

Comparisons of HBR Models with Human-in-the-loop Performance in a Simplified Air Traffic Control Simulation with and without HLA Protocols: Task Simulation, Human Data and Results

Yvette J. Tenney
Sandra L. Spector
BBN Technologies, A Verizon Company
10 Moulton Street
Cambridge, MA 02138
617-873-2907, 617-873-3730
ytenney@bbn.com, sspector@bbn.com

Keywords:

Human Performance, Human Performance Modeling, HBR, Human Behavior Representation, Simulation, HLA, Air Traffic Control, Controller

ABSTRACT: *BBN Technologies (BBN) served as moderator for the AMBR Program model comparisons and developed a simulated, but much simplified, air traffic control task environment that required the to-be-modeled controller to manage the transition of several aircraft from one traffic sector to another. BBN also generated task scenarios that emphasized multiple task management, task priority setting and attention management and collected human performance data exercising scenarios that were identical to those the model developers were required to exercise. This paper will describe the task, the scenarios, the experiment design for collecting data from human subjects, and the results of the comparisons between human performance and model-generated data, with and without the use of HLA infrastructure.*

1. Introduction

To advance human performance modeling, the Air Force Research Laboratory sponsored the Agent-based Modeling and Behavior Representation (AMBR) Modeling Comparison Project, an opportunity for multiple developers to design different models of the same human operator activity and to compare the results both from model to model and from model to human participants performing the same task. The participating models were Soar Technologies' EPIC-Soar, CHI Systems' COGNET/iGEN, AFRL's DCOG, and CMU's ACT-R; the human participants were skilled BBN video game players. This paper focuses on the data collection and results for both humans and models on a simplified air traffic control task simulated in a Distributed Operator Model Architecture (D-OMAR) environment.

In AMBR Round 1, D-OMAR native-mode connectivity was used to link the D-OMAR simulator

and the models. In AMBR Round 2, native-mode D-OMAR connectivity was replaced by the HLA RTI. This paper describes the outcome of Round 1 of the AMBR Modeling Comparison Project and concludes with some preliminary results from Round 2.

Other papers in these proceedings provide additional information about the research program [1], the D-OMAR and HLA architectures [2], and the four human performance models [3, 4, 5, 6].

2. The Air Traffic Control (ATC) Task

An Air Traffic Control task, developed by Macmillan, Deutsch, and Young was adapted for purposes of the present study [7]. This task was loosely based on the work of an air traffic controller, but did not require any prior expertise. It was chosen because it exhibited the characteristics typical of multi-task situations: information arrives at inconvenient or unexpected times, information interrupts an ongoing chain of

thought, information relevant to one task may be obscured by information from another, and information irrelevant to the current task may be salient and distracting.

This task, though simpler than its real-life equivalent, presented a host of challenges to the modelers. For example, they had to decide how the model will manage the scenario as a whole, choose when to shift between tasks, remember and update tasks awaiting

service, and prioritize among them. An additional challenge for the modelers was to ensure that the behavior of the model changed appropriately under different conditions, for example, with different display types and workload time pressures.

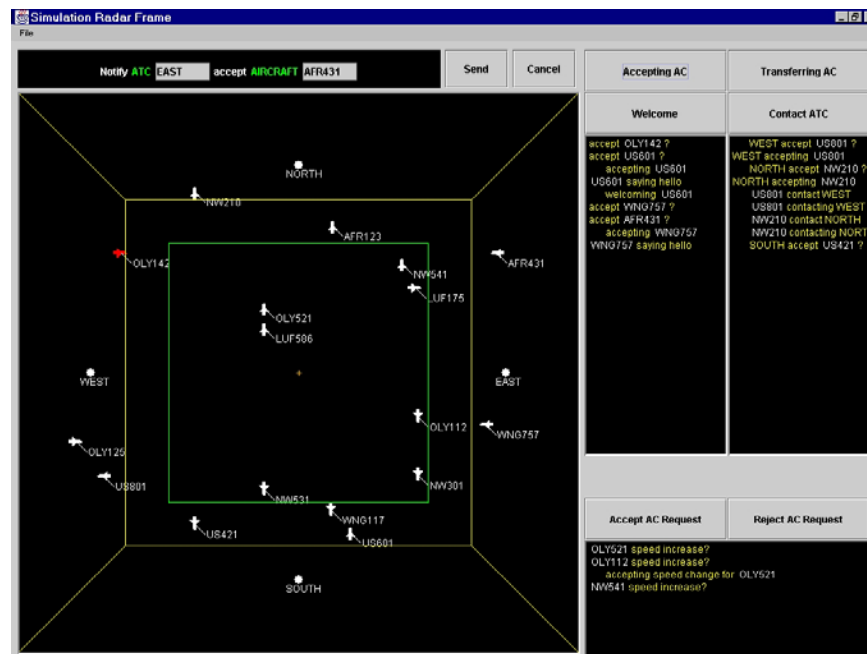


Figure 1: Display for ATC Task

2.1 Display Conditions

Two display conditions, adapted from MacMillan *et al.* were used in this study [7]. In the Text Display (see Figure 1), participants had to rely on the messages to determine what actions to take. In the Color Display, they could rely on the color-coding of the aircraft, which indicated the actions that were pending for each aircraft.

2.2 Workload Conditions

Workload was manipulated, in the present study, by keeping the number of aircraft constant and reducing the length of the scenario. The advantage of this procedure was that the maximum number of penalty points that could be earned remained constant across workload levels. Three workload levels were compared. In the Low Workload condition, 20 aircraft

were processed in 10 minutes (2 planes/min), in the Medium Workload in 7.5 minutes (3 planes/min) and in the High Workload condition in 5 minutes (4 planes/min). To allow all outgoing aircraft to reach the border an extra minute and a half of "play out" time was added to each scenario, bringing the duration to 11.5, 9, and 6.5 minutes for the three workload conditions.

2.3 Task Rules

In this task, the "player" (human participant or cognitive model) assumes the role of the Air Traffic Controller in the Central Sector bounded by the yellow, outer border (See Figure 1). The "+" in the center of the screen marks the location of the Central Controller. Adjoining controllers (simulated) are located at points labeled North, South, East, and West. Aircraft may enter the Central Sector from any direction, fly across

the sector on straight vertical or horizontal paths, and then proceed to the next sector. An aspect of this ATC task that sets it apart from the real one is that collisions are not of concern. Players are told that aircraft that appear to be colliding are actually at different altitudes. The objective of the task is to complete the required actions in a timely fashion and to avoid accumulating penalties for missed, delayed, or incorrect actions.

There are six actions that players can take by using the six labeled buttons (see Figure 1). For aircraft coming into the central sector, the player can ACCEPT and WELCOME each aircraft. For aircraft within the sector, the player can reply to a request for a speed

increase by using the ACCEPT/REJECT AC REQUEST buttons. For aircraft leaving the sector, the player can TRANSFER the aircraft to the next controller and tell the aircraft to CONTACT ATC. Each of these actions is described in further detail below and in Table 2:

ACCEPT. If the player does not ACCEPT an aircraft before it reaches the outer border of the Central Sector (yellow line), the aircraft will turn red and enter a holding pattern. The player can release the aircraft by doing an ACCEPT. There are penalties associated with an aircraft turning red and staying red (see Table 1). Aircraft will turn red in both the Text and Color Display conditions.

Table 1: Penalty Points in ATC Task

<i>Penalty Category</i>	<i>Player's Goal</i>	<i>Penalty</i>
Hold	Prevent aircraft from holding either while incoming or outgoing	50 points each time an aircraft turns red
Holding Delay	Get aircraft out of holding	10 points for each time unit* aircraft stays red
Speed Error	Respond to speed change requests correctly	50 points for an incorrect response to a speed change request
Speed Delay	Respond to speed change request in timely manner	2 points for each time unit* request not answered
Welcome Delay	Welcome aircraft in a timely manner	1 point for each time unit* aircraft not welcomed
Duplication	Avoid sending the same message twice.	10 points for duplication of a message
Extraneous Click	Avoid clicking on an air traffic control center when not required	10 points for an extraneous click
Incorrect Message	Avoid sending a message when proper trigger not present	10 points for an incorrect message

* The time unit was 60 seconds for Low Workload, 45 seconds for Medium and 30 seconds for High Workload scenarios, respectively, to keep the maximum number of penalty points that could be earned in each condition constant.

WELCOME. The Welcome is an optional action. Omitting it will not cause the aircraft to turn red. There are small penalties, however, associated with a delay in welcoming an aircraft (see Table 1).

ACCEPT/REJECT REQUEST. Aircraft (AC) within the bounds of the Central sector may, from time to time, request a speed increase. Players can respond with an ACCEPT AC REQUEST or a REJECT AC REQUEST action. Participants are instructed that the judgment of whether to accept or reject the request for a speed increase is entirely straightforward and does not require any calculation of speed or distance. If the aircraft requesting a speed increase has no aircraft traveling in a direct line in front of it, the players must ACCEPT AC REQUEST, otherwise they must REJECT AC REQUEST. An incorrect response to a

speed request carries a heavy penalty; a delay in responding to the request carries a lighter penalty (see Table 1).

TRANSFER. When an aircraft in the Central Sector reaches the inner border (green line), the player initiates a TRANSFER to hand the aircraft off to the controller in the next sector.

CONTACT ATC. When the next controller accepts the aircraft, the player tells the aircraft to CONTACT ATC. If both actions, TRANSFER and telling the aircraft to CONTACT ATC, are not completed by the time the aircraft reaches the outer boundary (yellow line), the aircraft will turn red and enter a holding pattern, with an ensuing penalty (see Table 1). The aircraft can be released by carrying out the missing

TRANSFER and/or CONTACT actions. The player may have to read the messages to determine if a TRANSFER, a CONTACT, or both, are required (assuming this information is not remembered). If the

player responds to a red outgoing aircraft by doing a TRANSFER and it turns out the TRANSFER had already been carried out, the player accrues a small penalty for a duplicate action (see Table 1).

Table 2: Action Triggers in ATC Task

<i>Trigger for Text Display</i>	<i>Trigger for Color Display</i>	<i>Action Button</i>	<i>Resulting Message</i>
Incoming			
“Accept TWA555?”	GREEN	ACCEPTING AC	“Accepting TWA555”
“TWA555 says hello”	BLUE	WELCOME	“Welcoming TWA555”
Within-sector			
“OLY112 speed request?”	MAGENTA	ACCEPT/REJECT AC REQUEST	“Accepting/Rejecting speed change for OLY112”
Outgoing			
[Nose of AC crosses green line]	BROWN	TRANSFERRING AC	“East accept US801?”
“East accepting US801”	YELLOW	CONTACT ATC	“US801 contact East”

2.4 Text Display

In the text display, messages appear over time on one of three message boards located beneath the appropriate action buttons (see Figure 1).

ACCEPT. Suppose an aircraft, TWA555, approaches the Central Sector from the East. When the aircraft is 25 miles outside the yellow boundary, a message appears on the left-most board: “ACCEPT TWA555?” The player must ACCEPT the plane as soon as possible after the message appears. Accepting involves four steps: 1) clicking the ACCEPTING AC button, 2) clicking on the TWA555 aircraft icon, 3) clicking on the relevant controller icon (in this case East), 4) clicking the SEND button. It is OK to reverse Steps 2 and 3.

Step 1 (pressing the ACCEPT button) brings up a message template in the upper left hand corner, above the radar screen (see Figure 1) with slots indicating the required information. While most actions require four steps, two actions (Welcome and Speed Request) require the player to omit step 3 (clicking on the controller).

WELCOME. Some time after TWA555 has been accepted, a message appears on the left-most board: “TWA555 saying hello.” The player clicks the WELCOME button. The message template reveals that only two additional steps are required: clicking on the aircraft icon (not on the controller icon) and the SEND button (see Figure 1).

SPEED REQUEST. A message appears in the bottom-most message area: “TWA555 speed increase?” The player decides to ACCEPT AC REQUEST (or REJECT AC REQUEST) depending on whether or not

there is another aircraft directly in front of the aircraft and traveling in the same direction. The response (like that for WELCOME) requires only 3 steps: 1) Click the appropriate ACCEPT/REJECT AC REQUEST button, 2) click on aircraft icon 3) SEND.

TRANSFER. This action is the only one that is not triggered by a message, but rather, by the position of the aircraft. As soon as the nose crosses the green line, the player can do a TRANSFER. (A TRANSFER cannot physically be carried out prior to that point because the button is disabled.) The transfer involves the usual four steps: 1) Click the TRANSFERRING AC button 2) click on aircraft icon 3) click icon of controller in next sector 4) SEND.

CONTACT ATC. To continue the illustration, suppose the player has sent the TRANSFER message to transfer TWA555 to the East Sector. After a delay of several seconds, a message appears on the right board: “EAST accepting TWA555.” The player tells the aircraft to CONTACT ATC by completing the usual four steps: clicking the CONTACT ATC button, the aircraft icon, the controller icon, and SEND.

2.5 Color Display

The color display is identical to the TEXT display, with all the same messages, but also includes a color coding of the aircraft that serves as a trigger for the actions that need to be taken (see Table 2). The icon for the aircraft turns color at the same time that the Trigger message appears on the board, making it unnecessary to read the message. The icon color turns back to white when the SEND is completed.

2.6 Scenarios

Each scenario involves twenty-two aircraft, each moving along a horizontal or vertical, straight-line path, visible on the radar screen shown in Figure 1. The entire Central Sector, bounded by the yellow lines, is a square with each side 200 nautical miles (NM). Aircraft enter and exit from all four sides. The distance between the outer and inner border (yellow and green lines) is 25 NM. The aircraft are traveling at a speed of 1,000 NM/hr or .28 NM/second. At that rate, it takes 3.6 seconds to travel one mile.

Each scenario consists of 20 aircraft that are either incoming and require an Accept or are outgoing and require a Transfer, but not both. The scenarios are constructed so that no entering aircraft reaches the opposite green border. Exiting aircraft are already within the sector at the time the scenario begins.

There are also three speed requests in each scenario, two calling for an ACCEPT AC REQUEST response and one calling for a REJECT AC REQUEST. In actuality, granting a speed increase had no effect on aircraft speed.

2.7 Scenario Sets

Two sets of scenarios (A and B) were initially constructed. Two additional sets (A* and B*) were generated by rotating each of the aircraft in the original scenarios 180 degrees. For example, an aircraft that would enter from the northern portion of the West sector in the A scenario, would enter from the southern portion of the East sector in the A* scenario.

3. Method for Data Collection

3.1 Participants (Human)

Sixteen participants, four females and twelve males, were tested. They were employees of BBN Technologies (now part of Verizon). They were all experienced at video games and not more than 35 years old.

Half the participants were randomly assigned to the Development group and the other half to the Comparison group. The data and scenarios from the Development group were released to the modelers during the model development phase. The data from the Comparison group was withheld from the modelers until after the runs.

The original plan was to run the models only on the Comparison scenarios and compare them only with the human Comparison participants. However, to achieve a more stable basis on which to compare the models,

the plan was changed. The human data from both the Development and Comparison groups were combined, and the models were run on both the Development and Comparison scenarios.

3.2 Participants (Models)

AFRL, CHI Systems, and Soar Technology provided model runs that produced data equivalent to that of four participants. CMU provided runs equivalent to sixteen participants.

3.3 Design

The design was a within-participants comparison of two display conditions (Text, Color) and three workload levels (Low, Medium, High).

There were four equivalent sets of scenarios A, A*, B, and B*, (see Scenarios, above). Half the participants were trained on the A and A* scenarios (e.g., A for Text and A* for Color) and tested on the B and B* scenarios. The other half were trained on the B and B* scenarios and tested on A and A*. Within each group, half the participants received the unstarred scenarios in the Text Display condition and the starred scenarios in the Color condition. The other half received the starred scenarios in the Text condition and unstarred in the Color condition. The starred and unstarred versions, mirror images of each other, were structurally equivalent and therefore, directly comparable in difficulty.

3.4 Procedure (Human Participants)

Each participant took part in two sessions scheduled no more than a few days apart. The first session, 2-1/2 hours in length, involved an initial phase of training and a "Practice Block" of six trials (covering 2 display conditions x 3 workload levels).

The second session, 2 hours in length, involved the actual "Test Block" of 6 trials. At the end of each trial, participants completed the Task Loading Index (TLX) workload rating sheet [8]. A debrief questionnaire concerned with the participant's strategies was administered at the conclusion of the test trials along with a simple motor reaction time test.

Data given to the modelers during the development phase included all the data (both qualitative and quantitative) collected from the eight "Development" participants.

3.5 Procedure (Model Runs)

The model runs took place at BBN. The presence of the modelers was not required, though some chose to

be present. No changes were made to the models between runs.

4. Results

The results were analyzed in terms of the three dependent measures that were available for both human and model runs: Penalty points, response time, and subjective workload. In addition, human responses to the Debrief Questionnaire were examined.

For the three measures of primary interest (penalties, response time, and workload), a bar graph was prepared showing the averages for the sixteen human participants in each of six conditions (2 displays x 3 workload levels). Error bars were used to indicate the confidence interval for each bar (see Figures 2, 4, and 6).

The model data were graphed in the following way. Multiple runs in a given condition (16 for CMU and 4 for the other models) were averaged together and plotted as a single point on the bar graph showing human performance. This display made it easy to see where the model averages fell with respect to the confidence interval for the human averages. One criterion for evaluating the models was whether the means fell within the confidence interval for the humans. A less strict criterion was whether the models showed the same general trends as did the humans.

The results for the humans and models are discussed below, for each of the measures of interest.

4.1 Penalty Points

An examination of the human results (six bars) in Figure 2 shows that more penalty points were accrued with the text display than with the color display, especially at higher workloads. These trends were supported by an analysis of variance, with Display and

Workload level as within-participant variables and Group (Development vs. Comparison) as a between-participant variable. The main effects of Display, $F(1,14)=23.27$, and Workload, $F(2,28)=10.80$, were significant ($p < .001$), as was the interaction between Display and Workload, $F(2,28)=9.76$, $p < .001$. The main effect of Group, $F(1,14) < 1$, was not significant ($p > .10$). Group did not interact significantly with any of the other variables ($p > .05$). All four models supported these trends. None of the models met the strict criterion of falling within the human error bars for all conditions. Deviations tended to be in the direction of too many penalties. CHI came closest to the human norms.

Penalty scores were explored in greater detail in the most demanding condition. Figure 3a shows the penalty points earned by humans and models in each of the penalty subcategories for the Text-High Workload condition. It is clear from the graph that the overriding source of points for humans was Holds (at 50 points each). All the models showed this same pattern, although no model was exactly within the human error bar limits on all subcategories. Again, deviations from the human norms tended to be in the direction of too many penalties

Figure 3b shows the actual number of occurrences of each type of error. (These occurrences resulted in the penalties shown in Figure 3a.) This figure suggests that participants prioritized their actions so as to minimize overall penalties. Thus, Welcome Delay, which carries the lowest penalty (1 point per minute), was the most frequent omission made by humans. The next largest category of errors, for humans, was Speed Delay (2 points per unit of time). The strategy of postponing actions carrying low penalties to focus on preventing aircraft from turning red, which carries a higher penalty (50 points), is a reasonable strategy for coping with high workloads. Although none of the models managed

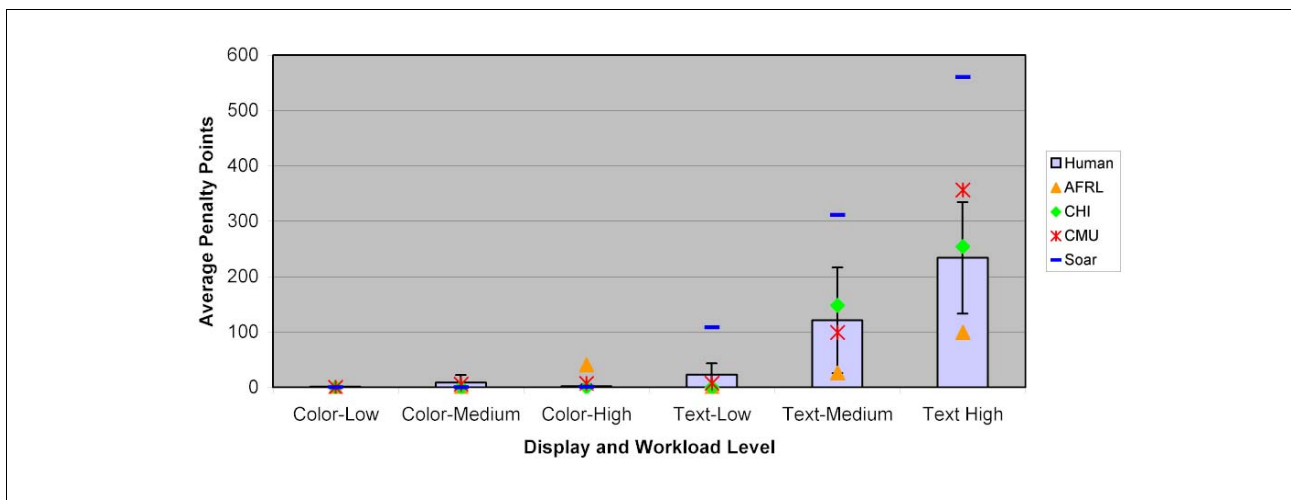


Figure 2: Penalties as a Function of Display and Workload

to fall within the error bar criterion consistently for all penalty categories, EPIC-Soar did show evidence of load shedding. EPIC-Soar was the only model to have more occurrences of Welcome Delays and Speed Delays than of Holds, resembling the humans.

4.2 Response Time

Figure 4 shows the average response time for each condition for the human participants. Response time was defined as the time interval between the appearance of the trigger for an action (see Table 2) and the activation of the SEND button to complete that action. All of the actions in Table 2 were included in the average.

As can be seen in the graph, participants responded to the triggers more quickly with the Color display than with the Text display, and workload effects were more pronounced in the Text than in the Color condition. These results resemble the Penalties results. An analysis of variance showed significant main effects of Display, $F(1,14)=60.78$, $p<.0001$ and Workload, $F(2,28)=19.69$, $p<.0001$, and a significant interaction between Workload and Display, $F(2,28)=12.73$, $p<.0001$. There was no main effect of group, $F(1,14)=1.15$, $P>.10$, and group did not interact with any other variable ($p>.10$). The models all showed the same trends. No model fell within the confidence interval for all conditions, but, again, CHI came extremely close. Where they deviated from the human norms, the models tended to be too slow. Two further analyses were undertaken to explore why some of the models may have been responding too slowly. Because the models tended to deviate most from the human reaction times when the task became most demanding, the Text High condition was selected for further study.

Number of Intervening actions. Figure 5a shows the frequency distribution, in the Text High condition, for responses that occurred after varying numbers of intervening actions, from no intervening actions to five. Although subjects occasionally carried out more than five intervening actions before responding to a trigger, such occurrences were rare. For this analysis, the number of intervening actions was defined as the number of other actions completed between the appearance of a particular trigger and its response. For example, if, after TWA555 turned blue, the player completed an action on another aircraft before welcoming TWA555, then the number of intervening actions would be counted as one. If no such actions occurred between the trigger and the response, then the number of intervening items would be zero. Results for the humans showed that, even in the difficult, Text High workload condition, participants were most likely to respond to a trigger immediately or after taking just one intervening action (see Figure 5a). The results for the models follow this same pattern. No model met the criterion of falling within the confidence interval in every case, although CHI came close. Where the models fell short, they tended to have too low a frequency of immediate responding and too high a frequency of responding after many intervening actions.

Response time as a function of number of intervening actions. Figure 5b shows response times as a function of the number of intervening actions for the Text High condition. The human data show a smoothly increasing response time as the number of intervening actions increases. The models all show this trend. None of the models fell within the confidence interval for all intervals, but CMU came close.

To summarize the data thus far, for both penalties and response times, all the models showed the appropriate

human-like trends. Some of the models were even able to duplicate the human averages exactly for some of the measures. Performance of the models tended to deviate in the direction of under-performance: too many holds, too many intervening items, and slow response times. The modelers clearly met the goal of “humanizing” the models and making them subject to human-like cognitive limitations.

4.3 Subjective Workload

An overall subjective workload rating was obtained for each participant (human or model) by averaging across the six individual workload scales that are part of the TLX (mental demand, physical demand, temporal demand, performance, effort, frustration). CHI was the only team to produce workload scores for all six TLX scales. The CMU Model and EPIC-Soar Model each

produced an overall workload score. AFRL did not calculate workloads.

The human results, as seen in Figure 6 showed that participants rated their workload as higher for the text than for the color display. There was also an increase in subjective workload as actual workload increased, especially for the text display. An analysis of variance showed significant main effects of Display, $F(1,14)=43.97$, $p<.0001$, and Workload, $F(2,28)=24.43$, $P<.0001$, and a significant interaction between Display and Workload, $F(2,28)=13.21$, $P<.0001$. The models showed the same trends and matched the human data quite well. CMU was within the human confidence interval for all conditions.

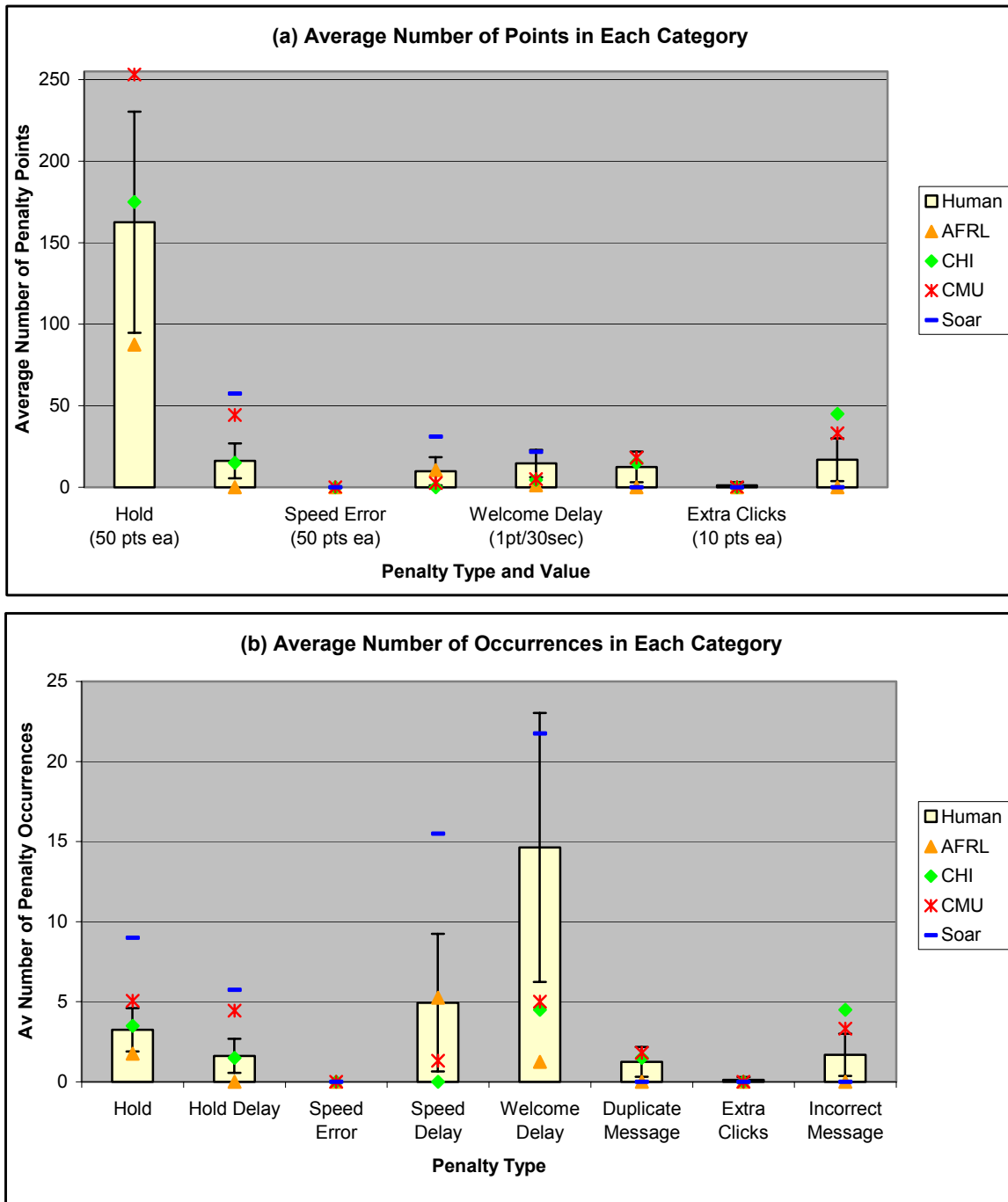


Figure 3: Detailed Analysis of Penalty Categories for Text-High Workload Condition

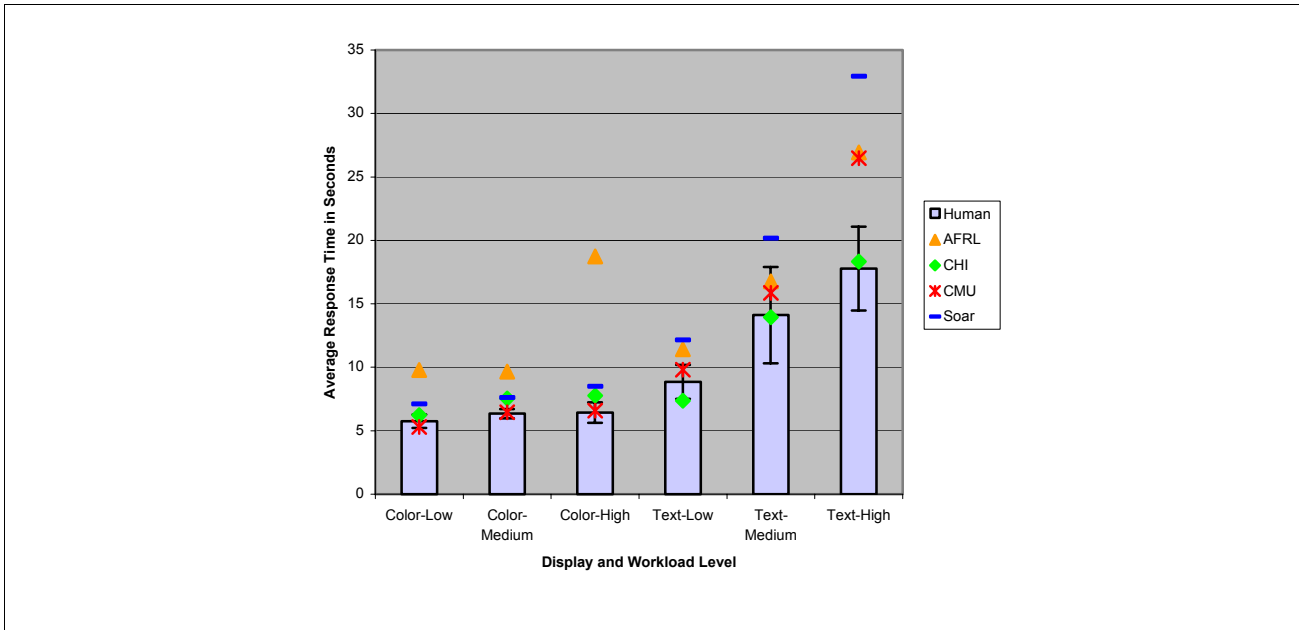


Figure 4: Response Time as a Function of Display and Workload

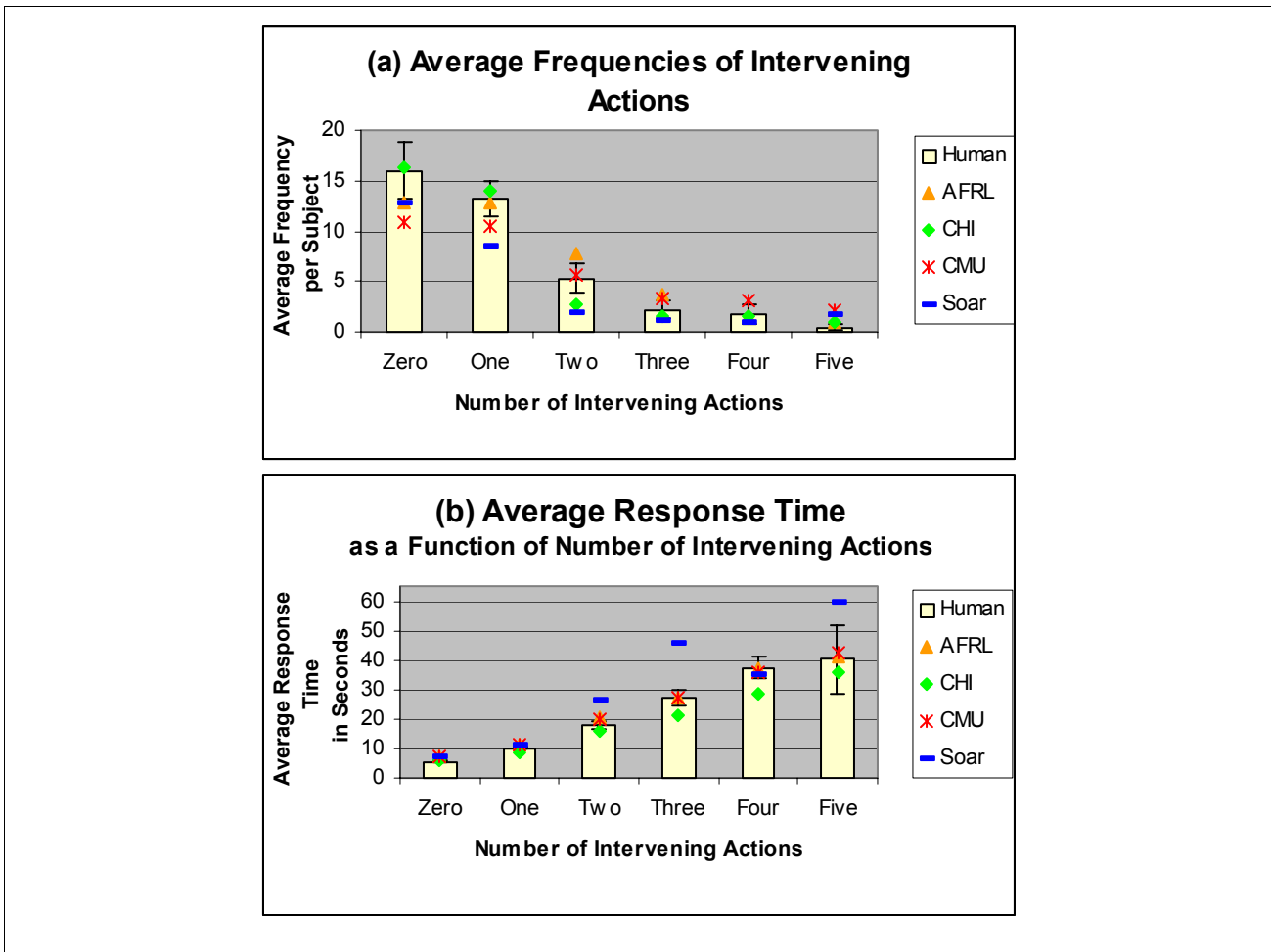


Figure 5: Detailed Analysis of Response Time in Text-High Workload Condition.

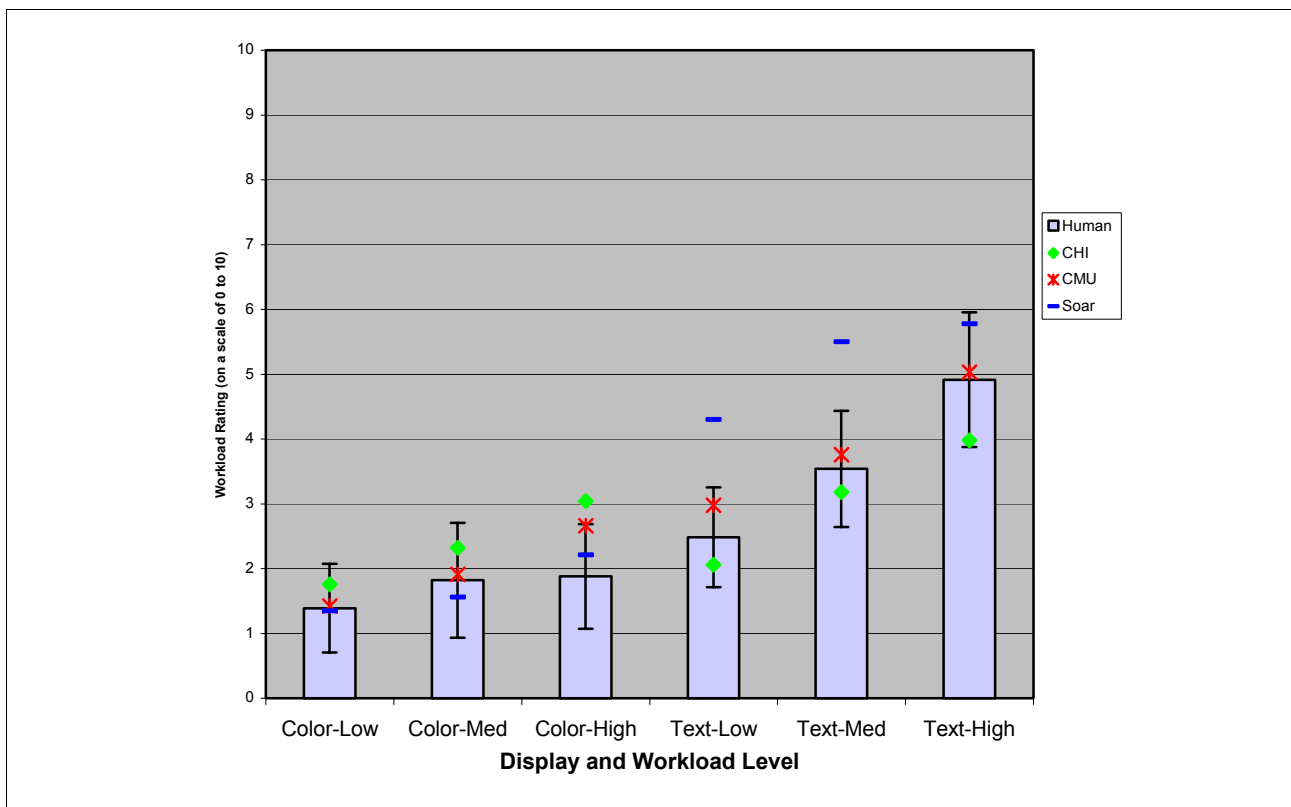


Figure 6: Subjective Workload as a Function of Display and Workload Condition

5. Round 2: Replication of Study Using an HLA and Non-HLA Implementation

Using the sponsorship of DMSO, a replication of the study was undertaken with two separate implementations, the Original and an HLA implementation. The purpose of the study was to determine whether comparable results could be achieved for both humans and models with the two different implementations. The study was part of DMSO's efforts to explore the feasibility of using HLA for future modeling efforts. For details regarding the implementations, see [2].

5.1 Method

Of the human participants who were trained in Round 1, twelve were available to participate in the Replication study. The same four human performance models from Round 1 (EPIC-Soar, COGNET/iGEN, DCOG, and ACT-R) participated. ACT-R produced runs for eight participants, the others for two participants each, for a total of 14 "simulated" participants.

The design called for participants to experience both the HLA and Original implementation, one after the other, across increasingly more difficult workload levels. The four workload levels, from easiest to

hardest, were the Color-High, Text-Low Text Medium, and Text High conditions from Round 1. Each participant completed eight scenarios (4 workload levels x 2 implementations). The order in which the two implementations were administered at each workload level was counterbalanced across participants, as was the assignment of scenarios to implementation conditions.

The same measures--penalties, response time, and subjective workload ratings--were collected as in Round 1. In addition, the human participants were asked after every pair of trials (one HLA and one non-HLA): "Did you notice any difference at all between the last two trials you just did and if so what was different?" Following this open-ended question, they were asked to compare the two trials in terms of the behavior of the airplanes, the actions of the buttons and mouse clicks, and the way in which the aircraft changed color (in the lowest workload condition). Finally, they were asked to decide "which of the last two trials was the original one you did a year ago" and to rate their confidence in their answer.

5.2 Results

The results suggest that performance of both models and humans was similar for the two implementations on all measures (Figure 7 shows the results for

response time). Responses to the subjective questions showed that the human participants did not perceive any reliable differences between the implementations. An analysis of variance with two repeated measures showed no significant main effect of Implementation (HLA vs Original) for any of the measures, $F(1,11)=4.715$, $p<.10$, for performance scores; $F(1,11)=3.396$, $p<.10$, for response times; $F(1,11)<1$, $p>.10$, for subjective workload ratings, and there was no significant interaction between Implementation and Workload on any of the measures ($p>.10$). The main

effect of workload, as expected, was highly significant, in all cases ($p<.0001$).

Similarly, an analysis of variance on the data from the 14 "modeled participants" (8 ACT-R, 2 EPIC-Soar, 2 COGNET/iGEN, 2 DCOG) showed no main effect of Implementation, ($F(1,13)<1$, $p>.10$, for performance scores; $F(11,13)<1$, $p>.10$, for average response time, and there was no interaction between Implementation and Workload on any of the measures ($p>.10$). The main effect of workload, as expected, was highly significant for all measures ($p<.0001$).

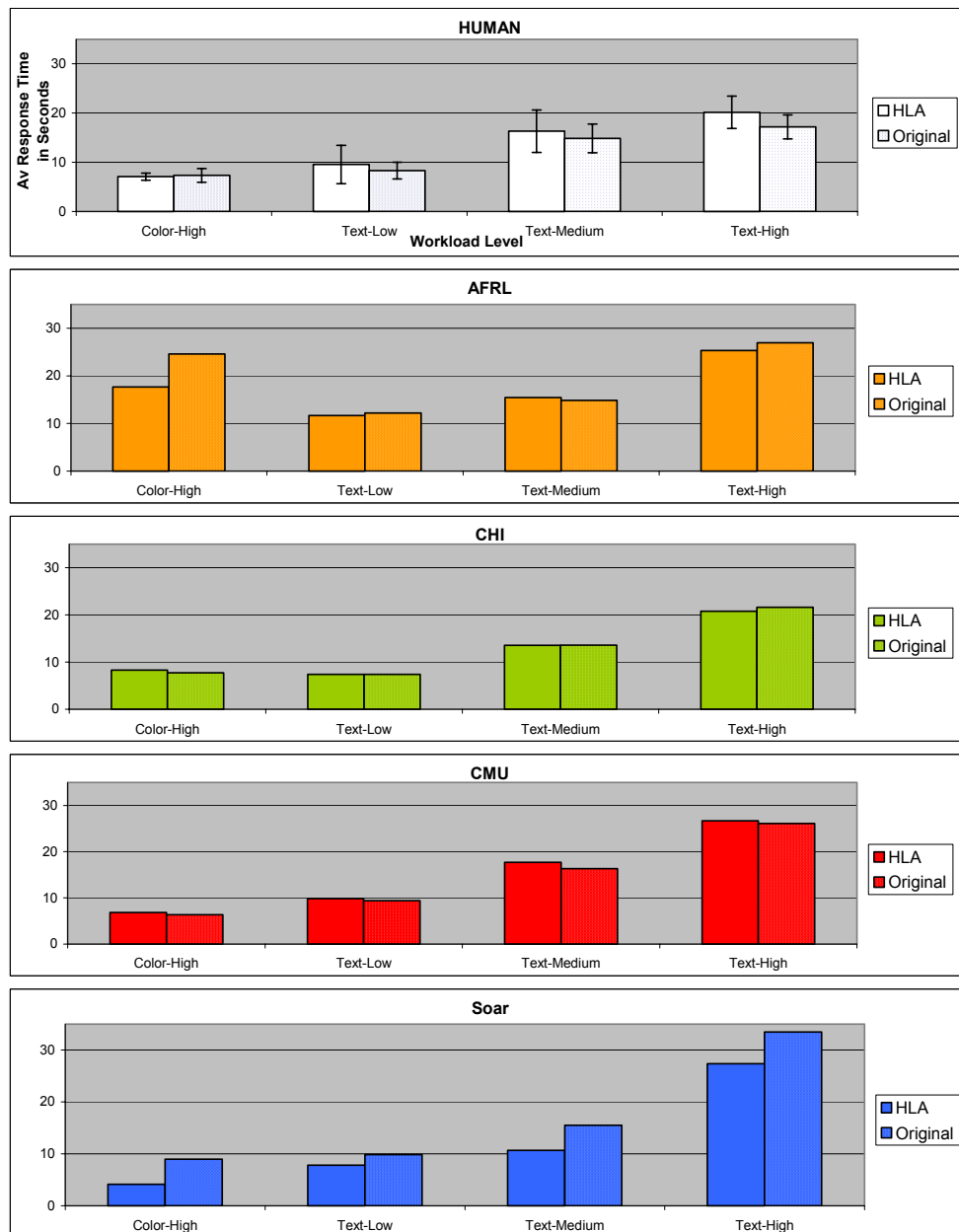


Figure 7: Response Time as a Function of Workload

6. Conclusion

The AMBR modeling comparison effort is alive and well and is producing results. The program holds the promise of advancing the state of the art in human performance modeling, distributed simulations, and understanding multi-task behavior.

7. References

- [1] K.A. Gluck and R.W. Pew: "Overview of the Agent-based Modeling and Behavior Representation (AMBR) Model Comparison Project" Proceedings of the 10th Conference on Computer Generated Forces and Behavior Representation, Norfolk, VA, 2001.
- [2] S. Deutsch, and B. Benyo: "The D-OMAR Simulation Environment for the AMBR Experiments" Proceedings of the 10th Conference on Computer Generated Forces and Behavior Representation, Norfolk, VA, 2001.
- [3] R. Chong: "Low-level-Behavioral Modeling and the HLA: An EPIC-Soar Model of an Enroute Air-Traffic Control Task" Proceedings of the 10th Conference on Computer Generated Forces and Behavior Representation, Norfolk, VA, 2001.
- [4] R. G. Eggleston, M.J. Young, & K.L. McCreight: "Modeling Human Work through Distributed Cognition" Proceedings of the 10th Conference on Computer Generated Forces and Behavior Representation, Norfolk, VA, 2001.
- [5] C. Lebiere, J.R. Anderson, & D. Bothell: "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, Norfolk, VA, 2001.
- [6] W. Zachary, T. Santarelli, J. Ryder, J. Stokes, & D. Scolaro: "Developing a Multi-tasking Cognitive Agent Using the COGNET/iGEN Integrative Architecture" Proceedings of the 10th

Conference on Computer Generated Forces and Behavior Representation, Norfolk, VA, 2001.

- [7] J. MacMillan, S.D. Deutsch, & M.J. Young: "A Comparison of Alternatives for Automated Decision Support in a Multi-task Environment" Proceedings of the 41st Annual Meeting of the Human Factors and Ergonomics Society, 1997.
- [8] NASA-Ames Research Center, Human Performance Group: Collecting NASA Workload Ratings: A Paper-and-pencil Package (Version 2.1), NASA-Ames Research Center, Moffet Field, CA, 1986.

Author Biographies

YVETTE J. TENNEY is a cognitive psychologist with interests in the areas of applied cognition, training, and human factors. She received her B.A. and Ph.D. in psychology from Cornell University. As a scientist at BBN Technologies, in Cambridge, MA, Dr. Tenney has been involved in developing training technologies and job aids for underwater exploration, harbor navigation, complex electronic troubleshooting, command and control, and commercial aviation. Her list of publications includes articles in *Cognitive Science*, *Human Factors*, and the *International Journal of Aviation Psychology*, as well as several book chapters. Her recent work has focused on problems of operators in semi-automated, multi-task environments, particularly on the maintenance of situation awareness and the management of cognitive workload.

SANDRA L. SPECTOR is an Associate Scientist at BBN Technologies. She received her B.S. from Umass, Amherst and M.A. in psychology from Northeastern University. Sandra is part of the Human Factors group responsible for supporting BBNT's psychological research and usability needs. She has worked on speech applications, human performance modeling, intelligent tutoring systems and Web development.