Towards Using Human Performance Moderator Functions in Human-Robot Teams

Caroline E. Harriott[1], Rui Zhuang[2], Julie A. Adams[1], and Scott A. DeLoach[2]

[1] Vanderbilt University, Nashville, TN
[2] Kansas State University, Manhattan, KS
caroline.e.harriott@vanderbilt.edu, zrui@ksu.edu,
julie.a.adams@vanderbilt.edu, sdeloach@ksu.edu

Abstract. Human-Robot peer-based teams are evolving from a far-off possibility into a tangible reality. Human Performance Moderator Functions can be used to predict human behavior by incorporating the effects of internal and external influences, such as fatigue and workload. The applicability of human performance moderator functions to human-robot teams is not proven; however, the presented research demonstrates the applicability of a workload human performance moderator function to a human-robot reconnaissance team. The research models the performance function for both human-human and human-robot teams, empirically validates the model, and creates a simulation for allocating tasks to human and robot teammates. The results show that this particular human performance moderator function is applicable to peer-based human-robot teams.

Keywords: human performance factors, human-robot teams, human-robot interaction, runtime models

1 Introduction

Robotic technology continues to develop and humans are beginning to be partnered with robots for peer-based tasks [15, 27]. It is known that individual human performance can impact human team performance [22]. Similarly, human performance will impact the task performance of human-robot teams. As human-robot team capabilities improve, it is necessary for the robotic team members to understand how the human's performance capabilities affect the task at hand. Future human-robot team task assignments will need to allocate tasks to the team entities based upon the predicted human performance. Thus, it is necessary to understand if and how existing human performance moderator functions apply to human-robot teams. However, the same human performance moderator functions that are used in human-human scenarios should not be used to predict human-robot team performance without a thorough test of the functions' applicability.

Human performance moderator functions are equations derived from empirical results that predict human performance due to specific performance factors, such as

adfa, p. 1, 2011.

fatigue, mental workload or temperature. These factors greatly influence human behavior. Cognitive processing speed, physical capability, and stress levels can be manipulated by the progression of time on task and changes to the environment; these elements feed into human performance moderator functions to create performance predictions. Approximately 500 human performance moderator functions [30, 31] are known to exist, and it is well known that a number of interactions exist across human performance moderator functions. For example, an individual's skill level, perhaps based on training, can reduce cognitive workload. In other words, tasks that become automatic can require lower cognitive workload. Specifically, our research investigates integrating humans and robots into a single team. We have examined how human performance moderator functions can be used to predict a human's performance in various team roles in order to be able to assign team members (humans or robots) to appropriate roles in real-time. The research has assessed the applicability of existing human performance data to human-robot teams; existing, modified, and new human performance moderator functions have been modeled, verified and validated [17, 18, 19]. If modeled correctly, these human performance moderator functions can inform predictions on how the human and robot can most successfully interact and assign tasks to the most appropriate team formations. In addition, we have been investigating and devising a design-time and runtime framework for human-robot team systems that predicts human and robot performance for various team tasks and assigns appropriate humans or robots to team roles.

Specifically, our research has focused on first response and hazardous materials scenarios. The first response scenario teamed an uninjured human with a robot that instructed the human as to how to triage injured victims in a contaminated area [17, 19]. The hazardous materials scenario assigned specific responsibilities to both the human and the robot, but also required joint team decisions. This paper specifically focuses on the hazardous materials scenario.

The usefulness of human performance models in making task assignments in human-robot teams was evaluated by modeling the human performance characteristics for each scenario (Section 2.1) and empirically validating the resulting human performance model (Section 2.2). The usefulness of such models at runtime was demonstrated by developing a simulation of the hazardous materials scenario, which used the human performance model to calculate the effect of the tasks on the human during scenario execution (Section 3). While the model outputs have yet to be used to inform the task assignment process, the simulation results show that the model outputs calculated at runtime were consistent with the results of the empirical experiment.

2 Human Performance Modeling and Evaluation

The hazardous materials evaluation required a reconnaissance of a single floor of an academic building after the receipt of a bomb threat. The team was responsible for collecting air samples, locating suspicious objects, and updating the incident commander. Fig. 1 provides the layout of the entire reconnaissance area, which included a hallway and two laboratories.

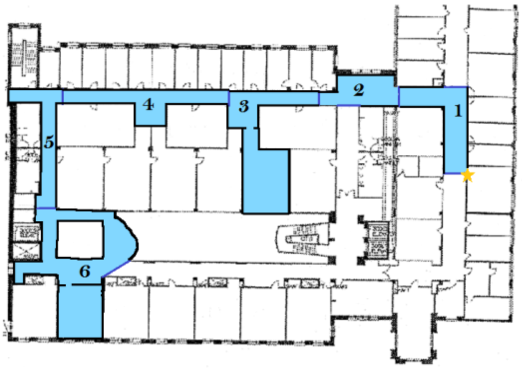

Fig. 1. A bird's eye view of the hazardous materials scenario reconnaissance area. The area was divided into six areas for purposes of manipulating and measuring cognitive workload.

The scenario assumes that the human partner works closely with the robotic partner to identify all potential suspicious items. The team was responsible for collecting air samples, locating suspicious objects, and updating the incident commander. If a suspicious object was found, it was not disturbed and information regarding its whereabouts was reported immediately to an incident commander [25]. This scenario requires an interactive relationship between the teammates, while also providing the human team member with some control over the direction of the investigation. Each team member had a few specific responsibilities, for example, the human was responsible for looking in trash and recycling bins, while the robot was responsible for investigating fire extinguisher cabinets that were located at the height of the robot's sensors and collecting chemical air samples. Although not discussed in this paper, the scenario also required the partners to make joint decisions regarding potential actions for each partner.

The entire area to be investigated was divided into six areas of similar size, shown as the numbered shaded areas in Fig. 1. While the areas were divided into similarly sized areas, the workload within each area was manipulated based on the number of objects that required investigation, the complexity of the objects in the area, etc. The current paper focuses on the first two areas, Area 1 and Area 2. Area 1 contained two suspicious items and four non-suspicious bulletin boards. Area 2 contained a fire extinguisher, two suspicious items and two non-suspicious items.

The human performance modeling and empirical validation research developed a model and validated the model for each team type considered. One model and validation were developed for the human-human teams (H-H) and another set for the human-robot teams (H-R). This is an important point, since it is unclear if the workload human performance moderator function that has been developed for other domains is applicable to humans working with a robot partner. The H-H team provides a baseline from which one can better understand for the applicability of the human performance moderator functions to the H-R team.

2.1 Human Performance Modeling

Independent computational human performance models for each team representation (H-H and H-R) were created to predict the workload levels experienced by the human teammate for each of the investigation areas. The performance models were developed using the Improved Performance Research Integration Tool (IMPRINT Pro) [1], which allows for the creation of a network of tasks to be completed by the team. Each task has an associated time and workload. The workload value is decomposed into seven workload channels, which include: Auditory, Visual, Speech, Tactile, Cognitive, Fine Motor, and Gross Motor. IMPRINT Pro offers guidelines for assigning workload values for each task and provides micromodels of human performance based on empirical data to aid in determining the expected task time. The tool's value ranges for the workload channels include different ranges for the different channels. For example, the cognitive channel has a range of values from 0 (little demand) to 7 (high demand), while the visual channel has a range of 0 to 6. The modeling tool provides an estimate of the overall task execution time and the predicted workload at each step of the task.

The IMPRINT models incorporate all tasks performed by the human during the reconnaissance, including walking between locations, making decisions regarding whether or not an item is classified as suspicious, listening to feedback from the partner, and participating in decision making tasks. This scenario incorporates uncertainty due to team decisions and individual differences, thus the models incorporated probability. The primary difference between the models is the incorporation of longer task durations. The robot spoke slower than the human and, even though the speed of the robot was increased, the robot was required to move more slowly than average human walking pace due to safety concerns when other humans (not part of the team) were present in the hallways.

As previously mentioned, this paper focuses on the first two reconnaissance areas. Fig. 2 provides an overview of the overall workload results as determined by the H-H and H-R team models for each reconnaissance area. Due to the uncertainty represented in the IMPRINT models, ten trials were run for each team model. The overall

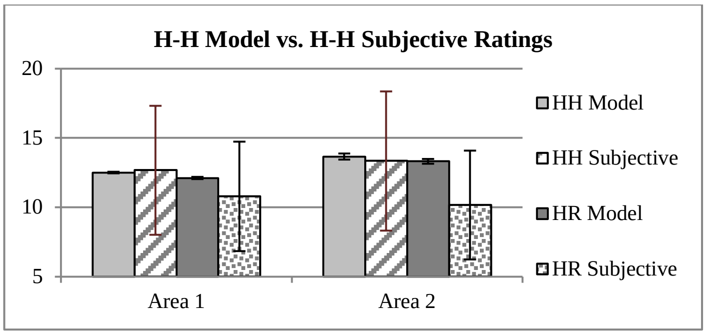

Fig. 2. Scaled mean workload for the model versus the mean subjective workload results by condition and reconnaissance area. Error bars represent one standard deviation above and below the mean.

Table 1. Mean workload values and standard deviation for the model and validation results by condition and reconnassiance area.

	H-H		H-R	
	Model	**Validation**	**Model**	**Validation**
Area 1	12.50 (0.06)	12.67 (4.64)	12.11 (0.09)	10.78 (3.95)
Area 2	13.65 (0.23)	13.33 (5.02)	13.31 (0.17)	10.17 (3.92)

workload value averages the workload values from each trial by investigation area. Similarly, Table 1 provides the means and standard deviation for the overall workload. Please note that the model workload results are scaled to a value between 1 and 5, further details and justification of this scaling are provided in Section 2.3.

2.2 Validation Evaluation

The IMPRINT models were verified via a user evaluation that included both H-H and H-R teams. A single evaluator played the role of a trained first responder in all H-H condition trials. The H-R condition paired the participant with a Pioneer 2-DX robot. Both evaluation conditions required the team to perform a reconnaissance of all six areas in a continuous fashion. Unknown to participants, a remote evaluator controlled the robot's speech and supervised the robot's movement. Thirty-six participants completed the evaluation with eighteen in each condition. The 19 male and 17 female participants were not experts in first response or robotics and ranged in age between 18 and 56 years old.

The teams in each condition followed the same path from Area 1 to Area 6 (see Fig. 1) and the same suspicious items were placed in the same locations for both conditions. The human (an experimenter) or robot partner was responsible for checking fire extinguishers and taking chemical air samples, and informed the participants that they were responsible for checking bulletin boards and trash cans, while looking for out of place items. At points, the team was required to make joint decisions regarding the appropriate action to take given the current circumstances.

After completing the reconnaissance of a specific area (e.g., Area 1), but prior to beginning the reconnaissance of the next area, the participants rated their subjective workload on six channels: Auditory, Visual, Speech, Motor, Tactile and Cognitive. The modeled Fine and Gross Motor channels, presented in Sec. 2.1, were combined in the subjective ratings in order to facilitate participant understanding and data collection. The rating process and each channel were defined during training. The participant's partner, either the human experimenter or the robot, asked the participant to rate the demands they experienced during that portion of the investigation on a scale from 1 (little demand) to 5 (extreme demand).

The mean overall workload by participant was calculated by averaging the total demand from each area. Fig. 2 provides the mean workload results for areas 1 and 2, while Table 1 provides the means and standard deviations. It is noted that the workload results for these two areas were lower for the H-R condition than for the H-H condition. This result mirrors the overall evaluation results. The mean overall workload for the H-H condition was 13.991 (St. Dev. = 5.215), and was 10.972 (St. Dev. = 3.957) for the H-R condition. An ANOVA indicated that H-H condition resulted in significantly higher overall workload, $F(1, 214) = 22.752$, $p < 0.001$.

2.3 Model Validation

It is necessary to compare the results from the validation evaluation to the model trial results. As indicated in Section 2.1, the IMPRINT model channel ranges are not identical to the subjective workload channel ranges, thus it is necessary to convert the model values to the same scale as the subjective ratings. The first step sums the model based results for the Fine and Gross Motor workload channels. The IMPRINT workload channels values range from zero to a value between four and seven. The mean model results by channel were scaled to a value between 1 and 5 in order to match the subjective workload range. The total workload was calculated by summing all six channel values.

As can be seen in Fig. 2 and Table 1, there is virtually no difference between the workload results for the model or the evaluation for the H-H team condition. T-tests indicated that the H-H model and validation workload values were not significantly different in either reconnaissance area. This result indicates that the model is a good representation of empirical workload in the H-H condition. A reasonable difference between the model results and the evaluation results existed for the H-R team condition. The H-R model and evaluations results were not significantly different for Area 1, but the model predictions were significantly higher than the subjective ratings for Area 2, $t(22) = 2.521$, $p = 0.02$. Even though this difference is significant, the H-R team result was within one standard deviation of the model results. In the H-R condition, the model was identical to the H-H condition model with the addition of adjustments for the robot's slower movement and speech. Overall, the model was a good predictor of workload in the H-H team condition in both investigation areas and for the H-R team condition Area 1, but it overestimated workload for the H-R condition in Area 2. The adjustments for predicted differences in movement time and conversation length in the H-R condition may have been insufficient to closely predict the difference in workload between conditions, but the model was able to provide an estimate within one standard deviation of the empirical results. The model can be adjusted to more accurately match the empirical results, but the goal of the presented research was to present the model as developed compared to the validation results.

3 Agent Simulation

The usefulness of human performance models at runtime was demonstrated by developing a simulation of the hazardous materials scenario. The simulation used the human performance model from Section 2 to calculate the effect of tasks on the human during scenario execution. While not complete, the purpose of the simulation was to demonstrate our envisioned use of human performance moderator functions in hu-

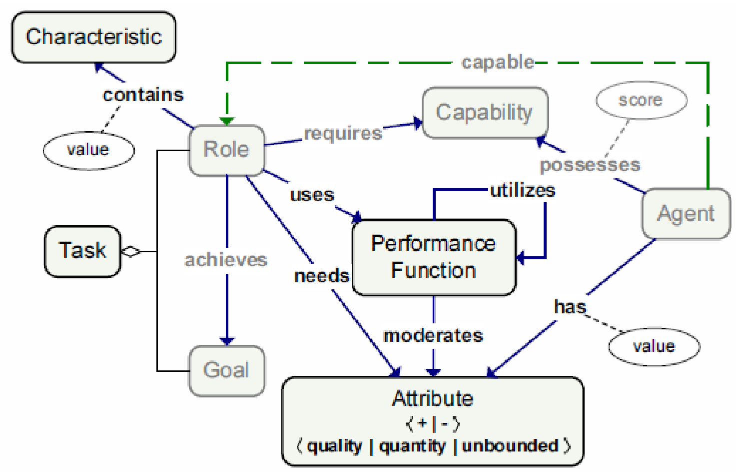

Fig. 3. Chazm Runtime Model

man-robot teams. The simulation is representative of the scenario described in Section 2 for Areas 1 and 2 and was created in the Cooperative Robotic Organization Simulator[1] (CROS). CROS is a multithreaded, grid-based environment for simulating multi-agent systems designed around the Organizational Model for Adaptive Complex Systems (OMACS) [11]. CROS supports grid-based environments with a variety of object types and simulates the behavior of a set of heterogeneous agents within that environment. The simulation of the human-robot teams in CROS required that the OMACS model be extended to incorporate human performance moderator functions, as described in Section 3.1. A CROS simulation of the human-robot hazardous materials scenario was created along with scripted versions of the human and robot agents. The resulting simulated scenario was employed to validate that the human performance values computed by the human performance model during the simulation were consistent with those in the real experiments (Section 2.2).

3.1 Chazm runtime model

The Chazm runtime model (see Fig. 3) was designed to support the incorporation of humans into multiagent and multi-robot teams. Chazm is an extension of the OMACS runtime model, whose main use is to dynamically allocate and reallocate tasks in complex adaptive systems. The key elements of OMACS are Agents, Goals, Roles, and Capabilities. The main use of OMACS is to allow Agents to be assigned to Roles to achieve specific organizational Goals where, in order to be assigned to a Role, an agent must possess all the required Capabilities. While humans can be viewed as OMACS Agents, OMACS does not have the ability to capture the notion of human performance factors that affect the performance of assigned Roles. Thus, Chazm added two main concepts to provide a means of capturing human performance, *Perfor-*

[1] See http://macr.cis.ksu.edu/cros

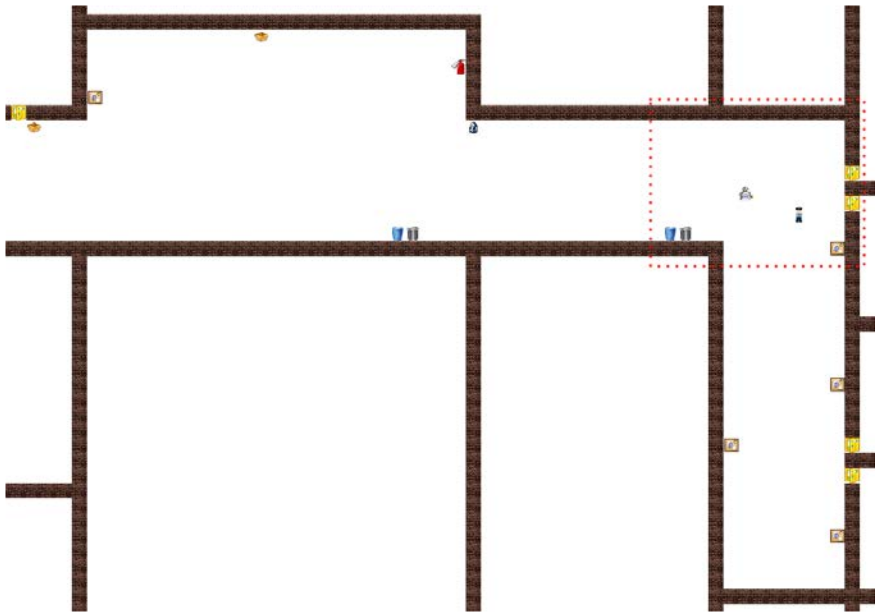

Fig. 4. Simulated environment map for reconnaissance Area 1 and Area 2 that includes the robot, teammate, and various objects.

mance Functions and *Attributes*. Attributes capture the values associated with human performance factors such as workload, fatigue, and experience, while Performance Functions determine how those Attribute values change as human agents perform various tasks. Thus, the ultimate research goal is to incorporate human performance moderator functions as Performance Functions into a runtime model (Chazm) that can be used to predict how a human's performance will degrade when assigned particular tasks in order to optimize overall human-robot team performance. While the simulation reported in this paper does not yet use Chazm to allocate/reallocate tasks, the objective of this paper is to demonstrate that Chazm can be used to accurately capture and compute such information.

3.2 Agent Simulator

The scenario modeled in IMPRINT Pro was recreated in the CROS simulator to demonstrate our proposed approach for using human performance moderator functions in human-robot teams. Fig. 4 shows a screenshot representing reconnaissance Areas 1 and 2 in the CROS environment. The dark brown lines represent the walls, while the various icons next to the walls represent the objects in those areas. The rectangle outlined by a dotted line in the upper right corner of Fig. 4 is shown in more detail in Fig. 5. All objects are positioned in the simulated environment as they were placed in the real world.

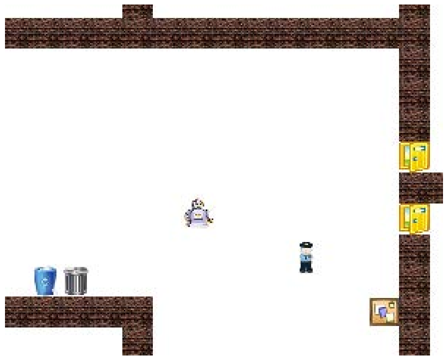

Fig. 5. A zoomed view of the upper right corner of Fig. 4, that better shows the robot, human, a recycling bin (blue), a garbage can (grey), and bulletin board (brown).

Two agents were simulated, one agent representing the human and another representing the robot. The robot (gray and white) and the human (black, blue and white) are shown in the middle of the hall in Fig. 5, while a recycling bin (blue) and garbage bin (gray) are shown to the left of the agents along the wall and a bulletin board is shown below and to the right of the agents (brown). The yellow blocks in line with the walls represent closed doors.

The human and robot agents possessed the same basic types of capabilities, including movement, localization, path planning, communication, vision, and air sampling. Although they had the same capability types, the actual capacities of those capabilities varied between human and robot. For example, the human could "see" further and higher than the robot. These differences in capabilities led to significantly different behavior in the simulation.

The first objective was to validate that the simulation produces similar workload results when compared to the real experiments, (Section 2). Thus, the initial simulation experiments defined the agent behavior in terms of a set of predefined scripts designed to match the actions from the real world experiment. As each task in the scenario was modeled separately and executed sequentially, it was straightforward to translate the modeled task behavior directly into simple scripted behavior.

Due to the large number of tasks modeled in the IMPRINT Pro (408 tasks in the six areas), the simulation used task categories to avoid the tedious process of creating a script for each individual task. The 408 tasks were categorized into 8 general categories: Walking, Listening, Speaking, Deciding, Reacting, Investigating, Taking Pictures, and Waiting. Scripts were created for each task type category.

The IMPRINT Pro model data for the scenario tasks was captured in six separate files, one for each reconnaissance area. An extra field was manually appended to the model data of each task to specify its category. A Java program was developed to convert the data files into script files that the simulator used directly. Essentially, the conversion program extracted five pieces of data from the files: the sequence number of the task, the task name, the task category, the total workload for the task, and the time taken to perform the task. Each task category, except Walking was extracted directly and reformatted into a new file. For example, the task to ‘Determine if it is necessary to go back to look at an unnoticed object’ from the scenario was converted into the following script.

```
Decide {
  index 19
  content Determine if going back is necessary
  workload 15.25
  timespan 3.0
}
```

The task category for this script is Decide, the sequence number (index) is 19, the name (content) is ‘Determine if going back is necessary’, the workload is 15.25, and the time taken to perform the task (timespan) is 3.0 seconds. The sequence number (index) is used to order the execution of tasks and synchronize the task executions between the human and the robot.

The Walk category required additional effort. The workload needs to be computed based on the time required to move from one location to another in the autonomous simulation (see Section 3.4), thus it was necessary to verify that the workload computed in the simulation (which is based on the time required to move the associated distance) was consistent with the workload from the IMPRINT Pro model. Since each grid in CROS represents 1 foot (or approximately 0.305m) and the walking speed modeled in IMPRINT Pro is 1.612 m/s, we computed the average time an agent would spend in each grid while Walking at 0.189 seconds. Thus, by knowing the agent’s start location and destination, the agent’s path and the total time spent Walking can be calculated. Unfortunately, since CROS uses square grids, agents may only move horizontally or vertically, which dramatically increases the distance and time required to move from one location to another. Therefore, given the objective of validating the values from the real world experiment, the time calculations use the realistic assumption that humans tend to move in a straight line when possible, which is easily computed in a grid-based system using the Pythagorean Theorem. Thus, for the task ‘Walk to BB1’ (BB1 means Bulletin Board 1) the following script was generated, where the workload parameter is interpreted as the workload per second.

```
Walk  {
  index 2
  content Walk to BB1
  workload 12.73
  timespan 2.28
  location (160, 71)
}
```

The scripts were parsed into the CROS simulator where each task was stored in a sequential list of *action task* objects. The human was simulated by executing each task in sequence from its list of action task objects. The robot was scripted by manually creating a similar set of data files based on the expected robot behavior from the real world experiments. Since workload is only related to the human agent, this information was omitted from the data for the robot. Scripting each agent in this manner enabled the simulated agents to perform the same basic tasks in the same sequence, as occurred in the real world experiments.

3.3 Validation of Simulated Results

All tasks in the scripted simulation, with the exception of Walking, are executed while the agents are stationary. Thus, the workload and time span values are used directly from the scripts, which were determined by the IMPRINT Pro model. This approach ensures that the IMPRINT Pro values match the values from the scripted simulation perfectly and no additional validation is required. However, since the Walk tasks required computing the time to move from one location to another, based on the actual simulator time, it was necessary to validate that, on average the times were consistent with those modeled in IMPRINT Pro.

Table 2 provides a comparison of the results of the Walk tasks between the scripted simulation and the values produced by the IMPRINT Pro model for Areas 1 and 2. The Workload Unit column represents the workload unit value assigned for this specific task in the IMPRINT Pro model. The Workload Units vary between tasks as the unit values are assigned to the Walking tasks based on the other tasks the human is doing. For instance, when walking to the bulletin boards, the human may have been looking at the bulletin board trying to ascertain what was on it, in addition to walking. The Script Time column represents the time calculated for the human to move to the next location, which is based on an estimate of the distance in the grid-based environment. The Model Time column represents the modeled time in IMPRINT Pro. The Scripted and Model Workload columns represent the total workload for each task, as

Table 2. Walk workload for Areas 1 and 2

Task Name	**Work-load Unit (w/s)**	**Scripted Time (s)**	**Model Time (s)**	**Scripted Workload (w)**	**Model Workload (w)**
Walk to bulletin board 1	12.73	1.86	2.28	23.70	29.02
Walk to bulletin board 2	12.73	1.74	3.32	22.18	42.26
Walk to bulletin board 3	12.73	1.89	0.76	24.06	9.674
Walk to bulletin board 4	12.73	1.70	1.24	21.65	15.79
Walk to recycling bin 1	11.0	1.71	3.61	18.83	39.71
Walk to backpack	13.88	3.05	2.85	42.38	39.56
Walk to recycling bin 2	15.25	1.21	1.52	18.46	23.18
Walk to white board	15.25	4.25	2.66	64.77	40.57
Walk to box	14.11	1.97	1.9	27.84	26.81
Walk to book box	16.11	2.88	4.18	46.38	67.34
Walk to WL2	15.25	0.95	1.9	14.41	28.98
SUM		**23.21**	**26.22**	**324.66**	**362.90**
AVG				**13.99**	**13.84**

estimated in the simulation and IMPRINT Pro, respectively.

Obviously, the time spent on each Walk task does not match the IMPRINT Pro model exactly. This imprecision is caused by the fact that the distance is estimated based on a one foot grid size, which causes several location and distance errors. Notice, however, that when aggregated, the average walking workload in the scripted simulation is within two percent of the overall model workload. We believe this result is sufficiently close to support the investigation of using human performance moderator functions and the Chazm runtime model in human-robot team systems to predict human performance and to assign appropriate humans or robots to team roles in order to increase overall team performance.

3.4 Future Work

In the next phase of our work, we will (1) add an *autonomous* (as opposed to scripted) mode to both the human and robot agents in the CROS simulator and (2) use runtime models to assign tasks to the agents. First we will use the OMACS runtime model, which will enable the team to assign tasks based only on the agent's basic capabilities. Next we will use the Chazm runtime model, which will allow the team to assign tasks based on the basic capabilities of the agent's as well as the human's current performance attribute values. While using Chazm, we expect that when the human's workload is too high, some tasks normally assigned to the human when using OMACS will be assigned to the robot instead. We will determine the effect of including the human's performance in the assignment process by measuring overall team performance

Next, we will extend our experiments by allowing the robot to monitor the human's performance using the Chazm runtime model. Based on Chazm human performance attribute values, we will enable the robot to autonomously adapt its behavior by either assuming the human's tasks or modifying how it executes its own tasks. Again, we will measure the effect of these changes on overall team performance.

In future work, we plan to investigate allowing the robot to adapt its behavior to better support the human when human performance attribute values exceed appropriate levels. First, we will investigate the use of adaptive automation/adjustable autonomy to allow the robot to take over tasks assigned to the human. We expect the robot will take over lower priority tasks when the human's workload reaches some threshold in order to improve overall team performance.

Finally, we plan to investigate allowing the robot to adapt the human-robot interaction in an attempt to reduce human workload even while allowing the human to continue performing high priority tasks or task only the human can perform. This type of adaptation requires monitoring the human in order to capture current human performance for comparison to the predicted human performance. Our current human performance results provide insights into how human performance can be monitored, but the future work will require the investigation on various monitoring capabilities. For example, wearable heart-rate monitors are easily displaced during periods of high activity, thus it will be necessary to devise other monitoring capabilities. We expect this form of human-robot interaction adaption, when coupled with adaptive automation/adjustable autonomy, will further improve team performance.

3.5 Related Research

Human performance modeling has traditionally focused on developing theories of human performance and predicting how humans will interact with systems in order to evaluate the system or inform system design [3]. A number of human performance modeling tools have been developed [1, 2, 10, 14, 16, 32, 33]. These human performance modeling tools have been applied to a broad set of domains including aviation, military, and nuclear power plants [3, 14, 26] and some work has focused on predicting robot operator capacity [7, 8, 9]; however, to date little research has focused on applying performance modeling and human performance moderator functions to peer-based human-robot interaction. Howard [20] focused on employing ACT-R to model and predict human performance for repetitive collaborative tasks in order to predict workload and allocate tasks between a human and a robot. The developed model was validated by requiring the task to be completed by a human teleoperating the robot and by an autonomous robot. Harriott et al. have focused on modeling and validating workload and reaction time for peer-based human-robot teams [17, 18, 19].

A natural model for human-robot teams is multiagent systems, which focus on adapting to dynamic environments, agent failure, and lack of global knowledge. While multiagent systems models are often based on human organizational concepts (e.g., [12]), few explicitly support humans as agents or support adaptation to changes in agent capabilities. Humans are usually viewed as external to the multiagent systems or are only considered in preset roles, such as supervisors [23]. When included in multiagent systems, most pair humans with proxy agents [6] or only consider humans as they relate to user interfaces [24]. Notable attempts at integrating humans into multiagent systems include a NASA space crew/spacecraft management system [28], some initial work in assigning tasks in human-robot teams based solely on workload [20], and DeLoach's previous work extending OMACS. OMACS supports robots teams in reorganizing in response to dynamic environments and changes in robot capabilities [11]; key entities include the team's goals, roles, agents, and capabilities. The Chazm runtime model presented in this paper extends OMACS with agent attributes to capture human performance factors, such as workload and fatigue.

3.6 Discussion

The overall goal of our research program is to develop a framework for human-robot peer teams that allow robots to support their human teammates while working to achieve shared tasks. Unfortunately, as human performance degrades over time, the human's ability to complete assigned tasks can also degrade. Thus, we plan to give the robot insight into this human performance degradation to allow the robot to respond appropriately to its partner as the human's performance degrades. Our first step is to develop a framework that allows tasks to be allocated and reallocated as appropriate based on team member capabilities and performance degradations. Next, we plan to extend the framework to support adapting the interactions between the human and robot during shared tasks to better support the human, As we move toward our goal, this paper presents potentially groundbreaking research in the field of human-robot teams. First, it is one of the first known validations of human performance moderator functions in the domain of human robot teams. An existing workload human performance moderator function was modeled and analyzed in order to determine its applicability to human-robot teams. While the results showed some difference for Area 2, the model was a good predictor of workload in Area 1, but overestimated workload for Area 2. While the workload human performance factor may require some additional work, what was clear was that workload is the human-human team was higher than that of the human-robot team. We have hypothesized that the robot's slower movement and speech are the main cause of this result.

This research also lays a solid foundation for demonstrating the use of human performance moderator functions in human-robot teams. We demonstrated that we can capture human performance moderator functions in the Chazm runtime model and compute attribute values that show a human's performance. We also validated that the values produced consistent with those from the IMPRINT Pro model. Making information related to human performance available to the human-robot team will allow the team to make more informed task (re)allocation decisions and will allow the robot to modify its behavior in response to changes in the human's performance. In addition, the use of runtime models will enable human performance moderator functions to be easily reused in new human-robot team applications. Runtime models also support the development of new methods and techniques to help designers integrate new human performance moderator functions into new and existing applications.

It is well known that human performance not only varies dramatically, but can degrade over the course of prolonged activity. Human teammates often adapt the tasks and interactions to accommodate teammates whose performance ebbs and flows. The ability of the human teammates to adjust their interactions and task responsibilities is often critical to the completion of the assigned task or mission. As robot technology improves, and the research platforms move closer to deployment with their human partners, it is clear that the teammates should be initially assigned responsibilities based upon their current and predicted performance capabilities and that the robot needs to be able to adjust and adapt to the human over time. The presented human performance modeling and validation activities allow for the establishment of human performance metrics appropriate for human-robot peer-based teaming. The extension of OMACS to Chazm in order to account for human performance lays the foundation for the autonomous delegation, and reallocation of tasks between the human and robot teammates based on the human performance capabilities and changes. The ability of the robot to adapt and ensure successful task completion, either by taking on tasks from the human or by modifying the interaction with the human are important components of the resulting systems.

4 Acknowledgements

This research has been supported by AFOSR award FA9550-09-1-0108.

5 References

1. Allender, L., Kelley, T. D., Salvi, L., Lockett, J., Headley, D. B., Promisel, D., Mitchell, D., Richer, C., and Feng, T. (1995). Verification, validation, and accreditation of a soldier-system modeling tool. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, pp. 1219-1223.
2. Archer, S., Gosakan, M., Shorter, P., and Lockett III, J. F., “New capabilities of the army's maintenance manpower modeling tool,” *Journal of the International Test and Evaluation Association*, 26, 1 (2005), 19-26.
3. Baron, S., Kruser, D., and Messick Huey B. (Eds.). (1990). *Quantitative Modeling of Human Performance in Complex, Dynamic Systems*. Washington, DC: National Academy Press.
4. Bonasso, P., Kortenkamp D., & Thronesbery, C. (2003). Intelligent control of a Water-Recovery System: Three years in the trenches. *AI Magazine*, 24(1): 19-44.
5. Bradshaw, J. M., Feltovich, P. J., Jung, H., Kulkarni, S., Taysom, W. & Uszok, A. (2004). Dimensions of Adjustable Autonomy and Mixed-Initiative Interaction. *Autonomy*, M. Nickles, M. Rovatsos & G. Weiss (Eds.), LNAI 2969, pp. 17–39.
6. Chalupsky, H., Gil, Y., Knoblock, C.A., Lerman, K., Oh, J., Pynadath, D. V., Russ, T. A., & Tambe, M. (2002). Electric Elves: Agent technology for supporting human organizations. *AI Magazine,* 23(2): 11-24.
7. Crandall, J. W. and Cummings, M. (2007). Identifying predictive metrics for supervisory control of multiple robots. *IEEE Transactions on Robotics*, 23(5): 1-10.
8. Crandall, J. W., Goodrich, M. A., Olsen, Jr., D. R., and Nielsen, C. W. (2005). Validating human-robot interaction schemes in multitasking environments. *IEEE Transactions on Systems, Man and Cybernetics – Part A,* 35(4): 438-449.
9. Cummings, M. & Mitchell, P. J. (2008). Predicting controller capacity in supervisory control of multiple UAVs. *IEEE Transactions on Systems, Man and Cybernetics – Part A.* 38(2): 451-460.
10. Dahn, D. and Belyavin, A. (1997). The Integrated Performance Modeling Environment A tool for simulating human-system performance. *Proceedings of the 41st Annual Human Factors and Ergonomics Society Meeting*, pp 1037-1041.
11. DeLoach, S. A., Oyenan, W., & Matson, E. T. (2008). A capabilities based model for artificial organizations. *Journal of Autonomous Agents and Multiagent Systems*, 16(1): 13-56.
12. Dignum, V., Vázquez–Salceda, J., and Dignum, F. (2004). Omni: Introducing social structure, norms and ontologies into agent organizations. *Proceedings of the Second International Workshop on Programming Multi-Agent Systems*, LNCS 3346, pp. 181–198, Berlin: Springer.
13. Fowles-Winkler, A. M. (2003). Modelling with the integrated performance modelling environment (IPME), *Proceedings of the 15th European Simulation Symposium*, A. Verbraeck and V. Hlupic (Eds.).
14. Foyle, D. C. and Hooey, B. L. (Eds.) (2008). *Human Performance Modeling in Aviation*. Boca Raton, FL: CRC Press.
15. Goodrich, M.A., and Schultz, A.C. 2007. Human-Robot Interaction: A Survey, in *Foundations and Trends in Human-Computer Interaction*. 1, 3, 203-275.
16. Gore, B.F., and Jarvis, P.A. (2005). New integrated modeling capabilities: MIDAS, recent behavioral enhancements. *Eighth Proceeding of the Annual SAE International Conference and Exposition - Digital Human Modeling for Design and Engineering*, SAE Paper #2005-01-2701.

17. Harriott, C. E., Zhang, T. & Adams, J. A. (2011). Evaluating the applicability of current models of workload to peer-based human-robot teams. *Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 45-52.
18. Harriott, C. E., Zhang, T. & Adams, J. A. (2012). Assessing Workload in Human-Robot Peer-Based Teams. *Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction*.
19. Harriott, C. E., Zhang, T. & Adams, J.A. (2011). Predicting and validating workload in human-robot teams. *Proceedings of the 20th Conference on Behavior Representation in Modeling and Simulation*, pp. 162-169.
20. Howard, A. M. (2007). A systematic approach to predict performance of human-automation systems. *IEEE Transactions on Systems, Man and Cybernetics – Part C*, 37(4): 594-601.
21. Kaber, D. B., Riley, J. M., Tan, K.-W. & Endsley, M. R. (2001). On the Design of Adaptive Automation for Complex Systems. *International Journal of Cognitive Ergonomics*, 5(1): 37-57.
22. Katzenbach, J.R., and Smith, D. K. Best of HBR 1993-The discipline of teams. *Harvard Business Review*. (July-Aug. 2005), 1-10.
23. Kostuik, K. & Vassileva, J. (1999). Free market control for a multi-agent based peer help environment. *Proceedings of the Workshop on Agents for Electronic Commerce and Managing the Internet-Enabled Supply Chain*.
24. Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM*, 37(7): 30-40.
25. Mahoney, P. F. Businesses and bombs: Preplanning and response. *Facilities.* 12, 10 (1994), 14-21.
26. Pew, R. W. and Mavor, A. S. (Eds.) (1998). *Modeling Human and Organizational Behavior: Applications to Military Simulations*. Washington, DC: National Academies Press.
27. Scholtz, J. 2003. Theory and Evolution of Human Robot Interactions, in *Proc. of IEEE 36th Int. Conf. on System Sciences*, 5, 125 - 134.
28. Schreckenghost, D., Martin, C., Milam, T., and Bonasso, R. P. (2004). Modeling humans for coordinating distributed human-agent teams in space operations. Agent Tracking Workshop at the International Joint Conference on Autonomous Agents and Multi-Agent Systems.
29. Silverman, B. G. (2004). Toward realism in human performance simulation. *The science and simulation of human performance*, J. W. Ness, D. R Ritzer, and V. Tepe (Eds). New York: Elsevier, pp. 469-498.
30. Silverman, B. G., Johns, M., Cornwell, J., & O'Brien, K. (2006). Human behavior models for agents in simulators and games: Part I: Enabling science with PMFServ. Presence: Teleoperators and Virtual Environments, 15(2): 139-162.
31. Silverman, B.G., Johns, M., Shin, H., and Weaver, R. 2002. Performance Moderator Functions for Behavior Modeling in Military Situations. *Human Behavior Program–Defense Modeling Simulation Office*.
32. Tyler, S., Neukom, C. Logan, M. and Shively, J. (1998). The MIDAS human performance model. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, pp. 320-325.
33. Zachary, W. Ryder, J. Stokes, J. Glenn, F. Le Mentec, J-C and Santarelli T. (2005). A COGNET/IGEN cognitive model that mimics human performance and learning in a simulated work environment. *Modeling Human Behavior with Integrated Cognitive Architectures*, K. A. Gluck and R. W. Pew (Eds.), pp. 113 – 176.