Predicting and Validating Workload in Human-Robot Teams

Caroline E. Harriott
Tao Zhang
Julie A. Adams
Vanderbilt University
400 24th Ave. South
Nashville, TN 37212
615-322-8481
caroline.e.harriott@vanderbilt.edu, tao.zhang@vanderbilt.edu, julie.a.adams@vanderbilt.edu

Keywords:
Human Performance Modeling, Human-Robot Interaction, IMPRINT Pro

ABSTRACT: Human Performance Moderator Functions (HPMFs) are equations derived from empirical results to predict human performance in specific conditions. These functions have been proven to apply to humans working with other humans, but have not yet been analyzed for application to human-robot peer-based teams. This paper presents research to understand if current workload HPMFs apply to human-robot teams. Specifically two human performance models of workload were created in IMPRINT Pro – one with a human-human team and one with a human-robot team – performing the same set of triage tasks. These models were validated in an experiment. The empirical workload results were compared to workload values predicted by the models. Results indicate that the model was able to predict human performance for both types of teams.

1. Introduction

Individual human performance can impact the overall performance of human teams (Katzenbach & Smith, 2005). Similarly, when humans are partnered with robots for peer-based tasks, human performance will impact the task performance of human-robot teams (HRTs) (Goodrich & Schultz, 2007). It will become necessary for the robotic team members to predict human performance capabilities in order to adapt their behavior, as humans do in human teams. This adaptation is necessary to mitigate and accommodate changes in the performance of the human partners. It is necessary to understand if and how existing human performance moderator functions apply to human-robot teams. Developing such understanding necessitates modeling human performance moderator functions (HPMFs) for such teams, conducting evaluations to gather the associated empirical results, and understanding how the empirical results relate to the modeled performance moderator functions.

HPMFs are equations derived from empirical results that predict human performance due to specific performance factors such as fatigue, mental workload or temperature. Over 500 HPMFs (Silverman, Johns, Cornwell & O'Brien, 2006) are known to exist. Our current research is focused on understanding workload. HPMFs have been evaluated for various domains (Gawron, 2008; Mumaw, Roth, Vicente & Burns, 2000), but have not been proven to apply to human performance in human-robot teams (HRTs). While the human-robot interaction community has focused on measuring human performance, little research has focused on integrating HPMFs. Methods exist for modeling human performance; however, it is necessary to examine whether the HPMFs used in modeling human performance are applicable to HRTs. Our solution was to model workload using IMPRINT Pro (Archer, Gosakan, Shorter & Lockett, 2005) and empirically validate the model results. Our initial research focused on modeling, evaluating and analyzing workload for human-human team tasks to human performance in peer-based human-robot teams. The reported research focused on teaming a human with a partner, human or robot, to complete a task. Both human-human and human-robot teams were modeled and evaluated. The intent was to validate that the HPMFs for the humans are accurate before analyzing the human-robot team. It should be noted that in the reported research, the partner (human or robot) instructed the primary human on task steps to complete. There are no shared decision making tasks.

This paper will discuss relevant background information regarding the assigned tasks, describe the IMPRINT Pro models of workload, detail the evaluation, and discuss the comparison of the predicted workload values to the empirical results.

2. Background

The presented research focused on a victim triage task for a mass casualty Chemical, Biological, Radiological, Nuclear, and Explosive (CBRNE) incident. CBRNE incident response procedures dictate that first responders are not to enter the contaminated incident scene until a decontamination site is established, the potential hazards are identified, and personal protective equipment is donned. However, any response delay can result in significant damage and civilian injury or death. Robots have the potential to enter the scene immediately in order to provide immediate feedback regarding contaminants, victim locations, injuries, etc. Such information can assist human responders in determining the appropriate response - including locating, treating, and transporting victims (Humphrey & Adams, 2009). While robotic technology is not yet capable of all these tasks, robots should be able to enter a scene, identify and recruit an ambulatory victim, with no or minor injuries, to assist with victim triage, hazard sampling, general search and hazard reconnaissance tasks. Such human volunteers can assist with locating and assisting victims and gathering information about potential hazards. The robot can relay provided information to first responders located outside the contaminated incident area. The employed triage scenario is representative of current robotics capabilities; however,. the robot's speech and movement capabilities are not equal to a human's and therefore make the interaction more cumbersome.

The START (Benson, Koenig & Schultz, 1996) triage system is commonly employed to triage victims during emergency incidents. Generally, the START steps require 60 seconds and focus on assessing the immediacy of care required for a particular victim. The START system steps the responder through a number of decision-tree style steps intended to classify a victim into one of four triage levels: Minor, Delayed, Immediate and Deceased/Expectant.

The presented research focused on analyzing workload for a CBRNE incident situation in which multiple non-ambulatory victims are triaged. An ambulatory, uninjured victim with no, or minimal first aid training, and no prior robotics experience forms an ad-hoc team with either a human first responder located outside of the contaminated area or the robot.

3. Human Performance Modeling

Human performance modeling simulates human behavior under various conditions and tasks. The models require inputs related to human performance and result in the likely actions. HPMFs can be incorporated into Human Performance Models (HPMs) in order to improve the model fidelity (Card, Moran & Newell, 1980). A large number of domains have incorporated HPMs in order to understand how system design, task assignments and environmental changes can impact human behavior and performance. For example, in the aviation field, the NASA Human Performance Modeling Project (Foyle and Hooey, 2008) involved using multiple modeling techniques on a common set of problems to investigate different aspects of human performance in aviation tasks. The effect of a secondary task on human performance while driving a car was analyzed (Salvucci, 2001) using ACT-R (Anderson & Lebiere, 1998).

PMFServ represents human decision-making processes based on a subjective utility ranking (Silverman et. al, 2006). PMFServ incorporates the effects of HPMFs to affect each agent's decision-making process. PMFServ has been used to model human behavior in many scenarios including hostile civilians in an urban military scene (van Lent et al., 2004).

IMPRINT Pro is a task network modeling tool intended to assess human and system performance in military missions. IMPRINT Pro has been used to model personnel on a United States Navy destroyer bridge, the Land Warrior System and the U.S. Army's Crusader System (Allender, 2000) all for improvements to existing U.S. military systems. IMPRINT Pro has also been used to model pilot performance while flying simulated unmanned air vehicles (Wickens, Dixon & Chang, 2003).

IMPRINT Pro permits the simulation of human behavior for various conditions via the representation of task and event networks. IMPRINT Pro includes a number of pre-defined HPMFs (e.g., workload) and permits the incorporation of other HPMFs via the User Stressors module. IMPRINT Pro has been employed in the reported research to model both the human-human and human-robot conditions.

3.1 The Model

The two conditions modeled represent a team-based scenario involving first responders (both robot and human) instructing an ambulatory, uninjured victim who is located in the contaminated incident area to perform triage on nearby non-ambulatory victims. The models are specific to the human-human condition and the human-robot condition. The models represent the task activities and the uninjured volunteer's workload. IMPRINT Pro's workload HPMF is divided into seven channels (Cognitive, Auditory, Visual, Fine Motor, Gross Motor, Speech, Tactile), with assignment guidelines provided by the IMPRINT Pro documentation (IMPRINT Pro User Guide, 2009).

The triage scenario requires the uninjured victim to perform the START triage steps on six victims with differing levels of required triage in Round 1 and repeat the triage steps on all victims who were not classified as Expired during the initial triage (five victims) in Round 2. Each of the six victims has an assigned triage level and a list of required steps to determine the triage level. The triage levels of two victims may be the same, but the sequence of steps required to reach this decision can vary depending on the initial tests of breathing and pulse rates. The modeled human-human scenario assumes that the uninjured victim has contacted 911 to report the incident and has volunteered to assist a remote (e.g., located outside of the contaminated incident area) first responder with the triage task. The scenario further assumes that the uninjured victim communicates with the remote first responder via cell phone and that the remote first responder provides step-by-step instructions that lead the uninjured victim through the triage steps. The uninjured victim provides responses that are recorded by the remote first responder to assist with incident response planning.

The modeled human-robot scenario assumes a robot deployed into a contaminated incident area that has discovered an uninjured human who has agreed to assist the robot with conducting the triage task. The uninjured victim executes the instructions provided by the robot and reports results to the robot. The robot reports this information, as well as the location of the injured victim to remote first responders. The robot communicates with the uninjured responder using voice interaction.

Both scenarios use the same task, which is to perform an initial triage assessment on and classification of the injured victims before conducting a follow-up triage. The victim order, provided triage instructions, and victim information are identical across the conditions except that the human-robot model takes into account the robot's slower speech pace and an extra step of placing a triage card on each victim with a color representing the triage level.

The IMPRINT Pro models iterate through each atomic task for the entire triage scenario. Tasks include each START triage step. For example, participants count the number of breaths a victim takes in one minute which is decomposed into discrete, atomic tasks. They are instructed to watch the victim's chest rise and fall for one minute, while counting the number of breaths and listening for the teammate to say "stop." When stopped, the participant will report the total number of breaths counted.

For each atomic step, the modeler specifies a running time, title and workload values. Activities such as speech, walking, reading and listening use IMPRINT Pro's built in calculator to estimate how long an average person takes to say a specific number of words, walk a certain number of feet, etc. These times are based on empirical data sets. Each atomic task associated with the uninjured victim has an associated workload value. A value is assigned to each of the seven channels, which results in an overall workload value associated with a particular atomic task. Each channel has an independent value scale and predefined guidelines for choosing an associated value. IMPRINT Pro provides guidelines for assigning these values. When the model executes, assigned workload values for each task are in effect during the entire execution for each atomic task.

Our current research focuses on the results for the six injured victim triage levels in two rounds – 11 time periods of measurement. Once the model completes execution, the model outputs the list of tasks completed by the uninjured victim when triaging the injured victims. Along with each atomic task, the results include the time required to complete the task and the associated workload value for each workload channel and an overall workload value. Figure 1 displays the total time taken to complete triage tasks for each victim by each of the models, human-human (H-H) and human-robot (H-R).

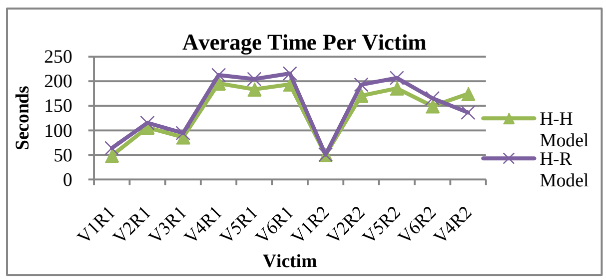

Figure 1. Predicted triage time by victim for each model.

The model output includes a graph of the total workload over the entire scenario and a breakdown of each action at each moment for every human and robot modeled. Model workload for each victim assessment is calculated via a weighted average of the workload by the amount of time spent on each atomic task. Figure 2 displays the total workload output for the H-H team scenario, while Figure 3 provides the total workload for the H-R team scenario, both using the same timescale. During each victim assessment, the workload changes correspond to the individual task demands and the black boxes indicate the time periods when the volunteer triaged the victim labeled at the top of the box. The number after the "V" indicates the victim assessed and the number after the "R" indicates the round, for example V1R1 Victim 1 during Round 1.

It can be seen that overall, the human-robot team took a longer time to complete all tasks. Both teams experienced very similar trends in workload with peaks

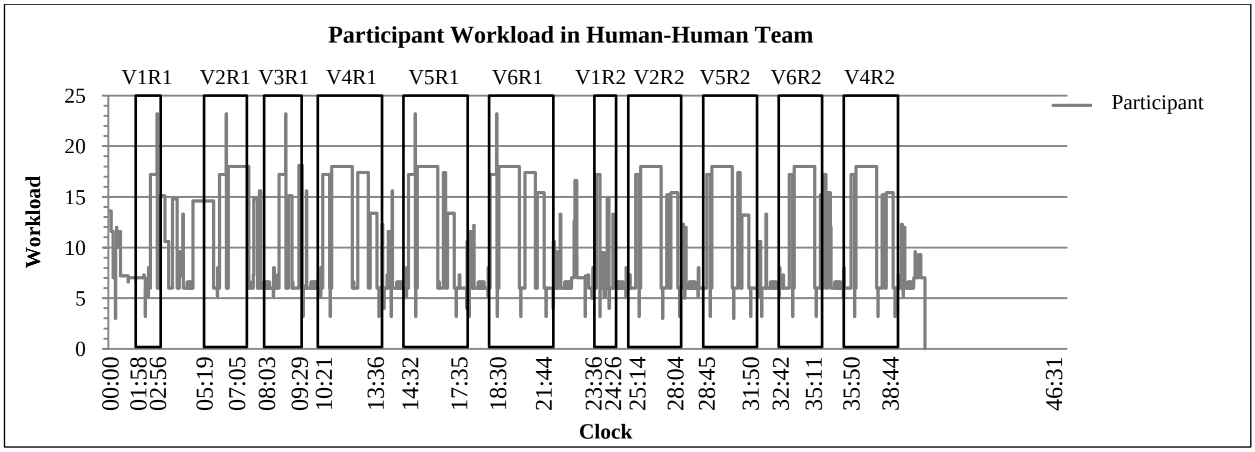

Figure 2. Total workload output of Human-Human triage model.

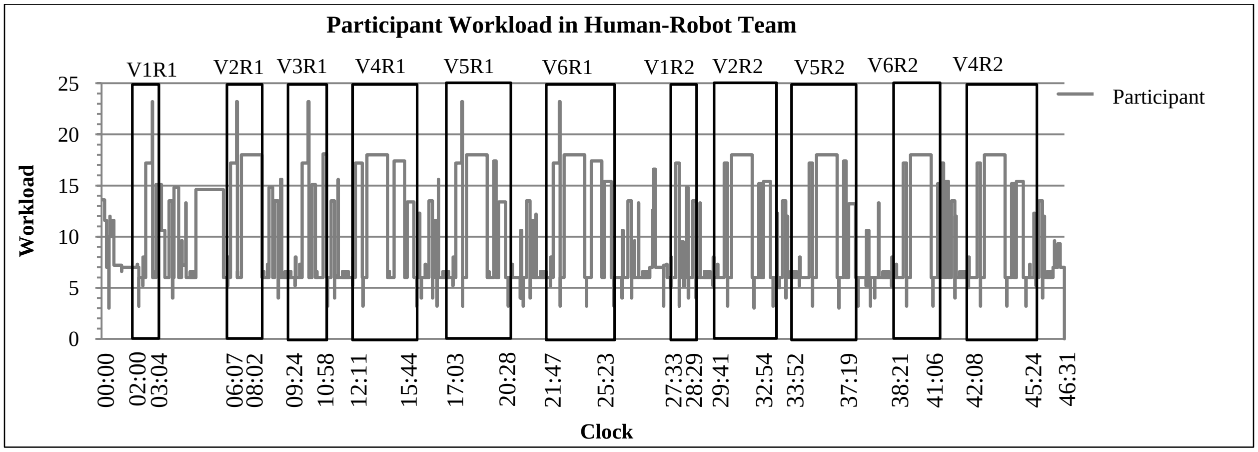

Figure 3. Total workload output of Human-Robot team triage model.

and valleys in similar spots within each victim assessment period, but the overall workload may be different due to the time weight for each task. Figure 4 shows the overall total workload score for each victim for the two models. Figure 5 displays the workload values from each individual workload channel for the H-H model, while Figure 6 provides the individual workload channel values for the H-R model.

The models predict changing workload values based on the specific victim assessed. The models also predict that total workload for both conditions will follow the same general trend independent of teaming partner, either human or robot. The human-robot team was expected to exhibit lower overall workload based upon the modeling results (Figure 4).

4. Experimental Validation of the Models

An empirical validation of the model results for both teams conducting the victim triage task is required. The modeled workload results were validated with workload metrics provided humans completing the modeled tasks. The evaluation was intended to understand the workload differences between the

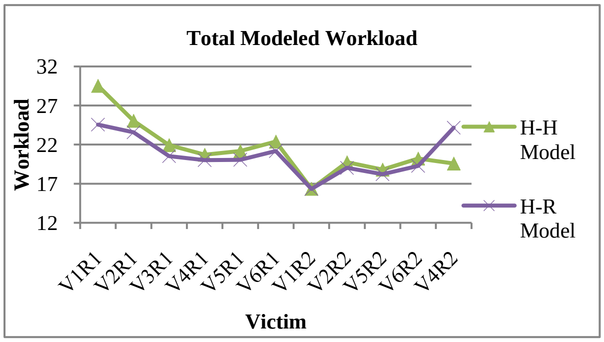

Figure 4. Total workload from the H-H model and H-R model.

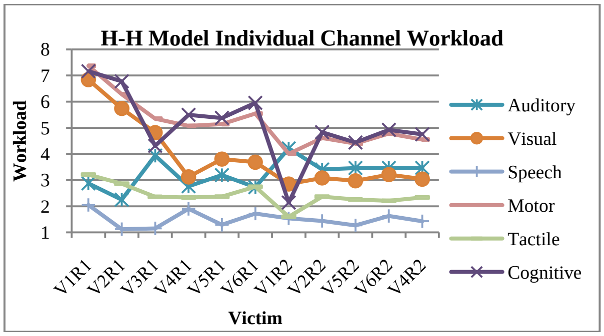

Figure 5. Individual workload channel values for H-H model.

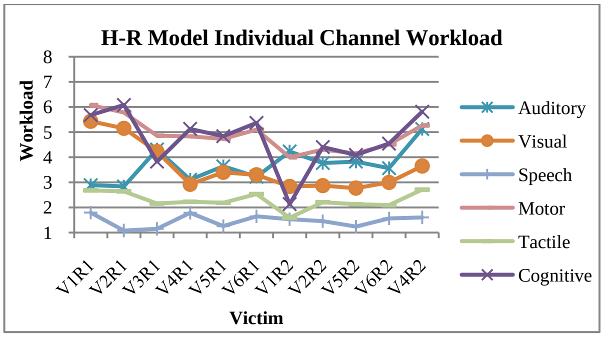

Figure 6. Individual workload channel values for H-R model.

conditions and the predictability of the models for actual human workload. Additional experimental validation results are available in (Harriott, Zhang & Adams, 2010; Harriott, Zhang & Adams, In Press).

4.1 Experimental Design

The experimental design was a split-plot design with the participants as a random element. The experimental condition differed between-subjects (human-human vs. human-robot) and the within-subjects element included the series of triage assessment tasks that each participant completed. The independent variables were the experimental condition, the victim triage levels, the triage round (e.g., either the initial triage – Round 1, or the follow-up triage – Round 2) and the participant age, gender, experience with robots and first aid training. The dependent variables include both subjective and objective measures of workload.

4.2 Participants

Twenty-eight participants completed the evaluation, fourteen in each condition. The participants were evenly split, with fourteen males and fourteen females. The participants were nearly evenly split by gender across the two conditions, with six males and eight females in the H-H condition and eight male and six female participants completing the H-R condition. The average age of all participants was 25.2, with age ranging between 18 and 57 years.

4.3 Evaluation Metrics

The evaluation metrics included physiological data from the heart rate monitor (heart rate variability, breathing rate, R-R (rest-rest) interval data, heart rate, respiration rate, skin temperature, posture, vector magnitude data and acceleration data), time spent on each victim, secondary task question correctness, and accuracy of triage assessments on the victims' breathing rates and ages. The subjective metrics included the subjective workload ratings, the post-trial questionnaire responses and the NASA-TLX (Hart & Staveland, 1988) workload questionnaire responses. All of the physiological data was collected via a portable BioHarness ECG monitor[1]. For the purposes of this paper, we only present the model data in comparison to the time spent on each victim and the subjective in-task workload ratings.

The time spent on each victim was determined by recording a start time for each victim as soon as the participant indicated that he or she had reached a victim. The ending time was determined by noting when the participant finished the last triage task, but before the responder (human or robot) announced the triage level of the victim.

After completing the triage steps for each victim, the participant's teammate asked the participants to rank six workload channels on a scale from 1 to 5 where 1 represented little to no demand and 5 represented extreme demand. Each channel was defined during the first round of questions after the participant triaged the first victim and the participants were informed that they could ask for clarification of questions at any time. The six workload channels were Cognitive, Auditory, Visual, Tactile, Motor and Speech. The questions were adapted from the Multiple Resources Questionnaire (Boles, Bursk, Phillips & Perdelwitz, 2007) and the channels were chosen so as to be easily compared to IMPRINT Pro's seven workload channels. In order to prevent confusion regarding the fine and gross motor channels, the participants were asked to rate a single Motor channel. When compared to the IMPRINT Pro predicted values, the Fine and Gross motor channels were summed.

Each IMPRINT Pro workload channel has a specified scale, starting from 0, as minimum workload and ranged from 4 to 7 as the highest workload. The first step in converting the two measures of workload (from the IMPRINT Pro model and the subjective results) to a unified scale added 1 to all IMPRINT Pro workload values. Depending on the specific channel's scale, the subjective workload ratings for that scale were multiplied by a fraction with the highest number on the IMPRINT Pro channel's scale on top and 5 on the bottom. Thus, the workload values are re-scaled for easy comparison. The conversion process for the IMPRINT Pro ratings is shown below where: CMR stands for Comparable Model Rating and IWR stands for a victim's overall IMPRINT Pro workload rating.

$$CMR = IWR + 1$$

The conversion equation for the in-task subjective workload ratings is shown below. CSR stands for

[1] www.biopac.com

Comparable Subjective Rating and HCV stands for the highest channel value possible in IMPRINT Pro. Both equations are performed for each individual workload channel and the total workload ratings.

$$CSR = Rating * (HCV / 5)$$

4.4 Experimental Environment

The evaluation occurred in the Center for Experiential Learning and Assessment at Vanderbilt University's School of Medicine[2]. The room lights were dimmed and a background sound track that included explosions, street noise, people coughing and screaming, construction noise and sirens was played during the experiment. Six medical mannequins were dressed as civilians. Victim 1 was an infant that emitted crying noises. Victim 2 was an adult female, with a pulse and breathing rate adjusted via a computer. Victim 3 was a young boy without any dynamic features. Victim 4 was an adult male with a breathing rate, pulse rate and speech capabilities. Victim 5 was a toddler with breathing and pulse rate. Victim 6 was an adult male with a breathing rate, pulse rate and eyes that blinked. A Pioneer 3-DX robot was the robot partner. The robot navigated the room autonomously, spoke to the participant with a digital voice through speakers mounted to the robot, and the participant spoke to the robot via a wireless microphone.

4.5 Procedure

The participant was briefed that he or she was playing the role of an uninjured, ambulatory and "contaminated" victim that is unable to leave the contaminated incident area until responders had set up the decontamination area. The general instructions provided to the participants in each condition were identical except for the explanation of how the interaction would occur and what led to the interaction between the participant and the human or the robot.

Independent of the evaluation condition, the participant's teammate guided the participant through the steps to identify a victim's triage level. The participants in both conditions started in the same position in the room and moved from victim to victim during the initial triage (Round 1). After completing the initial triage of all six victims, the participant was led back to the five surviving victims for a second triage check, creating eleven sets of triage steps. During the Human-Human condition, the next victim to visit during the second round was always specified by referring to the order in which they were visited during the first round. The robot led the participant back to the specific victim during the human-robot condition. Once the participant reached a victim during the second round, the teammate provided a short summary of the conclusions drawn from the first triage visit. The participant then performed the triage assessment again to update the results.

[2] www.mc.vanderbilt.edu/medschool/cela

5. Results

The validation included comparisons between the participant's subjective workload ratings for each victim and the average workload for each victim predicted by the models.

5.1 Total Workload

Total workload was compared between the model results and empirical results. Figure 7 provides a comparison of the H-H subjective data and H-H model. The calculated difference between the modeled workload values and the in-task subjective workload values demonstrate how effective the models were at predicting human behavior. The mean delta between the H-H subjective values and the model results at each time point was 0.88 (St. Dev. = 3.84). The mean absolute value of the difference between the model and actual data was 3.23 (St. Dev. = 2.03).

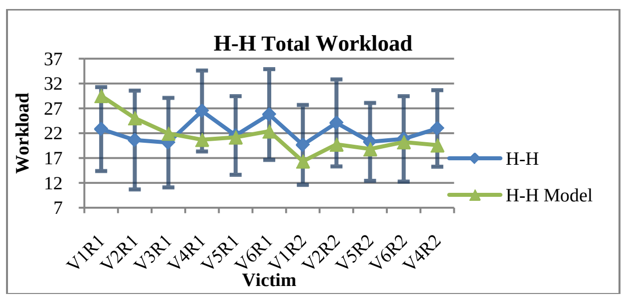

Figure 7. Total rescaled workload ratings for the H-H evaluation, versus the H-H model's total workload value. Error bars represent one standard deviation above and below the mean.

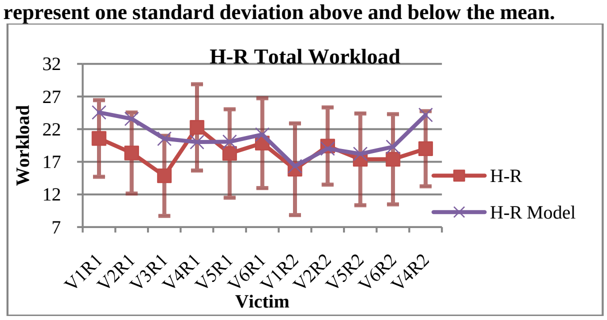

Figure 8. Total rescaled workload ratings for the H-R evaluation versus the H-R model's total workload value.

A comparison of the H-R validation Total workload and the IMPRINT Pro workload prediction for the H-R model is provided in Figure 8. The mean delta between the H-R condition and the model was -2.15 (St. Dev = 2.57). The mean absolute value of the difference between the data sets was 2.64 (St. Dev. = 2.01).

These results imply that the H-R model was a slightly better predictor of H-R workload than the H-H model.

5.2 Individual channels

The difference between the model and the empirical results was compared for each individual workload channel. A smaller average difference indicated that the model more closely predicted human performance. The H-R IMPRINT Pro model more closely predicted the actual Auditory, Visual, Speech and Tactile workload data. The two models produced virtually the same accuracy in predicting Cognitive workload.

IMPRINT Pro has two motor channels, fine and gross, that were combined in the evaluation. In order to compare this dual-motor channel to IMPRINT Pro's motor channels, all values were scaled to the fine motor channel scale from IMPRINT Pro and the two IMPRINT motor channels were then added together. The H-H IMPRINT Pro model more closely predicted the actual Motor subjective workload data.

5.3 Time Spent per Victim

The time spent triaging each victim for each condition was compared to the models. Figure 9 compares the H-H experiment timing data to the H-H model, which is within one standard deviation of the empirical data for each point except for V1R1 (1.14 seconds (s)), V1R2 (3.53 s) and V5R2 (1.10 s). The mean difference between the H-H empirical data and the model is -2.32 s (St. Dev. = 18.17). Some model data points are above the empirical data, therefore, the mean difference when taking the absolute value of the difference at each point is 13.30 s (St. Dev. = 11.89).

Figure 10 depicts the H-R timing data gathered from the experiment compared to the model, which is within one standard deviation of the empirical data except for V3R1 = 38.99 s, V1R2 = 16.20 s and V4R2 = 8.08 s. The mean difference between the human-robot empirical data and the model is 0.98 s (St. Dev. = 27.47). Some model data points are above the empirical data, thus the mean difference when taking the absolute value of the difference at each point is 21.95 s (St. Dev. = 15.02).

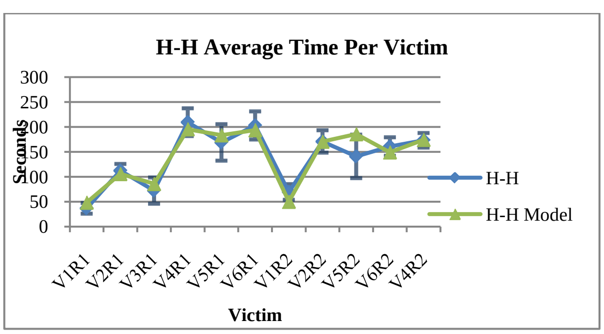

Figure 9. Comparison of average time to assess each victim for the H-H experimental data and predictive model.

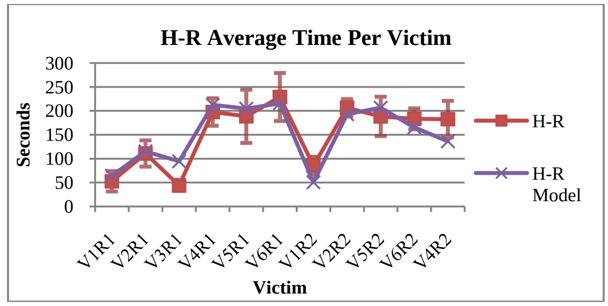

Figure 10. Comparison of average time to assess each victim for the H-R experimental data and predictive model.

While the model is not perfect, the predictions generally align with the subjective workload data. When comparing time spent for victim assessment, the model did a good job of providing a close picture of the required time to assess each victim. The model's predictions are within one standard deviation from the actual mean in 8 of 11 cases for the H-H condition and 8 of 11 cases for the H-R condition. While the number of data points outside one standard deviation is the same, the H-H model predicted the three points closer to the empirical data than the H-R model. The H-H model did a better job of predicting the time required to assess each victim.

6. Discussion

The predicted IMPRINT Pro model workload trends were very similar to the actual in-task subjective workload rating results. The models employed current workload models and were adjusted for the slight differences in task time between the H-H and H-R scenarios. The models predicted a slightly lower workload level for the H-R condition. Time spent assessing each victim was closely predicted by the models as well, with the H-R condition overall requiring more time.

Current workload HPMFs may be applicable to human-robot teams; however, additional analysis of more complex relationships and tasks are required. Overall, the models provide a valuable workload prediction tool for both the H-H and H-R teams.

7. Conclusion

In this study, human-human and human-robot teams were modeled and evaluated performing a series of medical triage tasks. The workload HPMF for both conditions was modeled using IMPRINT Pro. An empirical evaluation assessed workload for the corresponding conditions. The empirical results generally mirror the model results for each condition and the mean differences between the predicted workload values and experimental ratings were relatively small. This research provides initial support for the applicability of a current workload HPMF (for human-only teams) to human-robot peer-based teams. Further research is required incorporating more complex, shared decision making tasks before further generalization of the applicability of existing workload HPMFs to human-robot peer-based teams.

8. Acknowledgements

We thank M. Scheutz, P. Schermerhorn & D. Bender for the DIARC architecture. We thank M. Weinger, B. Immekus, R. Booker, A. Cross, T. Hoffman & S. Hayes. This research is supported by AFOSR award FA9550-09-1-0108, NSF Grant IIS-0643100, & ONR MURI Program award N000140710749.

8. References

Allender, L. (2000). Modeling human performance: impacting system design, performance, and cost. In *Proc. of the Military, Government and Aerospace Simulation Symp., 2000 Advanced Simulation Technologies Conf.* 139-144.

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

Archer, S., Gosakan, M., Shorter, P., & Lockett III, J. F., (2005). New capabilities of the army's maintenance manpower modeling tool, *Jour. of the Inter. Test and Evaluation Association, 26 (1).* 19-26.

Benson, M., Koenig, K.L., & Schultz, C.H. (1996). Disaster triage: START then SAVE – a new method of dynamic triage for victims of a catastrophic earthquake. *Prehospital and Disaster Medicine. 11(2*). 117-24.

Boles, D.B., Bursk, J.H., Phillips, J.B. & Perdelwitz, J.R., (2007). Predicting dual-task performance with the Multiple Resources Questionnaire (MRQ). *Human Factors, 49(1)*, 32–45.

Card, S.K., Moran, & T.P., Newell, A. (1980). Computer text-editing: An information-processing analysis of a routine cognitive skill. *Cognitive Psychology, 12(1).* 32-74.

Foyle, D.C. & Hooey, B.L (2008). *Human Performance Modeling in Aviation*. Boca Raton: CRC Press.

Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: North Holland Press.

Harriott, C.E., Zhang, T., & Adams, J.A. (2010). Applying workload human performance moderator functions to peer-based human-robot teams. Vanderbilt University, Technical Report: HMT-2010-04.

Harriott, C.E., Zhang, T., & Adams, J.A. (In Press). Evaluating the applicability of current models of workload to peer-based human-robot teams. In *Proc. of 6th ACM/IEEE Inter. Conf. on Human-Robot Interaction, 2011*.

Humphrey, C.M. & Adams, J.A. (2009) Robotic tasks for CBRNE incident response. *Advanced Robotics 23* 1217-1232.

Katzenbach, J.R., & Smith, D.K. (2005). The discipline of teams. *Harvard Business Review – Best of HBR 1993.* July-Aug 2005.

Mumaw, R.J., Roth, E.M., Vicente, K.J. & Burns, C.M. (2000). There is more to monitoring a nuclear power plant than meets the eye. I*Human Factors: Human Factors. 42 (1).* 36-55.

Salvucci, D.D. (2001). Predicting the effects of in-car interface use on driver performance. *Human-Computer Studies, 55*. 85-107.

Scholtz, J. Theofanos, M., & Antonishek, B. (2002). Theory and evolution of human robot interactions. In *Proc. of 36th International Conference on System Sciences.*

Silverman, B.G., Johns, M., Cornwell, J. and 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.

van Lent, M., McAlinden, R., Probst, P., Silverman, B.G., & O'Brien, K. (2004). Enhancing the behavioral fidelity of synthetic entities with human behavior models. In *Proc. of 13th Conference on Behavior Representation in Modeling and Simulation (BRIMS).*

Wickens, C.D., Dixon, S., & Chang, D. (2003). Using interface models to predict performance in a multiple-task UAV environment – 2 UAVs. University of Illinois, Urbana-Champagne, Technical Report: AHFD-03-9/MAAD-03-1.

Author Biographies

CAROLINE E. HARRIOTT is a graduate student in the Department of Electrical Engineering and Computer Science at Vanderbilt University.

TAO ZHANG is a post-doctoral research associate in the Department of Electrical Engineering and Computer Science at Vanderbilt University.

JULIE A. ADAMS is an Associate Professor of Computer Science and Computer Engineering in the Department of Electrical Engineering and Computer Science at Vanderbilt University.