Assessing Perceived Truth Versus Ground Truth in After Action Review

Randy Jensen
Stottler Henke Associates, Inc.
San Mateo, CA
Jensen@stottlerhenke.com

ABSTRACT

In simulation based training, the notion of ground truth corresponds to "real" events, states, or outcomes in the virtual or constructive environment. Perceived truth on the other hand involves the more complex combination of knowledge, recall, and inference that the training audience must continually apply in the course of their decision making during an exercise. Unlike ground truth, perceived truth relates to human cognition, with potentially incomplete or even self-contradictory information. While it is beyond the scope of any training system to collect and process a representation of perceived truth that would attempt to fully capture human thinking, there are training benefits to techniques that move in this direction. Where perceived truth errors can be identified for training audience feedback, they provide direct insight into not only the conditions where a problem occurs but also the precipitating factors and decisions. This paper describes an inferential approach for assessing perceived truth factors in an automated after action review (AAR) system for Marine Corps combined arms training. This approach involves the examination of simulation ground truth data combined with data from human-in-the-loop interfaces and radio voice communications. Because original human inputs are frequently not preserved in simulation data streams, these other sources are essential to an instructional perspective on perceived truth. In simulation based training, perceived truth errors can lead to hidden problem conditions that may not be manifested in the virtual "ground truth." For example, accidental good fortune in the timing of constructive fire events in relation to unit positions in the simulation ground truth may conceal the potential conflicts allowed by decisions based on poor situational awareness. In such conditions a capability to detect perceived truth failures is not only an additional source of training insight, but a necessary element of AAR.

ABOUT THE AUTHOR

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 natural language analysis methods. He is currently leading projects to develop an embedded training Intelligent Tutor for the Army, an after action review toolset for the Air Force, and an authoring tool for virtual training demonstrations for the Army. He holds a B.S. with honors in symbolic systems from Stanford University.

Assessing Perceived Truth Versus Ground Truth in After Action Review

Randy Jensen
Stottler Henke Associates, Inc.
San Mateo, CA
Jensen@stottlerhenke.com

INTRODUCTION

In simulation based training, the notion of ground truth corresponds to "real" events, states, or outcomes in the virtual or constructive environment. Perceived truth on the other hand involves the more complex combination of knowledge, recall, and inference that the training audience must continually apply in the course of their decision making during an exercise. Unlike ground truth, perceived truth relates to human cognition, with potentially incomplete or even self-contradictory information. While it is beyond the scope of any training system to collect and process a representation of perceived truth that would attempt to fully capture human thinking, there are training benefits to techniques that move in this direction. Where perceived truth errors can be identified for training audience feedback, they provide direct insight into not only the conditions where a problem occurs but also the precipitating factors and decisions.

This paper describes an inferential approach for assessing perceived truth factors in an automated after action review (AAR) system for Marine Corps combined arms training. This approach involves the examination of simulation ground truth data combined with data from human-in-the-loop interfaces and radio voice communications. Because original human inputs are frequently not preserved in simulation data streams, these other sources are essential to an instructional perspective on perceived truth. In simulation based training, perceived truth errors can lead to hidden problem conditions that may not be manifested in the virtual "ground truth." For example, accidental good fortune in the timing of constructive fire events in relation to unit positions in the simulation ground truth may conceal the potential conflicts allowed by decisions based on poor situational awareness. In such conditions a capability to detect perceived truth failures is not only an additional source of training insight, but a necessary element of AAR.

For this application, the objective for the AAR toolset is fundamentally to help human instructors in the tasks of detecting, reviewing, and collecting pertinent information relating to exercise events and trainee decisions. For an instructor, this is not only to further his own understanding of what happened and why, but also to present the information effectively in a debriefing.

BACKGROUND

Perceived truth is a broadly used term that applies in military contexts from the campaign level to urban dismounted operations, as well as non-military arenas. The Allies used misdirection to cause the German high command to expect an attack from Dover, based on an apparent massive troop buildup just prior to D-Day. While this is a widely recognized example of a perceived truth error, the challenges of acting on an accurate assessment of the situation arise equally in the context of individual decisions in a dismounted operation.

Perception of the facts greatly influences actions and outcomes, whether concerning enemy force size, capabilities, and locations, or concerning the positions of adjacent friendly patrol members. This is why it is a perennial training problem to develop situational assessment skills that can help warfighters at all echelons act upon an optimal perception of facts. But this remains a difficult problem because of the inherent requirement for diagnosis of cognitive states.

Situation assessment is commonly considered the process of building situation awareness, the "reception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1995). In practice, perceived truth refers to the dynamic product of situational awareness, more specifically in terms of something that can be theoretically compared with some form of ground truth reflecting reality.

Where the goal is to assess perceived truth factors in training events, a natural expectation is that more data means greater assessment fidelity. Essentially, there is increasing value as available data gets closer to the decision processes of the training audience through the artifacts of their interactions with the training system and with each other.

Indicators for Perceived Truth

Figure 1 shows a notional illustration of the increase in assessment capabilities that comes with increasing quantity and quality of indicators reflecting trainee states of mind.

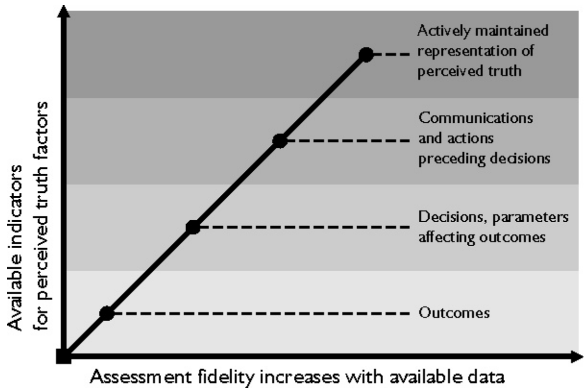

Figure 1. Links between Data and Assessment Fidelity

Outcome Data
At the lowest level, outcomes provide basic information about situations that *may* be attributable to perceived truth errors. For example, a fratricide occurring in a simulation is an outcome that commonly is associated with a perceived truth failure. However, where there is an instructional goal to train situational awareness skills, the detection of such a condition is really only a starting point for further investigation to determine what decisions and actions led to a given outcome.

Decision Artifacts
The next level indicator affording more insight includes decision artifacts, or other parameters that might affect simulation based training outcomes. For example, if a fratricide occurs as a result of a fire mission that is authorized, and if data is available about when and how this authorization was given, then this focuses the analysis. The question becomes, "why did this individual authorize the fire mission?" instead of the more open-ended question in the case based on outcomes, "why did this fire mission take place?" However, to understand the possible contributions of perceived truth in this decision, there are still several open questions. A dangerous fire mission may be approved even with accurate knowledge of friendly positions, when there is a shortcoming in knowledge of procedures or capabilities. For example, a decision maker who makes a calculation error in estimating the size of the danger area for an artillery mission may in fact approve this mission even while fully aware of the actual positions of ground units inside the space of what would be the proper danger area. In such a case, it still remains to be determined if the mission was approved *with* or *without* an accurate perception of ground truth.

Communications Data Preceding Decisions
Communications data can bring the analysis another step closer to determining how perceived truth factors contributed to decisions. Depending on the structure of the training environment, communications data may take the form of spoken interactions on virtual radios, or chat room threads, or other correspondence using command and control devices. The data from trainee actions and communications preceding a key decision can be mined for indicators of situational awareness. For example, high level inferences can be made from simple communications events such as the acknowledgement of a position report, without requiring natural language processing capabilities beyond the current state of the art. Unfortunately it remains true that when someone acknowledges a report, this still doesn't guarantee that the information will also be incorporated into a mental model and also recalled later. Still, as a source of data, this further increases the supporting indicators for what remains a probabilistic endeavor, to diagnose a cognitive state and determine its impact on real or simulated outcomes.

User-Driven Perceived Truth Representations
Possibly the highest fidelity indicator is the data available from maps or other operational representations that a training participant may be required to maintain. This may be impractical or simply unsupported in many training environments, but when present this provides a close approximation to a direct window into the trainee's perceived truth. It should be noted that there is a key distinction between the use of a blue force tracking device that automatically populates and distributes an operational picture, versus a command and control environment that specifically requires the operating picture to be maintained by the training audience. Although the former is coming into greater use in real-world operations, there are still many applications in both training and operational arenas where individuals must apply the skill of manually entering updates to maintain a perceived truth picture. In such applications, if the data can be available for assessment, it serves as a reliable, near real-time reference for the trainee's perception of ground truth, at a granularity level that can often be directly compared with exercise ground truth data. In such

cases, it is possible to have a high degree of confidence in assessments of perceived truth factors in training points.

Additional Assessment Challenges

Overlaid with the availability and forms of data for assessment, there are several additional challenges that complicate the task of arriving at conclusions regarding perceived truth factors in training.

Reliability of the Ground Truth Standard

Even if there is a wealth of data for determining trainee perceived truth states, the process of arriving at instructional conclusions can be problematic if a simulation is used as the only standard for ground truth, against which trainee perceptions will be evaluated. Thus, another challenge with attempting to assess perceived truth factors in simulation based training involves the level of confidence that the analyst can have in the "training version" of ground truth. This encompasses issues with the accuracy of constructive simulation models, the correct performance of roleplayers or training participants in controlling virtual artifacts as guided by the scenario or other participants, and also technical concerns such as network latency and loss.

With complex networked simulations supporting team training events, it is possible for simulation message packets to be lost or delayed on the network. This in turn can result in situations where the simulation version of ground truth published to federates that monitor the exercise actually varies from a "true" or even possible operational condition.

Similarly, simulation models by their nature must be approximations of the behavior of real-world operational entities. If the training audience is performing to a standard based on real-world operational behavior and capabilities, while the simulation model follows a different standard, this can again lead to disconnects. For example, indirect fire projectile models may be implemented in the simulation to follow approximated trajectories for reasons of technical expediency, without necessarily conforming to the same physics-based charts that trainees are instructed to apply. In this kind of case, a simulation may theoretically produce "ground truth" interactions between entity models, that do not correspond to the doctrinal calculations made by training participants in their process of making decisions based on perceived truth.

Perceived Truth for the Individual versus Team

In team training, situational awareness is a team product, both in terms of the way that an individual's perception is formed, and in terms of how a shared team understanding of the situation is maintained. If training assessment goals target individual or team level situational awareness skills in isolation, this can drive the direction for how data is best consumed and presented.

Individual and team situational awareness are linked by the inherent process of building shared models through communications. When an individual acts or makes a decision based on the inputs from another participant, then this impacts how perceived truth assessments should be traced. Although we may formally say that the decision maker has an inaccurate perception of ground truth, this is not so much an error that should be attributed to that individual in a team training context if it is the product of erroneous input from other team members. However, for team training objectives, it is appropriate to treat this kind of case as a failure of team situation awareness in terms of the construction of the shared perception through team process behaviors (Salas, Prince, Baker, & Shrestha, 1995; Stout, Cannon-Bowers, & Salas, 1996). So essentially, the problem becomes one of appropriately interpreting data for assessment in terms of the intended level of training.

Team situational awareness can be defined as an active construction by team members of "a situation model which is partly shared and partly distributed and, from which they can anticipate important future states" (Artman, 2000). The distributed nature of a team perception comes from the practice where individual members may not possess full knowledge of all elements of the present situation, but know which other members do have this knowledge. Teams that engage in cross checking of information and confirmation of communication are generally higher in team situational awareness than those that do not (Bowers, Jentsch, & Salas, 1998). Thus it becomes a key training objective in many team operational domains to assess team level perceived truth factors arising from communication processes. When communicating about problems, effective teams explicitly define the problem or goal, clearly articulate plans and strategies for solving problems, actively seek relevant information about the problem at hand, and explicitly communicate the rationale for selected courses of action (Orasanu & Salas, 1993).

TRAINING APPLICATION

Our training application is an after action review toolset for virtual and constructive Marine Corps combined arms training, developed under the Combined Arms Command and Control Training Upgrade System (CACCTUS) program. Combined arms exercises provide training and rehearsal for coordinating multiple supporting arms with maneuver. Training exercises may involve 100 or more participants at various stations in a single facility, or even more in distributed exercises across sites. There is an emphasis on providing experiential training by requiring the exercising force to perform responsibilities during training events which mirror those during operational actions. Training events require communication and coordination skills in the employment of tactics, techniques, and procedures in support of specific scenario goals.

The AAR toolset performs automated assessment to detect training points and also explore causal factors, where errors often involve perceived truth. With combined arms exercises, the perceived truth factors generally involve blue force positions, target locations, and current or planned fires and movement. Hand in hand with an awareness of fires and movement is an understanding of the proper calculations for danger areas in terms of time and space.

The training system primarily makes use of constructive simulation models, with optional virtual control mechanisms for selected vehicles and entities. Outcome data is available direct from the simulation for analysis. Missions are initiated and executed through a front-end designed to provide an operational interface consistent with Marine Corps operational methods. Actions and orders issued through this interface provide an additional level of analysis data, making decisions and mission parameters available for assessment beyond simulation outcomes. The training architecture also incorporates virtual radios which are recorded and processed in real-time. All of this data is available not only for automated analysis, but also for AAR playback. The radio traffic in particular provides further insight into the content of communications preceding key decisions, and therefore supports assessment of perceived truth factors in decision making.

Within this environment, exercising units may make use of equipment in facilities, or incorporate their own C4I devices. This may potentially be an additional source for perceived truth data in the future, but currently there is no C4I tools data available for assessment in the AAR tools, sourced from an operating picture maintained by the exercising force.

Conflict Categories

For combined arms training, battlespace geometry conflicts are one of the primary outcomes that must be detected and reviewed in AAR. Deconfliction skills are tightly interwoven with situational awareness, and the consequences of failure can be severe, potentially leading to fratricide.

Table 1 describes a set of categories for characterizing conflicts, where these categories are used to organize and refine the analysis of training points related to exercise outcomes.

Table 1. Conflict Categories

Type	Description	Example
Executed	Executed conflicts are the simplest form; these are battlespace conflicts that occur in the actual execution of fires and maneuvers in the simulation, as reflected by simulation data. [actual positions + actual fires] This category is concerned with real-time relationships between ***actual*** simulation ground truth positions and danger areas associated with ***actual*** simulation ground truth fire events.	A fixed wing aircraft is given a mission with instructions to stay above an altitude of 2000 ft. This will result in a conflict with an active artillery trajectory if flown at that altitude and time. As the aircraft mission is flown, it passes through the artillery trajectory in a time interval when a projectile is in the air in the constructive simulation (occurring between a simulation fire event and detonation event). The aircraft's actual executed altitude is inadequate to clear the trajectory, and this is detected as an Executed conflict.

Type	**Description**	**Example**
Approved	Approved conflicts are determined from a combination of advance information about scheduled fires and maneuvers, with real-time execution information about actual fires and movements. [actual positions + potential fires] This category is concerned with real-time relationships between ***actual*** simulation ground truth positions and danger areas associated with ***possible*** simulation fire events.	A fixed wing aircraft is given a mission with instructions to stay above an altitude of 2000 ft. This will result in a conflict with an active artillery trajectory if flown at that altitude and time. As the aircraft mission is flown, it passes through the artillery trajectory in a time interval when the artillery mission is active, yet ***no*** projectile is in the air in the constructive simulation at that moment. The aircraft's actual executed altitude is inadequate to clear the trajectory, but this is not an Executed conflict because of the absence of a projectile. However, because the conflict condition was only avoided by good fortune in the timing of constructive model fires, and did not escape the underlying decision error of the poor mission parameters, this is detected as an Approved conflict.
Predicted	Predicted conflicts are in advance of, and independent of the Executed conflicts. Predicted conflicts are determined from scheduled fires and maneuvers as a whole, considering the complete, *possible* space and time components based on orders issued in the constructive training simulation. [potential positions + potential fires] This category is concerned with one-time predictive relationships between ***current or planned*** simulation entity positions and ***current or planned*** simulation fire events.	A ground unit is already executing a maneuver when an artillery fire mission is executed. Although the current position of the ground unit is not threatened by the artillery fire mission, it corresponds to a constructive simulation model that has been assigned a maneuver route which ends inside the artillery fire danger area. This is detected as a Predicted conflict. If the ground unit is under virtual control with no assigned maneuver route, then the analysis for a Predicted conflict only considers the unit's current position. In Predicted conflict cases, there may or may not be an Executed or Approved conflict later, depending on whether anything changes, such as an intervention to halt the movement or halt the artillery fire.

A primary purpose for distinguishing between these forms of conflicts is to establish a framework within which to describe possible perceived truth errors, and their relationships to outcomes.

In terms of supporting data, Executed conflicts are almost entirely based on outcomes. Therefore, as roughly contemplated in the graph of assessment fidelity earlier in Figure 1, Executed conflict data in isolation gives little insight as to the perceived truth factors that may have led to the undesirable outcome. In the context of simulation-based training, Executed conflicts provide an instructor with information about what happened in the simulation, but frequently little accompanying information about how it came about. This is why the level of confidence in the simulation models can introduce ambiguity.

For example, a typical expectation would be that any Executed conflict always entails an Approved conflict. However, this relies on an assumption that events and outcomes in the simulation always correspond to the intentions and actions of the training audience. There are numerous circumstances where this assumption may not hold. A spurious or corrupted state message for a constructive simulation entity could report its position being inside a danger area, *after* the entity had in fact left that position. This could lead to the detection of an Executed conflict based on the simulation data, where in fact no conflict occurs in "true" ground truth. This would be a case where an Executed conflict appears without an Approved conflict. In such cases, this paired result provides information to an instructor that suggests that this may not be a critical event to debrief, since the absence of the Approved conflict suggests that there was in fact no perceived truth error.

The detection of an Approved conflict without a Predicted conflict may also reflect a unique situation, potentially a shortcoming in situational awareness at the team level as opposed to individual level. In team training events, especially those involving a combination of simulation entities under constructive and virtual control, it is possible for simulation "truth" to unfold in ways that do not correspond to the parameters approved by other participants. For example, using the situation described with the definition of Approved conflicts, suppose that a fixed wing aircraft is assigned a mission altitude staying above 3000 ft, which properly avoids conflict conditions with a simultaneously active artillery mission trajectory. However, in a given exercise, the aircraft is under manual virtual control by a trainee, who flies the route at 2000 ft, leading to an Approved conflict. The Approved conflict is a result of the actual aircraft positions and the full duration of possible artillery fires. An Executed conflict may or may not be detected, depending on the actual timing of the artillery fires intersecting with aircraft positions. But because the aircraft is under virtual control, there is no digital data transmitted for the flight mission instructions, other than the communications where the flight mission order was given. So there is no source data for detecting a Predicted conflict.

The best interpretation for a situation such as this depends on the training objectives. If training goals include team level situational awareness, then it can be concluded that there was a team level perceived truth error in flying the aircraft at the lower altitude, where the controller did not have an understanding of the reasons for the higher altitude requirement. If training goals are targeted for more individual feedback, this would lead to different results for the individual who authorized the mission at the higher altitude, versus the individual who controlled the virtual aircraft.

Applied Example

In order to show how these concepts are applied with assessment methods for after action review, the following discussion steps through an example from the combined arms domain.

Early in a combined arms exercise, a tank platoon begins a movement toward an enemy target, with instructions to give position reports at every 500 meters as they approach their objective position. In accordance with the planned coordination in the fire series, the tanks are to move to a position just outside the danger area for a planned Close Air Support (CAS) attack on the target, and then halt and report position. In this exercise, the tanks are under virtual control. The tank platoon leader makes the error of moving his unit past the halt position and into the danger area, without having given the last position report at the position just outside of this area. As the CAS aircraft approach their target, normally they must be given a "cleared hot" message as a final clearance, just seconds before releasing ordnance. The decision to give clearance relies heavily on the current perception of the battlefield and particularly blue force positions. In this case, both aircraft in the CAS mission are given clearance due to a perceived truth error. This results in a conflict when the tanks are inside the danger area at the time of CAS ordnance detonation.

Automated assessment mechanisms detect these conditions, and pre-package relevant information for instructors. The goal is not necessarily to automate the conclusions with regard to possible perceived truth errors, but rather to help instructors collect relevant information about this set of events for the purposes of debriefing playback, and also any further investigation that may be needed. Typical lines of inquiry for a training point involving a conflict as in this example include:

- **Nature of the training point**. What kinds of conflicts took place (i.e., Executed, Approved, and/or Predicted)?
- **Decision maker actions**. In this example, was the CAS mission cleared? And if so, what were expected versus actual positions of the tank platoon at the time of clearance?
- **Information passed to decision makers**. In this example, were the tank platoon positions reported? If not, we can reasonably conclude that there were perceived truth errors in the decision chain that allowed the tank platoons to come into conflict.
- **Information requested by decision makers**. In this example, if tank platoon positions were not reported when expected, were they elicited?

Representing the Nature of the Training Point

Examining this conflict condition further, it is determined to be both an Executed conflict and Approved conflict. There is no Predicted conflict because the tanks were under virtual control and therefore there was no mission data describing the tank maneuver, which could have been used in an initial Predicted conflict analysis. For this training point, the goal is for the tools to present information about the conflict in both the Executed and Approved contexts.

Once the conflict conditions occur, a training point is automatically prepared and pre-configured for review and playback. It is important to be able 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. Figure 2 below shows the debrief playback for this training point. The relevant vehicle symbols are depicted for both blue and red forces, along with the danger area for the CAS ordnance, and also a representation of the route followed by the tank platoon. At the current time in the simulation playback, the lead tanks can be seen well inside the red danger area at the moment when the second CAS detonation occurs, both in the 3D view and the corresponding timeline representation.

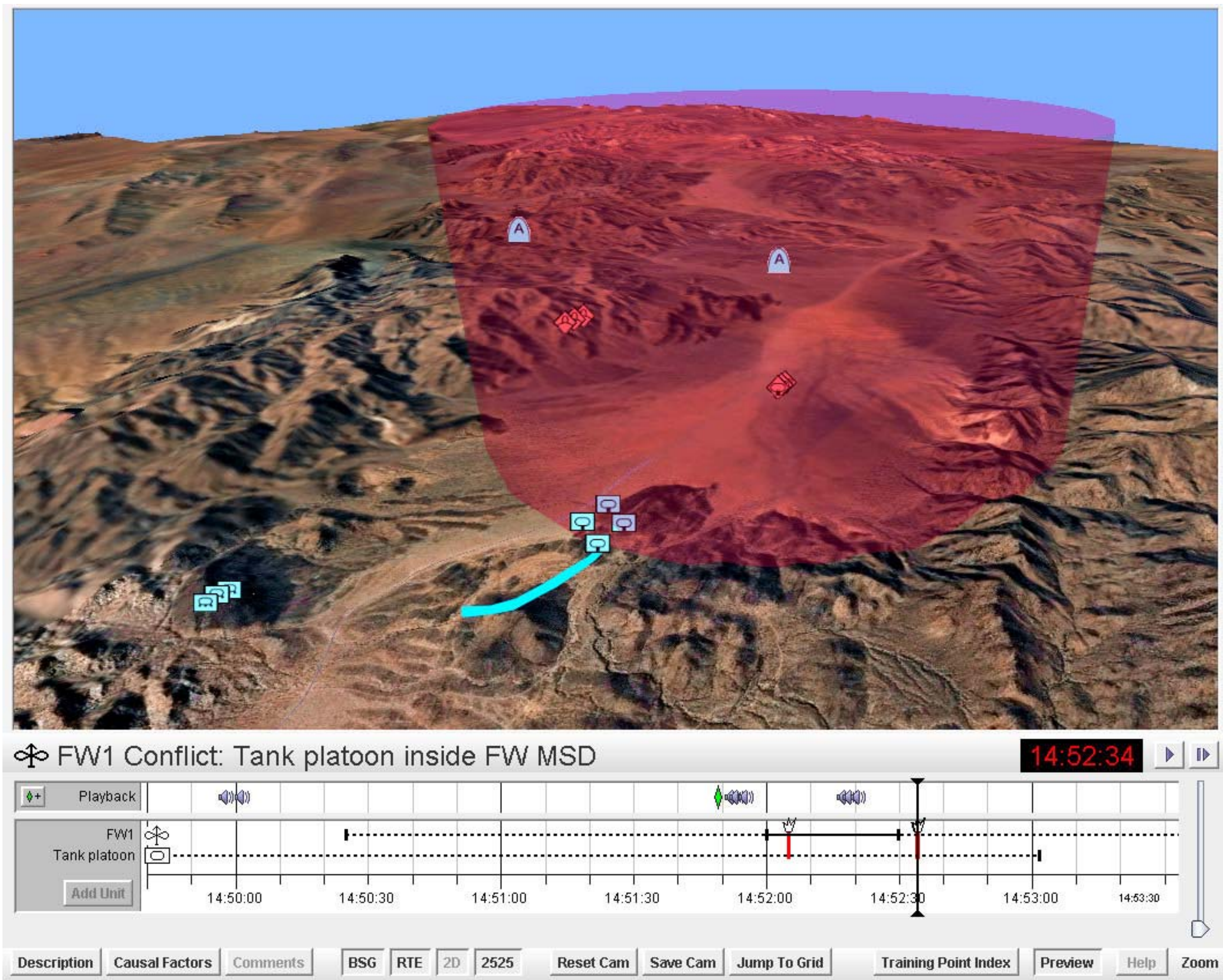

Figure 2. Visual Representation of a Conflict

One important implementation note concerning the detection of Executed conflicts is that there is a careful distinction between using the simulation ground truth standard for playback and using it exclusively for detection. As discussed earlier, there are potential pitfalls in using simulation data in isolation as the ground truth standard for assessment. In order to address this problem, the detection algorithms for Executed conflicts follow a two step sequence. For any detonation reported by the constructive simulation models, which might lead to a finding of a conflict, we make use of the availability of a second source of data from the Marine front-end to the simulation. Any detonation can only be valid if it in fact corresponds to one of the missions assigned via this front-end, so therefore any detonation that does *not* have a counterpart in the human interface data can be discounted. Thus, Executed conflicts are detected through a process of first listening to simulation data and filtering out any potentially invalid fire event data, and then detecting conflict conditions based on only the valid fires. This makes the conclusions about outcomes more reliable than without this filter.

In order to explore further into the nature of this conflict, an instructor can use the timeline representations to see details of when the Executed and Approved conflicts occurred. The following two figures show close-ups of the timelines in the Executed context first, and then the Approved context. The top row in both views is the playback timeline, which will be discussed later. The next two rows represent the simulation activity for the CAS aircraft on top, with the tank platoon underneath. Two CAS ordnance detonations (from two separate aircraft within the mission) are shown 30 seconds apart on the air unit row. The movement times for the tank platoon are shown on the bottom row. Figure 3 shows the representation of the Executed conflict, which occurs instantaneously at the times of the two detonations. Conflicts are indicated by a red link between unit rows on the timeline.

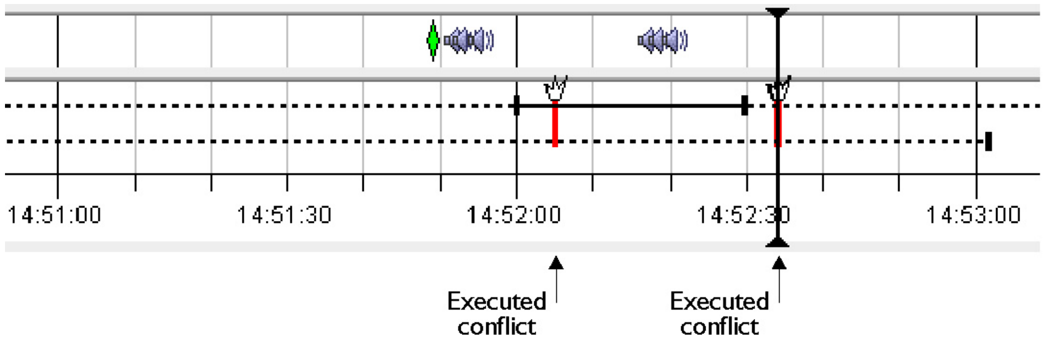

Figure 3. Closeup of Executed Conflict Representation

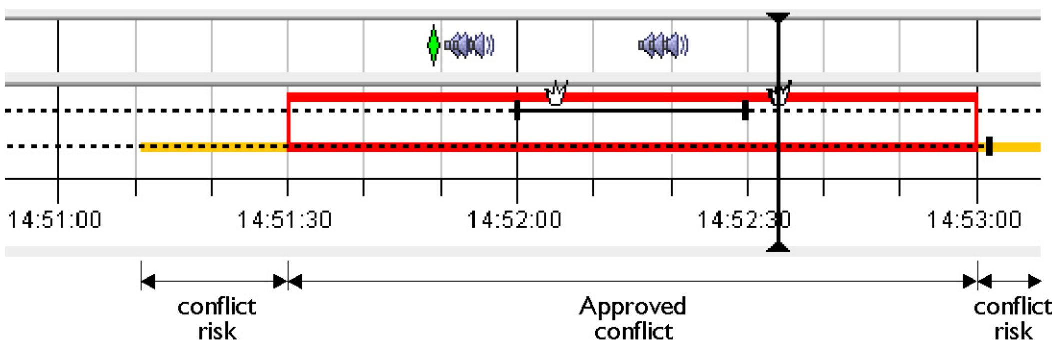

Figure 4. Closeup of Approved Conflict Representation

Figure 4 shows the same timeline view, toggled to the Approved conflict context. The instructor interface allows for quick switching between both contexts. This view contains several key pieces of additional information related to the Approved conflict. First, since the assessment logic has knowledge of all tank positions as well as the size and position of the danger area associated with the CAS mission, we can represent directly on the timeline when the tanks first entered this danger area. This is shown with a yellow highlight on the tanks movement row of the timeline, which indicates any time where their positions reflect a risk of a conflict with a known mission. Also, the depiction of the conflict interval takes on different meaning in this context. Approved conflicts are concerned with the times of possible fires, or more specifically with the doctrinal timing associated with danger areas for specific kinds of missions. These timing rules dictate the notion of when a danger area is considered *active,* irrespective of whether/when actual detonations occur. Therefore, the time interval for an Approved conflict, is represented as the entire duration in which the tank platoon occupied the active danger area.

We can see here on review that the tank platoon entered the danger area approximately 50 seconds before the first scheduled CAS ordnance time on target. This is important information for an instructor, who will also be interested in when CAS clearance was given, if any.

For this application, the notion of Approved conflicts is applied at the team level, as opposed to the individual level. So, an Approved conflict is detected in conditions where the end product from the team is that fires are authorized in the training system. This initial view shows that the performance of this team allowed the tanks to enter the danger area and allowed the outcome of the CAS ordnance detonation. This leads to the next level of review.

Decision Maker Actions and Communications

In this sequence of events, the tank platoon leader's individual error is compounded by situational awareness shortcomings in other decision makers who failed to anticipate the tank platoon positions and request a position report when none had come. In order to assess how a team level perceived truth failure may have come about, the instructor can review relevant decisions and communications using an expanded timeline view. Figure 5 below shows the visual representation of timeline information for this training point, with radio transmissions shown on communications timelines correlated with simulation events.

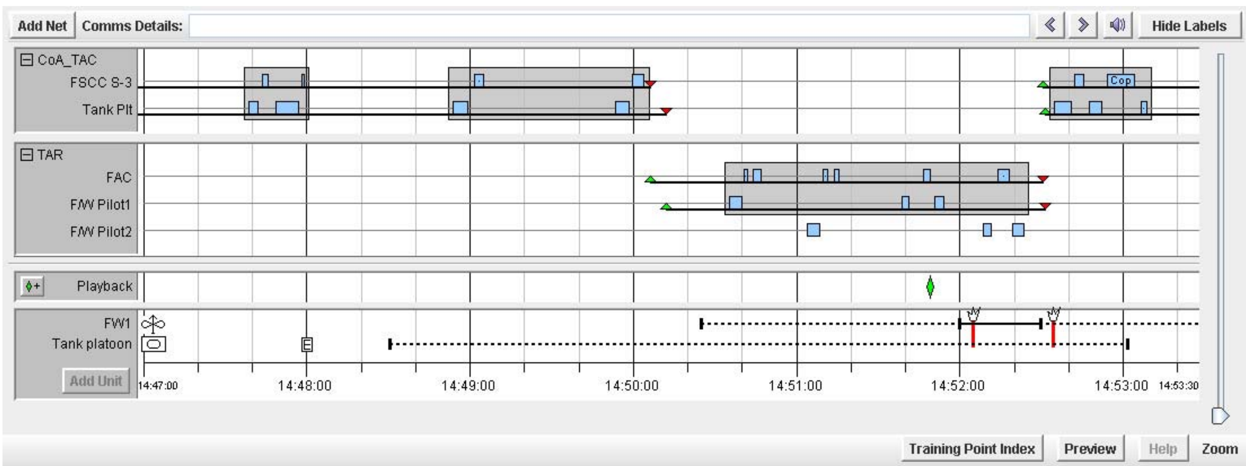

Figure 5. Training Point Timeline Display

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. 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. The timeline symbology represents individual transmissions and groups of transmissions that can be considered dialogs.

Transmissions are automatically analyzed for spoken content containing certain keywords that reflect decisions, with annotations added accordingly. Instructors use the timelines' visual representations and annotations to see what communications took place, and then directly play the audio for transmissions that may provide key insights. In this case, a quick review shows that the tank platoon was reporting position approximately every minute, which corresponds with an expected rate of speed for moving 500m. Their last position report was sent and acknowledged two minutes before the CAS time on target (TOT). Based on time-distance calculations, another position report should have been anticipated roughly one minute before the CAS TOT. However, the absence of *any* communications on the CoA_TAC radio net for a critical interval of over two minutes makes it apparent to the reviewer that this last report was neither initiated nor requested. Although the tank platoon continued their movement beyond the designated stopping point, this was unknown to other team members. This amounts to the fundamental team level perceived truth error in this example.

Reviewing the TAR net, the instructor sees (and hears) that final attack clearance was given to both pilots immediately preceding their respective scheduled ordnance drops. Both clearances were given after the tank platoon was already inside the danger area, as shown in the Approved conflict context of Figure 4. This is the final piece of data for this training point, and the recorded communications for the clearances will be included with debriefing playback. The playback timeline is provided as a place to arrange markers for the contents of the AAR playback for a training point, including radio transmission icons and bookmarks. An initial set of relevant communications to include in the playback is chosen automatically, but this can be customized. So for this training point, the clearance transmissions are simply dragged and dropped onto the playback timeline. Both Figure 3 and Figure 4 show a closer view of the playback timeline, with audio icons included for the CAS clearance communications.

CONCLUSIONS

With the combined arms training application, the available data is not limited to simulation states and outcomes, and therefore this affords a higher level of insight as to perceived truth factors in overall team performance. The goals for automated assessment are also eased by the fact that the objective is to provide tools to help human instructors with the judgments of cognitive states underlying key decision points, as opposed to attempting to automate these conclusions within a tool.

For other training domains that specifically have access to data from tools for command and control or other coordinated operational picture maintenance, there is the possibility for greater fidelity assessment. There is also the opportunity for additional playback features, such as a three dimensional battlefield visualization of perceived truth, based on a direct log of the tracking inputs as maintained by the training participants.

A similar opportunity for greater assessment fidelity is afforded in domains with more constrained communications methods. The radio communications in combined arms exercises do conform to a set of rules in terms of protocols and syntax with certain critical messages such as a 9 line order for an air mission. Yet, there is still a great deal of flexibility, and the practice of proper communications procedures is secondary to other training objectives during an exercise. This leaves considerable room for ambiguity in the communications. Under such conditions, it is possible for the human instructor to leverage an understanding of the domain to make quick diagnoses based on cues from the AAR toolset such as the absence of position reports during a key interval. But it is much more difficult to automate the assessment process with the end goal of reaching conclusions in a cross section of situations. However, in other applications where additional constraints regulate the communications syntax, it may be possible to take the next step with automated analysis. Chat room communications are also increasingly becoming common in team operational arenas, and therefore also in training, and this presents an opportunity in the sense that textual communications eliminate at least one level of ambiguity from the analysis of spoken radio transmissions. Additionally, the state of the art in natural language processing technology continues to advance and show promise for greater insight with automated assessment.

Although the concepts of Executed, Approved, and Predicted conflicts are highly tailored to the combined arms training domain, we hypothesize that these may in fact be useful to structure training points in other domains as well, where the objective is to assess situational awareness products in the training audience. Many simulation based training domains likely share the feature of a distinguishable difference between outcomes indicated primarily by simulation data alone (Executed), versus a combination of outcomes with plans (Approved), versus a pure analysis of planned actions as they are initiated (Predicted). The categorization of detectable exercise conditions into a structure such as this would likely afford similar insights as we found with the combined arms domain, where different combinations suggest different conclusions about perceived truth.

ACKNOWLEDGEMENTS

The CACCTUS effort described in this paper is sponsored by the USMC PMTRASYS through a contract with Cole Engineering Services, Inc.

REFERENCES

Artman, H. (2000). Team situation assessment and information distribution. *Ergonomics, 43*(8), 1111-1128.

Bowers, C. A., Jentsch, F., Salas, E., & Braun, C. (1998). Analyzing communication sequences for team training needs assessment. *Hum Factors, 40*(4), 672-687.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Orasanu, J., & Salas, E. (1993). Team decision making in complex environments. In G. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 327-345). Norwood, NJ: Ablex.

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors, 37*(1), 123-136.

Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (1996). The Role of Shared Mental Models in Developing Team Situational Awareness: Implications for Training. *Training Research Journal, 2*, 85-116.