Robotics Operator Manager ACT-R Model and Validation

Daniel N. Cassenti & Kristin E. Schaefer
U.S. Army Research Laboratory
Human Research and Engineering Directorate
Aberdeen Proving Ground, MD 21005
daniel.n.cassenti.civ@mail.mil

Keywords:
Cognitive Modeling, Network Science, Decision Making, ACT-R

ABSTRACT: With the United States military's proposed increase in activated robots in the field, and the subsequent mental workload demands associated with control of each robot, there is an opportunity to form a new role for the management and coordination of multiple robotic assets. The role of a Robotics Operator Manager (ROM) was studied using a Model Test Model (MTM) approach during a scenario in which robots were used to detect engineered explosives during a convoy mission. This work is the final phase in the MTM approach. We begin by: 1) discussing the ROM role; 2) provide a brief discussion of the Initial Model; 3) explore the Test phase, which showed promise to accomplish military missions with robots in light of the negative effects of time and bandwidth limits on performance and trust in network; and, 4) conclude with a modeling and simulation approach for the Final Model using Adaptive Control of Thought-Rational (ACT-R) modeling. An examination of the results of the Final Model indicated not only a close approximation of the mental processes required to fill the ROM role, but also that the Final Model can be used to establish a personnel selection process.

1. Introduction

The U.S. military increasingly uses robots in their operations (Chen, et al., 2006). Robots have the potential to provide a number of benefits from completing repetitive tasking to functioning in hazardous or life-threatening environments. In the latter, the inclusion of the robot may be used to provide additional close-to-real-time data that can be used to enhance the decision-making capability of the Soldiers with their "boots on the ground" (e.g., surveillance operation), or the inclusion of a robot in the mission may be the means to provide some additional space between the Soldiers and the potential threat (e.g., bomb disposal).

Regardless of the type of fielded robotic asset or task-based operation, human-robot interaction (HRI) within military operations requires a human-in-the-loop approach in which an individual (or set of individuals) interacting with or providing some degree of control over the robotic asset is referred to as an operator. One problem with the control of robots is the potentially high mental workload costs (depending on degree of automation) required to operate a robot (Cassenti, et al., 2012; Yagoda, 2010). Mental workload is the proportion of limited mental resources required to perform a task or set of tasks (Wickens, 2002) with performance generally decreasing when workload increases. At least two factors contribute to the high mental workload required to operate a robot – the demands on perceptual, cognitive, and motor skills of operating a robot and transferring situation awareness from body to robot (Riley & Strater, 2006).

Goodrich and Olsen (2003) claimed operators need a given proportion of time to neglect robots by allowing the robot to function autonomously while the operator performs other necessary tasks. Operators may neglect the robot without sacrificing performance to the extent of the robot's autonomous mode reliability. However, no robot is completely error free, so full neglect is not possible. Cassenti et al (2012) found that even single digit arithmetic with single digit answers is demanding enough as a secondary task to hurt performance when using a semi-autonomous robot to perform a search and find task. They concluded that robot operation is mental resource intensive and should be restricted to one robot at a time, with a minimum of simultaneous task performance.

This issue of mental workload and the operator's capability for multi-tasking takes on greater importance when we begin to consider the future goal of one operator for multiple robots. While the Department of Defense (DoD) goal has encouraged this one-operator-to-many-robots structure (Agah & Tanie, 1999), research has shown that simultaneous control of multiple robots is neither efficient nor effective to achieve mission success (Chen, at al., 2008; Schurr, 2007). More specifically, research has shown that within the current instantiation of HRI, safety drives the human-robot ratio, and thus recommends a 2:1 (N_{human}:$N_{vehicles}$) ratio for unmanned ground vehicles (UGV) or 3:1 ($+1_{safety\ officer}$) ratio for unmanned aerial vehicles (UAV), and 3:1 for mixed cooperative UGV-UAV teams (Burke & Murphy, 2007; Murphy, et al., 2006).

The issue of the human-robot ratio has been addressed in a few different ways. One of the key recommendations has been in the advancement of asset automation and subsequent autonomy. Murphy and Burke (2010) used view-point oriented cognitive work analysis to determine that it is the "perceptual foci of the team members as the key affordance in determining which roles can be safely merged or made autonomous." One limitation of this finding is that the human-robot ratio is then dependent on the specific capabilities of the human team members to collaborate with advanced automation, and may vary based on specific contexts. One continued recommendation has been the inclusion of an overarching "director" role to maintain adequate, overall awareness of the entire situation during multi-robot control (Murphy, et al., 2008). One previous instantiation to address this issue was the development of RoboLeader, an intelligent agent that assists human operators in controlling a team of robots through a single user interface (Chen, et al., 2011; Chen, at al., 2011). Here, we take a different approach, and look at the potential for incorporating an additional human role of the Robotic Operator Manager (ROM) to manage multiple operators for joint UGV and UAV missions.

2. Robotics Operator Manager

Cassenti et al. (2011) proposed adding a role to Army operations called the Robotics Operator Manager (ROM). The ROM would be filled by a military officer at the platoon level and tasked with coordinating and commanding squad member robot operators. The role of the ROM is designed to be restricted to decision making only (e.g., decide which robotic assets to deploy on missions, how to coordinate deployment of multiple robots, and the locations to send the robots). Operators would still be required to attend to perceptual activity, motor skills, and emphasizing situation awareness from the robot's environment, not the operator's environment. The ROM would decide whether to send an SUGV or SUAS, where and when the robots would go, and process the information gathered by the robot to make mission-critical decisions.

Implementing a ROM makes sense in theory; however, a concerted effort to research the role is required to examine the viability of the concept before possible implementation. The Model Test

Model (MTM) approach (Mansager, 1994) was adopted to research the potential benefits of incorporating a ROM into the command structure. The Initial Model phase required the development and assessment of a computational model to determine whether potential issues and problems would prevent the concept from being usable (see Cassenti, et al., 2011). After establishing computational viability, the Test phase was implemented an examination of the concept. In this case, the ROM role required a human-subjects experiment to empirically validate the viability of the concept against a realistic, practical scenario (see Cassenti and Chan, 2012). The Final Model, and the focus of this paper, used a modeling & simulation approach to use the results of the Test phase to revise the Initial Model to the point where it reliably simulates the empirical results. The Final Model phase strengthens the model to become an accurate representation of the concept.

3. Review of Phase 1: Initial Computational Model

This process began through the development of a cognitive computational model using an Engineered Explosives (EE) scenario (Cassenti et al., 2011). Within this scenario, a ROM commanded two robot operators with one operating a Small Unmanned Ground Vehicle (SUGV) and the other operating a small UAS. The robots were used to survey a future convoy path for EEs with requirements to maintain the path if no EE was detected and to divert the path if an EE was detected.

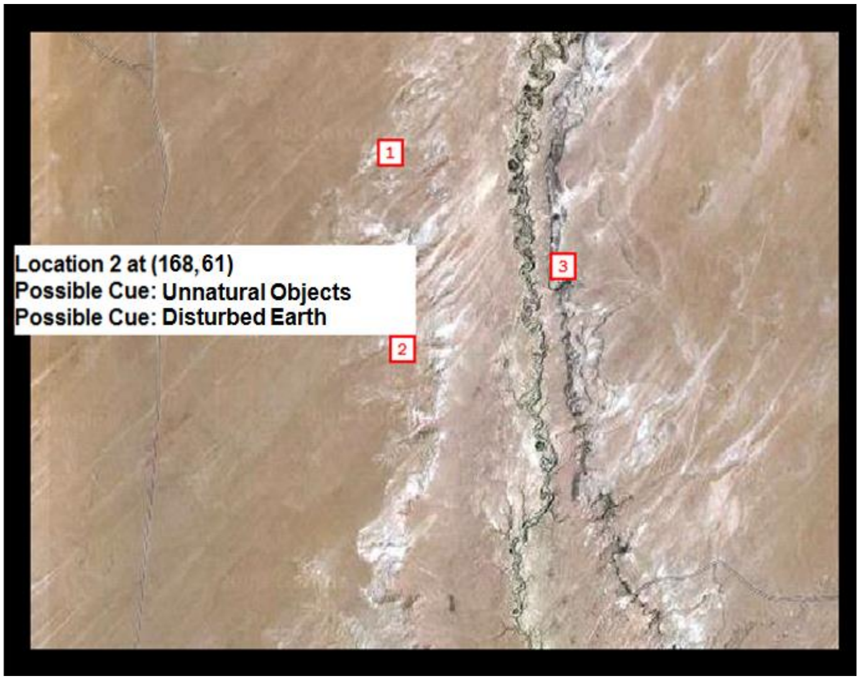
Cassenti et al. (2011) suggested that EE detection requires military personnel to attend to at least 19 cues in the environment. Examples of these cues are suspect tracks or disturbed earth. Out of these 19 cues, the five most visually salient cues were retained for the Phase 2 test (Cassenti & Chan, 2012): ground discontinuities, disturbed earth, exposed wires, suspect tracks, and unnatural objects. The initial computational model underwent a great deal of change in conversion to the final model presented here and was discussed in detail in Cassenti et al. (2011), therefore it will not be discussed further.

4. Review of Phase 2: The Test phase

After establishing computational viability, Cassenti and Chan (2012) addressed the test phase of the MTM approach. Human-subjects experimentation was used to empirically validate the viability of the concept against a realistic, practical scenario using 16 civilian employees as test subjects. Although, a goal of the test phase was to keep as closely to the initial model as possible, certain parameters needed to be changed to adapt the ROM into a practical scenario for experimentation.

Participants were instructed to adopt the role of a ROM to detect potential sites of EEs along a future convoy path. To do so, they interacted with a computer interface that provided trial conditions, ground reports on EEs, and the command interface. Using the Tactical Ground Reporting Network (TiGRNet; Gregory, 2010), the ROM would gather information about three or four possible EEs. The command interface displayed available robot operators, the type of robot they controlled, the potential EE locations, and the coordinates of the locations. After a robot reached a site, it transmitted a text of possible EE cues as well as access to a video or image from the EE. Figure 1 provides a visual display from the TiGRNet interface (Panel A) and the command interface (Panel B).

(Panel A)



(Panel B)

Figure 1. Panel A shows an example of the TiGRNet interface with the information for Location 2 displayed. Within the command interface (Panel B), participants could select an available robot operator and send mission commands. Following a delay, the ROM had access to a text description of the operator's perception of EE cues on site and could also click on an image or video link to view these stimuli (subject to properties of the bandwidth condition). Participants were instructed to make a decision using the available stimuli on whether to maintain or divert the convoy path.

The dependent variable retained for this phase was the accuracy of decisions. The new dependent variables included: 1) *signal detection* encompassing hits (diverting the path from an EE), misses (maintaining the path through an EE), false alarms (diverting path around a non-EE), and correct rejections (maintaining the path though a non-EE site); 2) *resource use* including frequency of robot selection of the SUAV or SUGV, image request, and video request; and, 3) *trust-in-network* (revised version from Jian, et al., 2000). The weights of the EE cues in the model's calculation of EE decision were not able to be measured in human-in-the-loop experimentation.

The study included two independent variables. First, a bandwidth limitation was active on half the trials and inactive for the other half. Participants were instructed that bandwidth limitations resulted from sharing resources; thus not all robot operators were available at once and the bandwidth was divided, resulting in slow transmission of images (i.e., delay following image requests) and transmission of distorted videos. The other independent variable was a time limitation. Time pressure is often a reality of critical military missions. On half of the trials, participants were limited to a five minute trial time, thereby creating stress through time constraints.

The results revealed several significant findings that were included in the Phase 3, ACT-R Model:

- an overall greater-than-chance accuracy;
- greater accuracy and hit rate for EE detection within the unlimited bandwidth condition;
- an overall greater false alarm rate than miss rate in EE detection;
- more requests for images in the unlimited bandwidth condition than limited bandwidth condition;
- more video requests for unlimited bandwidth than limited bandwidth;
- more image than video requests;
- more use of ground-based robots than air-based robots;
- positive correlation: ground-based robot use and accuracy; and,
- use of air-based robots decreased linearly with trial number (i.e., learning not to use air-based robots);
- greater trust with unlimited bandwidth than limited bandwidth;

In brief, the results indicated significant, wide-ranging effects of bandwidth limitations on several aspects of the ROM role, including reduced performance, cutbacks in use of resources, and diminished trust in networked assets. In addition, the participants expressed clear preferences in which assets were most useful. In particular, participants preferred viewing images over videos and using ground-based robots over air-based ones. We interpret these results as highly dependent on the scenario. Static images while looking for static EE cues makes more sense than a moving image, while, the small size of some EE cues makes the closer angle of the ground robot more desirable.

5. Phase 3: The Final ACT-R ROM Model

The focus of this paper is on the Final Model produced at the conclusion of this MTM approach. This approach dictates that in the final phase, experimental data must be simulated by adapting the first model to encompass the experimental conditions, then refining the model parameters until the model run results approximate the empirical data. Since the main goal of ACT-R is to represent empirical data (Anderson & Lebiere, 1998), ACT-R was a good choice for the second model stage of this project.

ACT-R was used because of its capability to model the type of empirical data provided through the previous test phase. ACT-R is a computational modeling system used to simulate cognitive processes. Users build models by inputting the chunks (i.e., declarative or fact memory), and productions (i.e., procedural or memory for perception, mental steps, or actions) that ACT-R needs to complete a task or set of tasks. Relative to the initial model, the ACT-R productions underwent the most amount of change.

5.1 New Model Considerations

The most difficult challenge of the Phase 3 model revision was including the independent variables of bandwidth and time pressure because the original model was not designed to include these variables. Not only did these variables need to be programmed into the model, but the changes they made to performance and resource allocation also needed to be represented.

The next most difficult change was recording the new dependent variables. In particular, visual stimulus selection and trust were not represented in the initial model, thus requiring methods for producing and recording them in the final model.

5.2 Advancements to the Model

5.2.1 Modeling robot selection

Before continuing, it is important to note that the description of the model is the final one after adjusting the parameters. This is the form of the model that we believe most represents human behavior in this task. One major change in the final model productions were related to the process of choosing a robotic asset. In the initial model, the three factors for selecting the robot were availability, elevation, and time pressure. However, in the final model, only availability remained. Time pressure in the initial model was a nebulous factor that favored the faster air-based robot, but following the test phase, robot speed was removed as a factor (note that this aligns with actual military scenarios where a ground-based robot may be first transported to the region and not necessarily get there just from its own movement, making the difference in robot speed inconsequential).

In the test phase, participants favored ground-based robots, a preference they gained during trials. This is not all that surprising because EEs are typically placed on the ground. To accommodate these results, the model began with an even odds probability of selecting a ground- versus air-based robot (assuming the equal availability of both). Whenever the model selected to view an image or video, the probability of choosing a ground- over an air-based robot increased. When the model viewed a ground-based image, the probability of selecting a ground-based robot increased by 3%. Since videos held less sway over decisions in the experiment, the probability that the model would select a ground-based robot increased by half that or 1.5%. The possibility also existed that the participants would not select any robot and they did not choose a robot 13% of the time in the experiment. The final model included the same 13% probability that the model would try to assess a site without sending a robot to it.

5.2.2 Modeling EE identification and cues

The primary tasking of this type of military operation was the identification of EEs on a given convoy pathway. To meet this end, the first computational model decided on the presence of an EE by assuming that the cues presented by the robot operators were true, then used an internal equation of the 19 cue weights to calculate whether the total evidence exceeded a threshold (i.e., present or not present). The final model used a similar equation of strengths of the five salient cues from the test phase, and also incorporated lines of feedback to sum the evidence to each of the cues. The model kept track of the total by using the location chunks from declarative memory. The final model accumulated evidence from four sources: TiGRNet, the robot operator text message, images, and videos. Of these, images and videos accumulated the most evidence, while TiGRNet and the text accumulated less evidence.

The accumulation of evidence followed a set structure. First, the model gathered information from a simulated TiGRNet. It added the location information to the location chunks; then it read the cue information to increase the cue evidence for each of those cues by 10 points. Following the return of each simulated robot operator, the model increased the cues from the operator's text message by 15 points while decreasing the unselected ones by 15. Seeing an image stimulus increased the correct cues by 45 points, whereas seeing the video increased the correct cues by 27 points. Participants could view stimuli multiple times, as could the model. However, repeated viewings did not increase or decrease evidence since it was assumed that looking at the same stimulus two times meant that the participant (and by extension, the model) did not view the evidence well enough the first time to make decisions about the cues.

Two algorithms determined whether the model thought an EE was present or absent. First, after summing up all the evidence for each cue, the model chose the cues that exceeded 30 points as the cues that were actually present at the location. In the experiment, there were never more than two cues at a site in either TiGRNet or the operator's text reports, so the model used an upper limit of two cues. Second, when three or more cues passed the 30 point threshold, the model narrowed down the cues to the two with the highest evidence totals. When zero cues exceeded the threshold, the model selected the strongest cue as the only one for the site.

The purpose of the second algorithm was to make a decision about the presence of an EE. In the final model, ACT-R had a list of strengths for each cue. Each cue began with a strength value of 22 points. Whenever the feedback at the end of the trial pinpointed a cue as part of an EE, the model added 5 points to the cue strength. Conversely, when a cue was at a site location with no EE, the model subtracted 5 points from the running total. When making the decision on presence or absence of EE, the model summed the strength of the one or two cues it selected. If the sum passed a threshold of 35, it considered the site to contain an EE, if not, it was deemed to not include an EE.

5.2.3 Modeling visual stimulus selection.

Stimulus selection also needed to be simulated in the revised model. The model included productions to view an image or a video once the robot operator returned. A function also allowed multiple viewings of the same image or video to match the opportunity offered to participants to repeat stimulus viewing. Based on the experimental data, the probability of requesting an image again was set to 56.2%; while probability of viewing a video again was set to 30%. These probabilities were obtained from the empirical data.

Bandwidth and time limits affected the probability of choosing image or video stimuli. First, video viewing was affected by the time limit. The time limit meant risking exceeding five minutes as the participant took the time to watch the video. Even if the model was on track to view the video, the time limit condition ran a 10% chance of ultimately deciding against requesting a video. The probability of requesting an image was unchanged by the time limit in the empirical data so was left unchanged in the model. The bandwidth limit affected both images (delays) and videos (distortion). To simulate bandwidth, the model decreased the amount of evidence that accumulated from videos, since videos were distorted (while images were only delayed). Videos in the limited bandwidth condition only provided 25% of the evidence (i.e., 7.75 points for correct cues) of the full bandwidth condition. The model used a 5.5% and a 19% chance of not requesting an image or video in the limited bandwidth condition, respectively even if the previous stimulus selection production decided to view the stimulus.

5.2.4 Modeling trust

Trust ratings were also added to the final model. Experiment instructions defined trust as the subjective judgment of the network's (i.e., all information sources) reliability in giving accurate information. Trust was included not just by getting the model to answer the questions from the trust questionnaire, but also transitioning the initial model from implicitly trusting the robot reports to needing to figure out for itself how much it should trust the information it received from multiple sources in the network. For example, the first model (Phase1) used TiGRNet to only get information about the location of potential EEs, but the inclusion of trust into the final model required TiGRNet to provide additional information specific to EE cues. This experimental finding from Phase 2 and inclusion in the final model supports the theoretical importance of cueing on trust development in HRI (Schaefer, 2013; Schaefer, et al., 2014).

The experimental results indicated less trust in the network when under bandwidth limitations. To account for this result, the model was set to a base rate value of 45 out of 100 for trust when in the no bandwidth limits condition and to 40 under bandwidth limits. Every time the model viewed an image, 4 points was added to the running total. When calculating trust, the model generated a random number of 0 to 100, and then added the constant to the random number. This total was standardized to be a whole number between 1 and 7, the Likert scale values from the experiment. This trust procedure permitted consideration of the factors that seemed to affect subjective attribution of trust (image viewing and bandwidth condition).

5.3 New Model Complexity

Even though the final model included much more complexity with two more complex independent variables and many additional dependent variables, the final model did not add that much code. Despite all the additions, the final model included just as many productions (i.e., ACT-R's mental steps) as the initial model (33 productions) and only included 350 more lines of code (2,421 versus 2,071). We hypothesize that the final model did not increase that much in complexity despite more independent and dependent variable because of the guidance from real world data lacking in the initial model. The data helped guide our thinking on how mental processes actually unfolded rather than trying to making educated guesses.

6. Model Results

The alterations resulting from the experimental data changed the initial model, but the question of whether the alterations resulted in positive change remains. This may be assessed by comparing the model data to the empirical data, the purpose of the second model phase of the MTM approach. This is not a property of the initial model, which could only be recommended by the ability of the model to learn.

Table 1 includes all the analyzed dependent variables from the experiment that were also recorded in the final model. The table presents the model and experimental data marginal means and also includes the Pearson correlation coefficients comparing the two. All correlations exceeded a critical p-value of 0.01.

Table 1. Means and correlations comparing Phase 2 (Test) and Phase 3 (New Model).

	LBW, TL			**LBW, No TL**			**UBW, TL**			**UBW, No TL**		
	Test	**Model**	**r**	**Test**	**Model**	**r**	**Test**	**Model**	**r**	**Test**	**Model**	**r**
Accuracy	0.61	0.63	0.97	0.63	0.60	0.98	0.72	0.66	0.97	0.64	0.60	0.98
Hit Rate	0.42	0.39	0.95	0.37	0.39	0.91	0.41	0.48	0.97	0.45	0.49	0.95
FA Rate	0.62	0.63	0.96	0.65	0.57	0.84	0.83	0.71	0.95	0.70	0.68	0.94
Images	3.91	4.07	0.95	3.92	3.40	0.96	4.61	4.47	0.96	4.71	4.38	0.98
Videos	1.91	2.1	0.97	2.12	1.93	0.97	2.18	2.45	0.97	2.81	2.38	0.91
SUGV	0.97	1.02	0.90	1.03	0.98	0.88	1.87	1.85	0.71	1.88	1.77	0.78
SUAV	0.92	0.98	0.90	0.82	1.00	0.92	1.18	1.23	0.94	1.13	1.23	0.94
Trust	6.11	6.08	0.97	6.00	6.09	0.97	6.45	6.25	0.97	6.48	6.21	0.93

Note. Marginal means covering the Bandwidth Limitation (LBW versus UBW) and Time Limit (TL, No TL) for both the final model and experimental (Test) results. Each set of Model and Test values includes a Pearson correlation coefficient (r) across the following variables: proportion correct (Accuracy), hit rate, false alarm (FA) rate, image request frequency, video request frequency, ground-robot frequency, air-based robot frequency, and trust score.

7. Model Discussion

The results from the Final Model indicated a strong match with the empirical results from the Test Phase. Not only were correlations between the final model and the experimental results all high and statistically significant, but the dependent variables that showed differences in the experiment results were replicated in the model for seven of eight calculated results. The replicated results (derived from observation) included better-than-chance accuracy, greater hit rate for unshared than shared resources, greater false alarm rate than miss rate, more requests for images and videos with unshared than shared resources, more image than videos requests, preference for ground-based robots over air-based robots, and greater trust when robots did not need to be shared. Greater accuracy with unshared than shared resources with a time limit was the only significant difference that did not replicate in the model. A t-test on the model results showed this difference too, however models generally have low variability due to high internal consistency of a model relative to humans (e.g., models do not get mentally distracted). A mere 3% proportion correct difference with the model results would not be high enough to show a significant difference in the empirical results. This raises the possibility that the result in Cassenti and Chan (2012) was a result of Type I error.

7.1 Modeling Cognitive Processes

There are two strong reasons to reinforce the belief that the model accurately represented the cognitive processes of the participants in the experiment. First, the high correlations between model and experimental results show that the underlying processes of both the computer simulation and the participants are likely parallel. Second, ACT-R is the most used and adapted modeling system for cognition. The goal of the ACT-R system is to be a unified theory of cognition, (see Newell, 1990; Anderson & Lebiere, 1998) the system that can explain all aspects of human cognition. Over its decades of use, ACT-R had undergone continuous updating so that it can encompass cognitive phenomenon that it could not model in previous versions. This process has met with a great deal of success and the ROM model adds to this legacy.

7.2 Practical Value of ROM Model

The model also has elements of practical value. First, the model (and experiment) isolates bandwidth limitations as an important consideration. A ROM serving in a military mission needs to have timely images and non-distorted video evidence to provide good performance and with the confidence of trust in information. The time limit condition did not change behavior much, but did moderately impact it so providing the time to perform tasks is also an important consideration. There is reason to suspect that the five minute time limit was much too short as well considering that moving robots would take much more time than actually seen in the experiment. The short time was necessary to structure a feasible experiment with multiple trials (the same number of trials was used in the second model), but one could reason how it did not accurately represent the time burdens experienced in real military missions involving robots, which would theoretically add importance to time limits.

The results also have practical considerations involving resource preferences. Deploying the ROM to the field should include a focus on making images available over videos. Images seem to have the added allure of being static so that the observer can more carefully examine the details of the scene, whereas videos include apparent motion and jitter that adds nothing but distraction.

Another resource preference is for ground-based robots, which was learned over trials. Ground-based robots allow closer examination of the scene than do overhead images. Although participants used air-based robots when they were available, they did not use them when they had ground-based robots available and had fewer locations left to check than robots available, especially in later trials. It may be that other scenarios would favor air-based visual evidence, but in the case of EE detection, results from this study suggest that ground-based robots should be available as much as possible.

7.3 Modeling Decision Making

The final model gave a good indication of the decision making processes that a ROM must go through. One could argue that the prescribed structured way that the model made decisions is an artifact of the way the model was built. However, from the observations of the participants making decisions, they appeared to follow the same pattern of always checking with TiGRNet, then sending all robots, and using the information to judge the presence or absence of EEs. Given the strong relationships between model and empirical data, the model’s productions likely matched the mental processes of participants with some degree of precision.

The final model also showed that the ROM role is laden with decision making. The ROM model and likely a real life instantiation of the ROM role must make the following decisions: 1) evaluate the area (e.g., using TiGRNet); 2) decide how much to trust the support systems; 3) select robots based on scenario conditions; 4) consider the importance of robot operator generated information; 5) choose whether to collect information and in what format (e.g., images, videos); 6) judge threats; and 7) incorporate feedback in a schema for judging future threats. The final model helps to focus on two measurable skills that could help select the best candidates for the ROM role: working memory capacity and visual detection.

7.4 Selecting a ROM

First, working memory capacity (i.e., the amount of mental workload an individual can handle at once) of candidates for the ROM should be measured. Those with higher working memory capacities would make better ROMs as they can manipulate more elements of vital information at once to make decisions. I recommend the Operation Span (OSpan; Unsworth, et al., 2005) Test because it measures retention during secondary task completion. So, instead of just measuring how many items a person can hold in memory at once (i.e., short-term memory capacity), the OSpan approximates dynamic capacity as other mental processing occurs. The OSpan is therefore more pertinent for assessing capability or capacity of a candidate for the ROM role.

Second, the model accumulated the most evidence from visual stimuli rather than from TiGRNet or the operator’s text messages. A second quality that would aid ROMs is the ability to attend to visual information, presuming that this result would carry into real world application. The most popular measure of this is the useful field of view test (Sanders, 1970), that gives a quantification of the spatial range that an individual can process to derive detailed visual information (e.g., Ball, et al., 1993). Those with higher unified field of view scores would be preferred candidates for becoming a ROM.

8. Conclusions

The Final Model concludes this MTM project for the ROM role. Not only does the ROM fill an Army need, but the study as a whole demonstrates that the ROM role can be useful for a practical scenario and mission. With a traditional optimal ratio of operators to robots of 2:1 or 3:1 (SUGV or SUAS, respectively), the ROM study shows that with a single ROM covering multiple operator-robot teams, these ratios may be reduced to more than one, but less than two operators per robot.

The Initial Model phase instantiated a practical scenario within an ACT-R model and demonstrated learning over model runs (Cassenti et al., 2011). The following Test phase introduced the model to human subjects and included practical independent variables and new dependent variables to make the scenario richer (Cassenti & Chan, 2012). The Final Model, discussed here, generated a beginning-to-end set of mental processes and decisions that resulted in high degree of relationship with the empirical results. Altogether, this project took an Army need and developed the role to the point where it could be examined on multiple levels. It helped isolate what assets would be required to implement this and which personnel would be best suited to handle the role. The findings and implications of this project may prove to be of value to further research on the use of robots on the battlefield.

9. References

Agah, A., & Tanie, K. 1999. Taxonomy of research on human interactions with intelligent systems. In S*ystems, Man, and Cybernetics, IEEE SMC'99 Conference Proceedings*, 6. 965-970.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Ball, K, Owsley, C., Sloane, M.E., Roenker, D.L., & Bruni, J.R. (1993). Visual attention problems as a predictor of vehicle crashes in older drivers. *Investigative Ophthalmology & Visual Science, 34*, 3110-3123.

Burke, J.L., & Murphy, R.R. 2007. RSVP: An investigation of remote shared visual presence as common ground for human-robot teams. In *ACM/IEEE Human-Robot Interaction*, Arlington, MD. 161-168.

Cassenti, D.N., & Chan, K. (2012). Network science for decision making: Performance, trust, and workload with information from a network. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, USA. Technical Note (ARL-TR-5890).

Cassenti, D.N., Dutt, V., Gonzalez, C., Pomranky, R.A., & Hunn, B.P. (2011). Modeling a robotics operator manager's responsibilities in ACT-R. In Proceedings of 20th Annual Behavior Representation in Modeling & Simulation (BRIMS) Conference (pp. 237-244).

Cassenti, D.N., Kelley, T.D., Yagoda, R.E. & Avery E. (2012). Improvements in robot navigation through operator speech preferences. *Paladyn. Journal of Behavioral Robotics, 3*, 102-111.

Chen, J.Y.C., Barnes, M.J., & Qu, Z. (2011). RoboLeader: Dynamic Re-tasking for Persistence Surveillance in and Urban Environment Using Robot-to-Robot Control (Report No. ARL-MR-0761). U.S. Army Research Laboratory.

Chen, J.Y.C., Barnes, M.J., Qu, Z., & Snyder, M.G. (2010). RoboLeader: An intelligent agent for enhancing supervisory control of multiple robots (Report No. ARL-TR-5239). U.S. Army Research Laboratory.Chen, J. Y., Durlach, P. J., Sloan, J. A., & Bowens, L. D. 2008. Human-robot interaction in the context of simulated route reconnaissance missions. *Military Psychology*, 20, 135-149.

Chen, J.Y.C., Haas, E.C., Pillalamarri, K., Jacobson, C. N. (2006). *Human-robot interface: issues in operator performance, interface design, and technologies*. Technical Report (ARL-TR-3834). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Goodrich, M., Olsen Jr., D. (2003). Seven principles of efficient human-robot interaction. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (pp.3943-3948).

Gregory D. (2010). Seeing red: Baghdad the eventful city. *Political Geography, 29*, 266- 279.

Jian, J.-Y., Bisantz, A.M., & Drury, C.G., (2000) Foundation for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*, 53–71.

Mansager B. (1994) *Model-test-model*. Monterey, CA: Naval Post Graduate School.

Murphy, R. R, & Burke, J.L. (2010). The Safe Human-Robot Ratio. In *Human-Robot Interactions in Future Military Operations*, M. Barnes & F. Jentsch, Eds. Ashgate, Burlington, VT, 31-49.

Murphy, R.R., Griffin, C., Stover, S., & Pratt, K. 2006. Use of micro air vehicles at Hurricane Katrina. *IEEE Workshop on Safety Security Rescue Robots,* Gaithersburg, MD.

Murphy, R.R., Pratt, K., & Burke, J.L. 2008. Crew roles and operational protocols for rotary-wing micro-UAVs in close urban environments. ACM/IEEE Human-Robotic Interaction, Amsterdam.

Newell A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Riley, J.M. & Strater, L.D. (2006). Effects of robot control mode on situation awareness and performance in a navigation task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (540-544).

Sanders, A.F. (1970). Some aspects of the selective process in the functional field of view. *Ergonomics*, 13, 101–117.

Schaefer, K. E. (2013). *The perception and measurement of human-robot trust*. Doctoral Dissertation. University of Central Florida, Orlando, Florida.

Schaefer, K. E., Billings, D. R., Szalma, J. L., Adams, J. K., Sander, T. L., Chen, J. Y. C., & Hancock, P. A. (2014). A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Human-Robot Interaction. Technical Report (Report No. ARL-TR-6984). U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

Schurr, N. 2007. *Toward Human-Multiagent Teams*. Doctoral Thesis. University of Southern California, CA

Unsworth, N., Heitz, R.P., Schrock, J.C., & Engle, R.W. (2005). An automated version of the operation span task. *Behavioral Research Methods, 37*, 498-505.

Wickens CD. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, *3*, 159-177.

Yagoda, R.E. (2010). Development of the human robot interaction workload measurement tool (HRI-WM), In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (pp.304–308).