Utility Assessments of Soldier-Worn Sensor Systems for ASSIST

Michelle Potts Steves
National Institute of Standards and Technology
100 Bureau Drive, Gaithersburg, Maryland, USA
msteves@nist.gov

***Abstract*—Utility assessments were performed for the Defense Advanced Research Projects Agency (DARPA) Advanced Soldier Sensor Information Systems and Technology (ASSIST) program [1]. This paper describes the field-based, formative methods used to assess utility for prototypical software designed to provide value to mission reporting for infantry and related intelligence operations. While results from these evaluations are not presented here, design considerations, evaluation procedures and metrics, as well as, lessons learned are described.**

I. Introduction

The Advanced Soldier Sensor Information Systems and Technology (ASSIST) program [1] is a Defense Advanced Research Projects Agency (DARPA) advanced technology research and development program. The objective of the ASSIST program is to exploit soldier-worn sensors to augment a soldier's recall and reporting capabilities to enhance mission reporting. The program consists of two parts, called tasks. Task 1, Baseline System Development, stresses active information capture and voice annotation exploitation. The resulting products from Task 1 are prototypical, wearable capture units and the supporting operational software for processing, logging, and information retrieval. Task 2 is Advanced Technology Research, it stresses passive collection and automated activity and object recognition. The anticipated results from this task are the algorithms, software, and tools that will undergo system integration in later phases of the program.

The National Institute of Standards and Technology (NIST), along with NIST subcontractors, Aptima and DCS Corporation, are funded to serve as the Independent Evaluation Team (IET) for Task 2. The IET is tasked with assessing the capabilities and developmental progress of the funded research systems. To that end, the IET identified two major assessment categories, technical performance and utility, that when combined would provide an overall assessment of value to the war fighter. Two very different types of tests were designed and administered to address these assessment objectives.

This paper focuses on the methods used by the IET to assess utility to the war fighter for these prototypical systems (Task 2). In a nutshell, utility was assessed by identifying soldiers' information needs with respect to mission reporting and intelligence product development, and then having the soldiers rate how well each information need was addressed by ASSIST contributions and the relative need for that information. To date, two evaluations for ASSIST Task 2 teams have been performed, one in November 2005, a baseline test, and one in May 2006, a go/no-go (program) assessment. Each of these evaluation periods combined technical performance tests as well as assessments of utility. Descriptions of the technical performance tests can be found in [7]. This paper discusses how the IET designed and executed the assessments of utility provided by the ASSIST systems.

First, the paper presents some background on the ASSIST program and on assessing utility in prototypical software applications. Section III discusses the development and vetting of mission scenarios and environments, called vignettes. Section IV provides an overview of the test procedures used, while Section V presents a sampling of the metrics used and types of assessments made. Section VI concludes the paper with lessons learned.

II. Background

A. Expected use of ASSIST systems

Soldiers perform missions of various types including presence patrols (where soldiers are tasked to make their presence known in an area), reconnaissance, apprehension of suspected insurgents, and so on. These missions can often be long and stressful. Regardless of the mission type, after a mission is complete, a unit, such as a platoon, typically provides a report or debriefing that describes any "events" encountered during the mission and collection for information requests that the unit was tasked with during the pre-mission briefing. Due to many factors, including human stress and fatigue, there are undoubtedly many instances in which important information is not captured in the report and thus is not available for use, such as planning for future missions.

The ASSIST program is addressing this challenge by supporting research in the instrumentation of a soldier with sensors that can be worn directly on a uniform. These sensors include microphones, video cameras, still cameras, Global Positioning Systems (GPS), Inertial Navigation Systems (INS), and accelerometers. The intent is to record continuously what is occurring around the soldier while on a mission. When soldiers return from a mission, the collected sensor data is run through a series of software systems which indexes it and creates an electronic chronicle of events that happened throughout the time that the ASSIST system was recording.

ASSIST, Task 2 system capabilities include:

- Image/video data analysis capabilities, including: object detection and image classification through analysis of video, imagery, and related data sources, Arabic text translation from image data, and change detection in related images over time
- Audio data analysis capabilities, including: sound and speech detection and recognition, and shooter localization and classification, e.g., origin of shots fired and identification of the weapon producing those shots
- Soldier activity data analysis capabilities, including mapping a soldier's path during a mission and activities along that mission path.

There is no single integrated ASSIST system at this point in the program's life-cycle. Instead, several university and corporate research and development organizations have formed into "research teams". Each organization is developing specific technology components, and these components are gradually being integrated as a "research team" system. See [5] for more detailed information on the program objectives and Task 2 system capabilities.

B. Evaluation and Utility Assessments

Several important benefits for research and development programs can be gleaned from evaluation. First, evaluations help program managers determine what progress is being made. These assessments can be used to obtain appropriate funding for programs that show progress and to determine alternative directions for programs that are less successful. Second, evaluations help researchers to see objectively how their software can help end users, and if necessary, help them refine their research goals. These more formative evaluations provide end-user feedback during pre-release software development phases geared towards showing current utility and utility enhancement opportunities.

In the early stages of software research and development, it would seem that effort spent on improving the utility of an application provides a greater long-term return than perfecting the overall usability of an application that may not provide increased utility, i.e., value, to the end-user. Given this premise for prototypical research software, it is appropriate to place an emphasis on assessing utility, i.e., the value the software application provides to the user, rather than focusing on training, user interface issues, and the like. However, *utility* and *usability* are certainly intertwined. Many have stated that usability today is multidimensional; it encompasses effectiveness, learnability, flexibility, and user attitudes towards the application, e.g., [2, 3, 4]. So, while we are certainly concerned with the eventual usability of the interface and usability issues that impact our assessments of utility of these prototypical applications, these other aspects of usability need only be "good enough" to assess the value the application provides currently and identify opportunities to increase the application's utility to the end user. With this in mind, the evaluations for this effort were designed to be more formative than summative. That is, the evaluations were performed during design and development of the software applications with the intent of informing the design rather than summative or validation evaluations that are performed at the end of development [6].

The IET was interested in assessing utility to the war fighter; as such, we were concerned with impacts on both their processes and products. To reflect this perspective, user-centered metrics were employed. We attempted to identify metrics that would help assess such questions as: What information do infantry soldiers want and/or need after completing a mission in the field? How well are information needs met, both from the soldier perspective and the S2 perspective? What were the ASSIST contributions to mission reporting with respect to user-stated information needs?

III. VIGNETTE DEVELOPMENT

The vignette tests were designed to assess the value of ASSIST systems in 1) infantry squad reporting of critical information, events, and intelligence encountered during a mission, and 2) S2/intelligence operations. Additionally, the following design requirements were considered:

- Exercising the ASSIST technologies within their operating constraints
- Incorporating some environmental characteristics beyond the current system capabilities (i.e., establishing baseline system performance for comparison in future tests)
- Providing an operationally relevant environment within which infantry procedures could be executed safely but in a reasonably realistic manner
- Establishing a rich information environment (e.g., population dynamics, multi-sensory stimuli, realistically evolving terrain, etc.) to support intelligence collection and analysis procedures for the current mission
- Defining methods for the evolution of the information environment over time (i.e., supporting comparability of current missions with future missions)

A. Mission Scenarios

To support these assessment concerns, two types of scenarios were employed. The first type engaged soldiers in realistic, albeit short, missions, conducted at a Military Operations in Urban Terrain (MOUT) site, where the ASSIST technologies were used to "shadow" soldiers as they performed their missions, and an S2 officer conducted debriefings post-mission. The second type of test used ASSIST data collected from prior "missions" to assess the ASSIST contribution to another aspect of S2 responsibilities, specifically, data-gathering for a strategic product (actual production was not the focus here).

To date five scenarios have been developed and employed in ASSIST evaluations. Four of those scenarios involved data collection on the MOUT site, while the fifth used previously collected data. The scenarios are as follows:

- presence patrol with deliberate search (2)

- presence patrol leading to a cordon and search
- presence patrol and improvised explosive device (IED) site reconnaissance
- assessment of local situation with respect to an upcoming election

B. Mission Environments

To further facilitate obtaining an operationally-relevant test environment, the in-play area of the MOUT site was set-up with objects, persons, and background sounds, whose placement, behavior and occurrences were scripted. The purpose of this was to provide an environment that would exercise the different ASSIST systems' capabilities as they detect, identify and/or capture various types of information. The IET included all of these elements in the vignettes: foreign language speech, Arabic text, shots fired, vehicular sounds, soldier states, soldier locations (both inside and outside of buildings), objects of interest including vehicles, buildings, people, etc. In contrast, the soldiers' actions were not scripted as they moved through each exercise, with the intent of having the soldiers act according to their training and experience.

C. Vignette Vetting

The vignettes (each scenario with its supporting environment) were developed with the intent of exercising the ASSIST technologies within their operating constraints, while maintaining operationally relevant procedures in execution of those scenarios. Various Subject Matter Experts (SMEs) were consulted to ensure that the scenarios would be as accurate and realistic as possible while still allowing for maximum opportunities for system data capture and assessment. The IET consulted SMEs who provided the following perspectives, all with foreign deployment experience: commander, S2 (intelligence officer), platoon and squad leaders.

D. Details for an Example Vignette

Space constraints do not allow a detailed enumeration of all the particulars for each vignette. However, to provide some insight into the various aspects of a vignette, an overview is provided for one of the vignettes which mimicked a presence patrol.

The presence patrol included leaving a forward operating base (FOB) to patrol a local village, make the military presence known, and collect intelligence on the village and/or villagers before returning to the FOB. In this vignette, the soldiers were instructed via their pre-mission briefing, to conduct a presence patrol in the market area of the village and then conduct a deliberate search of the factory area.

During the vignette, the following activities occurred in the mission environment:

- the market area was crowded with shoppers
- a group of locals was engaged in a soccer match in the open space
- two mechanics worked on a car at the auto shop
- a group of factory workers finished eating lunch at a café and returned to work loading boxes on a truck
- two electricians strung wires around the village
- one insurgent covertly monitored the activities of the squad of soldiers
- foreign aid workers attended to an ill villager at the clinic
- a delivery man delivered packages
- a 6-vehicle convoy traveled by the village area
- a local villager rode around the area on his bicycle taking interest in all the activities
- Iraqi music was played at various locations throughout the village and an Arabic documentary movie was shown at the village café

The environment was envisioned to reflect a typical village setting in an area where U.S. forces were providing stability and reconstruction support. The environment included:

- Population groups, including simulated local villagers and shopkeepers, outside businesspersons and workers, foreign aid workers and contractors, soldiers, and insurgents. Actors were assigned to groups and given specific roles, e.g., "you are a local villager who operates a small clothing shop", "you live in another town but were hired by a cousin to operate an auto repair shop in this village". Actors were instructed to maintain an appropriate attitude relative to other groups, e.g., "you are friendly with everyone", "you like other villagers but dislike all the outsiders, foreigners, and soldiers that have come into your village", etc. Since language identification was a specific ASSIST capability, great care was taken to simulate a (near-)realistic language environment. Actors that were able to speak a foreign language were assigned to an appropriate group, e.g., three German speakers were assigned as foreign aid volunteers who worked at a health clinic. Those actors who did not speak a foreign language, were taught several phrases in Arabic, e.g., "I don't speak English", and assigned to roles that required minimal speech, for example, "you are a playing soccer with a group of friends". All actors were instructed to speak in an manner consistent with their role.
- Terrain, consisting of a market area with several shops, a factory building, a clinic, an auto repair shop, an open space used as a forum for gatherings and/or make-shift athletic field, and a construction zone. The terrain was populated with objects, signage, vehicles, and other "set dressing" objects. For example, a vehicle was placed on ramps next to the auto shop, with tires and car batteries stacked outside the shop.

IV. Testing Procedures

This section describes the test procedures for assessing the utility provided by the ASSIST system in infantry mission reporting and S2/intelligence operations, as well as the supporting activities used to facilitate a successful test.

A. Test preparation

In addition to defining the scenarios and populating the environment with objects and actors, further preparations were required prior to the start of vignette tests. The IET took steps to familiarize test participants with the test procedures prior to participation in these tests, ensure data capture for systems with immature user interfaces, and to capture of "ground truth" of what happened during each mission. These preparations are described in the following subsections.

1) Test procedures familiarization: Due to time and resource constraints during each week of testing, it was impractical to schedule a "practice" vignette test, as these tests ran 7-8 hours start to finish. In an ideal world, the IET would have scheduled a practice vignette so that all participants would have had direct experience to inform their understanding of the sequence of events, along with their roles and responsibilities during each aspect of a vignette test. To avoid having the first vignette degrade into a practice run, the IET took steps to prepare the test participants. Prior to the first vignette test, several activities were conducted that were intended to familiarize participants with various aspects of the test. They included:

- Research teams were given the mission report (MR) template. This provided the teams with an opportunity to see what types of information the squad would be asked to provide at the completion of each mission.
- Research teams were briefed on the general flow of the testing procedures, with the time allowances for each segment.
- The S2, Squad and Fire Team Leaders were briefed on the testing procedures.
- Squad and Fire Team Leaders were briefed on the mission report template, e.g., what information they would be expected to provide in their mission report.
- Research teams briefed the soldiers on the capabilities and information their systems could be expected to provide to augment after-mission reporting.
- Researcher-soldier shadow assignments were made and each pair was given training and practice time on 1) how the soldier would move and give direction to the 'shadow' and 2) how the researcher should 'shadow' the soldier – see next subsection, IV.A.2.

2) Shadowing: The research teams prepared "soldier-worn" systems, allowing each system to capture data from the soldier's perspective. Some of these systems did not require any user interaction, but some did. To reduce data capture failure due to inadequate operator training for these prototype systems and possible system failure due to soldier-user of unhardened systems, a soldier-researcher shadowing tactic was employed to provide the best opportunity for the systems to gather their data during the mission. This procedure called for a researcher to "shadow", i.e., follow closely, their assigned soldier throughout each vignette. Soldiers directed their shadow's attention to pertinent data elements, allowing the researcher "shadow" to operate any manually-activated capture features of the system. Use of this tactic reduced the probability of user error contributing to poor data capture, and reduced the time and expense for the researchers to provide good user interfaces as well as the need to harden their systems for soldier use.

The IET carefully developed the shadow assignments of the systems to soldiers within the fire teams to ensure that systems would be afforded as much opportunity as possible for exposure to activities, events, and/or objects that their systems were designed to detect. The assignments were designed to allow for maximum opportunities for the systems to collect data while maintaining the integrity of the vignette scenario in an operational setting. As mentioned in the previous section, a training session was provided to work on the mechanics of shadowing for each soldier-shadow pair prior to the vignette tests. Additionally, each pairing assignment was maintained for the entire week of testing.

3) Ground-truth capture: The IET captured "ground truth" data of what happened during each vignette mission. This data was captured to document the actual events during each mission, with time information, in the event that any questions were to arise regarding what actually happened or data capture opportunities each research team had during each vignette exercise. Multiple methods to record the ground truth were employed.

On the MOUT site, ground truth data was gathered in the following ways:

- Observers (IET) provided targeted data collection on specific events such as time of gunfire or explosions, convoy passage, a fire team entering a building or interacting with villagers. These observers were trained before each test week on their individual collection responsibilities.
- Video and audio capture of the mission from each fire team's perspective.
- Elevated-perspective video capture of the squad's movement through the MOUT site and inside buildings.
- Maps and still images of the environment that captured object and person placement at the start of each vignette

During the mission-reporting sessions following each mission, data was collected in the following ways:

- Audio recordings of the discussion were collected throughout each mission reporting session.
- An IET observer performed targeted data collection including noting information needs, soldier reactions during discussion and the semi-structured interviews.
- An audio recording synchronized with screen capture was collected for each research team's interactions with the soldiers.
- All information needs identified by the soldiers were recorded by IET members.
- Each soldier was asked to rate each research team's contribution with respect to each information need identified, both during production of the "naked"

mission report and while interacting with the research teams to review the mission information.
- Observer notes from the semi-structured interviews.
- Ontology data: Members of the IET noted which ASSIST data elements appeared to address the soldiers' information needs. These data elements were then tagged for use with the ontology.

B. ASSIST Utility in Infantry Squad Operations

For the vignettes in which a mission report was completed, the following procedures were used:

- A "simulated squad" of soldiers, comprised of two fire teams, with researcher 'shadows', ran through an operationally-relevant scenario on the MOUT site. The squad leader was provided with a pre-mission briefing. The squad leader was instructed to conduct the mission in the manner he deemed most appropriate.
- Upon completing the mission, the squad produced a mission report.
- Soldiers were asked to identify their information needs with respect to producing their MR, e.g., information they would have preferred to include in their MR but did not recall. ASSIST research team members were permitted to observe the soldiers identify their information needs.
- Each research team shared its processed data with the squad. Each soldier was asked to rate the importance of each information need and how well each ASSIST technology addressed each need. Additionally, the soldiers were encouraged to ask questions of the researchers to explore if and how the ASSIST systems produced data that might meet their previously-identified information needs, as well as, any new information needs that were uncovered during the ASSIST information reviews, e.g., newly-identified things that the soldiers would include in their reports.
- The soldiers participated in a semi-structured interview to get at more overall impressions from the exercise and ASSIST systems. The interview facilitator focused discussion on assessing if and how the mission report produced by the squad would be different if the soldiers had been given access to ASSIST system functionality.

C. ASSIST Utility in Intelligence Operations (S2 Level)

Additionally, an S2 (intelligence officer) evaluated ASSIST systems using the following procedures for vignettes in which a mission occurred:

- Following the mission, the S2 was provided with the pre-mission briefing (as appropriate) and the mission report produced by the squad. (Note: the S2 was not allowed to observe the actual mission.)
- The S2 was asked to identify information needs, e.g., information that would improve situation awareness, information about critical events, individuals, or situations, etc.
- The S2 interviewed a member of the squad[1]. During this interview, the S2 was encouraged to discuss his information needs with the fire team leader. The S2 was asked to rate the importance of each information need and how well his interview with a soldier addressed each need.
- The S2 met with representatives of the research teams, and shared his information needs with the researchers.
- Each research team shared its processed data with the S2. The S2 was encouraged to discuss his information needs with the research teams and attempt to probe the ASSIST systems for relevant information. The S2 was asked to rate the importance of each information need and how well the ASSIST system addressed each need.

For vignettes in which a mission did not occur, rather the vignette revolved around an S2 tasking where the S2 is to use previously collected data, the following test procedure was used:

- The S2 was asked to give a description of his understanding (what he knew) regarding the content of the tasking.
- The S2 was asked to identify information needs, e.g., information that would improve situation awareness, information about critical events, individuals, or situations, and so on with respect to his tasking.
- The S2 met with representatives of each research team to address the identified information needs. This was a time-constrained review, appropriately 10 minutes were allowed to each team. Indexed methods of information retrieval were stressed as this is expected to be the preferred retrieval strategy in future use.
- The S2 was asked to rate the importance of each information need and how well the ASSIST system addressed each need.
- After having interacted with the ASSIST system, the S2 was asked to provide an updated description of his understanding regarding the content of the tasking.

V. ASSESSMENTS

A. Metrics

The vignettes were structured to allow the IET to assess the utility of the ASSIST technologies in enhancing operational effectiveness. Utility is assessed using the following categories: effectiveness, efficiency, and user satisfaction. More targeted assessments were then made within those categories, as noted below.

Effectiveness

- What information do infantry soldiers want and/or need after completing a mission in the field?
- How well are information needs met, both from the soldier perspective and the S2 perspective?

[1] In this case a fire team leader was used, as the squad leader was participating in the review of ASSIST contributions for the squad's information needs that was occurring simultaneously.

- What were the ASSIST contributions to mission reporting?
- What is the impact on situational awareness (after mission)?
- What are the users' perceptions of the impact ASSIST technologies will have on mission reporting (content & process)?
- What are the users' perceptions of the impact ASSIST technologies will have on soldier performance in the field?

Efficiency

- Was the post-processed data available when required?
- What are the users' perceptions of the impact ASSIST technologies will have on the time taken to produce mission reports?

User satisfaction

- How do ASSIST technologies impact information confidence levels?
- Overall
- User interface[2] and capability comments

Since the systems being evaluated under Task 2 of the ASSIST program are prototypes, several of the assessments deserve qualification. Some of the metrics above are meant to serve as guides for future development rather than attempting to inappropriately assess these prototype systems. For example, under the efficiency category, post-processing time was identified as an important metric. However, it was not timed how long post-processing took because these are prototype systems resulting in the researchers still exploring which algorithms produce the most effective results rather than having spent time optimizing a specific algorithm to run very efficiently. We used a much coarser measure of efficiency in this setting of whether the data was available when needed. This, more operational metric, is of course, situation-specific. However, we felt our approach identified important metrics while not inappropriately attempting to assess them.

B. Metrics-based Assessments

Use of the procedures outlined in the previous section provided the IET with quantitative and qualitative data that informed assessments of utility and opportunities to enhance utility for the ASSIST systems. These assessments were informed by questionnaire ratings, comments, and semi-structured interview discussions with the soldiers and S2. While actual results can not be reported here, the types of assessments made can. Assessments were made regarding product and process in the following areas:

1) Mission report content (product)

- Mission report content, soldier perspective – Assessments were made in the areas of effectiveness, efficiency and user satisfaction with respect to how ASSIST systems might change quality and content detail in mission reports. Additionally, opportunities for value enhancement were identified.
- Mission report formatting, soldier perspective – Assessments were made in the areas of effectiveness, efficiency and user satisfaction with respect to how ASSIST might change mission report formatting, e.g., automatic generation of an after-mission report with annotated imagery included for "events".
- Mission characterization, soldier perspective – Assessments were made regarding the soldiers' perception of the ASSIST systems impact on their observations about "what happened" on these missions. For example, how often did the soldiers modify the list of objects, events, and activities identified in their mission reports following very detailed reviews of the data provided by the ASSIST systems.
- Mission report content, S2 perspective – Assessments were made in the areas of effectiveness, efficiency and user satisfaction with respect to ASSIST systems might change the quality and content detail during the debriefing conducted by the S2. Having collected the S2's ratings of how well his information needs were met when he debriefed the fire team leader, the IET was able to compare these ratings with how well the ASSIST systems met the S2's information needs. Additionally, information confidence levels were assessed, and further opportunities to enhance value for an S2 were often identified.

2) Mission reporting (process)

- Mission reporting process, soldier and S2 perspectives – Assessments were made in the areas of effectiveness, efficiency and user satisfaction for both perspectives: the soldiers and the S2, with respect the predicted impact ASSIST will have on the mission reporting process, as well as opportunities to enhance utility.
- Intelligence analysis, S2 perspective – Assessments were made in the areas of effectiveness, efficiency and user satisfaction for the S2's perspective with regard to using ASSIST data for intelligence analysis.
- Pre-mission planning, soldier and S2 perspectives – Assessments were made in the areas of effectiveness, efficiency and user satisfaction for both the perspectives: the soldiers and the S2 with respect to the use of ASSIST technologies in planning for future missions.

It should be noted that, due to the sample size of soldier subjects in these evaluations, that the assessments drawn were not presented as being necessarily representative of soldiers in other infantry squads and environments and it would be inappropriate to generalize these responses to a great extent. That qualification given, the assessments made did provide formative feedback to the developers.

[2] Because the user interface was not the target of these evaluations, soldier comments on aspects of the user interfaces were reported only as a feedback to researchers to guide future development of the applicable systems.

VI. Lessons Learned and Conclusions

Formative evaluations for field trials of ASSIST technologies were designed and executed in this project. Challenges arose, some anticipated, some not, during this process. Provided below are some of those challenges and experiences.

- It was relatively difficult to obtain soldier participation for this study. Further, most of our SMEs recommended that we retain soldiers who were part of a cohesive squad with recent deployment experience.
- Due to time and resource constraints and the immaturity and lack of integration of the various ASSIST systems, we did not construct an evaluation that produced mission reports with and without ASSIST contributions. Therefore, no direct comparison of mission reports completed with and without the ASSIST technologies was possible. This might be possible in future evaluations.
- The soldier-shadowing tactic proved a useful device for reducing the chance of data capture failure and or system damage with these prototypical systems. We expect the effectiveness of this tactic to diminish as the systems mature and operational scenarios under which these systems are exercised become more intense and the 'shadow' has more difficulty following his assigned soldier.
- The vignette tests were designed to evaluate operational value as perceived by the war fighter rather than technical performance, therefore some soldiers answered questions while considering the *concept* of the technology not necessarily whether the technology worked or not, which clouded some of the data.
- The S2 took much longer to review data than the squad. In retrospect, this is because the S2 was attempting to put together pieces of information to understand relationships and societal dynamics as opposed to the squad's more straight-forward task of completing a report describing their mission and related events. Since we did not fully anticipate the S2 time requirements, we did not schedule sufficient time for the S2 to fully explore all the possible contributions the ASSIST data might have been able to make towards filling his information needs. As a result of the time constraints, the S2 employed a strategy typical to analysts in time-pressure situations of exploring topics that would yield the greatest increase in situational awareness using data sources that seemed to have the most potential for return.
- Following the first evaluation period in November 2005, SME feedback recommended increasing the environmental complexity significantly. The IET took steps to achieve greater environmental complexity for the May 2006 evaluation in the following areas: increased number of people in the environment with more realistic and intertwined relationships, more visual and audio clutter, more vehicles and movement, and a more developed storyline. While this did increase the realism in the environment in which the ASSIST systems were exercised, it does make comparisons of system performance difficult to assess between the two assessment periods.

In summary, the NIST IET designed and executed formative evaluations to assess utility for ASSIST technologies in mission reporting. This paper presents the design considerations, the field-based evaluation methods used, types of assessments that were drawn, and lessons learned from the evaluations perform to-date.

Acknowledgements

This work was supported by DARPA.

Additionally the author would like to thank the following:

- Aberdeen Testing Center staff
- The soldiers who participated as subjects
- Those who contributed their subject matter expertise, especially with respect to military operational and tactical relevance.
- Other members of the IET, as they spent many hours procuring props, setting up equipment and environments, ensuring data capture, acting as observers, and the countless other tasks that contributed to the success of these evaluations.

Disclaimer

Certain commercial software and tools are identified in this paper in order to explain the work performed. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the software tools identified are necessarily the best available for the purpose.

References

[1] DARPA ASSIST, "Advanced Soldier Sensor Information System and Technology (ASSIST) Proposer Information Pamphlet," http://www.darpa.mil/ipto/solicitations/open/04-38_PIP.htm, 2006. (accessed 29 June 2006).

[2] J. S. Dumas and J. C. Redish, *A Practical Guide to Usability Testing*. Intellect Publishing Co. Portland, OR USA. 1999.

[3] G. Grinstein, A. Kobsa, C. Plaisant, B. Shneiderman, and J. T. Stasko, "Which comes first, usability or utility?" Proceedings of the IEEE Visualization 2003 Conference. Pp 11-12.

[4] B. E. John, Evaluating usability evaluation techniques. ACM Computing Surveys 28(4es). 1996.

[5] C. Schlenoff, "Overview of the First Advanced Technology Evaluations for ASSIST", Proceedings of Performance Metrics for Intelligent Systems (PerMIS) 2006, IEEE Press, Gaithersburg, Maryland, USA, August 2006.

[6] M. Theofanos, and W. Quesenbery,. Towards the Design of Effective Formative Test Reports. Journal of Usability Studies 1:27-45. 2005.

[7] B. Weiss, M. Shneier, and A. Virts, "Technology Evaluations and Performance Metrics for Soldier-Worn Sensor Systems for ASSIST", Proceedings of Performance Metrics for Intelligent Systems (PerMIS) 2006, IEEE Press, Gaithersburg, Maryland, USA, August 2006.