2004 Command and Control Research and Technology Symposium
THE STATE OF THE ART AND THE STATE OF THE PRACTICE

Proposed Topic Area: Cognitive Domain Issues

Evaluation of Intelligent Agent Technology for C2 of Human and Robotic Entities

Dr. Scott D. Wood, swood@soartech.com
Jack Zaientz, jzaientz@soartech.com

Soar Technology, Inc.
3600 Green Court Suite 600
Ann Arbor, MI 48105
734-327-8000

Dr. Carl Lickteig, Carl.Lickteig@knox.army.mil
U.S. Army Research Institute for Behavioral and Social Sciences
Bldg. 2423
Fort Knox KY 40121-5620
USA
(502) 624-6928

POC: Dr. Scott D. Wood, swood@soartech.com
Soar Technology, Inc.
3600 Green Court Suite 600
Ann Arbor, MI 48105
Voice: 734-327-8000 x207
Fax: 734-913-8537

Evaluation of Intelligent Agent Technology for C2 of Human and Robotic Entities

ABSTRACT

This paper presents an evaluation of an intelligent user interface prototype for controlling mixed elements of manned and robotic forces. Our goal was to create an environment where we could research, develop, and test intelligent interface technologies to improve combat effectiveness, increase span of control, and reduce training requirements for future battlefield commanders. Our approach centered on providing intelligent control and decision aids to allow the commander to focus on primary military objectives rather than on technology. To evaluate this approach we developed a C3 framework of cooperative interface agents that reflect roles found in military command staffs. The intention is to create a virtual staff for the commander of robotic forces by embedding these military functions within the C3 interface. This functional organization allows for improved scalability, knowledge management, and a natural decoupling of tasks from platforms.

Wood, et al., (2004) focused on the architecture and technology for developing the prototype system; this paper focuses on the evaluation of the system using U.S. Army officers running a simulated scenario. Metrics included successful mission completion and ability to maintain situation awareness. Post-evaluation questionnaires and interviews were used to obtain additional feedback. The prototype allowed participants to successfully complete the evaluation tasks in a simulated scenario. A key observation was that the prototype system appeared to be a good platform for training and performing unmanned asset management: participants were able to effectively manipulate the position of multiple unmanned sensor assets in a way that maximized sensor coverage for the mission.

Evaluation of Intelligent Agent Technology for C2 of Human and Robotic Entities

As robotic and automation technology improves, fundamental complexities in human-system interaction remain. It is clear that significant progress must also be made to improve the means by which human commanders interact with this new technology before its full benefit can be realized. This paper presents final results from the evaluation of a Phase II SBIR project to address the operational need to effectively command and control mixed teams of human and robotic elements (Wood, et al., 2006). There were three main goals for this project:

- Understand the requirements for human-system interaction at the company-command level in a realistic military scenario.
- Develop technology to enable improved human-system interaction of mixed human and robotic elements for a company-sized unit.
- Evaluate the developed technology with respect to effectiveness, usability, and training requirements.

Our basic approach to addressing the problem was to research, design, and develop intelligent user interface technology to assist battlefield commanders using the paradigm of intelligent software agents as a unifying concept (Wood, et al., 2004). A graphical user interface was developed in a simulated environment using OneSAF (One Semi Automated Force) Testbed as the underlying simulation, and a formative evaluation was conducted with U.S. Army officers using a scenario derived from an FCS (Future Combat Systems) vignette as the overall evaluation task. While Phase I was demonstrated in a relatively simplistic context (Wood, 2003), demonstrating viability of the Phase II Technology under more realistic conditions required significant scientific progress in agent technology, agent-team collaboration, knowledge representation, and human-system interaction.

User Interface Analysis and Prototype

To address the objective of developing an effective user interface for robotic control, the research team first had to determine warfighter information needs. The target user for this system was a company commander or a subordinate who would be responsible for commanding and coordinating human and robotic forces, but not necessarily directly controlling them. The researchers started by first developing a detailed system usage scenario based on current doctrine and equipment. Then the new platform and weapons capabilities were projected onto the FCS scenario to determine how this might change or affect the target user's command task.

The prototype interface was then designed to support the resulting task. This involved two key assumptions:

- Irrespective of new technologies, fundamental tenets of command and control are unlikely to change dramatically.

- To keep from imposing an additional workload burden on the user, human-robot interaction should be at least as easy as human-human interaction.

Usage Scenario

The following usage scenario was developed to analyze how a warfighter might use the prototype system while conducting the scenario mission within the evaluation environment. The scenario is divided into distinct phases including staging, pre-operation, operational, and post-operation:

Staging Phase tasks begin with receipt of Operational Orders (OPORD) and mission briefing, and analysis of data from numerous sources including intelligence reports, maps, and other available information. From this data, information is developed, correlated and displayed on system displays. This will enable more accurate situation awareness to be developed and maintained regarding friendly, enemy, and civilian positions and courses of action. Pre-Operation Phase tasks then follow with analysis of mission goals, plan development, and plan approval. The Operational Phase commences using the system GUI (graphical user interface) to issue commands, communicate, receive reports, and make tactical decisions as necessary. The initial plan in this Operation places three UAVs at recon points with the expectation that there may be UAV losses. Each loss triggers a notification that is matched against a pre-set loss threshold. When this threshold is in danger of being crossed, the user is warned. The user can choose to change the ratio, move or delete recon points, or ignore. Operations continue with the user issuing orders to subordinate units via the GUI to conduct movement, breaching, and assault tasks to successfully accomplish the mission. The Post-Operation Phase includes debriefing and an after-action review.

From the usage scenario, eleven general-purpose GUI tasks were defined to enable a user to perform the necessary tasks using the prototype system. For each GUI task, assumptions were listed and corresponding user and system behavior was specified. Using the set of GUI tasks from this list, the user could execute all of the evaluation tasks.

GUI Screen Design

From the usage scenario and GUI task definitions, a two-screen user interface was designed. The first screen, the Map Display (Figure 1), was designed around a simulated view of the battlefield, in this case using the OTB simulation environment. It includes mission control widgets for starting and stopping the simulation, map navigation controls, and a scrolling message window where the system and simulated entities can communicate with the user.

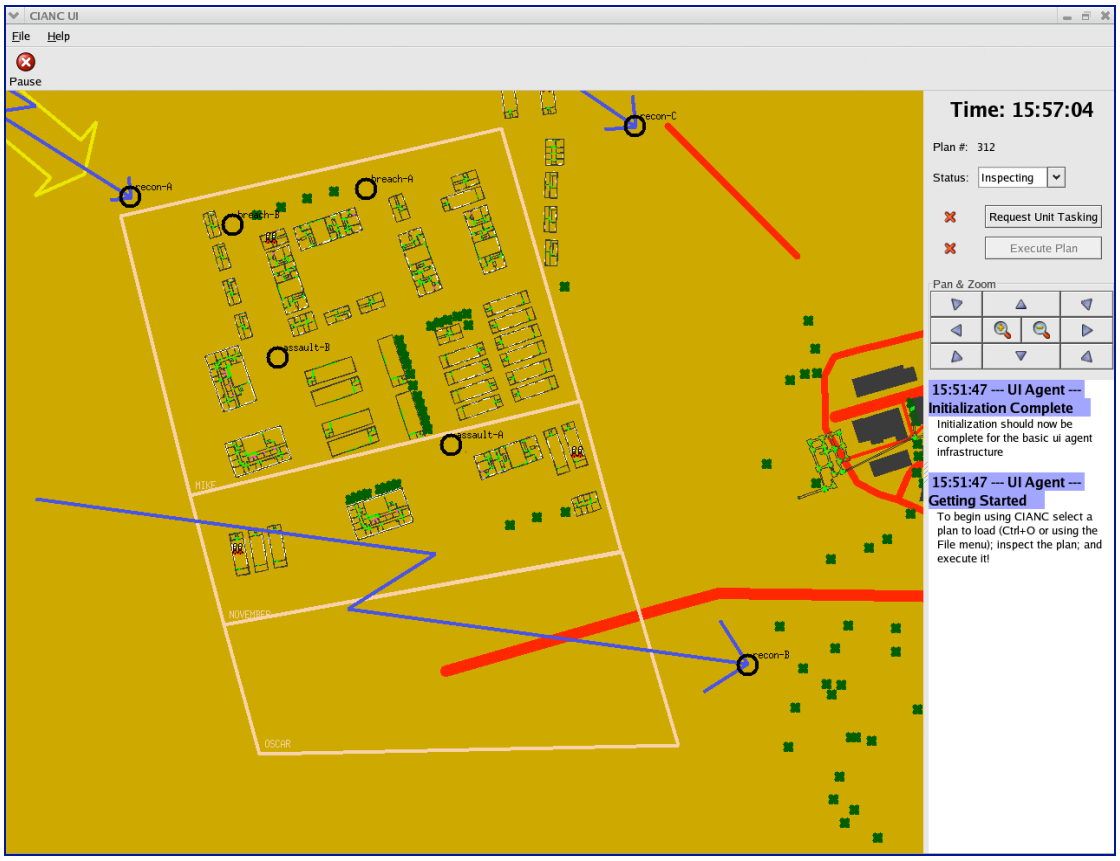

Figure 1. Map Display screen.

The second screen, the Plan Display (Figure 2) was designed around four information panes: objectives, decision points, points, and units. The objectives pane is used to display all of the mission objectives as specified in the OPORD. As objectives are completed, the list items status is changed as an indicator for the user. The Decision Points pane lists all of the decision points necessary to complete the objectives. For each decision point listed, the criteria for making a decision is indicated, and branch points are described if the decision cannot be made positively. As the user makes a decision, he or she tells the system to either continue with the mission, branch to a contingency plan, or halt the mission completely by pressing the appropriate check box. The Points pane is a list of waypoints used for mission planning. The Units panes are information only panes that allow the user to see the status and composition of all subordinate forces.

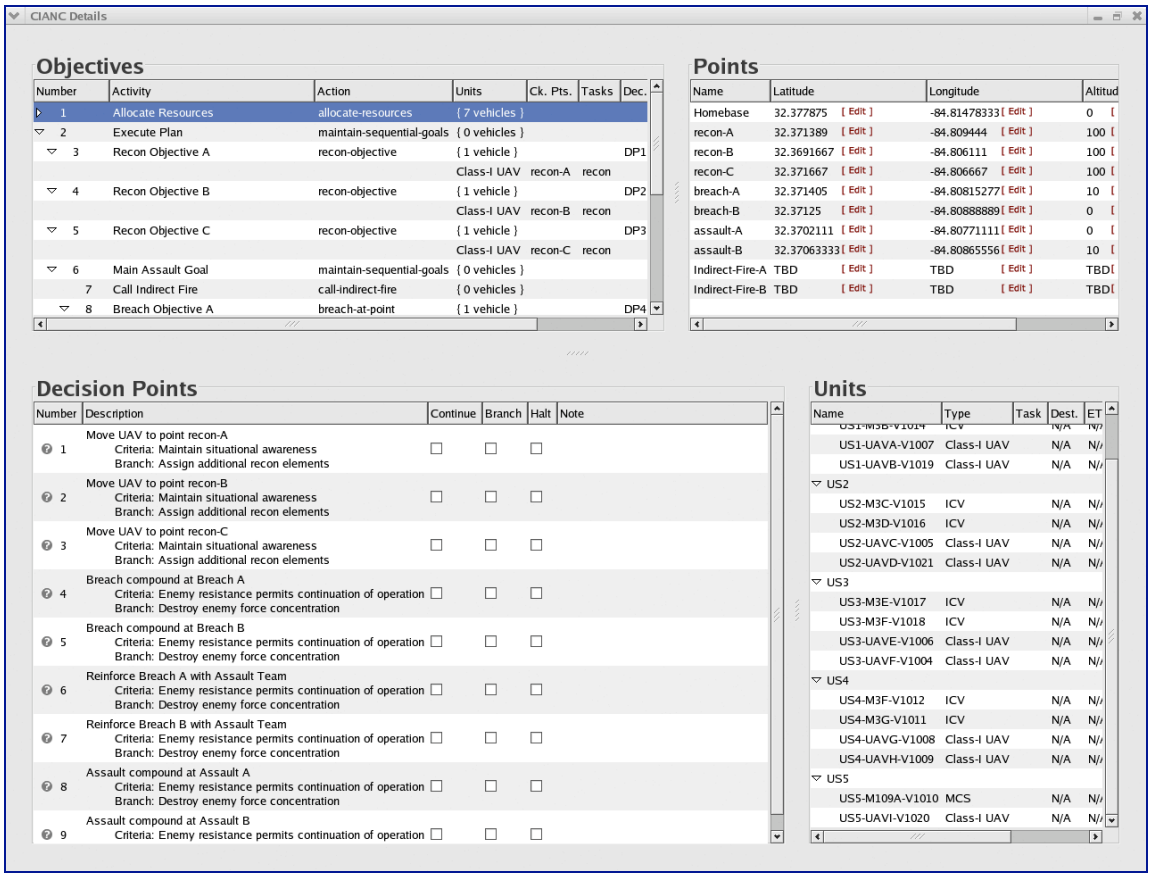

Figure 2. Plan Display screen.

System Evaluation

The purpose of the evaluation was to assess the usability of the Phase II CIANC[3] prototype, particularly with respect to human interaction with the agent-based automation. The overall goal was to determine how and to what extent the research concepts and developed system components have the potential to reduce warfighter workload, reduce training requirements for human-robotic interaction, and improve mission effectiveness. The evaluation criteria were:

- Ability to successfully complete mission
- Performance, such as task accuracy and completion time
- Error rate, error type and error-inducing task methods
- Situation awareness
- Potential impact on mission performance

The evaluation focus was predominantly on functionality provided by the underlying agent system, and how this functionality could improve operator performance and effectiveness. As will be further discussed later in this section, a late change in the evaluation procedure involved the use of a “puckster” (a surrogate or assistant to input or

implement the commands given by the user). This change in the evaluation procedure reduced participant training time significantly, enabling the evaluators to focus on the core question of whether intelligent user interfaces could help at a deep level, rather than be distracted by the more superficial implementation details of the prototype interface.

A total of nine active duty U.S. Army officers with training and experience at company-level operations participated in the evaluation. By design, all officers were either Captains or 1st Lieutenants to most closely match the intended user population. As indicated in Table 1, all were male and all but one had an Army Officer Area of Concentration (AOC) of 12A (Armor, General) or 19A (Armor). All had substantial experience participating in simulation exercises, but only half as computer system operators. Most of the participants reported using computers daily inside and outside work, but there was a wide range of game-playing experience for both training and entertainment. Only one participant listed any experience with simulated Unmanned Air Vehicles (UAVs) or Unmanned Ground Vehicles (UGVs).

Table 1. Participant Background and Experience.

Participant	General Background			Computer Experience		Simulation Experience	
	Age	Rank	AOC	Training Games	Personal Games	Military Sims	Sims Used
P1	26-29	0-3	19A	Never	Never	No	
P2	34+	0-3	19A	Never	Some	Yes	JANUS
P3	26-29	0-2	19A	Never	Never	No	
P4	34+	0-2	12A	Some	Some	Yes	TACOPS
P5	34+	0-3	12A	Some	Daily	Yes	JANUS
P6	26-29	0-2	12A	Some	Some	No	
P7	30-34	0-3	19A	Never	Never	Yes	CCTT
P8	30-34	0-2	42A	Some	Some	No	
P9	30-34	0-3	12A	Some	Daily	Yes	JANUS, BBS

Note. TACOPS = Tactical Operations, CCTT = Close Combat Tactical Trainer, BBS = Brigade//Battalion Battle Simulation.

Apparatus

The Phase II CIANC^3 evaluation system consisted of two standard 1.5GHz PC's running standard Red Hat Enterprise Linux 3.0 Workstation operating systems, with two 19" CRT displays set at 1600 x 1200 resolution and 32-bit color. User commands were issued via standard three-button mouse and keyboard. The hardware was instrumented to collect user keystrokes and menu selections. Participant actions and speech were video-recorded from an over-shoulder angle as shown in Figure 3.

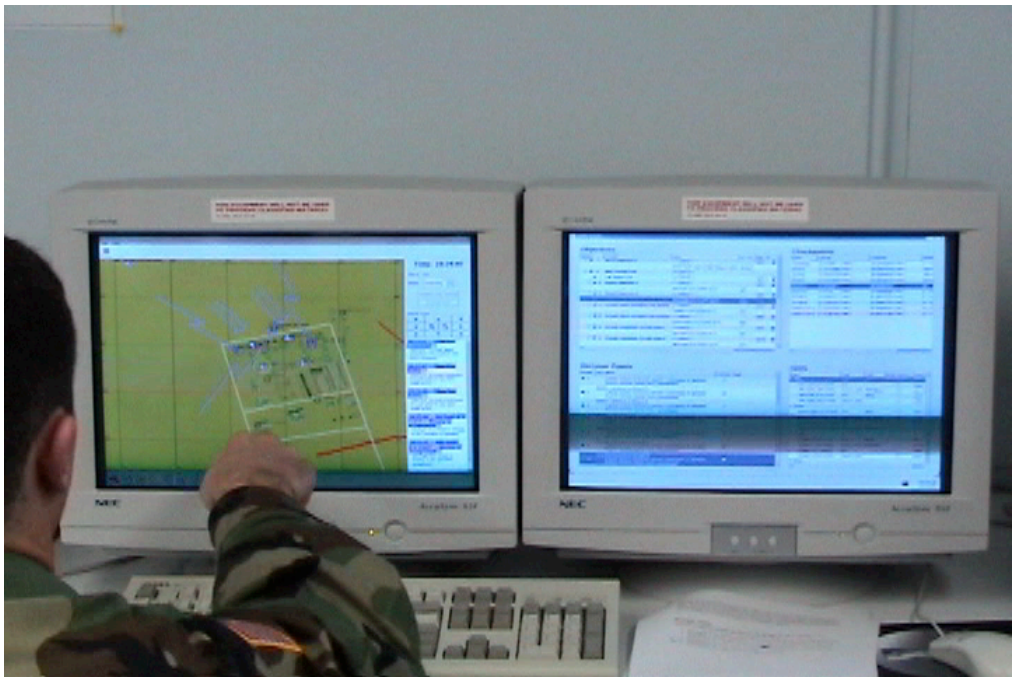

Figure 3. An evaluation participant using the CIANC^3 interface.

Data Collection Instruments

Several techniques were use to capture usability data and other relevant information. The objective for the use of multiple instruments was to seek convergence on key usability issues. Furthermore, since each technique addresses evaluation from a different perspective, using multiple instruments allows the collection of a broader set of data that should reveal more usability issues than any single technique alone. This was seen as especially important given the relatively sparse number of participants. The instruments used are available as appendices in the project final report (Wood, et al., 2006).

Background Questionnaire. A questionnaire was used to gather data relating to participant background. A key issue the questionnaire data was used to address was whether experience, training, or affinity for computer games affected participant performance with the CIANC^3 prototype. It was anticipated that participants with substantial experience with gaming and simulation would more readily accept computer automation, more easily grasp the skills necessary to complete the evaluation tasks, and would perform better than those without extensive computer experience.

Evaluator Observation. The CIANC^3 system was also evaluated according to the ability of participants to perform the evaluation task, the time it took to complete the task, and the type and severity of human errors and confusions experienced by participants. Evaluators observed each participant during performance of the conduct mission exercise, noting completion of tasks and apparent difficulties experienced. Participants were asked to perform using a “think aloud” protocol that helped evaluators infer usage concerns and difficulties. Actions and speech were recorded using two video cameras, and the user interface was instrumented to capture keystrokes. While these techniques

supported measurement of task completion and difficulty, they did not provide a more objective measure of system capability in terms of users' situation awareness (Endsley, 1988).

SAGAT (Situation Awareness Global Assessment Technique). The SAGAT method (Endsley, 1995) was used to help assess the systems' ability to support users in attaining and maintaining situation awareness. The SAGAT technique provides a measure of situation awareness by comparing participants' perceived assessment of the situation with the actual situation. Measurement is accomplished by freezing the evaluation task at randomly selected times, suspending the simulation scenario, and blanking the user interface display screens while the participants answer questions about their understanding of the current situation. These perceptions are then compared to the actual situation based on system data or evaluation by a subject matter expert. If the performance interruptions are relatively few (3 or less) and the duration of each SAGAT measurement is kept relatively short (5 minutes or less), the SAGAT method can provide a relatively unbiased assessment of participants' situation awareness without adversely affecting overall performance (Endsley, 2000).

Post-Evaluation Questionnaire. A post-evaluation questionnaire was administered to individual participants to gather subjective feedback regarding system functionality, ease of use, and other issues regarding system utility. Participants answered one set of questions by rating the system using a 7-point Likert scale. A second set of questions allowed the participants to write specific suggestions and comments regarding the system.

Focus Group Questionnaire. A structured survey instrument was used to inform and guide group discussion for the final, focus group session. Where the post-evaluation questionnaire was intended to gather feedback on the evaluation prototype that was developed, the Focus Group was to discuss how intelligent command and control tools might be used in the future, based on their prior experience and their experience with the CIANC^3 system. This instrument concentrated on soliciting battle command tasks and situations that were exceptionally cognitively demanding, such as maintaining situation awareness and synchronizing actions.

Procedure

Session Design. The evaluation process consisted of six sessions over the course of three days as shown in Table 2. The sessions were divided into three types and included: three single-participant sessions with think aloud protocols; two single-participant sessions with SAGAT; a final four-participant guided-discussion focus group.

Table 2. Evaluation Schedule by Session Type and Duration.

	Day 1	Day 2	Day 3
9:00 -- 12:00	Think Aloud	SAGAT	Think Aloud
13:00 -- 16:00	Think Aloud	SAGAT	Focus Group

All sessions took approximately three hours to complete. The approximate schedule for each session was: 30-minute in-brief, 30-minutes of training, 60-minute conduct of mission, 30-minute survey and discussion, and 30-minute debrief. The key segments for each session segments were as follows:

1. Participant completed background questionnaire to determine military experience in company-level operations and simulation software.
2. Evaluator conducted in-brief to explain evaluation purpose and procedure.
3. Military subject matter expert (SME) provided mission brief and rehearsal.
4. Evaluator described and demonstrated system interface and functionality.
5. Participant conducted mission and provided supporting data through Think Aloud or SAGAT measurement techniques.
6. Evaluator served as "puckster" to perform system interactions during the mission, as directed by the participant.
7. All participants completed post-evaluation questionnaire and group participants also completed the Focus Group questionnaire.
8. Evaluator and military SME led debrief or group discussion.

During the single participant sessions, each participant completed the same mission twice, as discussed later. SAGAT and think aloud measures were collected during the participant's first conduct of the mission. One version of SAGAT questions was administered at the start of the mission immediately after mission rehearsal and training. The second version instrument was administered just after the participant had successfully initiated the breach, at approximately the 7-minute mark into the mission. For each administration, the prototype screens were blanked and the simulation suspended while the participant answered the SAGAT questions.

Immediately after both mission runs were complete, Think Aloud and SAGAT participants were asked to complete the post-evaluation questionnaire. Immediately after the completion of the questionnaire, the evaluation team reviewed participant responses to identify any additional interview questions based primarily on response outliers, as is common practice in usability evaluations. For example, since most responses were expected to be within a relatively narrow range of values, responses that indicated either exceptionally poor- or high-usability of some aspect of the system were further clarified. In some cases, such as when evaluators noted confusions during mission exercise execution, participants were asked to clarify what they were trying to do and what they were expecting the system to do. Such events indicate clear mismatches between system models and user models (e.g. expected system behavior).

During the group session, the four participants were divided into two 2-person teams. Each team conducted their missions separately while the other team was temporarily excused from the evaluation setting. The members of each team worked together as they completed their two repetitions of the mission exercise, using puckster

assistance as with single-participant sessions. Group participants were asked to complete the Focus Group questionnaire immediately after the second team completed their last run through the mission exercise. Evaluators reviewed participant responses and prepared questions to lead the group through discussions. A single group discussion was conducted with both 2-person teams. Questions were asked about the relative importance of battlefield command tasks, which tasks were the most difficult to perform, how a CIANC^3-like system might be able to help in difficult situations, and what additional features and functionality would improve utility and likelihood of Soldier acceptance.

Evaluator Role. Evaluators recorded audio and video for all sessions with an emphasis on each participant's conduct of the mission and individual interview or group discussion. Evaluators observed participants conduct of the mission and took notes on participant behaviors, difficulties, questions, and comments. Evaluators timed performance tasks and elicited participant commentary when there were extended lulls in think aloud commentary or when participants appeared confused or frustrated. A typical evaluator elicitation took the form of, "What are you thinking about now?" Otherwise, evaluators did not interact with participants during the evaluation or help with system interactions unless it was requested by the participant or it was clear that mission progress had ceased.

Mission Tasks

The evaluation focused on the ability of each participant to use the CIANC^3 interface to conduct a simulated assault on an urban compound with an FCS company of predominantly unmanned systems. The participant was provided a pre-established mission plan and was responsible for executing the plan as quickly as possible. The mission, CCIR, and decision points were also pre-encoded into the prototype to reduce scenario complexity and to constrain participant actions. This was done to reduce participant training time and improve the ability to compare results across participants. To further constrain participant behavior, the simulated opposing force was intentionally restricted to a static, defensive posture. Additionally, the Map display always reflected simulation ground truth, so participants always saw the enemy forces (although they did not know that they were seeing all of the enemy). The mission entailed a set of key tasks and decision points with specified criteria for continuing to the next task. The mission tasks are listed in Table and a detailed mission brief can be obtained from ARI.

Table 3. Usability Evaluation Mission Tasks.

Tasks & Decision Points	Criteria
Recon Objectives A, B, C	Maintain SA
Call for Effects (indirect fire)	Resistance permits mission continuation
Breach Objectives A, B	Resistance permits mission continuation
Reinforce Objectives A, B	Resistance permits mission continuation
Assault Points A, B	Resistance permits mission continuation

The usage scenario described earlier to analyze how a warfighter might use the prototype system included several distinct phases including staging, pre-operation, operational, and post-operation. A detailed analysis of tasks by phase was performed for the evaluation's urban assault mission, and is available from ARI. Across all phases of the urban assault mission, a common set of generic tasks was developed that summarize the participants' performance requirements during the usability evaluation. The generic tasks are listed as follows:

- Use prototype to inspect and approve plan
- Use prototype to request initial asset assignment
- Evaluate assigned assets & asset routes
- Approve assigned assets & asset routes
- Use prototype to initiate battle sequence
- Use displayed information and markers to maintain awareness of current battle progress
- Interact with prototype to react to decision points as they arrive
- Respond to prototype-generated CCIR notifications
- Change the "Acceptable UAV Loss Ratio"
- Move recon points
- Delete recon points

Mission Brief and Rehearsal

A military subject matter expert provided each participant a mission brief and rehearsal to clarify mission requirements, tasks, constraints and success criteria. This brief addressed key and relevant aspects of the FCS unit the participant was to command and control with an emphasis on the capabilities and limitations of the unmanned systems and the network nature of FCS information and communication. A poster sized wall map of the mission setting was used to illustrate and rehearse mission tasks and decision points prior to using the CIANC^3 system.

Participant Training

After the mission brief and rehearsal, the evaluators provided scripted training to participants on the usage and functionality of the CIANC^3 user interface. The training provided an introduction to the system that was read by an evaluator to each participant while seated in front of the user interface. During this familiarization training, interface features were demonstrated by the evaluator and performed by the participant with clarification provided by the evaluator, as requested. In addition, participants were provided a copy of the CIANC^3 Training Manual as reference material available any time during the evaluation session. After the scripted introduction, participants completed a

set of example tasks to assess their basic familiarization with the user interface. Participants were required to use the interface to find the information necessary to answer the questions correctly. Participants who had problems with the training questions were provided additional guidance from evaluators until all questions could be answered correctly.

Conduct of the Mission

During the mission exercise the participants were seated facing the two CIANC^3 display monitors, as shown in Figure 1. At the participant's side sat a supporting researcher who served as "puckster" to assist in performing computer interactions with the CIANC^3 system, as requested by the participant. Participants were given several reference sheets to assist them in conducting the mission and the evaluation tasks. Evaluators observed from behind the participant and recorded completion times for evaluation tasks and the mission and noted exceptional events, such as participant confusion or mistakes. Participants performed the same urban assault mission twice, as discussed below. Each mission was completed when all scripted tasks were performed and the mission objective was accomplished.

The use of an assistant puckster to help participants directly manipulate the CIANC^3 interface is a notable aspect of the evaluation procedures. The reasons for this assistance are discussed here and potential impacts on results are examined in the Discussion section. The primary reason for using a puckster was to focus the participant and the evaluation on the major concepts and functions represented by the interface agents rather than more minor and modifiable implementation issues. Participants faced a considerable challenge already in learning and employing the novel and complex FCS assets, particularly unmanned systems, provided for their urban assault mission. Requiring the participants to also acquire proficiency in manipulating the CIANC^3 interface would have increased the training load and perhaps impeded their ability to employ and asses more basic concepts and functions. It was also anticipated that the use of a puckster might increase the quantity and at best quality of each participant's verbalization of thought, intention, and action during the conduct of the mission exercise.

The procedure of having each participant complete the same urban assault mission twice also bears explanation. The primary rationale for mission repetition was to allow participants to spend the first trial better learning the mission and the CIANC^3 system and the second trial exploring alternate courses of action. Such repetition mimics a standard military training technique used for example in a STX (Situational Training Exercise) Lane that allows units to run the same scenario repeatedly to assess and explore different tactics and alternate courses of action. While multiple scenario runs are not standard practice for usability evaluations, issues that continue to surface even after experience tend to be more severe procedural errors that indicate a need for system refinement rather than training workarounds (e.g. Wood & Kieras, 2002).

After repeating the urban assault mission a second time, participants completed the remaining evaluation activities as previously described. These included completing the post-evaluation questionnaire, participating in a group discussion for the participants

in the group condition, and receiving an evaluation debrief from an evaluator and military SME.

Results

Data was collected and analyzed for each of the three session types: Think Aloud, SAGAT, and Focus Group. All participants completed background questionnaires. Think Aloud and SAGAT participants also completed post-evaluation questionnaires and observations were noted regarding mission exercise execution. SAGAT participants were also evaluated using the SAGAT instrument to assess situation awareness. Focus Group participants completed group questionnaires instead of the normal post-evaluation questionnaires. All groups participated in either individual debrief sessions or a group discussion.

Observation Results

The primary criterion of success for the evaluation task was whether the participant could successfully complete all mission tasks. The results were mostly positive with only one subject unable to complete the entire mission exercise. This participant was the only one with a non-Armor Area of Concentration and it seemed that most of the difficulty concerned tactics and decision making rather than difficulty with the interface. All other participants completed the mission exercise in roughly the same time; the critical path was determined more by the simulation underlying the evaluation task scenario rather than by user actions. Other non-critical difficulties (e.g. those not affecting the ability to complete the task) were either observed by evaluators or taken from the think aloud protocol.

Usability issues were grouped into four areas: (a) automation design, (b) user interface design, (c) information design and, (d) miscellaneous. Although most participant interaction with automated aspects of the interface were positive, there was some confusion about why particular automations were happening, and some frustration at not being able to override the automated actions. Although these issues can mostly be seen as an artifact of how the CIANC^3 user interface was implemented, the ability for the automation to explain its actions and the capability for the human user to inspect and override any automated action seems to be a critical design feature for future development. Despite the difficulties, most participants wanted more automated-support rather than less.

Difficulties relating to the graphical user interface (GUI) design mostly centered on insufficient integration of display elements. For example, participants sometimes had difficulty relating information on the text-based Plan Display to graphical representations on the map-based Map Display. This was especially apparent when participants attempted to spatially relate the text-based decision-point information to a specific location on the map. Apart from these issues, participants reacted favorably to the information that was presented and how it was organized. Automated display of CCIR and other IR types was called out as being particularly helpful.

Issues relating to information design focused mainly on desired information that was not presented or information that was displayed in a non-standard or unfamiliar manner. For example, participants requested terrain information, structure elevations, line-of-sight information and other information that is typically combined from maps, photographs, human reports, and satellite imagery. They also wanted real-time data on items such as fuel status, ammunition available, unit capabilities and unit health. There were also difficulties with non-standard symbols and graphical controls used on the map display. In general, participants were pleased with automated display of sensor information and other intelligence information, but also wanted access to the raw sensor feeds and reports.

The other main issue, classified as miscellaneous, relates to the realism of the evaluation environment. Participants noted that in actual command situations, much time and effort is spent communicating information both to upper echelons and laterally to other commanders. Since such communication is a major source of cognitive and performance workload, they felt it difficult to accurately assess a system that did not consider that factor.

SAGAT Results

Table contains the results from the SAGAT evaluations. Overall the SAGAT results were positive and informative with 10 of the 16 questions answered correctly. There was a mix of SA errors between the two participants. Both participants answered the following questions correctly:

- *Indicate the locations of each element on the map*. Although this result could have been influenced by planning and training time with map display, or by individual abilities to maintain an accurate mental representation of the battle area, this result also indicates that the participants were actively using the graphical display for problem solving and attending to the data presented as part of that display.
- *What do you expect the enemy to do in the next 5 minutes?* By maintaining a real-time view of the exercise battle area, and understanding the capabilities of the reconnaissance elements, participants were able to project their awareness of the current situation at least 5 minutes into the future. While this result may be influenced by the relatively straightforward mission scenario, having real-time intelligence information is likely a key enabler for accurate predictions of enemy behavior.
- *Which enemy element is your highest-level threat?* This indicates that enemy units are clearly indicated on the Map display and that the participants could differentiate between enemy unit types. It also indicates a consistent level of training regarding threat assessment (from prior combat training).

The one question that both participants answered incorrectly is a fairly important one: *How many casualties have you suffered?* As friendly unit assessment is a critical element of situation assessment for commanders, this result points to a need for better display of unit status and aggregate company strength. Although unit status is available,

the necessary information is only available by selecting individual units from the Plan display.

Table 4. SAGAT Evaluation Results.

Instrument SAGAT 1	Participant PC	Participant PE
Indicate location(s) of each element on the map.	Correct	Correct
Which of the following assets are available to support you?	Incorrect	Correct
Where are the principal enemy concentrations?	Correct	Incorrect
What do you expect the enemy to do in the next 5 minutes?	Correct	Correct
Instrument SAGAT 2	**Participant PC**	**Participant PE**
Which friendly forces are currently exposed to enemy fire?	Correct	Incorrect
Which enemy element is your highest-level threat?	Correct	Correct
How many casualties have you suffered?	Incorrect	Incorrect
Indicate which threats are currently under reconnaissance. Indicate those that are not.	Incorrect	Correct

Post-Evaluation Survey Results

Participants were asked to complete a survey and provide feedback regarding their use of the prototype interface. The survey consisted of 31 questions rated on a 7-point Likert scale and several free-response questions that allowed for discussion and comments. Analysis of the post evaluation survey indicated considerable similarity between participant responses, as indicated in Table 5. The median score across all questions was 6, which indicated participants generally answered positively to all questions.

The focus of this type of early development-stage formative evaluation is not necessarily to confirm that the design was done correctly; it is assumed that good design must be an iterative process that depends on continual user input. While it is always good to get confirmation that a design is on the right track, the focus for this evaluation was to find design flaws and other usability issues within the system concepts and functionality that could compromise mission or prevent task completion. Perhaps more important was the second-order question of whether and to what extent did the application of intelligent agent technology contribute to successful or flawed system design. One important aspect of the survey results considered was the range of response values. When all responses for a particular question are uniformly negative or positive, the interpretation is typically clear. However, when the range of responses is wide, even if the median or mean value is within acceptable norms, it indicates that individual differences can play a significant role in the system's ability to support the task. While there will always be those who excel at particular tasks, one goal of good system design is to minimize the risk of complete task failures, irrespective of individual differences.

Table 5. Post-Evaluation Questionnaire Results.

Question	Participant									Range		
	P1	P2	P3	P4	P5	P6	P7	P8	P9	Low	High	Median
System Behavior												
… understandable	4	6	6	5	3	5	6	6	6	3	6	6
… predictable	5	5	6	6	5	5	6	6	7	5	7	6
… controllable	2	6	5	6	2	5	6	7	4	2	7	5
… appropriate	4	5	6	5	1	5	4	6	5	1	6	5
System Concepts												
… familiar	6	7	7	7	3	5	3	6	5	3	7	6
… extended well	6	7	6	6	4	5	4	6	5	4	7	6
System Terminology												
… familiar	4	7	5	6	4	5	3	6	6	3	7	5
… extended well	4	7	6	6	4	5	4	6	6	4	7	6
Work procedures												
… familiar	4	5	5	6	5	5	4	6	7	4	7	5
… extended well	4	5	6	6		5	4	6	7	4	7	5.5
System organization supported task	6	5	5	7	5	5	4	7	6	4	7	5
Information Display												
… clear	6	7	7	5	5	5	6	6	6	5	7	6
… sufficient	2	7	7	6	1	5	5	6	2	1	7	5
… relevant	3	6	7	6	6	5	5	6	5	3	7	6
… satisfying	5	6	7	5	5	3	5	6	5	3	7	5
Learning to operate												
… easy	7	7	6	7	7	NA	7	6	7	6	7	7
Controls												
… easy	7	6	6	6	7	NA	6	5	7	5	7	6
Locating functions & information easy	5	7	5	6	7	NA	6	5	7	5	7	6
System messages helped learning	4	6	6	6	2	5	6	5	5	2	6	5
Reference materials												
… clear	4	7	7	5	1	3	6	7	6	1	7	6
Training Time												
… sufficient	6	6	7	6	4	1	6	6	7	1	7	6
System Speed												
… fast	3	6	6	6	6	6	4	6	6	3	6	6
System Reliability												
… reliable	3	5	6	7	6	NA	6	6	6	3	7	6
Overall Reaction												
… positive	4	7	6	6	5	6	5	6	5	4	7	6
Using System												
… easy	6	7	7	7	6	6	5	5	6	5	7	6
… satisfying	5	7	6	7	5	6	5	5	6	5	7	6
… engaging	4	7	5	7	5	6	4	6	6	4	7	6
System												
… powerful	3	5	5	6	1	5	2	5	3	1	6	5
… flexible	2	6	6	6	1	4	4	2	2	1	6	4
… appropriate	4	7	6	6		6	4	6	4	4	7	6
… clear	5	7	6	7	3	6	4	5	6	3	7	6

Questions where the median value was relatively high (at 6 or higher) and had a narrow range of values were considered to have a potentially positive impact on usability. Questions where the median value was below the overall median (lower than 6) and had a wide range of response values were considered to have a potentially negative impact on usability. For each question that fit in these categories, post-evaluation interviews sought to clarify the responses. These responses were then grouped into several categories and described further in the following section on usability findings.

Usability Findings from Post-Evaluation Survey

Survey questions whose responses indicated a large potential for effecting system usability were classified into 5 categories: (a) automation functionality, (b) automation implementation, (c) information design, (d) ease of use, and (e) ease of learning. These categories are further described in their respective sections.

Automation Functionality: The system needs to provide flexible and doctrinally correct automation. Automation functionality is defined here as the capability of the underlying agent-based automation system. Good automation functionality can often go unnoticed because it doesn't distract the user from their task. Poor automation functionality typically results in system behavior that is incorrect, unexpected, or undesired. In many cases, participants asked for more automation, but in some instances, they wanted more manual control over the automated system actions.

There were two specific concerns about system behavior that participants noted: automated behavior that was incorrect, and ability to override automated behavior. First, the way the system assigned units to objectives did not always seem correct to the participants. It also offered nothing to explain its behavior. This made the participants question the system's competence. This was a well-deserved criticism, since the logic the system used was limited and did not take into account a number of important criteria.

Second, the system's flexibility was limited in a manner that made it behave incorrectly in some circumstances. There are two primary examples of this. First, the participants were not able to control the movement of a number of their units. These units were pre-positioned and could not be moved. Second, the system would not allow the participant to substantially change the plan, either in terms of decision points or objectives. This led the system to follow a narrow set of steps that the participants would have changed if they had the opportunity.

Automation Implementation: The system needs to provide manual access to results of automation. Automation implementation is defined as the design of the user's interaction with the automation functionality (i.e. human-automation interaction). The category of automation implementation relates to how users were to interact with the underlying automation. This category is closely related to, and perhaps hard to distinguish from, the category of automation functionality. However, automation implementation relates more to the overall perception of how the automation fits within the user's task rather than specifics of what can be automated.

The participants gave low scores to the system's power and flexibility. Understanding that this was an evaluation prototype, the participants' reaction is not surprising. There were a number of features that the participants insisted were critical to system usage that were not present in the evaluated system. Most commonly requested features were:

- Ability to override automated task assignment and to manually assign units to objectives when appropriate.
- Ability to modify Decision Points: adding new ones, removing or editing existing ones, adding new branches and sequels as appropriate.
- Ability to modify Objectives, adding new ones, removing or editing existing ones.

Information design: The system needs to provide raw data or quick access to details to support aggregate information displays. Information design describes the integration of information elements within the CIANC^3 prototype graphical user interface. The category of information design was composed of a single question whose answers were substantially below median. This category relates to what information was available to make decisions and how it was presented.

Participants made a number of suggestions and design requests relating to information design, similar to evaluator observations noted earlier. The participants' general request was to provide additional data on the information display to improve situation awareness and decision-making. Specific requests that they made included:

- Display geographic locations of decision points on Map Display.
- Display all units on Map Display, including dismounts and UGVs.
- Improve clarity of unit ID and status displays.
- Improve use of standard graphics, and add legend for non-standard graphics.
- Add control for visualization layers, allowing commander to see different sets of details as needed, including terrain features.

Ease of Use: Task-centric design and application of automation contributed to user performance and system acceptance. Ease of use indicates that the procedures necessary to utilize the implemented functionality were straightforward to perform. It also indicates that the type and means of information that was displayed, was presented in a manner consistent with standard operating procedures. Specifically, the graphical information was displayed in a way that made perception and understanding natural for Soldiers trained to operate with paper-based geospatial artifacts. Additionally, the objective and decision point information corresponded well to the types of written orders and battle plans currently conducted primarily with non-digital methods. One participant noted that the objective and decision point information was very similar to information he currently manages by strapping a notepad to his thigh and updating it manually. System

automation that is designed using task-centric or other user-centered techniques can have a strong impact on usability.

Ease of Learning: Designing to the user's mental model reduced learning time. Similar to ease of use, responses relating to ease of learning indicate strong congruence between CIANC^3 system design and the participants' mental model for maintaining situation awareness and making battle command decisions. In general, computer-based skill training requires a combination of procedural learning that maps computer procedures to operational needs as well as declarative knowledge that maps system-implementation concepts into operational concepts. Minimizing the amount of knowledge necessary to make such system-operational mappings can dramatically reduce learning time. Notably, the participants had a puckster to help manipulate the CIANC^{3}controls. This meant that the participant did not necessarily need to learn the specific controls required for manipulating the interface elements. Although this discounts the high responses to some degree with respect to procedural learning, the results do support the contribution that matching system-implementation concepts and operational concepts can have on system usability.

Open-Ended Question Results

Participant comments to the open-ended questions on the Post-Evaluation Survey provided many useful suggestions for improving the CIANC^3 system as well as positive support for selected system concepts and functions. Features regarded as most useful included participants' strong endorsement of the CIANC^3 system's use of military schemas and a decision-centric approach to GUI design. As noted, the Plan Display (see Figure 6) presented users dedicated information panes for mission objectives and decision points. Participants reported that this aspect of the design provided the commander with relevant information in an understandable and actionable format that explicitly linked agents' activities with humans' decision-making processes. Participants also indicated this design provided an intuitive means for controlling subordinate activities at a more macro level.

One aspect of the CIANC^3 system that participants regarded with ambivalence was that of automatic tasking, such as re-tasking UAVs when one has been destroyed. Underlying there responses was a general concern that the pace and demands of future warfare (as characterized in the evaluation mission) would be difficult to manage without some assistance. Specifically, the idea of commanding unmanned systems in addition to human forces seems to increase workload on what is already a very demanding task. In this respect, the participants universally agreed that systems such as CIANC^3 would be welcome, if not essential for future warfare. What did not seem natural for participants was giving up control or trusting battlefield decisions and actions to a machine. While such concerns are normal, it should be noted that unconditional acceptance of battle command automation cannot be taken for granted. Furthermore, participants seemed to agree that systems that do provide task automation must be able to explain their actions or decisions when requested. Thus, participants simultaneously stressed the need for the assistance provided by the CIANC^3 system as well as the need to constrain that assistance to non-mission critical tasks.

Focus Group Results

The focus group discussion centered on how an intelligent battle-command system, as represented by the CIANC^3 prototype, might be used in a real-world environment. The written questions, and subsequent group discussion, were designed to guide participants along a chain of reasoning that included problem characteristics, problem definition, possible solutions, and ideas for further extensions and applications. Giving the participants some experience with the CIANC^3 system seemed to help solidify the abstract nature of intelligent automation systems and provided them with a command and concrete example on which to base discussion.

The participants' characterized situation awareness challenges mostly as expected. They described the challenges in terms of standard battle command tasks such as, determine enemy and friendly force location and status, assess and prioritize enemy threats and, in general, to "know the situation your Soldiers are facing." One unexpected challenge (given the nature of the mission scenario) was that of maintaining SA within a building or other structure. While this challenge seems specific to a narrow range of urban warfare missions, the problem can be generalized to any situation where visual and verbal cues are not available to the commander. This becomes very important to the design of future battle command systems where use of the system for SA may not be optional.

Participants also discussed situations and decisions where some form of automated assistance would be especially useful. The group initially focused on conditions when their normal cognitive abilities would be impaired, such as after conducting combat operations for multiple days with little sleep or rest. The group's desire however was not to give up control to the automated system, but rather to have the system help provide "sanity checks" on the decisions they would be making under stress and impairment. Other tasks discussed for which automation would be useful included clarification and display of adjacent friendly units and other available combat multipliers and assets. More mundane tasks included determining route feasibility, refuel point planning, and other logistical planning.

When asked to assess the value of automating specific activities, participant responses varied greatly. Several activities, however, were ranked highly by multiple participants. These included:

- Checkpoint placement – automate route planning, specify discrete points along the route for status checks and automatically processing and displaying status reports related to movement.

- CCIR and other reporting – automate and make explicit the linkage between observed world data and information requirements by superiors. Automate as much as possible the content of CCIR reporting and other situation reports.

- Movement and hazard-avoidance – automate the display of known hazards, obstacles, and alternate movement routes.

- Logistics and resupply – track fuel and ammunition levels and automate scheduling of resupply and maintenance, especially when commander is engaged in combat.

In general, the Focus Group participants were very supportive of the research pursued in this project and the prototype that was developed. Furthermore, they universally agreed that in the future much more automation would be useful and necessary. However, as with other participants, the Focus Group cautioned that too much automation, or poorly designed automation would be quickly rejected. Again these concerns seemed primarily focused on ability to control and predict system behavior, and to be able to inspect system reasoning when necessary.

Discussion

This section briefly summarizes the CIANC^3 system successes and issues associated with the Phase II evaluation. The issues identified based on participants' responses provide useful recommendations for refining the CIANC^3 system and adjusting the balance of human-machine control and functions in future development. The section closes with a discussion of how providing an assistant puckster may impact the evaluation of novel and complex systems.

System Successes and Refinement Issues

The CIANC^3 system as designed and evaluated showed a great deal of promise, but also exposed a number of critical issues with using agent-based technologies to support mixed teams of human and robotic forces on a dynamic battlefield.

Among the successes of the evaluation were:

- Demonstrated the ability of intelligent agents to control and coordinate robotic entities, allowing commanders to focus on higher level objectives.
- Demonstrated the use of intelligent agents to maintain and respond to changes in the battle-plan, helping the commander maintain SA and mission tempo.
- Demonstrated the use of a schema/decision-centric approach to GUI design, which presented the commander with relevant information in a form that linked the agent system with their decision making processes.

As expected, there were some critical issues raised by the participants during the evaluation. In particular, these issues centered on the need to balance human-machine control and functions with a decided emphasis on machine support and human control. Based on this evaluation, critical system requirements for such systems should include:

- The commander to be able to override decisions at any time.
- A system with sufficient knowledge to produce doctrinally-correct suggestions.
- The system to be able to explain and justify suggestions.
- The commander to be able to reject and/or improve upon system suggestions.

- Visual thinking support.
- Complementary display forms to 'snap together', highlighting common information across displays.
- The system to provide enough information for the commander to maintain SA and be able to confidently make his or her own decisions.
- Flexibility.

In sum, the evaluation was a positive step toward demonstrating that an intelligent agent system can support the commander's management of human and robotic teams. Future CIANC^3 research efforts, however, must increase system competency, trustworthiness, and supervisory control. Refinements should also stress improving system flexibility so that commanders can more readily accept, modify, and improve system suggestions as well as better support the situational awareness of the commander.

Usability Evaluations for Complex Systems

Much of the literature and practice on usability testing assumes that the system being evaluated should be as close to "walk up and use" as possible. This is often not the case, however, particularly when the evaluation focuses on a novel and complex system such as a futuristic military command and control system. For such systems a substantial amount of training and experience is required to master the tactical and technical skills required to complete the mission and supporting tasks most germane to system objectives (Lickteig et al. 2003).

As noted, the primary reason for using a puckster was to focus the participant and the evaluation on basic system concepts and functions rather than more minor and modifiable implementation issues. At this early stage of research and development there is more interest in how the warfighter reacts to the core functions of the system than in small usability details. It was also anticipated that the use of a puckster might increase the quantity and at best quality of each participant's verbalization of thought, intention, and action during the conduct of the mission exercise.

Post hoc, it seems that use of a puckster was a mixed blessing. It did help the warfighter focus on understanding and applying the CIANC^3 concepts and functions through all phases of the scripted mission and supporting tasks on successive trials. However, it also meant that the warfighter could multi-task more easily by assigning the puckster to interact with one task or operational concern while the warfighter moved on to assessing a different task or decision point. Such multi-tasking is inappropriate for a system intended for use by an individual warfighter. Having a puckster also may have altered think aloud verbalizations. Many of participant verbalizations requested specific system interactions by the puckster at a cost perhaps to verbalizations related to situation assessment and decision-making.

Method refinements might at least partially overcome some of the negative impacts of providing an assistant puckster. Probes or queries might be inserted during the mission requesting the participant to provide ongoing assessments of the situation to

surrogate higher commanders. In addition, usability evaluators might develop methods for relating micro behaviors such as human-computer interactions, or participant requests for such interactions, to more macro command and control functions (see Lickteig, Sanders, Durlach, Lussier, & Carnahan, 2004). Most importantly, the greater the investment in participant training and experience early in system design and development, the greater the return on that investment.

Conclusions

The CIANC^3 system developed during this project allowed evaluation participants to successfully complete several clearly-defined performance tasks in a simulated FCS scenario. The feedback received from participants was supportive and constructive, providing an empirical base for further research and development. User satisfaction with the potential for the new system was demonstrated by a majority of user requests being for more of the types of automation provided by the CIANC^3 system. While some participants wanted more control and less automation, this was possibly due to intentionally limiting the complexity of the performance tasks. The prototype system appeared to be a good platform for training and performing unmanned asset management. Participants were able to effectively manipulate the position of multiple unmanned sensor assets in a way that maximized sensor coverage for the mission.

The most important results may prove to be Phase II advances in mixed-initiative technologies at the command, versus the operator, level. A triad of intelligent agents, Tasking, Coordinating, and Monitoring, has proven to be able to form the nucleus of an intelligent user interface for command and control. These agents, which roughly correspond to command, control, and communications, respectively, can work as a virtual command staff for users to reduce workload and simplify complex tasks. Well-defined protocols for inter-agent communications were developed as well as the establishment of responsibilities, permissions, and prohibitions for those agents. Finally, this project proved to be a key enabler for future knowledge-rich intelligent systems via the development of bridge technology that connects ontologies with agent systems. Although these results are encouraging, future work should explore scalability issues; such as how cooperative agent clusters can operate and coordinate across echelons, in more complex scenarios, and under more realistic conditions.

Acknowledgements

The work described here was conducted under a Department of the Army Small Business Innovation Research Program 2000.2 contract for topic A02-024 (contract #DASW01-03-c-0019). We gratefully acknowledge the sponsorship of this research by the United States Army Research Institute.

References

Endsley, M. R. (1988). *Design and evaluation for situation awareness enhancement.* Paper presented at the Human Factors Society 32nd Annual Meeting, Santa Monica, CA.

Endsley, M. R. (1995). *A taxonomy of situation awareness errors*. Lubbock, TX: Texas Tech University.

Endsley, M. R., Garland, Daniel J. (Ed.). (2000). *Situation awareness analysis and measurement*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Lickteig, C. W., Sanders, W. R., Durlach, P. J., Lussier, J. W. & Carnahan, T. J. (2004). Human performance essential to battle command: Report on four future combat systems command and control (FCS C2) experiments (ARI Research Report 1812). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Lickteig, C., Sanders, W., Lussier, J., & Sauer, G. (2003). Human Systems Integration for Future Command Systems: A Focus on Battle Command. *Proceedings of IITSEC 2003*.

Wood, S. D. (2003). Cooperative Interface Agents for Networked Command, Control, and Communications (CIANC^3): Phase I Final Report (No. 1134): United States Army Research Institute for the Behavioral and Social Sciences.

Wood, S. D., Kieras, David E. (2002). Modeling Human Error for Experimentation, Training, and Error-Tolerant Design. Ann Arbor, MI.

Wood, S., Zaientz, J., Beard, J., Frederiksen, R., Lisse, S., & Huber, M. (2004). An Intelligent Interface-Agent Framework for Robotic Command and Control. Paper presented at the Command and Control Research and Technology Symposium, San Diego, CA.

Wood, S., Zaientz, J., & Lickteig, C. (2006). Cooperative Interface Agents for Networked Command, Control and Communications: Phase II Final Report (No. TR-1179): United States Army Research Institute for the Behavioral and Social Sciences.