

ASSESSING INTELLIGENT SOFTWARE AGENTS FOR TRAINING MARITIME PATROL AIRCRAFT CREWS

Stuart C. Grant
DRDB Department of National Defence
Toronto, Ontario Canada

ABSTRACT

Use of intelligent software agents to play the role of supporting personnel in simulations has the potential to reduce the requirement for support staff and increase the instructor's control of the simulation. The research presented here examines the suitability of intelligent software agents to aid training of individual crewmember skills and team skills. In this evaluation, human crews, intelligent agent crews, and mixed human-agent crews performed a simulated antisubmarine mission by a CP140 Aurora maritime patrol aircraft. Mission performance was recorded and crew communications were observed and rated to determine whether the intelligent software agents could perform individual crewmember functions and whether they could provide the interaction necessary for crew coordination training. The results indicate that (1) the intelligent software agents can perform individual crewmembers' functions adequately; and (2) the intelligent software agents did not interact in a way suitable for crew coordination training. The paper concludes with a discussion of the generalizability of the results and the growth potential of intelligent software agents in crew coordination training.

ABOUT THE AUTHOR

Stuart C. Grant is a Defence Scientist with the Defence Research and Development Branch of the Canadian Department of National Defence. He conducts human factors research on the use of emerging technologies for training, including virtual reality technologies and intelligent agents. Current projects include simulation for dismounted combatants and the acquisition of team skills in virtual environments. He received his Ph.D. in cognitive psychology from the University of Toronto in 1994.

ASSESSING INTELLIGENT SOFTWARE AGENTS FOR TRAINING MARITIME PATROL AIRCRAFT CREWS

Stuart C. Grant
DRDB Department of National Defence
Toronto, Ontario Canada

This paper reports an evaluation of the suitability of intelligent software agents for training team skills. The evaluation was conducted in the context of antisubmarine operations by Canada's CP140 Aurora maritime patrol aircraft, a system that demands expert teamwork from its operators.

Increasing attention is being paid to team skills as military systems become more complex and joint operations the norm. As a result, there is greater recognition that mission success requires skilled individuals working cooperatively in an organized fashion towards common goals (Hill 1982; Kanki and Palmer 1993) (Salas, Dickinson et al. 1992). In the case of the Aurora, the complexity of the systems necessitates a crew of 10. The tactical navigator (TACNAV) leads the crew in tactical situations, integrating information and directing courses of action. The information exchanges amongst the TACNAV, acoustic sensor operators (ASOs), non-acoustic sensor operators (NASOs) and the navigator communicator (NAVCOMM) as well as the flight deck crew must be appropriate, accurate and timely.

Achieving skilled team performance can be resource-intensive. Simply assembling a team for training can tie up a significant number of people whose time is increasingly scarce. This, of course, is essential for those that require training, but it provides little return for the team members who are already trained and whose presence is required only to train others.

A potential solution is suggested by recent simulations using the emerging technology of intelligent agents to play the role of supporting staff and players in training simulations. Intelligent agents are promising because the nature of their design means that they can capture some of the strengths and weaknesses of humans that are desired in simulations while retaining the controllability, repeatability, and affordability of

computer software. Like expert systems, intelligent agents are programmed to solve problems within a specific domain using whatever algorithms are appropriate. In addition, agents are reactive in the sense that their behavior changes in response to ongoing changes in their environment, and they are proactive in that they seek out information and take actions without prompting in order to achieve their goals. Furthermore, agents are social. They interact with humans, other agents, or conventional software to perform their functions (Wooldridge and Jennings 1995). The simulation and training community is beginning to exploit these capabilities to play the role of individuals (Glenn, Stokes et al. 1998; Howard and Lee 1998) and entire teams (Hill, Chen et al. 1998; Zachary, Ryder et al. 1998). The Virtual Crewmembers Simulator (VCMS) evaluated here attempts to go one step further by allowing a human student to obtain team skills by interacting with a team of intelligent agents.

THE VIRTUAL CREWMEMBER SIMULATOR

This evaluation considered an early prototype of an agent-based simulator built by IntelAgent R & D. The Virtual Crew Member Simulator (VCMS) prototype is a partial simulation of the tactical compartment of the CP140 Aurora. It can be operated by human crewmembers, by software agents, or a combination of humans and agents. The sensors, systems, and weapons available to the TACNAV, ASOs and NASOs are simulated, and they interact with a commercial scenario generator that controls the environment and other entities in the simulation. At this stage of development the sensors, systems, and weapons simulations are only developed to the degree required to perform passive tracking and attack on a nuclear submarine. They operate in the same way and have the same functionality for human and agent users, who interact with them through a

mock-up of the crew console or an application programming interface, respectively.

Intelligent agents are available to play the roles of the ASO and TACNAV. The agents interact solely with the sensor and systems simulations. They do not have privileged access to scenario data and have no capabilities or knowledge that are unavailable to human users of the VCMS. One ASO agent and three TACNAV agents (standard, novice, and slow) are available. The agents are programmed using SOAR, with 369 and 533 rules used for the ASO and TACNAV, respectively. Their rule bases allowed them to localize and track submarines using passive acoustics. Rules covering the attack aspect of the mission were not available for evaluation.

Crew communication in the VCMS is accomplished through a simulation of the aircraft's intercommunication system (ICS). The ICS simulation is text based, where the human or agent crewmember assembles a message by selecting message fragments from a menu and typing the values for variables, such as speeds or sonobuoy numbers. The text output of the system is supplemented with voice production so that the human crewmember need not monitor the ICS text window. Other than that, the communication system is the same for both humans and agents.

THE EVALUATION

The evaluation of the VCMS required qualified TACNAVs to work with an agent ASO (denoted as a human-agent crew) to perform antisubmarine missions in the VCMS and then perform a similar mission with a human crew (denoted as a human-human crew) in the Operational Mission Simulator (OMS), the existing tactical simulator for the Aurora that accommodates the entire tactical crew. In addition, crews consisting of an agent TACNAV and an agent ASO (denoted as an agent-agent crew) performed the same mission in the VCMS (without the attack component). The evaluation sought answers to two questions:

- Can the agents competently play the role of crewmembers?
- Do the intelligent agents support the acquisition of team skills without requiring a full team?

Given that the purpose of the agents is to play the role of human actors in the simulation, they must be able to perform the tasks required of these actors. Thus the first point was addressed by having a subject matter expert blindly rate the performance of human TACNAVs and agent TACNAVs performing ASW missions in conjunction with an agent ASO. How closely the agent-agent performance approaches the human-agent performance served as a measure of the agent's competence.

The second question was addressed by examining team communications in the VCMS and OMS. The focus was on communications because it is identified as both the basic medium of team processes as well as being the behavior that best demonstrates the team skills of the team members (Foushee 1982; Oser, McCallum et al. 1989; Kanki and Palmer 1993). Communications in the VCMS and OMS were categorized using the Crew Communications Behaviour Categories (Banks, Hendy et al. 1996) and the distributions of utterances in human-agent and agent-agent crews compared those of human-human crews.

Method

The experiment used a repeated measures design. TACNAVs first worked with the agents in the VCMS and later worked with their own human crews in the OMS. All testing was conducted in the facilities of Canadian Forces 14 Wing, at Greenwood, Nova Scotia.

Subjects

Fourteen TACNAVs and their crews acted as research participants in the evaluation. Their experience serving as qualified TACNAVs ranged from 6 to 96 months. Unfortunately, equipment failure and operational requirements prevented some crews from completing the study, thereby reducing the number of complete data sets to nine.

Apparatus

The crew station in the Aurora provides a CRT display and a keyset with a trackball to the TACNAV. Identical devices are used in the OMS, which faithfully reproduces the full interior of the Aurora's tactical compartment. In the VCMS one computer monitor presents the same information as the Aurora's CRT as well as the menu interface

to the ICS. A computer keyboard is used to type in the numbers needed to complete some ICS messages. A touchscreen monitor emulates the functionality of the Aurora's keypad while a mouse takes the place of the trackball.

Awareness: Declarative statements concerning present states – e.g. On heading for E1
Preplanning: Explicit statements of future intentions or plans – e.g. I'm putting in an intercept. I'd like you to stay back so you can come down for the attack.
Task Prioritization: Overt direction or request for direction concerning task prioritization – e.g. Pilot, we want E1 then E2.
Criticism: Negative statements concerning performance – e.g. All you have to do is ask, eh? (sarcasm)
Positive Reinforcement: Positive statements concerning performance – e.g. Well put
Crew Coordination: Explicit statements concerning the apportioning of tasks between crew – e.g. Roger that's fine go to the next one. I'll destroy this one.
Cross Checking: Explicit noting of deviations with respect to another crewmember's actions – e.g. How are you doing on that one?
Implications: Explicit recognition of the consequences of present system states – e.g. Buoy 19 is going to die and we'll lose attack criteria
Queries: Questions or soliciting information – e.g. Is he showing higher on 9 and 10?
Proactive: Explicit proffering of information, not in response to a query – e.g. Tac, do you want a time slice right now?
Directive: Commands or directives – e.g. OK pilot, let's descend to 300 feet
Input: Information provided in response to a query – e.g. Your strongest contact is 10.
Acknowledgement: Statements to acknowledge actions or statements – e.g. Roger.

Table 1. CCBC categories used to categorize utterances.

To characterize the teamwork functions served by the communications amongst the crews two judges categorized each utterance made in the VCMS or OMS was categorized using the CCBC (Banks, Hendy et al. 1996). The keying of the microphone in the OMS or the contents of the text string in the VCMS defined the beginning and end of an utterance. An utterance could be assigned to one or more categories. Agreement between the

judges was substantial, $\kappa = 0.65$. The subset of the CCBC categories used in this evaluation is presented in Table 1.

The Standards and Training Officer responsible for evaluating TACNAVs served as the subject matter expert. He rated mission performance using the TACNAV B Cat Checkride Assessment Form, an instrument used to qualify TACNAVs. The form provides a rating scale, from 1 to 5, for each task performed by the TACNAV. Higher numbers denote superior performance and the form provides behavioral anchors for assigning ratings. The distribution of ratings generates a letter grade as well as a pass or fail assessment. A TACNAV is deemed to fail when more than 20% of the tasks are rated 3 or less, or where the deployment of weapons or search stores is rated 2 or less. The subject matter expert completed only the sections of the form relevant to the scenarios used, including the section provided for unstructured comments.

Procedure

TACNAVs first completed two ASW missions in the VCMS working with an agent ASO. Next, they completed another ASW mission in the OMS working with their usual human crewmates. The time between the VCMS session and the OMS session was approximately 12 weeks.

At the beginning of the VCMS session the experimenter explained to the TACNAVs that their participation was sought to assess the utility of the agent technology, and not to assess their skills. After the TACNAVs consented to participate in the study they were familiarized with the VCMS system. The limitations of the VCMS were explained and they practiced using the touchscreen controls and interacted with the ASO via the ICS until they were comfortable with their knowledge of the system, normally about 5 minutes. At that time the experimenter briefed the TACNAVs on the first scenario and answered any questions about the VCMS and the scenario.

In the scenario briefing, the TACNAVs were told the location of the scenario and that their mission was to localize and track a hostile submarine known to be somewhere in their area of responsibility. They were told that there were no other aircraft, ships, or submarines in the area,

only themselves and the target. The rules of engagement at the start of the scenario did not allow them to attack the submarine immediately, but they were told that authorization might come during the scenario.

The TACNAVs then began the first scenario. It begins with a pattern of 12 sonobuoys in the water and the submarine located so that one sonobuoy quickly obtained direct path contact with the submarine. The submarine followed a preset course away from the initial pattern of sonobuoys, changing its speed and heading periodically to make tracking challenging and make difficult the planning and placement of new patterns of sonobuoys. The initial contact was sufficient to enable the ASO to provide a rough bearing to the submarine, but a new pattern of sonobuoys had to be quickly laid before the submarine moved out of range. Twenty five minutes into the scenario the experimenter informed the TACNAVs that they were authorized to attack the submarine. The scenario ended when the TACNAV hit the submarine with a torpedo or the when 60 minutes elapsed. All TACNAVs were able to hit the submarine before the deadline.

The second scenario was then briefed and played out. The second scenario was similar to the first, except that it began with a pattern of two sonobuoys and the authorization to attack came in 20 minutes. Obviously, the position and course of the submarine were different. All TACNAVs successfully attacked the submarine before the end of the scenario.

At the end of the VCMS session, the TACNAVs completed a questionnaire soliciting their opinions on various aspects of the VCMS.

In the OMS session, the TACNAVs performed one ASW mission that was similar to the ones they performed in the VCMS. The whole crew was briefed on the scenario and the scenario continued until either the submarine was sunk or four hours elapsed. Data collection ended, however, when the first torpedo hit the submarine or the four-hour time limit was reached. All crews were able to successfully attack the submarine before the deadline.

The all-agent crew also performed the two VCMS scenarios. Because the agents did not have the

rule base required to carry out an attack, they continued to track the submarine until the time limit expired.

Results

Analysis of the data from the VCMS began with replaying the recorded missions by all the crews and observing them for notable characteristics. This revealed both similarities and differences between the human and agent TACNAVs. Figure 1 is a view of the tactical situation as it appeared on an agent TACNAV's CRT during the first scenario. Moments before this point in the scenario the ASO obtains contact with the submarine just long enough for the TACNAV to estimate its course.

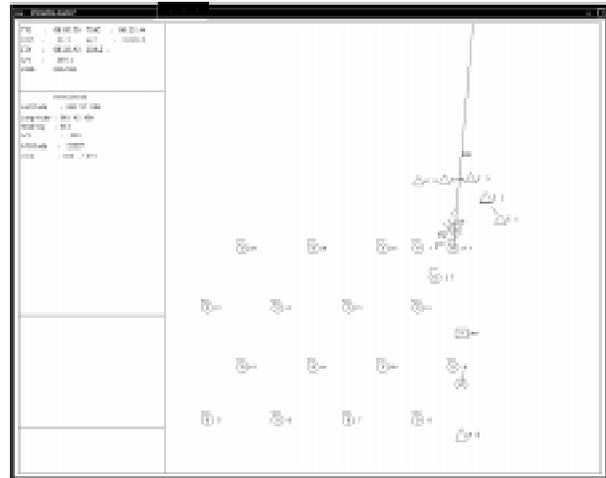


Figure 1. Tactical display during scenario 1, as performed by an agent TACNAV. Sonobuoys in the water are represented by δ = sonobuoy in the water; Δ = planned sonobuoy positions; \vee = estimated position of the submarine; \odot = true position of the submarine.

When contact is then lost and the submarine changes course from 015 to 090, the behavior of the agent TACNAV is much like that of the human TACNAVs. They both rely on the existing pattern of sonobuoys to cover possible movement to the south, and they plan another pattern of sonobuoys ahead of the submarine based on the assumption that the submarine maintains the same course.

The granularity of their actions differ, however. The agent plans a five buoy wedge pattern, based on one of the preset patterns available in the aircraft mission system, whereas human TACNAVs typically drop less than a complete pattern and

thereby conserve buoys and freedom of movement. Furthermore, although contact is regained shortly after this point in the scenario, and the new course of the submarine obtained, the agent remains committed to completing the previously planned operation, the five bouy wedge pattern to the north, despite its updated knowledge of the submarine. Human TACNAVs, of course, are quick to interrupt their previous pattern at whatever stage of completion and immediately adapt their behavior to the new data.

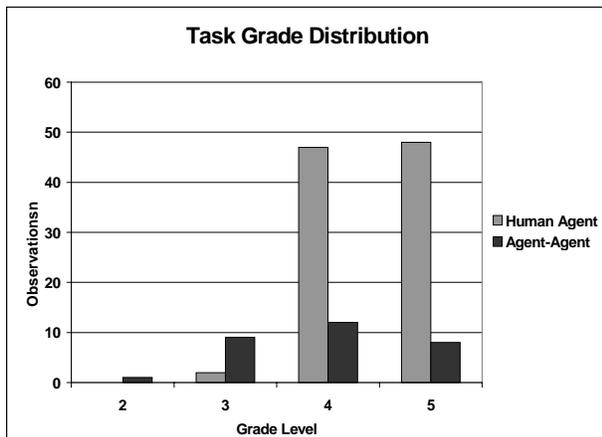


Figure 2. Distribution of grades assigned to crew task performance.

Subject Matter Expert Ratings

Distributions of the ratings assigned by the Standards Officer to each of the tasks performed by the TACNAVs in the VCMS are presented for Human-Human crews and Human-Agent crews in Figure 2. Considerable overlap in the distribution of ratings earned by the two different types of crews is apparent, although the higher ratings tend to be earned by the Human-Agent crews. Using a Mann-Whitney U test comparing the modal ratings earned by the human-agent crews and the agent-agent crews is inconclusive because of the large number of ties in the data, $U = 7.00$, $p > .1$. This is unavoidable due to the deterministic nature of the agent crews, the limited number of different agent crews, and the limited behavioral repertoire of the agent crews. Turning to the pass/fail data, only the Agent-Agent crews performing the first scenario warranted failing grades. Second scenario performances by the agent-agent crews and all missions by the human-agent crews received passing grades.

Communications Analysis

The distributions of utterance types with a frequency of at least 1% are presented in Figure 3 for the three different crew types (human-human, human-agent, and agent-agent). The utterance types not appearing in the figure (task prioritization, positive reinforcement, cross-checking, crew coordination, criticism, and proactive) were used only in the human-human crews. It is apparent that the communications involving agents were largely confined to awareness statements and directives. Not surprisingly, when χ^2 goodness of fit tests were used to compare the utterance distributions of each of the human-agent and agent-agent crews to the utterance distribution that characterized human-human crews, the result of each test was significant, χ^2 (df=12) > 26.22, $p < .01$.

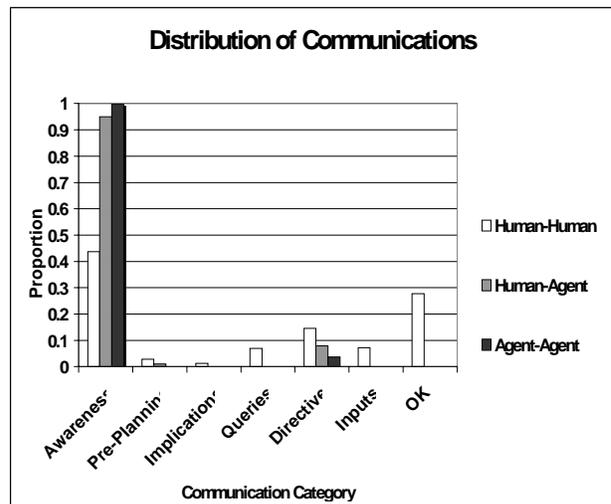


Figure 3. Distribution of utterance types by crew type.

Conclusions

Observing the performance of human-agent crews and agent-agent crews in contrast to human-human crews revealed both strengths and weaknesses of the agent paradigm that suggest how the technology might best be used.

Agents for OPFOR

The performance of the agent-agent crews was reasonably good. They earned passing grades in one scenario and failing grades on the second. Furthermore, two of the agent-agent crews actually outscored one human-agent crew in one of the scenarios. It is clear that competent performance

from simulated crews can be achieved via the intelligent agent paradigm. Such a result from an early prototype is encouraging and suggests that intelligent agents are well-suited to playing the role of an opposing force.

Agents for Individual Training

The passing grades obtained by the human-agent crews demonstrate that the agent ASOs were competent enough to allow human TACNAVs to exercise their individual skills. The rudimentary ICS available at this stage in the VCMS development provided enough communications between the human and agent for them to jointly achieve the mission objectives. This reveals that intelligent agent technology can support the task performance and interactivity required for training individual skills.

Agents for Crew Training

The communications analysis revealed clear differences between the human-human crews and the crews with one or two agents. Some of the discrepancy arises from rather minor shortcomings, such as the absence of acknowledgements in the crews containing agents. Although acknowledgements are important communications acts, their absence is merely reflective of the prototype status of the VCMS and they could easily be added to the agent's behavioral repertoire.

Other discrepancies pose a greater challenge. Only the human crews made crew coordination and cross-checking utterances. The immediate, facile explanation is that the ICS interface did not offer these kinds of utterances. The ultimate reason, however, is that the agents do not have the means of incorporating into, or deriving from, their knowledge bases, these kinds of metacognitive statements. To meaningfully use these kinds of utterances requires the agent to have a model of the current functioning of the other crewmembers.

For this same reason, human TACNAVs rarely made preplanning utterances to agents, and agent TACNAVs never did. This was so despite there being such an utterance (the attack briefing) available for broadcast on the ICS. They did not give an attack briefing because doing so had no effect on the agents; the agents did not have a shared mental model of the situation that could be updated. Coupled with this is the difficulty of

programming agents that can interpret such linguistically complex statements. These statements are more difficult for artificially intelligent systems of any kind to handle than directives and awareness statements owing in part to their more varied syntax and less constrained vocabulary.

The inability for these agents to support some of the kinds of communications that are central to effective teams limits their usefulness for training team skills. It is important to note however, that this shortcoming does not apply in principle to all agents, or even these agents in particular. Dedicated programming effort may remedy this shortcoming. The point is simply that the solution does not automatically or elegantly follow from the intelligent agent paradigm in the same way that other attributes do.

Recommendations for using Intelligent Agents for Virtual Role Playing

Current intelligent agents are suitable for playing the role of individual actors in simulations, as demonstrated here and in the work of others (Glenn, Stokes et al. 1998; Hill, Chen et al. 1998; Howard and Lee 1998; Zachary, Ryder et al. 1998). They are best used where their interactions with humans will occur through either standardized communications protocols or through non-linguistic means (e.g. agents as OPFOR). Humans and agents should not be mixed where their interaction is unconstrained and not formalized; rather, greatest success will be enjoyed when agents represent all of the actors within a boundary defined by formal communications procedures.

REFERENCES

- Banks, R. D., K. C. Hendy, et al. (1996). Human Factors Study of CC-130 Operations. Toronto, ON, Defence and Civil Institute of Environmental Medicine.
- Foushee, H. C. (1982). "The Role of Communications, SocioPsychological, and Personality Factors in the Maintenance of Crew Coordination." *Aviation, Space, and Environmental Medicine* **53**(11): 1062-1066.
- Glenn, F., J. Stokes, et al. (1998). *A Simulation Model of the Operator of an Air Defense Weapons System*. Seventh Conference on Computer

Generated Forces and Behavioral Representation, Orlando, FL.

Hill, G. W. (1982). "Group Versus Individual Performance: Are N + 1 Heads Better Than 1?" Psychological Bulletin **91**(3): 517-539.

Hill, R. W., J. Chen, et al. (1998). Soar-RWA: Planning, Teamwork, and Intelligent Behavior for Synthetic Rotary Wing Aircraft. Seventh Conference on Computer Generated Forces and Behavioral Representation.

Howard, M. and C. Lee (1998). Architecture of a Generic Command Entity. Seventh Conference on Computer Generated Forces and Behavioral Representation, Orlando, FL.

Kanki, B. G. and M. T. Palmer (1993). Communication and Crew Resource Management. Cockpit Resource Management. E. L. Wiener, B. G. Kanki and R. L. Helmreich. San Diego, CA, Academic Press: 99-137.

Oser, R., G. A. McCallum, et al. (1989). Toward a definition of teamwork: An analysis of critical team behaviors. Orlando, FL, Human Factors Division, Naval Training Systems Center.

Salas, E., T. L. Dickinson, et al. (1992). Toward an Understanding of Team Performance and Training. Teams: Their Training and Performance. R. W. Swezey and E. Salas. Norwood, NJ, Ablex Publishing Corporation: 3-29.

Wooldridge, M. and N. R. Jennings (1995). "Intelligent Agents: Theory and Practice." The Knowledge Engineering Review **10**(2): 115-152.

Zachary, W., J. Ryder, et al. (1998). Simulating Behavior of Tactical Operators and Teams Using COGNET/GINA. Seventh Conference on Computer Generated Forces and Behavioral Representation.