A Diagnostic Human Workload Assessment Algorithm for Human-Robot Teams

Jamison Heard
Vanderbilt University
jamison.r.heard@vanderbilt.edu

Rachel Heald
University of Kansas
rachelmary15@ku.edu

Caroline E. Harriott
Vanderbilt University
ceharriott@gmail.com

Julie A. Adams
Oregon State University
julie.a.adams@oregonstate.edu

ABSTRACT

High-stress environments, such as a NASA Control Room, require optimal task performance, as a single mistake may cause monetary loss or the loss of human life. Robots can partner with humans in a collaborative or supervisory paradigm. Such teaming paradigms require the robot to appropriately interact with the human without decreasing either's task performance. Workload is directly correlated with task performance; thus, a robot may use a human's workload state to modify its interactions with the human. A diagnostic workload assessment algorithm that accurately estimates workload using results from two evaluations, one peer-based and one supervisory-based, is presented.

ACM Reference Format:
Jamison Heard, Rachel Heald, Caroline E. Harriott, and Julie A. Adams. 2018. A Diagnostic Human Workload Assessment Algorithm for Human-Robot Teams. In *Proceedings of HRI'18 Companion.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3173386.3176983

1 INTRODUCTION

Human-robot teams may be deployed in environments that require optimal task performance, but high stress environments can place considerable workload on the human, which can decrease task performance [4]. Task performance may be augmented by developing an adaptive teaming system that uses the human's workload state to determine how the robot interacts with the human.

The human's overall workload state can be decomposed into five components: cognitive, physical, auditory, visual, and speech [3]. Decomposing overall workload into components allows for an adaptive teaming system to determine if a robot interaction impacts the human, i.e., overload state, and how the interaction affects the workload state, i.e., will the cognitive component be overloaded. An adaptive teaming system can detect the human's workload state by using a workload assessment algorithm, which relies on objective workload metrics, e.g., heart-rate varability [2]. Workload assessment algorithms rely on machine-learning techniques to classify a subset of the overall workload state; typically, only cognitive workload [2]. Detecting a subset of the overall workload state limits a workload assessment algorithm's viability for use in an adaptive teaming system, as the robot will be unable to determine how an interaction may impact the human. Thus, there is a need for the development of a workload assessment algorithm that can identify the overall workload state and that of each workload component.

HRI'18 Companion, , March 5–8, 2018, Chicago, IL, USA

ACM ISBN 978-1-4503-5615-2/18/03...$15.00
https://doi.org/10.1145/3173386.3176983

2 WORKLOAD ASSESSMENT ALGORITHM

The workload assessment algorithm assesses overall workload and each workload component every thirty seconds using knowledge of the task being completed. IMPRINT Pro workload models [1] of the tasks are used as the desired estimates during training.

The workload assessment algorithm relies on heart-rate (HR), heart-rate variability (HRV), skin-temperature (ST), respiration rate (RR), posture-sway (PS), and posture magnitude (PM) metrics of human workload. Each metric is filtered using adaptive exponential smoothing [5] and fed into a neural network. Each workload metric has a corresponding neural network that produces noisy estimates of a workload component, which are subjected to a weighted aggregation. The result is a workload component value, where the weights are determined by multivariate regression. The resulting equations for cognitive (W_C) and physical workload (W_P) are:

$$W_C = w_{HRV} * HRV + w^c_{ST} * ST + w^c_{HR} * HR + b_{cog}, \tag{1}$$

$$W_P = w_{PS} * PS + w_{PM} * PM + w^p_{ST} * ST + w^p_{RR} * RR + w^p_{HR} * HR + b_{phys}, \tag{2}$$

where w_x represents the weightings for the corresponding metric x and b_y represents a bias term. The visual, speech, and auditory workload component values are determined by an IMPRINT Pro workload model. The overall workload value is the aggregate of the component values and is mapped to a workload state based on thresholds, which are the midpoint between the IMPRINT Pro workload models for each workload condition. For example, if 20 is the threshold value between low and high workload, then any value below 20 is classified as low workload.

3 EXPERIMENTAL DESIGN

Physiological data from two human-robot teaming evaluations, each focused on a specific interaction domain: peer and supervisory, were used to evaluate the algorithm. The workload models for each evaluation were generated using IMPRINT Pro.

Eighteen participants completed the peer-based evaluation, which required completing four fifteen-minute first response tasks with a Pioneer 3-DX robot. Each task elicited low or high workload.

The tasks required searching photographs for suspicious items, e.g., hazardous chemicals, on a Google Nexus 7 tablet; exhaustively searching an academic building hallway for hazardous or suspicious items; and sampling solid and liquid contaminants stored in containers based on instructions audibly given by the robot.

Thirty participants completed a supervisory evaluation, that required participants to monitor and control a simulated remotely piloted aircraft using the NASA Multi-Attribute Task Battery. Participants completed three fifteen-minute trials corresponding to the underload, normal load, or overload workload states. Four concurrent tasks required participants to maintain a target in the middle of the cross-hairs; monitor two lights and four gauges and click on any out-of-range light or gauge; control fuel pumps to maintain the fuel levels of two tanks; and respond to air-traffic control messages.

4 RESULTS

The results are presented for three trained workload assessment algorithms: trained solely on supervisory data (SUP), trained solely on peer data (PEER), and trained on equal proportions of both types of data (BOTH). Each trained algorithm was evaluated using the respective evaluation test data set, which was separate from the training data set. The average classification accuracies by evaluation and algorithm type are presented in Table 1. The algorithm achieves the highest accuracy when trained and tested on the corresponding data set, while classification accuracy decreases when the algorithm is trained on one data set and tested on the other. Specifically, the PEER algorithm is unable to classify the overload state for the supervisory evaluation, which results in a low average classification accuracy. The SUP algorithm achieves a high cognitive and overall workload classification accuracy for the peer evaluation, but achieves a low physical workload classification accuracy. The BOTH algorithm performs exceptionally well for both evaluations, which demonstrates that the algorithm can be trained on multiple data sets with minimal impact on performance.

Table 1: Average Classification Accuracy (%) by Evaluation.

Workload	Training	Evaluation	
		Peer	Supervisory
Cognitive	SUP	88.19	**97.91**
	PEER	**95.78**	63.12
	BOTH	95.21	97.76
Physical	SUP	63.40	**98.51**
	PEER	**90.42**	65.54
	BOTH	81.40	97.95
Overall	SUP	90.44	**100**
	PEER	**95.67**	66.66
	BOTH	95.44	**100**

Bold represents highest accuracy per column

It is important that a workload assessment algorithm be able to track the direction of change in workload. The Pearson's correlation coefficients between the algorithm's estimates and the IMPRINT Pro model's values are presented in Table 2. The highest correlations are achieved when the algorithm is trained and tested on data from the same evaluation. Additionally, each algorithm produced significant correlations on data that the algorithm was not trained on, which illustrates that when workload changes; the algorithm's estimates respond accordingly, regardless of the algorithm's training data set.

Table 2: Correlation Coefficients between the Algorithm Estimates and the Workload Models by Evaluation.

Workload	Training	Evaluation	
		Peer	Supervisory
Cognitive	SUP	0.73*	0.99*
	PEER	0.94*	0.86*
	BOTH	0.74*	0.97*
Physical	SUP	0.74*	0.98*
	PEER	0.92*	0.83*
	BOTH	0.79*	0.98*
Overall	SUP	0.82*	0.99*
	PEER	0.97*	0.98*
	BOTH	0.83*	0.99*

* indicates $p < 0.001$

The presented diagnostic workload assessment algorithm is an essential step to realizing an adaptive teaming system. The classification results demonstrate that the workload assessment algorithm can accurately classify multiple cognitive, physical, and overall workload levels in different human-robot teaming paradigms. Accurately classifying overall workload and its components allows for an adaptive teaming system to identify the distinct components contributing to the overall workload state and potentially why the human is in the current workload state. Identifying the distinct workload components permits targeted adaptations to the components in order to normalize the workload state. An adaption based solely on the overall workload state cannot target a specific workload component; thus, the adaptation may be ineffective.

5 CONCLUSION

The presented diagnostic workload assessment algorithm relies on objective workload metrics to accurately estimate overall workload and each workload component. Data collected from two human-robot teaming evaluations was used to evaluate the algorithm's capabilities, which showed that the algorithm accurately classified and tracked changes in cognitive, physical, and overall workload. The diagnostic workload assessment algorithm is an initial step to realizing an adaptive teaming system that is capable of determining a robot's interactions based on a human's workload level in order to augment task performance.

ACKNOWLEDGMENTS

This work was supported by NSF award CNS 1659746, IIS-0643100, NASA Cooperative Agreement Number NNX16AB24A, AFOSR Award FA9550-09-1-0108, and ONR Award N00014-12-1-0987

REFERENCES

[1] S. Archer, M. Gosakan, P. Shorter, and J. Lockett. 2005. New capabilities of the Armys maintenance manpower modeling tool. *Journal of the International Test and Evaluation Association* 26, 1 (2005), 19 – 26.
[2] J. Heard, C. E. Harriott, and J. A. Adams. 2018. A Survey of Workload Assessment Algorithms. *IEEE Transactions on Human-Machine Systems* PP, 99 (2018), 1–18.
[3] J. McCraken and T. Aldrich. 1984. *Implications of operator workload and system automation goals.* Technical Report ASI-479-024-84B. U.S. Army Research Institution.
[4] C. D. Wickens, J. D. Lee, Y. Liu, and S. E. Gordon Becker. 2004. *An Introduction to Human Factors Engineering* (2nd ed.). Pearson Education, Inc.
[5] Z. Yin and J. Zhang. 2014. Recognition of mental workload levels by combining adaptive exponential feature smoothing and locality preservation projection techniques. In *IEEE Chinese Control Conference.* 4700–4705.