Developing Team Performance Models: From Abstract to Concrete

Dr. Sowmya Ramachandran, Randy Jensen
Stottler Henke Associates, Inc.
San Mateo, CA
sowmya@shai.com
jensen@shai.com

Dr. Eduardo Salas
Institute for Simulation and Training
University of Central Florida, FL
esalas@ist.ucf.edu

ABSTRACT

As the military has moved increasingly towards distributed networked environments for Command and Control Intelligence, Surveillance, and Reconnaissance (C2ISR) missions, teams often operate remotely, and decision-making is distributed. Traditionally team training involved human observers for performance assessment, diagnosis, and after-action review and other training intervention. However, with much of the communication and coordination happening electronically, key aspects of the interactions between team members are no longer accessible to these trainers. Analyzing these communications involves poring over high volumes of raw electronic data. This is infeasible in all but the smallest scales of operation. Intelligent automated performance assessment tools can be valuable cognitive aids to trainers and assist them by warehousing and analyzing team interaction data, and presenting it to them in a user-friendly manner for real time coaching and after-action review. In order to build such a system, it is important to first define a concrete model of team behavior for the domain and to define rules to assess team performance dimensions from observations of team behavior in training exercises. Research literature is rich with different models of team performance; however, these models are defined at a very abstract level and not directly useful at the level of specificity that would be needed by a rule-based artificially intelligent assessment tool. This has always been the challenge of artificial intelligence. In this paper, we will present a case study that shows the process of translating an abstract team performance model into a concrete model and the resulting performance assessment rules that can be used by an automated tool. The model is being developed to serve as a basis for an automated after-action review tool to support large team training exercises within the Marines in the area of combined arms. The paper will also discuss the lessons learned along the way.

ABOUT THE AUTHORS

Dr. Ramachandran is research scientist at Stottler Henke Associates, a research and development firm dedicated to providing innovative Artificial Intelligence solutions to real-world problems. Dr. Ramachandran's interests focus on intelligent training and education technology including intelligent tutoring and intelligent synthetic agents for simulations. She is also interested in issues of motivation and metacognition. Experience with military and private industry gives Dr. Ramachandran a unique perspective on the needs and requirements of the ultimate end-users and their constraints. She contributes expertise in AI, instructional systems, probabilistic reasoning, and knowledge management. She has developed ITSs for a range of topics including reading comprehension, high-school Algebra, helicopter piloting, and healthcare domains. She has participated in workshops organized by the Learning Federation, a division of the Federation of American Scientists, to lay out a roadmap for critical future research and funding in the area of ITSs and virtual patient simulations. She has developed a general-purpose authoring framework for rapid development of ITSs, which is currently being used to develop an intelligent tutor training Navy Tactical Action Officers. She has also developed tools for training emergency first responders.

Randy Jensen is a group manager at Stottler Henke Associates, Inc., working in training systems since 1993. He has developed numerous Intelligent Tutoring Systems for Stottler Henke, as well as authoring tools, simulation controls, after action review tools, and assessment logic routines. He is currently leading projects to develop automated after action review for Marine Corps combined arms training, a framework for ITS interoperability with distributed learning architectures for the Joint ADL Co-Lab, and an authoring tool for virtual training demonstrations for the Army. He holds a B.S. with honors in symbolic systems from Stanford University.

Eduardo Salas is Trustee Chair and Professor of Psychology at the University of Central Florida. He also holds an appointment as Program Director for Human Systems Integration Research Department at the Institute for Simulation & Training. Previously, he was a senior research psychologist and Head of the Training Technology Development Branch of NAVAIR-Orlando for 15 years. During this period, Dr. Salas served as a principal investigator for numerous R&D programs focusing on teamwork, team training, advanced training technology, decision-making under stress, learning methodologies and performance assessment. His expertise includes helping organizations on how to foster teamwork, design and implement team training strategies, facilitate training effectiveness, manage decision making under stress, develop performance measurement tools, and design learning environments. He is currently working on designing tools and techniques to minimize human errors in aviation, law enforcement and medical environments. Dr. Salas is a Fellow of the American Psychological Association and the Human Factors and Ergonomics Society. He received his Ph.D. degree (1984) in industrial and organizational psychology from Old Dominion University.

Developing Team Performance Models: From Abstract to Concrete

Dr. Sowmya Ramachandran, Randy Jensen
Stottler Henke Associates, Inc.
San Mateo, CA
sowmya@shai.com
jensen@shai.com

Dr. Eduardo Salas
Institute for Simulation and Training
University of Central Florida, FL
esalas@ist.ucf.edu

OVERVIEW

This paper will first discuss a general model of team performance and describe each of its dimensions in separate subsections. It then describes a specific domain that this model was applied to, US Marine Corps (USMC) Combined Arms Training. How this general model maps onto the specific domain is presented. A prototype based on this mapping is described including examples of specific types of team errors and the rules that are used to find them. Examples of the visual debriefing produced by the prototype are shown and explained. Finally, future work is presented.

TEAM PERFORMANCE MODEL

For the purposes of training teams, it is necessary to diagnose differences between individual and team performance (i.e., between the presence or absence of teamwork and taskwork competencies) in order to provide remediation and corrective feedback of the appropriate type and at the appropriate level. Feedback on individual level knowledge, skills, and attitudes can be useful in building individual level expertise, but feedback on teamwork is necessary to leverage that expertise in expert teams by developing teamwork competencies. Motivated by this need, Smith-Jentsch and colleagues (Johnston, Smith-Jentsch, & Cannon-Bowers, 1997; Smith-Jentsch, Zeisig, Acton, & McPherson, 1998; Smith-Jentsch, Johnston, & Payne, 1998) engaged in a research effort aimed at producing a valid, reliable, and diagnostic measure of teamwork as a part of the Tactical Decision Making Under Stress (TADMUS) program (Cannon-Bowers & Salas, 1998). The measurement development approach involved the use of subject matter experts (SMEs) who reviewed videotapes of teams performing an anti-air warfare command information center task (Smith-Jentsch et al., 1998). SMEs then identified the behaviors that distinguished high from low performing teams (Johnston, et al., 1997). The resulting measure, the Anti-Air Teamwork Observation Measure (ATOM), consisted of eleven teamwork behaviors grouped into four dimensions of teamwork. To make use of the ATOM, observers give ratings of each of the eleven behaviors and then of the entire teamwork dimension. The structure of the ATOM has been supported through a confirmatory factor analysis (Smith-Jentsch, 1995). These dimensions are discussed below.

Information Exchange

Information exchange allows the team to develop and maintain a shared understanding of the situation as each member communicates critical information. The information exchange teamwork dimension consists of three specific behaviors. First, team members should *seek information from all available sources.* By doing this, team members ensure that they do not exclude information that may be vital to the decision making process. Second, team members should *proactively pass information to the appropriate team members*. By doing this, team members can compress the time involved in the perception → decision → action cycle of the team and ensure that each team member is operating with all of the available and needed information. Third, team members should provide one another situation updates that communicate the 'big picture' of what is happening. Essentially, these *'big picture' updates* enable team members to cross check their current situational understanding. If a team member's 'big picture' communication is inconsistent with another teammate's understanding, this provides an opportunity to remedy the conflict and ensure that everyone is operating under the most current and accurate representation of the environment.

Communication

The focus of the information exchange dimension of teamwork is on what information is passed to whom. The focus of the communication dimension is on how that information is exchanged. The communication dimension of teamwork in the ATOM consists of four specific teamwork behaviors. First, team members should *use the proper 'phraseology.'* Teams that speak with a specialized communication terminology are able to pass large amounts of information very quickly (Klein, Feltovich, Bradshaw, & Woods, 2005). Second, team members should *provide complete internal and external reports.* This completeness helps to minimize ambiguity associated with communicating only partial information about the situation (i.e., minimizes inferences made about missing information by other team members). Third, team members should *minimize unnecessary communications* (e.g., chatter). This minimizes the workload inherent in team communication and coordination by focusing only on the essentials of interaction necessary for team performance (MacMillan, Entin, & Serfaty, 2004). Fourth, team members should make sure that their *communications are clear and audible*. This minimizes the chance of misinterpretations and misunderstandings of communications as well as reducing the communication related workload involved in clarifying communications of initially low quality.

Supporting Behavior

Supporting behavior involved team members compensating for one another in order to reach the team goal. The supporting behavior dimension of the ATOM consists of two specific teamwork behaviors. First, team members should *correct the errors of other team members*. This process, of course, reduces the number of errors in the team's performance and leads to higher levels of team outcomes. Additionally, this practice also helps to develop the skill levels of team members as they receive feedback on poor performance. Second, team members should *provide and request assistance and backup when it is needed*. This involves team members monitoring each other's performance, identifying when their team members need assistance or they themselves need assistance, and stepping in to resolve the unbalanced workload situation.

Team Initiative/Leadership

The team initiative/leadership dimension of the ATOM consists of two specific teamwork behaviors. First, team members should *provide guidance and suggestions on improvements to one another.* This facilitates team learning and development of skill, which leads to higher levels of performance in future performance episodes. Second, team members should *clearly identify team and individual level priorities.* This ensures that the aspects of the team and individual tasks that are most critical for team outcomes are given the most attention.

Team Dimensional Training

Team Dimensional Training (TDT) is a training program designed with the aim of developing the four teamwork dimensions discussed above. It is a guided self-correction approach wherein a team leader guides the cycle of pre-brief → performance → diagnosis of performance → debrief. The team evaluates its own performance on the four teamwork dimensions and develops the teamwork related knowledge and skills by analyzing and learning from the team's past performance. TDT has been shown to yield significant improvements in team performance (Smith-Jentsch, Zeisig, et al., 1998).

This model of teamwork dimensions was used as a guide to knowledge elicitation from SMEs. Later on in Scenario Debrief Section we describe the details of how the automated after action review system analyzes causal explanations. Conflict trees serve as a causal model for the various conflicts and other interesting events that are anticipated in the domain. These concept trees are based on the idea of fault trees that is commonly used in systems engineering for troubleshooting complex systems and reasoning about their reliability (Vessely, 1987). The conflict trees represent knowledge of how team actions and individual actions (or lack thereof) can lead to such events. The above model of team performance will guide the knowledge elicitation process of developing these conflict trees.

USMC COMBINED ARMS TRAINING DOMAIN

Marine Corps Combined Arms operations fundamentally involve coordinating multiple supporting arms with ground maneuver. Marines are trained in these team coordination skills in several successive tiers of instruction, one of which involves major simulation based exercises at Combined Arms Staff Trainer (CAST) facilities. The Marines have 5 sites around the world with CAST facilities, where they've traditionally conducted low-tech simulated exercises of combined arms operations ***at the staff level***. Because the training objectives center on team coordination and staff level decision making, a low tech solution has historically been sufficient. Decision makers are limited to paper maps, while their "eyes" (reconnaissance teams or forward observers) are positioned around a terrain board where enemy positions and engagement effects are simulated in a very approximate manner (e.g., small metal tank models and cotton balls). Sample photos of a CAST facility are viewable on the web at: http://www.29palms.usmc.mil/dirs/ont/mands/cacctus_pictures.asp

The objective is not to train technical individual skills like adjusting artillery fire onto a target, but rather to practice the coordination between observers, approval authorities, and liaisons with air and other elements of the battle. Combined arms exercises train and rehearse the exercising force (EXFOR) personnel in the tasks of coordinating multiple supporting arms with maneuver. Training exercises may involve 100 or more participants at various stations in the facility, carrying out their respective operational responsibilities. There is an emphasis on providing the EXFOR the experience of performing a role in scenario-based combined arms exercises, so that training is experiential. As a result, EXFOR responsibilities during training mirror those during operational actions, including the employment of skills in communication and coordination, and tactics, techniques, and procedures in support of specific training goals.

The Marines currently have a program to upgrade the CAST facilities to computer simulation based training; the program is called CACCTUS (Combined Arms Command and Control Training Upgrade System) and is under USMC PM TRASYS. CACCTUS includes an intelligent after action review component. With the upgrade, the AAR system is able to present 3D playback of exercise events with synchronized playback of audio from virtual radios used during exercises. This also enables a communication analysis capability using automated speech recognition on the radio communications during the exercise to diagnose causal factors for possible errors.

For USMC CAST exercises, one of the primary decision making groups is the staff of the Fire Support Coordination Center (FSCC). Typically the exercises at the CAST facilities are oriented toward the battalion level FSCC. The FSCC operates as a central hub for combined arms operations within a defined space on the battlefield. In a real world operation, they are 8 or so people, typically situated together inside a command vehicle. They rely heavily on the paper maps they maintain during the operation, and therefore a major requirement is that they have good situational awareness coming from the elements reporting to them. The FSCC reviews all requested fires and approves, denies, or modifies them to make sure that units and friendly fires are separated by either time or space.

As an example, the battalion level Fire Support Coordinator (FSC) is responsible for clearing requests for fire missions and air strikes from subordinate units, and therefore must maintain a clear and accurate operational picture. This also requires coordination with senior and adjacent units, with a constant flow of requested and disseminated information. In a specific example, if a forward air controller requests a Close Air Support (CAS) mission, the FSC must have a correct operational picture in order to make a determination that is both timely and consistent with the scheme of maneuver, rules of engagement, and safety constraints.

TEAMWORK BEHAVIORS AND ERRORS

The abstract team performance model described first mapped into the above domain in a number of a concrete ways. Several examples, sorted by the team model dimensions are presented below.

Perhaps the most important dimension in this domain is Information Exchange. Friendly units should periodically report their position, every 500 meters of movement, for example. The Company Tactical networks can be easily monitored for verbal position updates from the companies to the FSCC. This can be compared against the actual positions of the companies in the simulation for accuracy and frequency. Similarly scout teams should be reporting the location and movement of enemy units. The Surveillance and Target Acquisition network can be monitored for reports from the scout teams on enemy units within their line of site. Again this can be compared against ground truth from the simulation. Very important information to exchange relates to Calls For Fire (CFF). CFFs initiate the events that ultimately cause a fire series to be executed. In particular, whether the mission is approved, denied, or approved with modified parameters is important information to get to both the requester of the CFF and the battery that would fire the series. In addition to friendly and enemy positions, an additional important big picture item is the set of upcoming fire series and missions that have been approved and will soon be fired. It's important that everyone in the FSCC be aware of these for approval of additional requests for missions and movement, for the maneuver companies to be aware of the upcoming fire missions in their area because it supports their tactical activities, and, of course, for the firing batteries to be aware so that they can begin preparations.

Because the Information Exchange dimension is so important in this domain, the Communication dimension is also important and in many cases there is a very specific vocabulary and syntax that should be used. Calls for fire and requests for close air support have a very specific vocabulary and syntax that allows the verbal utterances to be automatically checked by software. Similarly, mission approval and denial should be clearly indicated with the phrases "is approved" and "is denied", respectively. In particular, the phrase "not approved" is a common mistake that should never be used and can be easily checked for, automatically, by software. Other Communication issues that are important are to clearly differentiate between different enemy units when reporting their position and movement and speaking clearly (since sometimes the level of stress and excitement is high). The latter can often be detected by failure of the automatic speech recognition to discern meaningful words. The former by checking that the enemy positions and unit types input into their tactical situational awareness tool match the ground truth from the simulation.

The most important Supporting Behavior is correcting the errors of other team members. Common errors that are currently detected by automated software relate to conflicts between fire missions and friendly units. The most important example is incorrectly clearing a fire mission or air strike which violates safety constraints. Commonly this occurs because of a change in the timing or location of the mission or friendly unit involved in the conflict. Another common error relates to the fact that the artillery calculations are done in metric units, but the stay above or stay below altitudes must be reported to the relevant aircraft in feet. Most conflicts detected represents an error that a team member could have corrected. In the FSCC there are also substantial opportunities to provide or request assistance. This is because there are basically three functions that the FSCC's individuals perform. They communicate over the radio, they maintain an updated map of friendly and enemy units and the graphics associated with current and upcoming fire missions, and they approve or deny fire missions and air strikes. Because the first function flows into the second which flows into the third which flows back to first, the individuals involved in one function can easily take on additional responsibilities of an adjacent role.

PROTOTYPE DESCRIPTION

The prototype serves as an operational demonstration of general design concepts for team training AAR tools. An objective for the prototype is to demonstrate functionality with a specific domain, and use this as a platform to generalize from, for application to other team training domains. In light of this objective, we chose the USMC Combined Arms training domain described above. The central functionality of the toolset can be generalized beyond the specific training requirements of the subject domain, as one of the key features of the collection of team performance measures is its applicability to different domains. Also, by virtue of using an example training environment where many of the pieces (simulation, visualization, virtual radios, data architecture and protocols, etc.) are already in place, the prototype development effort could focus primarily on implementing an example with the more general team performance models rather than addressing trainer-specific engineering issues.

Demonstration Scenario

The demonstration scenario involves a sequence of events that could be taken from a full-scale training exercise, and it contains examples of both individual and team performance errors that occur in the course of a combined arms operation. The operational situation involves an offensive scenario where blue forces are moving to contact against red force tanks which are also protected by red force anti-aircraft assets. This would be considered a “suppress long attack short” scenario, where the enemy anti-aircraft assets must be suppressed to allow for a fixed wing Close Air Support (CAS) attack in advance of the ground assault on the target. As a result, the personnel in the Fire Support Coordination Center (FSCC) are responsible for oversight of ground movements, indirect fire suppression missions, and the CAS mission against the target. The scenario unfolds over 20 minutes, and involves 10 participants occupying roles in either the FSCC or the distributed elements in communications with the FSCC. Roles and communication nets are listed in the following table.

Table 1. Exercise information

Nets	Surveillance and Target Acquisition (STA) Artillery Conduct of Fire A (ArCOFA) Tactical Air Request (TAR) Company A Tactical (CoA_TAC)
Roles	FSCC S-2 FSCC S-3 FSCC Air Officer (AirO) FSCC Artillery Liaison Officer (Arty LNO) Forward Observer (FO) Forward Air Controller (FAC) Artillery Fire Direction Center (FDC) Tank platoon Fixed Wing Pilot1 (FW Pilot1) Fixed Wing Pilot2 (FW Pilot2)

The exercising force task in this scenario is to properly employ a pre-planned fire series which coordinates the suppression, attack, and ground movement missions. The friction points in the scenario range from the potential for battlespace geometry conflicts, to the potential for coordination errors that could lead to delays in execution. Delays can be extremely costly not only in terms of battle tempo but also because of specific windows of opportunity such as the limited availability of CAS aircraft on station before the necessity of refueling requires them to return to base.

Scenario Debrief

A sample exercise was conducted with the demonstration scenario, with events logged and a resulting debrief generated. In this particular exercise run, there are two primary training points, tied to team performance metrics involving Information Exchange and Supporting Behavior.

First Error – Delayed Fire Series Execution
In order to execute the pre-planned fire series, the FSCC personnel must consider the intel that one of the enemy targets is currently moving. Since the series involves suppression on this target, FSCC personnel must make reasoned time-distance calculations which consider the enemy position and also the capabilities of indirect fire assets to set up the mission against the target. In this specific scenario run, the decision maker in the FSCC (the Fire Support Coordinator) makes an error in establishing initial parameters for the fire series, by failing to consider the workload of the Artillery battery and thus planning on a mission to be fired at a time before the battery will be able to be laid onto the target. As a result, the battery is unable to execute the mission when the forward observer sends a command for adjusting fire onto the target, and the enemy unit is already well past the designated target grid by the time the Artillery battery is ready to fire. This is a team performance error in two ways: first because the battery fails to correct the Fire Support Coordinator's (FSC) error when he commands a fire series at a time when the battery will not be ready to fire it (a Supporting Behavior error) and, second, it is an Information Exchange error because the battery does not proactively tell the FSC their workload situation or that they will not be able to meet the time of the fire mission.

The result of these errors is an 8 minute delay in the execution of the fire series, which comes very close to moving the series outside the time period when aircraft are available with time on station to support the CAS missions included in the series.

As a training point, this is unique because of the absence of overtly negative simulation outcomes associated with it. Detection of this training point is keyed off from the “check firing” command sent to cancel the initial series execution. This causes the system to investigate the communications related to the initial fire series and provide suggestions to the instructor for those that should be included with the debrief. Notable communications include those where the series is initially established with the Artillery battery (with no objections back regarding the timing), and the communications where the battery ultimately answers, “Negative Dragon, no guns laid in on that target yet” when prompted to fire the mission.

Second Error – Tanks Inside CAS Danger Area
Early in the scenario, a tank platoon begins a movement toward the enemy target, following the convention of giving position reports at every 500 meters as they approach their objective position. Following the planned coordination in the fire series, the tanks are to move to a position just outside the danger area for the planned CAS attack on the target and then halt and report position. The tank platoon leader makes the error of moving his unit past the halt position and into the danger area, and also makes the additional error of not having given the last position report at the position just outside of this area. This results in a conflict when the CAS mission proceeds and the tanks are inside the danger area at the time of CAS ordnance detonation.

In this sequence of events, the tank platoon leader’s individual error is compounded by a team error in that the FSCC failed to anticipate the tank platoon's position and request a position report when none had come. From a formalized team performance metrics standpoint, this constitutes an Information Exchange deficiency when the position report is not proactively provided, and a Supporting Behavior deficiency when the S-3 in the FSCC fails to ask for the position report. Furthermore, within the FSCC itself, the other roles such as the Air Officer and the FSC should have been aware of the situation and also expecting confirmation of ground unit positions before giving the final clearance to aircraft on dropping ordnance. Either of them also could have initiated the process prompting the S-3 to request a position report if they were also performing effective Supporting Behavior as a team.

Visual Debrief

The discussion so far has described a general model of Team Performance and given concrete examples of how this model maps into a specific domain. However to have any training impact, the system must convey to the team and its individuals what problems occurred and how team work needs to be improved. Because in this domain teamwork is primarily carried out through verbal communication over the radio, these radio transmissions are very important and therefore prominent in the debrief. Figure 1 below shows the visual representation of timeline information for the second training point, with radio transmissions shown on communications timelines correlated with simulation events.

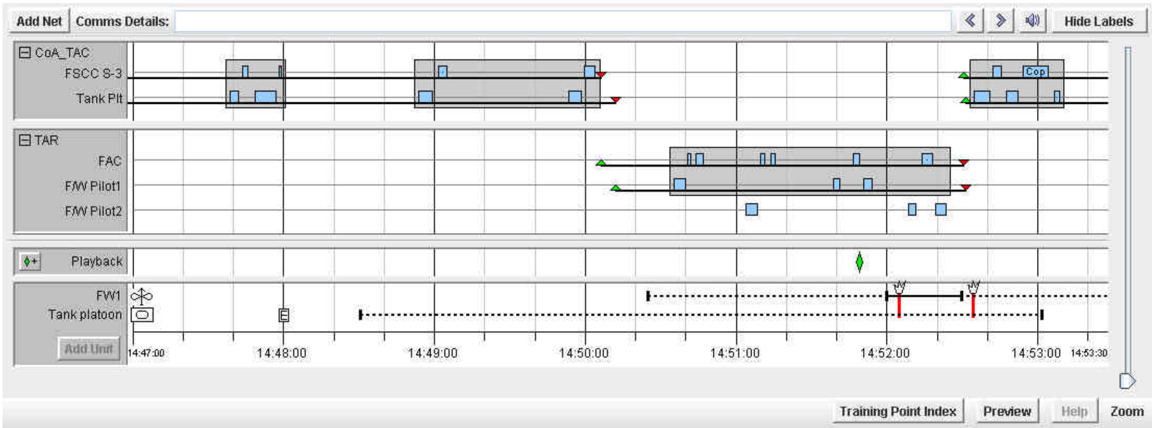

Figure 1. Training point timeline display

Based on the attributes of the Training Point (what kind of event, what units were involved, when it took place), it is automatically determined which radio nets may contain relevant communications. In retrieving this information, the system also disregards any nets that were inactive (no speakers or listeners) for time periods relevant to the Training Point. For this training point, it is relevant to show the position report communications on the CoA_TAC net, and also the air communications on the TAR net. Rows on the timelines represent individual stations that were listening to the net or sending transmissions. For this implementation, blue boxes represent individual transmissions and gray bounding boxes designate groups of transmissions, corresponding to dialogs between two individual stations; see Figure 2.

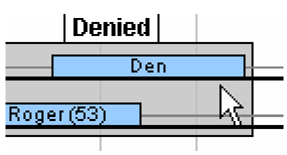

Figure 2. Transmissions and labels

Transmissions are automatically analyzed for spoken content containing certain keywords that reflect decisions. When such content is found, transmissions are labeled with the relevant keyword. The labels for individual transmissions are represented inside the blue boxes (grey in black and white printouts) for transmissions. When applicable, summary labels are also created for groups of transmissions to reflect the overall contents of a dialog between multiple roles.

The Playback Timeline arranges markers for the Playback contents that will be available for the Training Point in Debrief, including radio transmission icons and bookmarks for 3D playback. An initial set of relevant communications to include in the Playback is chosen automatically, but this can be customized. The Playback Timeline serves to show what is currently included in the planned Playback, as well as to allow for specific transmissions or additional bookmarks to be added or removed.

The Simulation Event Timeline presents a timeline-based visual representation of simulation events related to the current training point. The objective of this timeline for this domain is to show the relative times when movements and fire missions took place, and especially highlight critical times, such as a conflict, or the risk of a conflict. The Simulation Event Timeline is specifically designed to visually correspond to the representations in a fire series worksheet as used by the Marines. With this training point, the movement times for the tank platoon are shown below the timeline representation of the CAS mission, which also specifically shows the two CAS ordnance detonations. The fact that this represents a conflict condition is highlighted with red marks (appearing as dark lines in black and white printouts) linking the active danger area associated with the detonations to the tank platoon, as shown in Figure 3.

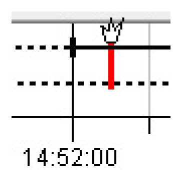

Figure 3. Timeline conflict representation

The generation of a training point for this error is triggered by the conflict conditions that result with the CAS mission and the tank platoon position. The resulting analysis simply notes the lack of a position report, whether proactively provided by the tank unit or elicited by the FSCC S-3, and concludes that this can be attributed to team coordination errors in the debrief.

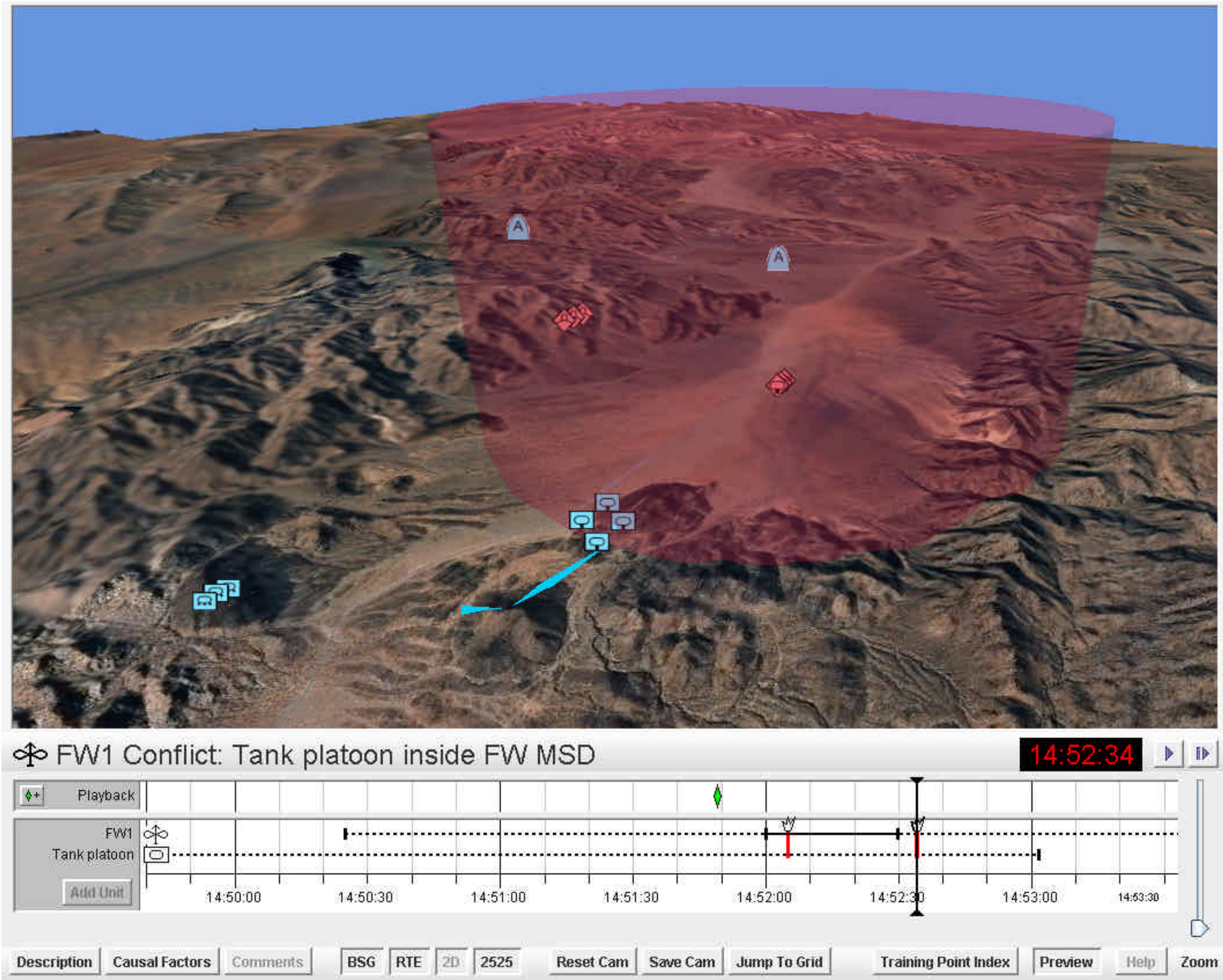

Figure 4. Debrief playback for ground conflict with CAS mission

For debrief playback, it is important to show not only the conflict itself, but also when the tanks moved past the original designated halt position, when they entered the danger area, and how these relate to the timing of the CAS mission in terms of the time available to abort the mission or pursue other remedies. Figure 4 above shows the debrief playback for this training point. At the current time in the simulation playback, the tanks can be seen inside the red danger area at the moment when the second CAS detonation occurs, both in the 3D view and the corresponding timeline representation.

The playback is automatically configured based on templates and rules, to prepare the camera position and viewing angle for 3D playback based on the nature and location of the associated simulation events. In addition to automated template-driven training point preparation, the system provides the ability to create custom training points as a way to access and present nearly any event from the exercise, whether associated with an automatically generated training point or not.

CONCLUSION AND FUTURE WORK

Currently, development is continuing on an automated, intelligent performance assessment tool to assist trainers by warehousing and analyzing team interaction data and presenting it to them in a user-friendly manner for real time coaching and after-action review. The tool will perform automatic event detection and causal explanation generation. Speech recognition, text processing, and causal techniques will be used to analyze and generate diagnostic explanation of teamwork performance events. This information will be sent to desktop, laptop or handheld clients as alerts. A visual tool will facilitate the rapid construction of the debriefing. The system will include both a standard

reusable library of teamwork performance assessment functionality and specific implementations for two domains involving large team training exercises. Since radio and voice communications are an integral part of this chain, a natural language processing capability is required to detect and parse key spoken transmissions and to apply reasoning to establish causal relationships with detected simulation events. Traceable data from distributed operational/simulation tools and direct simulation commands also serve as additional inputs to the explanation rules. By incorporating causal explanations into an automatically generated performance assessment, the result is meaningful feedback that the training audience can directly apply.
Focusing the reasoning in interesting events, and reasoning back from events to causes restricts the hypothesis space and therefore offers a scalable model of reasoning. The second domain will be distributed air operations in Distributed Missions Operations Center (DMOC) Virtual Flags exercises at Nellis AFB.

There is a strong interest across a wide range of military training programs to not only develop abstract theoretical models of human performance, but also apply these models to develop operational systems for performance assessment and training. Applying a theoretical model to a domain is a challenging task. This paper describes a case study that maps a specific domain to an abstract model of team performance in order to develop an after-action review construction tool. The lessons learned from this effort are that, even using established team performance models, identifying team performance assessment criteria for a particular domain requires careful detailed analysis involving SMEs. The theoretical models benefit the analysis by providing an organizing framework for the analysis. By generalizing over several domains, these models provide the dimensions of analysis and thus constrain the analysis space and lead to resource efficiencies. We hypothesize that these abstract psychological models of team work used to define abstract data and visualization models for automated performance assessment that can then be instantiated for each domain to create a domain-specific AAR tool. These would provide an organizing framework for developing systems and enable an AAR architecture that can be reused across domains. We are currently testing this hypothesis by abstracting the existing system to extend it to apply to the DMOC Virtual Flag exercises.

REFERENCES

Cannon-Bowers, J., & Salas, E. (Eds.). (1998). *Making Decisions Under Stress: Implications for Individual and Team Training.* Washington, DC: APA.

Johnston, J.H., Smith-Jentsch, K.A., & Cannon-Bowers, J.A. (1997). Performance measurement tools for enhancing team decision-making training. In M. T. Brannick, Salas, E., & Prince, C. (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 311-327). Mahwah, NJ: Erlbaum.

Klein, G., Feltovich, P.J., Bradshaw, J.M., & Woods, D.D. (2005). Common ground and coordination in joint activity. In W. B. Rouse, & Boff, K.R. (Eds.), *Organizational simulation* (pp. 139-184). Hoboken, NJ: Wiley-Interscience.

MacMillan, J., Entin, E. E., & Serfaty, D. (2004). Communication overhead: The hidden cost of team cognition. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 61-82). Washington, DC: American Psychological Association.

Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 271-297). Washington, DC: American Psychological Association.

Smith-Jentsch, K. A., Johnston, J.A., & Payne, S.C. (1998). Measuring team-related expertise in complex environments. In J. A. Cannon-Bowers, and Salas, E. (Ed.), *Making decisions under stress: Implications for individual and team training* (pp. 61-87). Washington, DC: American Psychological Association.

Smith-Jentsch, K. A. (1995, May). *Measurement and debriefing tools refined and validated at SWOS.* Presentation at the meeting of the TADMUS Technical Advisory Board, Moorestown, NJ.

Vessely, W.E. (1987). Fault Tree Handbook. Nuclear Regulatory Commission.