Automated After Action Review: State-of-the-Art Review and Trends

Cynthia Johnson
School of Electrical Engineering and Computer Science
University of Central Florida
Orlando, FL 32816-2450
cynthiajohnson@sprintmail.com

Avelino J. Gonzalez
School of Electrical Engineering and Computer Science
University of Central Florida
PO Box 162362 HEC 346
Orlando, FL 32816-2450
gonzalez@ucf.edu

Military training relies upon the after-action review (AAR) to provide feedback and instruction to the trainee. After-action review can refer to any number of activities including verbal feedback, review of audio and video recording, and playback of the training session. Training can be directed at an individual or group of individuals. The effectiveness of the review is dependent upon timely and accurate expert assessment. Traditionally, this assessment is made by a human expert; however, the need for this type of expertise far exceeds its availability. Modern intelligent computer systems cannot completely replace human expertise, but there are certainly places where an intelligent system can aid and add to the after-action review process. This paper examines the current state of the research and development in this area. In addition, some recent research in intelligent systems is reviewed for the possibility of use in the automation of military AAR.

Keywords: After Action Review, intelligent systems, assessment

1. Introduction

Timely and accurate feedback is essential in effective training. The United States military uses the after action review (AAR) as the primary method of communicating feedback to trainees. The origin of the modern AAR is believed to be historian S.L.A. Marshall. [1] In his efforts to document war time events, Marshall spoke with many soldiers in theater immediately after combat. His attempts to describe the actual events of World War II, the Korean War and the Viet Nam War from the memories of the soldiers with whom he worked became the first after-action reviews. Modern AAR started with observers scoring army unit performance during training. As training expanded to include virtual and constructive simulations and instrumented live training, it became possible to collect more accurate statistics than those provided by observers with clipboards and a pair of binoculars. [1] Today's AAR is a much more structured event than its historical counterpart. Training simulations have become more sophisticated and so have the tools used to collect data and analyze the training.

As training simulations become more computer-based and automated, the tools to assist AAR are also becoming more automated. This paper reviews the current state of the art in computer assisted and automated AAR based upon the available literature. In addition, research into the use of intelligent systems to further improve AAR automation is explored and examined for opportunities where further automation could be applied to improve the AAR process.

Before applying automation, however, the theory behind the current process of the after action review needs to be understood. It is important to understand that the AAR process is rooted in behavioral science [1] and has a social role in addition to its pedagogical role. As described in "A Leader's Guide to After Action Reviews**"** [2}, the goal is provide feedback on performance during the training exercise and guide the trainees through self discovery of their strengths and weaknesses. Ideally, it is a non-threatening, non-critical report of what happened and how to improve future performance. The trainee(s) should work with their leader to create their own plan for fixing training weakness and emphasizing their strengths. The actual AAR is intended to be an important learning experience for the trainees. The intention is for all those involved to discuss and learn from the information presented at the review. Typically, the training objectives are reviewed along with the intended plan of execution for the training. The goal is to identify areas needing improvement along with identifying areas of strength. In order to aid future training, it is not uncommon for the AAR to be recorded for the trainees to take back to their home station for review. If recording of the meeting is unavailable, the presentation material is often placed in a CD or DVD and made part of their take-home package.

Just as there is no one correct way to conduct an AAR, there is also no one correct way to automate an AAR. There are common requirements to be met in all types of training and AAR. The most important requirement is that the correct lesson be conveyed to the correct trainee at the right time. The steps to achieve this can be further subdivided as follows: [3]

1. The performance must be diagnosed as correct or incorrect.
2. The trainee must then be able to recall the performance.
3. The trainee is presented with examples of good or expert performance.
4. The trainee then must be guided into generalizing the behavior into similar situations.
5. The final division is the assessment of the trainee's competence, which will guide further training. [3]

Current systems automate all or some of these requirements, but no system exists which completely automates the entire process.

Interviews with army observer/controllers at various combat training centers [4] [5] clearly indicate that automation that improves training recall and diagnosis is desirable. Dyer et al [4] noted that “AAR aids should assist the trainer, and should be used when they are “value added”.” [4, p. 51] The statement indicates that automated AAR should be geared toward assisting the trainer and not as a replacement for the trainer. The automated tools currently in use focus on trainee movement, shooting, and aspects of force protection and communication. The biggest lack of automated AAR tools was deemed to be in the area of mission planning and mission preparation. This was not expected by the authors, given that these are important AAR topics in live training and operational environments. [4] Another area where automation is deficient is in determining cause and effect relationships and connections between events.[4] Available AAR automation easily list events, but do not assist the trainer in linking these events to the mission plan.

Intelligent systems may provide a technique for providing automated AAR tools which can fill the gaps left by current tools. This paper reviews the current trends in automated AAR tools, focusing on a sampling of the currently deployed or soon to be deployed systems in use. Also examined is recent research into using intelligent systems to further automate AAR. We explore the current (and available) literature into automated AAR tools and comparing them to the needs and desires of the actual trainers. Through this we identify the areas where new application of intelligent systems would be most beneficial to the AAR process.

Computer Assisted AAR Tools

An effective AAR relies on accurate perception of the training events. As often happens when humans are involved, perceptions can be quite diverse among people viewing the same thing. Because computers and software are available to track or simulate battle participants, it is possible to present the ground truth of the situation in addition to the perceived truth during the AAR session. Once data about participants are available for collection, instructional conclusions can be drawn that are more insightful than what human perception alone permits. These data can be used to point human experts to key events and collect information on less critical training objectives, thereby freeing the human experts for the more critical ones. [6] The AAR systems discussed in this section allow for the presentation of ground truth during and after the event. By doing so, they assist in the first two steps of the AAR process. They allow diagnosis of the performance as correct or incorrect and assist in the trainee(s) recall of the performance. By allowing the trainee and the trainer to see the ground truth of the event rather than just their own perceived truth, the first steps of the learning process are begun.

There are almost as many data collection and playback systems to support AAR as there are simulations. Many simulations log data internally and have added some software to allow access to the exercise events after the simulation is completed. This is not particularly helpful when multiple simulations are linked together for larger exercises. In these cases, some standards for data format and interaction do exist. The two most commonly used are Distributed Interactive Simulation (DIS) as documented in IEEE-1278 standards, and High Level Architecture (HLA) as documented in IEEE-1516 standards. DIS provides an open source standard allowing simulations to communicate over a network using pre-defined protocol data units (PDUs). PDUs are exchanged through existing transport protocol layers. The most commonly used transport protocol layer is Broadcast User Datagram Protocol. HLA, on the other hand, is a more formal architecture where simulations using HLA communicate via a Real-Time Infrastructure (RTI). The RTI provides an application programming interface and a programming library for the simulations wishing to use HLA to communicate. In spite of the commonality provided by DIS and HLA, there are still a large variety of data collection and analysis systems in use today. Both HLA and DIS are primarily used to communicate data necessary to enable a ground truth version of the training event be it virtual or live. They were developed with the intent of passing data between simulations rather than supporting AAR. The most common type of data passed is information on participant location and health. In addition, information on weapons fire and detonation is also communicated. The data are primarily directed towards presenting a common operating picture. This common operating picture can be used in evaluating trainee performance and after the event, in helping trainees recall performance, but can make it difficult to automatically detect more complex behaviors.

Despite the proliferation of AAR tools developed or under development, their functionality is usually very similar. Almost all the available tools provide a basic capability to collect data from the exercise, to display collected information about the exercise on a two or three dimensional map, to play back recorded information on the map, and to generate data reports from the collected data. Most also offer the ability to bookmark or time stamp a particular point of interest during the exercise. While these tools provide an objective viewpoint of the training exercise, they do not provide training feedback or assessment on their own. That portion of the after-action review is still provided by the human instructors and observers. This section details some examples of what is currently available and under development. Examples of AAR systems from live, virtual and constructive training are reviewed. These systems primarily provide support for enhancing the diagnosis of the training and enhancing the recall of the actual events.

An excellent example of a currently deployed data collection and AAR support system is the Distributed After Action Review (DAAR) system developed for the Joint Training Experimentation Program (JTEP). [7] JTEP is a National Guard Bureau program chartered to integrate and improve live, virtual and constructive training systems. The system is deployed in Exportable Combat Training Capability Exercises (XCTC) throughout the US at various National Guard training ranges. The concept of DAAR is to collect data from the exercise, share data amongst distributed exercise sites and support AAR. In a distributed exercise, the AAR processes are led by local observer/controllers. The DAAR provides logging, viewing and playback of data from live instrumentation systems and virtual and constructive simulations linked into the exercise. The playback controls are very similar to those of a common video recorder. The controls include play, pause, rewind, fast-forward and jump to a specified time. The log files can also be captured to a CD-ROM along with playback software to provide a take-home package to the training units. The National Guard continues to work with the developers of the system to improve and add capabilities to the AAR. [7]

The US Army has several AAR systems currently in use and in development for live and virtual training. Although still in limited use, it is important to mention the Live Training Transformation (LT2) product line, whose purpose is to develop a common architecture upon which components and capabilities can be developed. The Common Training Instrumentation Architecture (CTIA) provides the framework upon which a group of reusable components can be built. The goal is to avoid duplication among live training products. Each program using CTIA participates in developing the requirements for the core capability known as CTIA services and then contributes to the library of components which operate within the CTIA framework. The components and core service code are hosted on an LT2 web portal and access is granted to all government programs participating in live training transformation. [8] The AAR capabilities are components built to work with CTIA services data collection capability. The current capabilities of the AAR system include playback on a two-dimensional map, integrating with digital audio/video recording, generating reports from collected data and support for generating Microsoft Office presentations. Several Army programs are currently in development using CTIA as a basis. These include the range instrumentation program, Combat Training Center - Objective Instrumentation System (CTC-OIS), Homestation Instrumented Training System (HITS) and OneTESS. These programs are part of the Army's training transformation plan to instrument their large training ranges through CTC-OIS, allow for portable home station training (HITS) and add more sophisticated training tools for the individual soldiers (OneTESS).

The Army has also developed an AAR tool for use with many of its virtual simulations. Initially developed in the 1999-2002 timeframe, the Dismounted Infantry Virtual AAR System (DIVAARS) caters to the unique training requirements of small dismounted units. [9] It continues to be updated to interface with the latest virtual simulations. The challenge of working with small dismounted units in a virtual environment is primarily visual. Each participant has a unique view of the battlefield that drives their actions. DIVAARS capitalizes on the fact that the training is within a virtual environment and is able to replay the exercise from any viewpoint. In addition to playback and variable viewing modes, DIVAARS is able to provide movement tracks, entity identifiers, digital recording and playback of participant audio program, bullet lines and event data collection and display. All of these capabilities assist in achieving the AAR goal of facilitating trainee understanding of the events of the exercise. [9]

An example of an after-action review tool for a constructive simulation is the CACCTUS AAIRS. [6] The Marine Corps CACCTUS (Combined Arms Command and Control Trainer Upgrade System) is designed to support a combination of live, virtual and constructive simulations working together. The purpose of the simulation is to provide staff level training in the planning and execution of combined arms maneuvers. It works in conjunction with the OneSAF Objective System (OOS) constructive simulation. The AAR portion of CACCTUS is known as After Action Intelligent Review System (AAIRS). The AAIRS consists of two databases and a suite of tools for analysis and playback. As of November 2007, a prototype of the AAIRS toolset implementing a subset of the desired data collection and some simple analysis has been delivered to a Marine Corps training site to serve as basis for feedback from the end users. [6]

The core architecture of AAIRS separates the function of collecting data from the automated analysis of the data. AAIRS collects data from a variety of sources into its two databases. The database known as the playback database collects simulation data from OOS and recorded audio data from the trainees. These data are stored in a format meant to optimize live capture and playback with cueing of the simulated entities. The second database known as the CACCTUS AAIRS Runtime Database (CARD) collects data for analysis. The data collected include unit and entity names and hierarchical structure from OOS, information about missions from the various training devices and concise representations of simulation and communication data. The data are collected and analyzed by components of AAIRS. The separation of the data collection portion of the AAIRS from the analysis portion allows incremental development of the product using agile methodology. The currently deployed prototype collects DIS data from the simulation and speech recognition results from virtual radio communication. [6]

DAAR, LT2, DIVAARS and CACCTUS all have the same basic components of automated data collection, with various capabilities for exercise playback and report generation. For the most part, human input is limited to setup and requests for information in various formats. However, the instructor/observer is still an important factor in creating useful feedback to military trainees, particularly in live training exercises where there are often a number of participants and observers. Unfortunately, it is often difficult to take notes quickly enough or remember key aspects of the training exercise. Therefore, in support of Motorized Patrol Operations (MPO) training for the Marine Corps, a group of performance measures have been developed and defined that would adequately assess the MPO training objectives. [10] These are programmed into an existing assessment tool, SPOTLITE.

SPOTLITE [10] was originally developed for the Air Force to assist in training F-16 pilots. This tool runs on a tablet PC or Ultra-Mobile PC and is intended to assist the instructor/observer in capturing real-time performance data. For each type of training event, predetermined performance measures are loaded based upon the training objectives. The instructor/observer can simply check a predefined rating of the measure or annotate the rating with written comments. The tool groups together measures relating to the same event eliminating the need for the user to search for the measures. These ratings and comments can then be loaded into a central database or viewed on the handheld PC for the after-action review. While the use of PDA's and handheld computers is common for evaluating live training, the SPOTLITE tool gives a common input format to all the assessors and is specifically designed to make the process easier. The tool was designed to provide an interface for entering performance measures and entering pre-exercise data as well as serving as a real-time data entry. This means that it can be configured for any number of training environments where human assessment is used. The pilot test conducted by the Marine Corps showed that both trainers and trainees alike felt that having this computer-based tool enhanced the AAR. [10]

2. Intelligent Systems Research in AAR

The examples above as well as other similar systems have all contributed to increasing the effectiveness of after-action reviews in military training. However, the integration of intelligent systems and automated assessment is still primarily in the experimental stage. This section presents a review of some of the research that has been done towards creating an automated AAR tool which provides training feedback. The systems reviewed in section 2 primarily assist the trainer and trainees in the recall of the events of the exercise. Obviously this is a very important part of the AAR. Without an accurate record of the events, further automation would be worthless. Each of the following systems needs this basic functionality in order to further automate the AAR process.

An experiment was conducted to examine the effectiveness of various training strategies using feedback.[11] The goal was to determine the implications of these strategies on simulation-based training similar to that used by the military. The solution named the Benchmarked Experiential System for Training (BEST) implemented three separate strategies. First, they used mathematical optimization techniques to optimize problem solutions for feedback. Second, a feedback strategy focused attention on principles in the optimized solution rather than the solution itself. The third and final strategy was to leverage observational learning. All the participants performed the same air defense command and control task in a moderate fidelity simulator, but received feedback on their performance using only one of the three strategies. The task required the participant to defend a no-fly zone from enemy intrusion. [11]

Prior to the experiment, near-optimal solutions to the problem presented were developed using BEST. The process used mathematical optimization of scheduling and task assignment to create solutions to be used as examples of expert performance of the task at hand. The problem had enough boundaries to allow the development of near-optimal solutions. The problem was divided into three phases to develop the solution. The first phase allocated targets to assets. The initial solution assigned each target to the closest asset capable of prosecuting the target. The second phase determined the order in which the targets would be attacked. This was a feasible algorithm because there were a limited number of task sequences available in the simulation. The final phase determined a task schedule stating the exact times when targets would be prosecuted by the assigned assets. Each asset was considered a separate entity and scheduled over time. The algorithm used to find optimal task times and associated launch schedules was based on a dynamic programming problem. While developing an optimal task schedule is an intractable problem, iteration among the three phases allowed a near-optimal solution to be found. [11]

During the experiment, all participants were given the same instruction in completing the command and control task. All participated in a planning session that reviewed general strategies of asset allocation. The only difference among participants was the type of post-mission feedback received. The treatment group saw the near-optimal solutions generated by BEST in an animated visualization tool. A human recorded voice-over explained the principles underlying the expert strategies presented. The control group received only a general list of strategies as their feedback. The participants repeated the task after feedback and scores were compared between the two groups. Overall, the groups started out with similar scores, but those viewing the near-optimal solutions as examples improved their scores significantly throughout the training. [11] This experiment clearly showed the importance of step 3 in the AAR process as it presented the trainees with examples of good or expert performance. The results of the experiment showed the value of this step. In existing automated AAR, the only way to do this is by discussion or replaying a previously recorded exercise. In the BEST system, a computer algorithm devised the expert performance shown.

The Explainable AI (XAI) module is a tool developed to enhance learning through reflection during the after-action review. [12] Typically in live training, the trainees are allowed to ask questions of the opposition forces (OPFOR) to fill in the gaps of learning not provided by mission statistics and a list of accomplished and failed objectives. However, in many simulations the opposition forces are automated constructs in a simulation rather than actual people. In this situation, the learning gained through the dialogue with the OPFOR is lost. The XAI allows the user to ask questions of the simulated entities and get natural language answers based on the data generated from the simulation, thus giving the trainees an additional avenue for learning. The XAI has been prototyped for use with the One Semi-Automated Forces Objective System (OOS) and Virtual Humans simulation. [12]

One goal of XAI was to keep the tool as generic and modular as possible so as to avoid building a separate system for each simulation that wished to use it. This required importing the logged data from the simulation into an XAI relational database. It was unclear whether the data were imported during the exercise or at the conclusion of the exercise. Regardless, once the data were in the database, it is inspected for the occurrence of key events. The type of events designated as *key events* must be predefined by domain experts. For the OneSAF Objective System prototype, a key event was defined to be anytime an entity fired a weapon. Once the events are determined, this information is fed into the graphical interface for the user to choose events and questions. The chosen questions generate a query into the relational database and the resulting data are fed into a natural language generator. The language generator formulates an English response using XSL templates. The challenge of building XAI comes in defining and inputting the behaviors of the simulation into XAI. Rule-based representations are easiest to use, but not always available, so the builders had to allow for hand building of the behaviors. [12]

In addition to their work with OOS, a version of XAI for another simulation known as Virtual Humans was built. Virtual Humans is a simulation designed for teaching soft skills such as teamwork, negotiation and cultural awareness. The core architecture from the OOS XAI was used to build the one for Virtual Humans, but new graphical interfaces had to be built to accommodate the differences in type of events and questions between the two simulations. The researchers intend to continue their work towards making XAI more generic and modular to encourage reuse on many different types of simulations. [12] The contribution of this system in the AAR process is to allow the trainee to understand the differences between their performance and that of an expert. By understanding that difference, the trainee is aided in proceeding through steps 3 and 4 of the AAR process, understanding expert performance and generalizing the behavior into similar situations.

The Office of Naval Research sponsored an experiment known as The Debriefing Distributed Simulation Based Exercises (DDSBE) project which was designed to test the effectiveness of various technologies for collecting and integrating performance information for use in the AAR. [13] The experiment was designed around a distributed simulation combining constructive and virtual simulations for analyzing the performance of two E-2C Naval Flight Officers and an F/A-18 Sweep Lead during an air-to-air engagement. The experiment combined the feedback from an automated data collection component and human observations entered into a hand-held tool. While the goal of the experiment was to measure the effectiveness of automated feedback, it also succeeded in developing some innovative tools for automated assessment. [13]

The DDSBE automated collection and debrief system focused on specific, pre-identified events. When the event was detected, the automated system and human evaluators were alerted to watch for particular behavior. Two tools were used for detecting events and evaluating performance. The human observers were given a hand-held tool known as the Virtual Communication Assessment Tool (VCAT) to record their observations and assessments. The devices were networked to Automatic Performance Assessment (APA) software. The APA was also networked to the F/A-18 and E2-C simulations. In addition to receiving status data from the simulations, the APA was capable of recording the trainee's tactical audio communications and taking periodic screen shots of the simulations' tactical displays. Observers equipped with VCAT could also request additional screenshots. The DDSBE system monitored 28 performance measure data items. An example of the monitored items is "Sweep Lead makes external Shot call". Eleven of the measures were assessed automatically, but the remainder relied upon assessments made by the observers on VCAT. [13]

The AAR tool of the DDSBE system had a unique graphical interface. A portion of the screen contained a two-dimensional map showing aircraft positions. Alongside the map was a textual listing with training objectives interleaved with assessments in a chronological order. The training objectives were marked with "traffic light" red, yellow and green symbols indicating performance scores. The red symbol indicated that objective was not met; yellow indicated the objective was partially met and green indicated the objective was met fully. When an objective or related event was selected, any textual comments entered by the observers were available in another portion of the screen. The conclusion of the researchers was that presenting the information graphically as well as textually enhanced the feedback experience and integrating automated assessment with observer feedback was well-received. [13]

One of the roadblocks encountered during the development of the performance measures to be assessed was the lack of certain types of data from the simulations. The backbone of the collected data from the simulations was DIS PDU's. While providing the data necessary to evaluate the position and health of a simulated entity, DIS PDU's do not provide a great deal of information about the steps the trainee's took to cause a change in entity position or health. It was enough of an issue that additional research into automated assessment of the F/A 18 pilot performance was done in a separate but related experiment with a different F/A-18 flight simulator. The flight simulator was modified to output additional data about the pilot's actions that are not part of standard DIS PDU's. These additional data allowed the automated assessment of additional performance measures. [13]

The DDSBE APA and VCAT were not used in the second experiment. Additional data collection and assessment software was developed specifically for the specially constructed simulation. [13] This illustrates a problem with many of the automated assessment tools being researched and used. Very few of these tools are reusable. While the APA and VCAT software frameworks are probably quite reusable, much software would need to be written to do new performance measures. It is doubtful that this type of assessment software will be used heavily before a method for inputting new performance measures is developed that does not require software engineering interface. A second problem illustrated is the need to specially instrument simulations or interface directly with simulation software to get enough data to make certain types of evaluation. While it is possible to automatically assess certain measures with standard DIS PDU data, others require additional data that may not be available without changes to the simulation software. The DDSBE system did prove that it is possible to use an automated system to diagnose a performance as correct or incorrect which is step 1 of the AAR process. While existing systems contribute to this step by providing an accurate picture of what happened, the final diagnosis of a particular task as correct or incorrect is still performed by human experts.

Other research has investigated the possibility of using software agents to provide expert opinions for use in after-action review. In this approach known as SmartAAR, the performance of the trainee is matched against a software agent trained to be an expert in the skills being tested. [14] The personal expert agents were trained using a machine learning algorithm called Genetic Context Learning or GenCL. Once the expert agents are trained, they can be used to extend AAR systems. The expert agent basically "plays" the simulation at the same time as the trainee, and its actions are assumed to be the benchmark for correctness. Then, the actions undertaken by these expert agents can be compared to the trainee and note any discrepancies in behaviors. On the surface, this approach appears similar to the mathematically optimized solution used by the BEST researchers. In BEST, the near-optimal or expert solution was merely displayed to the trainees and did not change based upon trainee performance. If a trainee drastically deviated from the expert solution in BEST at some, the basis for comparison is lost making it difficult to do an effective AAR using that type of expert solution. The SmartAAR approach involves synchronizing the expert agent to the trainee's position periodically in order to keep the compared behaviors relevant. The comparison between trainee and expert agent involves the use of two types of discrepancies. The first is a significant difference in position and movement at a particular time. The second is a difference in tactic. The first is rather easy to determine, but because there are so many different ways to move, the result is often a very coarse filter with many detected discrepancies. Furthermore, a difference in position may not always be of particular importance. Thus, much of the work was centered on determining whether the discrepancy was of importance. [14]

The second type of discrepancy was termed *contextual discrepancy* by the authors. These types of discrepancies are more difficult to detect, but may be of much greater significance than a positional difference. In order to do contextual discrepancy, the system must be able to determine the trainee's intent without asking the trainee. This is where this solution is unique. The authors based their contextual discrepancy discovery on prior research in Context-based Reasoning. [15] In order to perform contextual discrepancy, the expert agent must be modeled in a context-based representation. When this is done, the context of the agent is easily known and the problem then becomes how to infer the trainee's context in a particular situation. The prototype SmartAAR inferred the trainee's behavior by inserting what the authors call *context agents* into the simulation. Built similarly to the expert agents using a context-based representation, these context agents differ in that they are assigned to behave in only one fixed context. One context agent is created for each possible context that the trainee could plausibly be in. In the prototype, no attempts were made to limit the number of context agents based upon what makes sense in a given situation, but there is room left for improvement in this area. The trainee's behavior is compared to each context agent and it is inferred that his/her context is the same one as that of the context agent whose behavior is most similar to the trainee's. The SmartAAR prototype was tested on a tank platoon simulation and was successfully able to determine some contextual discrepancies between the expert agent and the trainees. [14]

The SmartAAR system is an excellent fit for automated AAR because it does not grade the trainee. It points out differences in the performance of the trainee versus that of the expert agent. The trainee can then assess those differences and make an evaluation of his/her behavior during the simulation. This approach is the best fit for assisting the trainers in diagnosing behavior as correct or incorrect which is step 1 of the AAR process. In addition, it provides the ability to recall the performance (step 2) and presents an example of good or expert performance (step 3). Additionally, the comparison with the expert performance can be mapped exactly to the trainees. In systems where the expert performance is prerecorded, small differences in the exercise may obscure the ability to learn from that performance.

One area not currently addressed in automated systems is the ability to automatically translate training objectives into assessable performance measures. These objectives are typically referred to as Measures of Performance or Measures of Effectiveness. One problem is the very high level of these measures and many different ways to accomplish them. An objective in an Army exercise might be "Conduct Tactical Road March". [16] This cannot be assessed by merely tracking the live or virtual trainees to a particular spot. It involves tracking a number of complex behaviors. Currently, tools such as Reactive Information Propagation and Planning for Life-Like Exercises (RIPPLE) for the US Marine Corps [17] assist training developers by breaking down the training objectives into scenarios containing event threads that the observers can assess individually. Typically, the after-action review agenda is based around this list of events and threads. A feature currently lacking is the ability to translate these lists into events that can be detected and analyzed by the automated data collection systems that provide material for the after-action review. Automated detection available is fairly simplistic and is closely tied to the type of data being collected. In a system using DIS, it is easy to detect weapon fire, as one of the commonly used data PDUs is dedicated to reporting weapon fire. DDSBE did automate some detection of events, but still relied heavily on manual input to detect many of the events. [13]

Some research has been done into the development of an approach to breaking down training objectives into events which can be automatically detected .[16] Abbott et al have investigated the development of a Battle Management Language (BML) for use in analyzing training tasks. BML is a suggested grammar and vocabulary for use in task analysis. While not a complete or even an automated step, it does provide an important step in linking training objectives to scenarios to automated event detection. An intelligent system designed to translate BML into event detection would greatly aid automated diagnosis and provide computer assistance for a task that is currently very reliant upon human expertise.

3. Conclusion

The steps necessary to learn from a training event are diagnosis of task performance, recall of task performance, understanding of good or expert task performance, generalization of skills and assessment of the task performance. This paper reviewed literature describing existing and experimental systems for AAR. No one system performed all of the listed steps in an automated fashion, but all assisted in some way with automating one or more of the listed steps. In order to determine where intelligent systems could most effectively be applied to the automating of these steps, it is beneficial to see which steps the reviewed systems automate successfully.

The first step of effective AAR is to diagnose the task(s) performance. In most of the systems reviewed, this is achieved with playback functionality. This is the method used in the DAAR, CTIA, DIVAARS, CACCTUS and DDSBE. Playback is not only useful in diagnosing the task performance, but also assists in the recall of the task performance. If the task was performed satisfactorily, it also provides an understanding of good or expert performance. If the task was performed incorrectly or inadequately, playback still assists in the diagnosis and recall, but does not provide an understanding of good or expert performance. This understanding of expert performance must be conveyed by the instructors or leader of the trainee(s).

The SPOTLITE tool also contributed to the diagnosis and recall of the task performance. By allowing the instructors to enter their observations and assessments in a more efficient, easy-to-digitally-integrate manner, the instructor diagnosis and recall is greatly improved over handwritten notes or memory recall. DDSBE provided not only playback but automated diagnosis through the combination of instructor input into a hand-held computer and automated assessment of pre-defined performance measures. The automation provided by the XAI tool is also primarily rooted in diagnosis and recall. By giving the chance to question automated forces and examine the events that led to certain types of behavior, it gives the trainee an opportunity to place their own behavior in a different context and provides an additional learning environment.

The BEST system and SmartAAR both provide an example of expert performance, but in drastically different ways. The BEST developers came up with a mathematically optimized solution to the simulation they were using. The goal of BEST was to examine methods of feedback, not to derive a generic solution to providing expert feedback. The method used was effective for the experiment, but not easily reused for a different simulation. Each situation would need to be approached individually and could be quite time consuming. The SmartAAR is a better choice for a solution that could be more generally applied to providing expert performance and diagnosis in an AAR situation. While developing expert agents could in some cases be difficult, their approach to building them automatically goes a long way to minimize the effort. This is definitely an area where intelligent systems could be applied to the automation of AAR

It is a bit more difficult to provide automation in helping the trainee to generalize the knowledge gleaned from an AAR. This step in the AAR process is primarily supported by the group discussion portion of the AAR. The trainee is given the opportunity to discuss the simulated or live training event with peers and hopefully, an instructor/observer with some expertise in the area. Other than the XAI, none of the reviewed systems really automate this portion beyond possibly presenting playback of expert solutions to similar problems. Since an important part of military operations is the ability to effectively communicate and collaborate with team members, it is probably not worthwhile to attempt an automation of this portion of the AAR. The goal of the automated assessment should be focused on easing the workload of the instructor as well as enhancing their ability to guide the trainees into learning the lesson. The same could be said of assessment. The DDSBE and the SmartAAR provide partial assessment of the trainee's performance, but it is the human instructor who determines in which areas the student should receive additional instruction and training.

Existing AAR systems have mastered the ability to record the ground truth of an exercise and replay it in support of recall and diagnosis. However, systems which can further automate diagnosis and present an expert performance are still in the experimental stage. Improving performance diagnostics and presenting expert performance are where intelligent systems research could have the most significant impact on automating the AAR process. In the diagnostic area, an automated means of linking training objectives to exercise events is something that is currently lacking. For providing an expert performance, SmartAAR has the most potential for reuse. Because SmartAAR uses trainable intelligent agents, the research done is expandable and can be transferred to different simulations. XAI was also designed such that the framework is applicable to more than one type of training event. It is critical that future research into automating AAR include this flexibility to transfer the research to a variety of different types of training.

Future research into automating AAR should concentrate on providing a means to tailor an expert performance to each individual event. Of the reviewed systems, SmartAAR is the only system which met this goal. The BEST system provided an example of expert performance, but it was inflexible. The expert agents in SmartAAR respond to the exact same stimuli as the trainee. Unfortunately, SmartAAR research is currently limited in scope and more experimentation is needed to ensure that the solution is scalable to large scale live and virtual training simulations. Similarly, DDSBE showed that it is possible to automate diagnosis of a particular type of performance, but further research is needed to see if this type of automation can be expanded to other types of training.

Research [4] clearly indicates that the trainers support further automation of AAR. However, it is clear that the automation should support not replace the trainer. The most glaring lack in current systems is the inability to automatically link exercise events with predefined training goals.[4] Of the systems reviewed, only Abbott's BML research [16] has even addressed this issue. This type of linkage is not a simple task. Typically, the training goals are high level statements of complex mission goals. One such goal, "Conduct Tactical Road March", involves a series of small goals which must be met before the overall goal is met. BML is a start at breaking down these high level goals into detectable behaviors. Additionally, research into detecting complex behaviors during the event is needed to achieve this task.

Further research with SmartAAR context agents could be useful in this area. In addition to comparing detected context to an expert agent for diagnosing performance, it may also be possible to detect complex behavior based on context.

The content of modern AAR meetings has become much more automated since Marshall's work during and after World War II. Automated data collection, playback tools and multimedia presentations give the trainers advantages and insights that their earlier counterparts did not have. However, the goal of understanding and learning from experience remains the same. It is possible to continue to increase the level of automation in the AAR process without sacrificing the human touch necessary to achieve the social goals of the AAR. This increased level of automation could be achieved with further research into using intelligent systems to provide expert solutions and automate detection of key training events. This type of research can only improve the quality of military training and the AAR process.

4. References

[1] Morrison, J.E., & Meliza, L.L. (1999). "Foundations of the after action review process (ARI Special Report 42)". Alexandria, VA. U.S. Army Research Institute for the Behavioral and Social Sciences. AD 368651.

[2] Department of the Army (1993), "Training Circular 25-20, A leader's Guide to After-Action Reviews", available on 7 July 2008 at: http://www.au.af.mil/au/awc/awcgate/army/tc_25-20/table.htm

[3] Wiese, E., Freeman, J., Salter, W., Stelzer, E.., & Jackson, C. (2008). Distributed After Action Review for Simulation-based Training. In D. A. Vincenzi, J. A. Wise, M. Mouloua & P. A. Hancock (Eds.), Human Factors in Simulation and Training.

[4] Dyer, J.L., Wampler, R.L., &Blankenbeckler, P.N. (2005). "After action reviews with the Ground Soldier System. (ARI Research Report 1840)". Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. ADA438040.

[5] Salter, M.S., & Klein, G.E. (2007). "After Action Reviews: Current Observations and Recommendations (ARI Research Report 1867)". Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

[6] Chen, D. , Jensen, R., Bascara, O., Harmon, N. (2007). "Enabling Automated AAR Development by Abstracting Data Collection from Analysis"**.** Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2007 Paper No 7410, Dec 2007.

[7] Ford, R., Shockley, J., Kobryn, C. (2007). "A Proposed Distributed After Action Review (DAAR) Standard Based on the Joint Training Experimentation Program (JTEP) DAAR". JTEP Technical Report 07F-SIW-096, September 2007. Available on 10 July 2008 at: http://www.jtepforguard.com/pubs/07F-SIW-096.pdf.

[8] Rivera, J., Samper, W., Clinger, B. (2007). "Applying the Live Training Transformation (LT2) Software Reuse Strategy to the Homestation Instrumentation Training System". Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2007 Paper No 7200, Dec 2007.

[9] Meliza, L., Goldberg, S., and Lampton, D. (2007). "After Action Review in Simulation-Based Training"**.** Technical Report (ARI Research Report). U.S. Army Research Institute for the Behavioral and Social Sciences, July 2007.

[10] Wiese, E., Nungesser, R., Marceau, R., Puglisi, M., Frost, B. (2007) “Assessing Trainee Performance in Field and Simulation-based Training: Development and Pilot Study Results”. Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2007 Paper No 7160, Dec 2007.
[11] Shebilske, W., Gildea, K., Freeman, J., & Levchuk, G. (2008). “Optimizing instructional feedback strategies: A Benchmarked Experiential System for Training (BEST)”. Theoretical Issues in Ergonomic Science, Special Issue on Optimizing Virtual Training Systems.
[12] Core, M., Lane, H., van Lent, M., Gomboc, D. Solomon, S., Rosenberg, M. (2006). “Building explainable artificial intelligence systems”. Proceedings of the 18th Conference on Innovative Applications of Artificial Intelligence (IAAI06), 2006, pgs 1766-1773.

[13] Radtke, P., Johnston, J., Biddle, E., Carolan, T. (2007). “Integrating and Presenting performance Information in Simulation-Based Air Warfare Scenarios”. Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2007 Paper No 7199, Dec 2007.
[14] Fernlund, H. Gonzalez, A., Ekblad, J., Rodriguez, A. (in press). “Trainee Evaluation through After Action Review by Comparison”. Journal of Defense Modeling and Simulation.
[15] Gonzalez, A. J. and Ahlers, R. H. (1998) "Context Based Representation of Intelligent Behavior in Training Simulations” Transactions of the Society for Computer Simulation International, volume 15, no 4, 1998.
[16] Abbott, J., Hieb, M., Gustavsson, P., “BML Development by a Task Analysis Methodology”. IEEE Spring Simulation Interoperability Workshop, 2007.
[17] US Marine Corps (2008). “Program Manager for Training System (PM TRASYS)”. Available on 21 July 2008 at: http://www.marcorsyscom.usmc.mil/trasys/trasysweb.nsf/All/D51714977CEEE7F3852573B10054F11C.

Author Biographies:

Cynthia Johnson *is a doctoral student in Computer Engineering at the University of Central Florida. She has a BSEE from the University of Miami and an MSE in Computer Engineering from the University of Central Florida. She has worked as a systems software engineer on several modeling and simulation projects including JSIMS and OneTESS. Her research interests include intelligent systems, software engineering and simulation.*

Avelino J. Gonzalez *is a Professor in the School of Electrical Engineering and Computer Science at the University of Central Florida (UCF). Dr. Gonzalez received his Bachelor's and Master's degrees in Electrical Engineering from the University of Miami, in 1973 and 1974 respectively. He obtained his Ph.D. from the University of Pittsburgh in 1979, also in Electrical Engineering. He spent 12 years with Westinghouse Electric Corp. in Pittsburgh, PA and Orlando, FL before joining the faculty at UCF.*