders of important information flowing in the network as well as to indicate perceived information bottlenecks. Overall, the TeamViz technologies permit knowledge management of large amounts of communication as well as improved cognitive interoperability in distributed operations where communication among ad hoc teams is critical.

#### Competence Assessment and Alarms for Teams

Convoy operations require effective coordination among a number of vehicles and other elements, while maintaining security and accomplishing specific goals. However, in training for convoy operations it is difficult to monitor and provide feedback to team members in this complex environment. The DARPA Automated Competence Assessment and Alarms for Teams (DARCAAT) program was designed to automate performance assessment and provide alarms for live and virtual convoy operations training. As part of the program, communication data and SME-based performance measurements were collected and then specialized tools to assess and visualize performance in convoy operations were developed.

The DARCAAT program collected voice communication data during convoy training operations and then collected SME-based performance measurements on that data. From these, the DARCAAT program developed specialized tools to assess and visualize convoy operation performance. Two sources of data were used: one from teams in a virtual environment and one from teams in live training environments. The goal was to evaluate how well performance assessment tools could be applied to a single domain across both virtual and live training. For the virtual environment, communication data was collected from the Fort Lewis Mission Support Training Facility, which uses the DARWARS Ambush! virtual environment for convoy training. DARWARS Ambush! is a widely used game-based training system and

has been integrated into training for many brigades prior to deployment in Iraq (Diller, Roberts, Blankenship & Nielson, 2004; Diller, Roberts & Wilmuth, 2005). DARWARS Ambush! provides an excellent environment for team training and performance analysis because it provides reasonably controlled scenarios and environment, and has the ability to instrument teams for voice communications, video, and environmental event data collection. In this environment, up to 60 soldiers can jointly practice battle drill training and leader/team development during convoy operations. Figure 5 shows the training environment for DARWARS Ambush! and Figure 6 shows a typical user's view during training.



**Figure 5.** DARWARS Ambush! at the Fort Lewis Mission Support Training Facility



**Figure 6.** Screen from DARWARS Ambush! training scenario.

In addition to the virtual environment DARWARS Ambush! data, the DARCAAT program collected live convoy STX lane training data from the National Training Center (NTC) at Fort Irwin. The data included digital audio recordings of FM radio communication among the convoy team members as well as videos of the convoy operations. Using the virtual and live convoy communications data, Subject Matter Experts rated team performance on a number of metrics (Battle Drills, adherence to Standard Operating Procedures, Situation Understanding, Command and Control, and overall team performance) as well as indicated places in the scenario in which a critical event occurred (i.e., “an event that significantly alters the battleground”). Prediction models were then built by analyzing the communication data using the full team communication analysis pipeline shown in Figure 1.

The results indicate that the DARCAAT toolset is able to accurately match SME ratings of team performance as well as detect critical events (e.g., performance alarms). Using the DARWARS Ambush! data, the system could automatically detect 87% of the SME-rated critical events with a false positive rate of 19%. Thresholds for detecting critical events can be adjusted to allow them to be used as performance

alarms enabling a commander to set lower thresholds which provide alerts for any case in which a team might be having performance problems at the cost of only a slightly higher false alarm rate. The DARCAAT model also predicted the SME ratings of team performance on each of the performance metrics. Table 1 shows the correlations between the SME ratings of overall team performance and the predictions generated by the DARCAAT toolset from analyzing the teams' communications based on 45 Ambush! missions and 6 NTC missions (all significant  $p < .01$ ). It should be noted that the correlations between the SMEs and toolset were equivalent to those found between multiple SMEs rating the same missions.

TABLE 1. Correlation between SME ratings and DARCAAT predictions for overall team performance

Metric	NTC & Ambush! (n=51)	Ambush! (n=45)
Battle Drills	0.74	0.73
Command and Control	0.71	0.70
Situation Understanding	0.83	0.81
SOPs	0.73	0.79
TEAM	0.78	0.72

As a demonstration of the application of the DARCAAT toolset, an After Action Review application was developed that could be integrated into a training program to allow observer controllers (OCs) and commanders to monitor teams and receive feedback on the team's performance. The application provides efficient automatic augmentation of AARs assisting the OCs in choosing the most appropriate segments of missions to illustrate training points. Figure 7 shows one screen from the AAR tool.

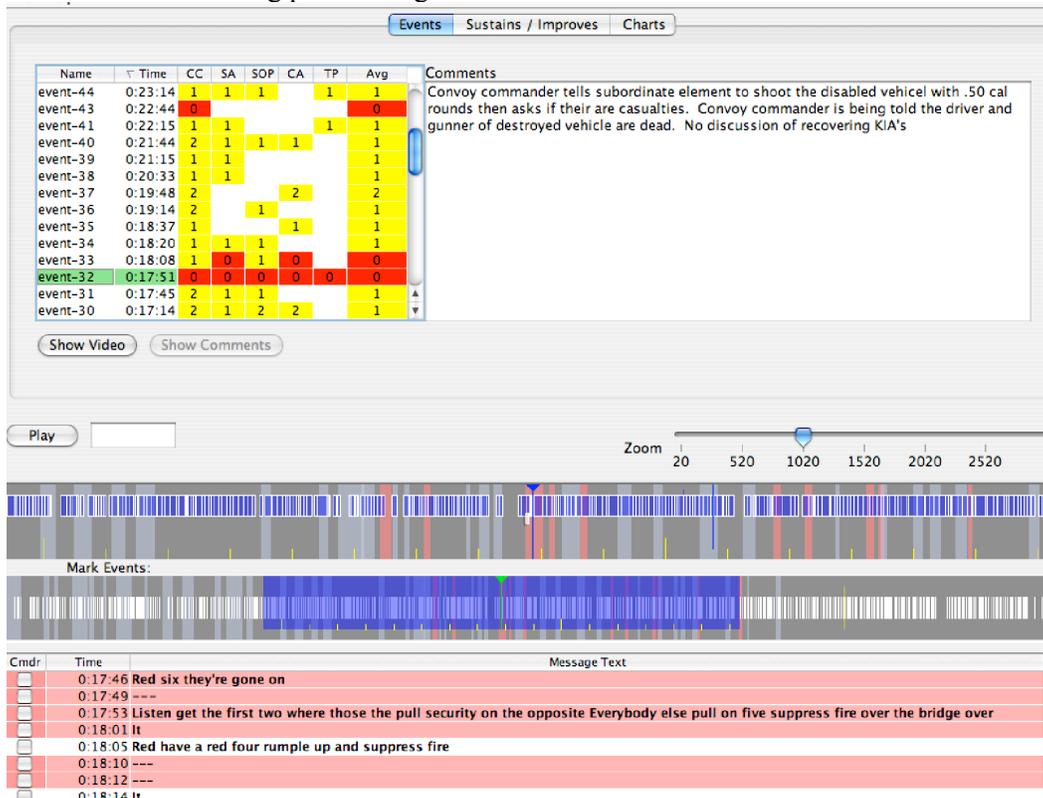


Figure 7. Visualization of team performance scores from the AAR tool.

The tool processes the incoming communication data from a team and then allows an OC or commander to load any mission and provides immediate access to several critical pieces of information:

- The top left portion of the AAR tool, displays the mission divided into a list of sequenced events with highlighted critical events;
- Each event in the list is scored on a series of metrics: CC (Command and Control), SA (Situation Awareness), SOP (adherence to Standard Operating Procedures), CA (Combat Action/Battle Drills), TP (overall Team Performance);
- The event list can be sorted by score, allowing rapid identification of the most serious issues;
- The lower portion of the AAR tool shows a mission timeline linked to the event list, with facility to play audio files and view an ASR transcript of each event.

Overall, the results from the DARCAAT project illustrate that performance measures can be automatically and accurately generated from communication in teams performing in multi-user virtual and live environments. These performance measures can then be incorporated into visualization and training tools which permit trainers to monitor and assess team status in real-time.

### **Conclusions**

Communication is the glue that holds teams together in networked virtual environments. It is also one of the richest sources of information about the performance of the team. The content and patterns of a team's communication provide a window into performance and cognitive states of the individuals and the team as a whole. Analysis of the complex cascades of communication requires tools that can assess both the content and patterns of information flowing in the network. The approach described in this chapter can automatically convert the communication into specific metrics of performance thereby permitting a better picture of the state of teams in virtual environments at any point in time. The tools use language technologies to analyze the content of communication thereby permitting characterization of the topics and quality of information being transmitted

The toolkit allows the analysis and modeling of both objective and subjective performance metrics and it is able to work with large amounts of communication data. Indeed, because of its machine-learning foundation, it works better with more data. The toolkit can automatically extract measures of performance by modeling how subject matter experts have rated similar communication in similar situations as well as modeling objective performance measures. Further, because the methods used are automatic and do not rely on any hand-coded models, they allow performance models to be developed without the extensive effort typically involved in standard task-analysis or cognitive modeling approaches. Notably, the approach can be integrated with traditional assessment methods to develop objective and descriptive models of distributed team performance. Overall, the toolset has the ability to provide near real-time (within seconds) assessment of team performance including measures of situation understanding, knowledge gaps, workload, and detection of critical incidents. It can be used for tracking teams' behavior and cognitive states, for determining appropriate feedback, and for automatically augmenting After Action Reviews.

### New Directions

There remain a number of challenges to incorporating automated analysis of the content of communication into full-scale virtual environments for training venues. First, virtual environments must provide technology to allow easy collection of communication data to allow analysis by toolsets. In addition, virtual environments need to make log files of participant actions, locations and movements easily accessible so that tools can derive and analyze additional performance measures. Second, while the results described in this chapter use teams ranging in size from 3 to 70 soldiers, it is important to understand the challenges of scaling up to even larger operations. Finally, a number of other technologies can be included to improve and help generalize modeling performance. These include better modeling of network structures, incorporation of additional modalities of information (e.g., event and action information), improved computational modeling tools, and leveraging of other advances in measuring

performance in complex virtual environments.

The automated analysis of communication can be applied in a wide range of virtual environment applications beyond those described here. This approach can be integrated into and make possible adaptable training systems which automatically adjust the level of difficulty of the training based on performance of the team. Finally the overall approach helps in understanding the role of communication in complex human networks. Results from analyses of teams in realistic situations can help clarify both how communication affects team performance and how performance is reflected through communication.

### **Acknowledgements**

This work was supported in part by grants and contracts from DARPA, ARI, ARL, ONR, and AFRL. The authors are grateful for the contributions of Terry Drissell, Marita Franzke, Brent Halsey, Kyle Habermehl, Tim McCandless, Chuck Panaccione, Manju Putcha, and David Wroblewski, for development and data analyses.

### **Biographies of authors**

Peter W Foltz is founder and Vice President for Research at Pearson Knowledge Technologies and Senior Research Associate at the University of Colorado, Institute of Cognitive Science. His work focuses on cognitive science approaches to measuring individual and team knowledge. He has published a range of articles on Information Retrieval, Natural Language Processing, Educational Technology, Clinical Diagnosis, Cognitive Science and Team Assessment. Peter has served as principle investigator for research for the Army, Air Force, Navy, DARPA, and NSF, Contact information: Pearson Knowledge Technologies. 4940 Pearl East Circle, Suite 200, Boulder, CO, 80305. [pfoltz@pearsonkt.com](mailto:pfoltz@pearsonkt.com).

Noelle Lavoie is a founder of Parallel Consulting, LLC where she is the lead Cognitive Psychologist, and a former Senior Member of Technical Staff at Pearson Knowledge Technologies. Her work includes studying online collaborative learning, visualization tools to support multinational collaboration, tacit knowledge based assessment and development of military leadership.

Rob Oberbreckling has worked with Pearson as a Senior Member of Technical Staff and currently performs research and engineering in natural language processing, cognitive science, and machine learning systems.

Mark Rosenstein is a Senior Member of Technical Staff at Pearson applying machine learning and natural language processing techniques to problems involving understanding and assessing language and the activities connected with the use of language.

### **References**

- Brannick, M. T., Salas, E., & Prince, C. (1997). *Team performance assessment and measurement: Theory, methods, and applications*. Mahwah, NJ: LEA.
- Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 330-380). San Francisco, CA: Jossey-Bass.
- Curtis, M. T., Harper-Sciarini, M., DiazGranados, D., Salas, E., & Jentsch, F. (2008). Utilizing multiplayer games for team training: Some guidelines. In H. F. Oneil & R. S. Perez (Eds.), *Computer games and team and individual learning* (pp 145-165). Oxford UK: Elsevier.
- Diller, D. E., Roberts, B., Blankenship, S., & Nielsen, D. (2004). DARWARS Ambush! – Authoring lessons learned in a training game. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference*. Arlington, VA: National Training Systems Association.
- Diller, D. E., Roberts, B., & Willmuth, T. (2005, September). *DARWARS Ambush! A case study in the adoption and evolution of a game-based convoy trainer with the U.S. Army*. Paper presented at the Simulation Interoperability Standards Organization, Orlando, FL.

- Foltz, P. W., Laham, R. D., & Derr, M. (2003). Automated speech recognition for modeling team performance. *Proceedings of the 47<sup>th</sup> Annual Human Factors and Ergonomic Society Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Foltz, P. W. (2005). Tools for enhancing team performance through automated modeling of the content of team discourse. *Proceedings of the HCI International Conference*. Saint Louis, MO: Mira Digital Publishing.
- Foltz, P. W., Martin, M. A., Abdelali, A., Rosenstein, M. B., & Oberbreckling, R. J. (2006). Automated team discourse modeling: Test of performance and generalization. *Proceedings of the 28<sup>th</sup> Annual Cognitive Science Conference*. Bloomington, IN: Cognitive Science Society
- Freeman, J., Diedrich, F. J., Haimson, C., Diller, D. E., & Roberts, B. (2003). Behavioral representations for training tactical communication skills. *Proceedings of the 12<sup>th</sup> Conference on Behavior Representation in Modeling and Simulation*. Scottsdale, AZ.
- Gorman, J., Weil, S. A., Cooke, N., & Duran, J. (2007). Automatic assessment of situation awareness from electronic mail communication: Analysis of the Enron dataset, *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting* (pp. 405-409). Baltimore, MD,.
- Gorman, J. C., Foltz, P. W., Kiekel, P. A., Martin, M. A., & Cooke, N. J. (2003). Evaluation of Latent Semantic Analysis-based measures of communications content. *Proceedings of the 47<sup>th</sup> Annual Human Factors and Ergonomic Society Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Hussain, T. S., Weil, S. A., Brunyé, T. T., Sidman, J., & Alexander, A. L. (2008). Eliciting and evaluating teamwork within a multi-player game-based training environment. In H. F. Oneil & R. S. Perez (Eds.), *Computer games and team and individual learning* (pp 77-104). Oxford UK: Elsevier.
- Jurafsky, D., & Martin, J. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. New York: Prentice Hall.
- Kiekel, P.A., Gorman, J. C., & Cooke, N. J. (2004). Measuring speech flow of co-located and distributed command and control teams during a communication channel glitch. *Proceedings of the Human Factors and Ergonomics Society 48<sup>th</sup> Annual Meeting*.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes*, 25(2&3), 259-284.
- LaVoie, N., Psocka, J., Lochbaum, K. E., & Krupnick, C. (2004, February). *Automated tools for distance learning*. Paper presented at the New Learning Technologies Conference, Orlando, FL.
- LaVoie, N., Streeter, L., Lochbaum, K., Wroblewski, D., Boyce, L., Krupnick, C., & Psocka, J. (in press). Automating expertise in collaborative learning environments. *Journal of Asynchronous Learning Networks*.
- Lochbaum, K., Streeter, L., & Psocka, J. (2002, December). *Exploiting technology to harness the power of peers*. Paper presented at the Interservice/Industry Training, Simulation and Education Conference, Orlando, FL.
- Orsanu, J., & Salas, E. (1993). Team decision making in complex environments. In G. A. Klein, J. Orsanu, R. Calderwood, & C. E. Zambok (Eds), *Decision making in action: Models and methods* (pp. 327-345). Norwood, NJ: Ablex Publishing.
- Paris, C. R., Salas, E., & Cannon-Bowers, J. A. (2001). Teamwork in multi-person systems: A review and analysis. *Ergonomics*, 43(8), 1052-1075.
- Pierce, L., Sutton, J., Foltz, P. W., LaVoie, N., Scott-Nash, S., & Lauper, U. (2006, July). Technologies for Augmented Collaboration. Paper presented at the CCRTS, San Diego, CA.
- Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology*, 52, 471-499.