HPMI: INTEGRATING SYSTEMS ENGINEERING AND HUMAN PERFORMANCE MODELS

Karen Gery
Jeff Doyal
Bryan Brett
Science Applications International Corporation, Inc.
4031 Colonel Glenn Highway
Beavercreek, OH 45431

Christian Lebiere
Eric Biefeld
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Edward A. Martin
Human Effectiveness Directorate
Air Force Research Laboratory
2210 Eighth Street
Wright-Patterson Air Force Base, OH 45433-7511

The Air Force Research Laboratory's Human Performance Model Integration (HPMI) program is exploring the merit and feasibility of combining multiple human performance models and tools to create hybrid models of performance that address application-specific requirements for model fidelity while controlling cost. The first exploration of HPMI feasibility involves development of a human performance model using task network modeling and ACT-R cognitive modeling. Under the Combat Automation Requirements Testbed program, a detailed task network model of strike fighter pilot performance was developed and integrated with a flight simulator. In the initial implementation the task network model managed prioritization of objects and made decisions about the order in which the targets would be examined. This original task network model of prioritization then was replaced with one implemented in ACT-R (Anderson and Lebeire, 1998). The cognitive model provided improved representations of decision-making and memory, as well as confusion and errors committed by humans. We will describe the task domain, model and interface, as well as, an informal verification study that compared the performance of the hybrid model with that of humans.

Introduction

In its *Modeling and Simulation (M&S) Master Plan*, the Defense Modeling and Simulation Office (DMSO) identified the capability to robustly represent individual and group behaviors as a critical need (DoD 5000.59-P, 1995). In a study commissioned by DMSO, the National Research Council's (NRC) *Panel on Modeling Human Behavior and Command Decision Making: Representations for Military Simulations* reviewed a number of architectures that support the representation of various aspects of human behavior (Pew & Mavor, 1998). The panel pointed out that the architectures reviewed can be viewed as useful, promising, and a good starting point – but are only very early steps. The panel went on to say, "It is not likely, even in the future, that any single architecture will address all modeling requirements." (ibid.) The panel expressed the opinion that a fruitful hybrid approach would be interfacing architectures via communication protocols – rather than reimplementing features of one architecture in another.

Human Performance Model Integration (HPMI) Program

The Air Force Research Laboratory's Human Effectiveness Directorate initiated the HPMI program to investigate the feasibility of using a hybrid approach to performance modeling, such as that suggested by the NRC Panel. This program is exploring hybrid human behavior representations (HBR) that exploit available, proven modeling technologies as a means for providing more realistic representations of operator behavior faster and more cost-effectively. The approach, illustrated in Figure 1, is to integrate high-fidelity, first principle representations[1] of perceptual, cognitive, or

---

[1] First principle representations describe the basic elements or laws determining the intrinsic nature or characteristic behavior.

Report Documentation Page

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2003	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
HPMI: Integrating Systems Engineering and Human Performance Models	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Air Force Research Laboratory Science Applications International, Inc. Dayton, OH 45433	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	**UU**	**6**	
unclassified	**unclassified**	**unclassified**			

psychomotor components of behavior into models built using more flexible modeling architectures such as task network models (TNMs). Rather than attempting to represent the underlying processes, TNMs represent human behavior at the higher level of task performance time and accuracy. At this level, fairly complex task scenarios can be readily modeled and understood by practicing engineers and computer scientists, instead of requiring detailed, theoretically validated models developed by experts in cognition or physiology. An additional incentive is that some TNM architectures have been extended to a point at which the models of human task performance can be independently developed, and then – by means of standard, open protocols – be integrated with existing constructive representations of systems operating in their envisioned mission environment.[2]
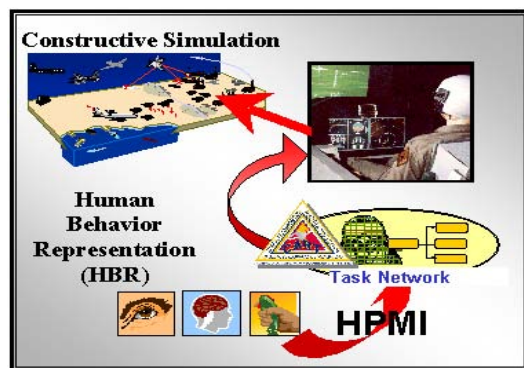

Figure 1. Vision of the Human Performance Model Integration Program

The HPMI vision is that while higher-level architectures (such as TNM) will enable the relatively fast and cost-effective development of models of human performance and their integration with constructive system models using DOD High Level Architecture (HLA) protocols, HPMI will demonstrate the viability and utility of using a hybrid approach to enhance the higher-level representation with first-principle fidelity that is focused on the critical behavioral component(s) where this level of fidelity is required.

Integrating a Model of Cognition and a TNM

The work described in this paper focuses on developing a theoretically-based model of human cognition, and

integrating it with a higher-level representation of a human conducting a cognitively-intensive task as one of a much larger set of tasks required to carry out an operationally realistic, complex, highly dynamic mission scenario. This HPMI project is creating a hybrid human behavior representation that integrates an Adaptive Control of Thought - Rational (ACT-R) model with an existing human performance TNM developed earlier under the Combat Automation Requirements Testbed (CART) program. The TNM represents the behavior of a strike fighter pilot conducting a complex, operationally realistic mission. Using HLA protocols, the TNM had been integrated with a constructive simulation of a strike fighter aircraft that operates inside a constructive mission environment. The goal of this HPMI effort is to explore the feasibility and utility of developing and applying a hybrid modeling approach.

ACT-R Cognitive Modeling Environment

ACT-R (Anderson & Lebiere, 1998) is a highly detailed cognitive theory that makes precise predictions about the course of human cognition at a very fine-grain scale. ACT-R has been validated by hundreds of psychology experiments and more than 75 ACT-R models have been published covering a wide range of cognitive domains. ACT-R combines a classical symbolic system with a neural network-like subsymbolic system. At the symbolic level, ACT-R implements a production system theory that models the steps of cognition through a sequence of production rules that fire to coordinate retrieval of information from the environment and from memory. At the subsymbolic level, every step of cognition implements parallel pattern matching that is tuned statistically to the structure of the environment. It is a cognitive architecture that can be used to model a wide range of human cognition – tasks as simple as memory retrieval (Anderson, Bothell, Lebiere & Matessa, 1998) and visual search (Anderson, Matessa & Lebiere, 1997) to tasks as complex as learning physics (Salvucci & Anderson, 2001) and air traffic control (Lebiere, Anderson & Bothell, 2001). In all domains, it is distinguished by the detail and fidelity with which it models human cognition.

ACT-R is structured around the concept of goals. A goal in ACT-R is a declarative structure that encodes a particular objective (e.g., perform a sequence of actions, or find an answer to a question) that is the current focus of attention. Each production rule applies to a specific type of goal. When a goal is solved, it is stored in declarative memory as a structure (called a chunk) that encodes the result of that goal. Thus a type of goal, together with the production rules that apply to it and

---

2 The Combat Automation Requirements Testbed (CART) program has developed a TNM architecture that includes the capability to interface its TNM representations to simulations that are High Level Architecture compliant (Martin, Brett, & Hoagland , 1999).

associated declarative chunks, can be thought of as a modular piece of knowledge. Models of complex tasks can be built around the assembly of multiple goals (Lebiere, Anderson, & Bothell, 2001). Declarative knowledge is stored in chunks. Chunks are structures composed of a small number of slots, each of which holds a single piece of information. Each chunk has an associated activation, which reflects environmental factors such as frequency of rehearsals and time-based decay. When attempting to perform a memory retrieval, the chunk with the highest activation is retrieved.

The Shootlist Management Task

As part of the CART program, case studies are being conducted to demonstrate the concepts and viability of human performance modeling for supporting the weapons system acquisition process. The CART program's first case study involved the development and validation of a TNM of a single-seat, strike fighter-aircraft pilot conducting a complex, operationally realistic mission. For this mission, the pilot's task is to search for a target using onboard sensors and to destroy it. A cognitively-intensive component of the pilot's activity in conducting this mission is the development of a 'shootlist' (simply stated, a shootlist is the pilot's prioritized list of the objects to be examined). Since our interest dealt with cognitive modeling and the integration of an ACT-R model with a TNM, the shootlist management task was selected for the modeling efforts.

To support HPMI hybrid modeling efforts, this shootlist management task environment was ported to a virtual simulation testbed. Empirical data are collected on this testbed both for parameterizing ACT-R models and for validating hybrid model predictions.

HPMI Shootlist Management Testbed

HPMI developed a virtual simulation testbed consisting primarily of a computer monitor and a BGSystems Joystick (Martin, 2002). Experimental subjects are 'flown' through various scenarios on 'autopilot' and instructed to find an actual target among a number of distracters. Icons displayed on the monitor represent objects moving on the ground. These icons are displayed along with a presentation of the projected ground path of the simulated aircraft. If a subject clicks on an icon to place it on the shootlist, its symbol changes accordingly. The icon representing the object of immediate interest is further highlighted. Once within a predetermined range, a visual depiction of the object of immediate interest is displayed to the subject.

The subject must sort through the objects in the shootlist priority order until the target picture is found, then signal 'target found' by squeezing a trigger to end the trial.

At the beginning of each trial, subjects are given the approximate location of the actual target. However, all objects are constantly moving, so the target will not be at the given location and the subject has to search for it.

There are two primary challenges for the subject in managing the shootlist. The first is remembering which objects have already been identified as not the target. There are two possible consequences of remembering incorrectly. Objects that have already been identified and rejected may be reevaluated wasting time and effort. Also a subject may overlook objects that need to be examined. The second challenge is to be aware of when the shootlist slots are full. Adding another item causes the oldest item to be 'bumped' off the list. This can result in excluding objects from the examination process.

Hybrid Model Development

At a conceptual level, the integration between a TNM and ACT-R is very natural. Goals are a central concept in ACT-R that corresponds directly to individual tasks in a TNM. When the TNM selects a task, the corresponding goal is made active in ACT-R. Inputs to the task correspond to the initial values of the goal in ACT-R; outputs of the task correspond to the new values in the goal. For each goal, ACT-R returns the time taken to perform that goal -- which can then be used to populate the task duration in the TNM. The overall objective of this interaction is to develop a hybrid model that exploits the unique strengths of both a TNM and ACT-R.

CART Shootlist Management Model

In the original CART strike fighter simulation, a shootlist management application was written that managed the shootlist based on a normative set of search criteria. In that implementation, shootlist management was handled simplistically. It managed the mechanics of shootlist development perfectly in accordance with the normative criteria, and it did not represent underlying cognitive processes that could produce errors and other effects (e.g., forgetting, or prioritizing the list inappropriately). This complex and dynamic information-processing task provides ample opportunity for human errors due to confusion, forgetfulness, or inappropriate prioritization of the

shootlist. None of these aspects of cognitive performance were represented in the original CART TNM. ACT-R was chosen to provide this representation.

The ACT-R Shootlist Model

An ACT-R model was developed for evaluation using the shootlist management testbed. It replaced the original CART shootlist management model. The ACT-R model executes the shootlist management task using the same basic object prioritization scheme as the original CART model. However, it also incorporates the cognitive processes and effects that would potentially degrade operator performance. This degradation is represented in terms of sub-optimal prioritization (the subject will not always perfectly implement prioritization rules) and forgetting (the subject may re-assign an object to the shootlist that he has previously identified as not being the actual target, or fail to identify it at all).

The roles of ACT-R mechanisms modeling these human cognitive limitations are described below.

The shootlist management task is implemented in ACT-R using three 'goals.' The first goal in the ACT-R model updates the Objects Of Interest (OOI) list. That list in ACT-R typically held six or fewer objects, since memory chunks in ACT-R –- as with humans -- are constrained to hold only a small, fixed number of items (Boff, Kaufman & Thomas, 1986) (Van Cott & Kinkade, 1972). Instead, a set of memory chunks is created that encode -- for each target -- its basic characteristics (id, latitude and longitude) and whether it had been previously detected and/or identified by the model. The CART TNM passes that information to the ACT-R model whenever this goal is called.

The second goal in the ACT-R model is to filter the Image List resulting from the processing displayed icons by the task network. For each object, given its description (id, latitude, longitude), ACT-R attempts to remember whether that object has been previously detected and/or identified. To do that, ACT-R simply attempts to retrieve from memory a chunk created by the goal to update the OOI list that states that the object has been detected or identified. If the retrieval fails, then the model assumes that the object has not been detected or identified. However, memory retrievals in ACT-R, like human memory, are far from perfect. Through ACT-R sub-symbolic level processing, it is possible that the retrieval of an object that has been encoded in memory as identified might fail. As a result,

the model might decide to examine that object again. Moreover -- unlike other production systems in which matching chunks in memory to production conditions is a perfect process -- in ACT-R all chunks of the same type compete for any given retrieval, with chunks that only partially match the desired pattern having their activation penalized by an amount that reflects the difference between pattern and chunk. This partial matching mechanism in ACT-R reproduces the confusion that may occur between objects in close proximity to each other. Thus probabilistic retrieval from memory can lead to occasional errors in which a target is examined multiple times or not at all.

The third main goal corresponds to the prioritization of the shootlist. ACT-R starts by attempting to retrieve objects near to the originally provided target position that the model has no prior memory of selecting. Some confusion can result because this retrieval is modeled using a probabilistic partial matching process. Since an object is selected only if there is no prior memory of it being identified (as was discussed for the previous goal) both omitted and repeated identifications are possible. Finally, after an object is selected, its position becomes the current focus of attention around which the search for the next object will start. Thus, a tendency toward selecting targets in clusters arises. This is compatible with the memory requirements of the task, since remembering that a cluster of objects has been detected and identified is much easier than for the same number of scattered points.

Data Interface

The interface between the virtual simulation cockpit environment, the CART task-network human performance model (HPM), and ACT-R is illustrated in Figure 3. The CART HPM and ACT-R communicate by means of the CART TNM External Model Call (EMC) Interface. When the CART model encounters an EMC during execution, it pauses, sends and receives the appropriate external variables, and then continues to run. This interface utilizes the Microsoft Common Object Model (COM) link to transmit data and control between the CART TNM and ACT-R. All of the COM methods utilize a remote procedure call type syntax, where the return value is used to indicate success or failure. Advantages of this approach, especially compared to HLA, include efficient communication, relative simplicity and the lack of dependency on additional third-party software.
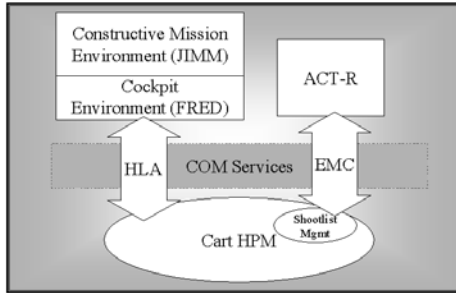
Figure 2. Virtual Simulation—CART—ACT-R
Interface

Hybrid Model Verification

While the knowledge representation and parameters of the shootlist model are strongly constrained by the ACT-R theory and the prioritization scheme specified, strategic variations and individual differences are inescapable realities of modeling human cognition. To enable ACT-R to accurately represent these components, model parameterization values were derived using the results of a prior HPMI study where human subjects performed the icon search task in the HPMI virtual simulation testbed (Martin, 2002). In this research, human-in-the-loop (HITL) performance data were collected to determine how effectively a subject could prioritize objects and the extent to which the subject re-examines objects that have already been identified.
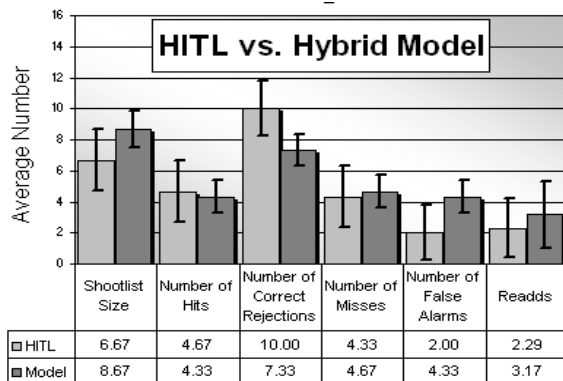

Table 1. A comparison of HITL and Hybrid Model
performance on key measures

The model was run and parameterized to match the results of the HPMI HITL study and the results are compared in Table 1. A detailed description of the data compared can be found in Martin (2002). Overall the model behaved similar to human subjects. One aspect of the data that the model does not currently reproduce is a subject's decision to use less than the full capacity of the shootlist, despite explicit instructions to the

contrary. One possible explanation for this behavior is the need to examine targets in clusters to facilitate the recall of which targets have or have not been identified. Such behavior will be included in the next iteration of model development.

Current Status

An investigation similar to the first case study is planned to compare the performance of the hybrid model to that of humans. The goal of this study is to compare the ability of a model that calls on ACT-R to represent a complex decision *process* to a model that calls a function representing the *outcomes* of the decision process within the context of a cognitively taxing icon search task. To meet these goals we will collect three sets of performance data: one each from a linear-outcome and an ACT-R-process model, and one from live subjects performing the same icon search task. The HITL data will serve as the performance standard against which the model representations will be evaluated. A revised ACT-R shootlist model and a simple linear model will be created. Both models will reflect insights gained in the HITL study and each will be integrated with a baseline TNM. Testing will be conducted under an extended set of scenarios that provide greater numbers of moving objects to be examined. This will allow us to evaluate the sensitivity of the shootlist management model to variation in number of moving objects. In addition, process equivalency of the model's acquisition behavior will be examined. Also of interest is the relative level of modeling effort associated with representing performance in each of the two modeling approaches. An accurate accounting of project resources will be performed to permit a cost comparison of the two levels of model fidelity.

Conclusions

The integration of an ACT-R shootlist model with a CART task network model demonstrates the feasibility of creating hybrid human performance models. We believe the ultimate result will be a robust human performance model that accurately predicts subject performance over a greater range of scenario conditions. Development and integration of the ACT-R shootlist management model required only a few months of labor and at this point we believe this was an acceptable cost for the degree of enhancement provided. The final criterion, however, will be the extent to which model performance correlates with that of humans.

References

Anderson, J. R. & Lebiere, C. (1998). *The Atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Anderson, J. R., Bothell, D., Lebiere, C. & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38,* 341-380.

Anderson, J. R., Matessa, M. & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human Computer Interaction, 12*, 439-462.

Boff, K. R., Kaufman, L. & Thomas, J.P. (1986). *Handbook of perception and human performance: Vol. II*. New York, NY: John Wiley & Sons.

DoD 5000.59-P (1995). *Modeling and simulation (M&S) master plan*. Alexandria, VA: Defense Modeling & Simulation Office. http://www.dmso.mil.

Lebiere, C., Anderson, J. R., & Bothell, D. (2001). Multi-Tasking and cognitive workload in an ACT-R model of a simplified air traffic control task. *Proceedings of the 10th Conference on Computer-Generated Forces and Behavior Representation*, 91-98, Norfolk, Virginia, May 15-17.

Lee, F. J. & Anderson, J. R. (2001). Does learning a complex task have to be complex?: A study in learning decomposition. *Cognitive Psychology, 42*, 267-316.

Martin, E. A. (2002). *Cognitive probe project: Development of a testbed for collecting cognitive model parameterization and validation data* (US Air Force Research Laboratory Technical Report AFRL-HE-WP-TR-2002-0156). Wright-Patterson Air Force Base, OH 45433-7022.

Martin, E.A., Brett, B.E., & Hoagland, D.G. (1999). Tools for including realistic representations of operator performance in DoD constructive simulations. *Proceedings of the AIAA Modeling and Simulation Technologies Conference,* 59-67. Portland, OR. (AIAA-99-4027)

Pew, R. W. & Mavor, A. S. (Eds.) (1998). Modeling human and organizational behavior: application to military simulations. Washington DC: National Academy Press.

Salvucci, D. D., & Anderson, J. R. (2001). Integrating analogical mapping and general problem solving: The path-mapping theory. *Cognitive Science, 25*, 67-110.

Van Cott, H. P. & Kinkade, R. G. (1972). *Human Engineering Guide to Equipment Design*. Washington, DC: American Institutes for Research.