

# Objective-Analytical Measures of Workload – the Third Pillar of Workload Triangulation?

Christina Rusnock<sup>(✉)</sup>, Brett Borghetti, and Ian McQuaid

Air Force Institute of Technology, Wright-Patterson AFB, Fairborn, OH, USA  
{christina.rusnock, brett.borghetti,  
Ian.mcquaid.ctr}@aft.edu

**Abstract.** The ability to assess operator workload is important for dynamically allocating tasks in a way that allows efficient and effective goal completion. For over fifty years, human factors professionals have relied upon self-reported measures of workload. However, these subjective-empirical measures have limited use for real-time applications because they are often collected only at the completion of the activity. In contrast, objective-empirical measurements of workload, such as physiological data, can be recorded continuously, and provide frequently-updated information over the course of a trial. Linking the low-sample-rate subjective-empirical measurement to the high-sample-rate objective-empirical measurements poses a significant challenge. While the series of objective-empirical measurements could be down-sampled or averaged over a longer time period to match the subjective-empirical sample rate, this process discards potentially relevant information, and may produce meaningless values for certain types of physiological data. This paper demonstrates the technique of using an objective-analytical measurement produced by mathematical models of workload to bridge the gap between subjective-empirical and objective-empirical measures. As a proof of concept, we predicted operator workload from physiological data using VACP, an objective-analytical measure, which was validated against NASA-TLX scores. Strong predictive results pave the way to use the objective-empirical measures in real-time augmentation (such as dynamic task allocation) to improve operator performance.

**Keywords:** Workload measurement · Machine learning · VACP · IMPRINT

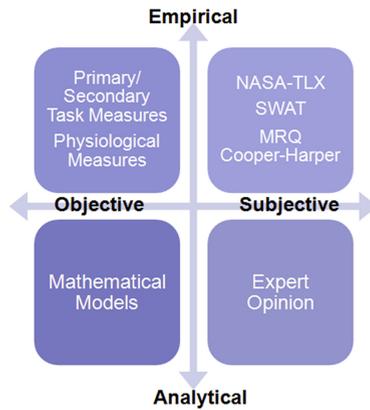
## 1 Introduction and Background

With the rising use of automation and recent interest in adaptive automation (e.g. Kaber and Kim 2011; Sheridan 2011; Parasuraman et al. 2007; Scerbo 2007), the human factors community has become highly motivated to find an effective and accurate means for measuring operator workload. The ability to measure operator workload is critical for the dynamic task allocation envisioned in adaptive systems, because workload is typically the impetus that determines whether an operator should be allocated a more or fewer tasks (e.g., Bailey et al. 2006; Parasuraman et al. 2009; De Visser and Parasuraman 2011). In this context, operator workload refers to the amount of attentional resources required of a specific person to perform a specific task (Hart and Staveland 1988). Due to the personal nature of workload, workload measurements

are difficult to accomplish and verify, and human factors professionals must choose from numerous workload measurement tools that often provide incompatible measurements.

### 1.1 Workload Measurement Taxonomy

These workload measurement tools can be categorized across two dimensions: objective–subjective and empirical–analytical (Fig. 1). Objective workload measurements are gathered from facts; subjective workload measurements are gathered from individual opinions. Thus, objective workload measurements use information and data about the real world, and are independent of the person gathering the measurement. Subjective workload measurements, on the other hand, are highly dependent on the person gathering the measurement.



**Fig. 1.** Dimensions of workload measurement

Orthogonal to the objective-subjective dimension is an empirical–analytical dimension. Empirical workload measurements are derived from experience; analytical workload measurements are derived from analysis. Thus, empirical workload measurements are a posteriori; they entail an actual data collection process, and are often used in conjunction with laboratory experiments or field observations. Analytical workload measurements can be thought of as a priori measurements, since they rely heavily upon the analytical reasoning of the human factors specialist. While they do not necessarily rely upon empirical data, in practice analytical workload measurements are often derived from extensive task analyses. The distinction is that empirical measures aim to measure workload directly, whereas analytical measures infer workload based on knowledge of the task, operator, and environment. These two orthogonal dimensions form natural axes which facilitate categorizing various workload measurement tools.

*Subjective-empirical* measures assess workload directly by gathering individual opinions—frequently in the form of self-report questionnaires. Due to the nature of these

tools, the measurements do not provide real-time feedback on the subject's mental workload, do not capture changes in workload over the course of the task, and may be subject to memory biases. Widely used subjective measurement tools include the NASA-TLX (Hart and Staveland 1988), SWAT (Reid and Nygren 1988), Cooper-Harper (1969), MRQ (Boles and Adair 2001), Overall workload (Jung 2001), and workload profile (Tsang and Velazquez 1996).

*Subjective-analytical* measures rely upon subject matter experts or experienced users to provide estimates for anticipated workload. These estimates are essentially expert opinions which can vary widely between raters. As can be expected, this method of estimating workload is not utilized in academia; however, it is commonly used in the public and private sector for product and system development, especially when making early design decisions early on in a program's life cycle. Before initial prototypes are developed, empirical methods are not feasible, thus workload evaluation must rely upon analytical methods.

*Objective-analytical* measures combine knowledge of task, environment, and individual as inputs to mathematical models which quantify estimated workload. Objective-analytical models are often used during the task design or re-design process, when empirical measurements are difficult to obtain. These models are built from detailed task analyses and can incorporate individual behavior when the information is available. Examples of objective-analytical measures include time-line analysis and prediction (TLAP) (Parks and Boucek 1989), visual auditory cognitive psychomotor (Aldrich and Szabo 1986), and W/INDEX (North and Riley 1989). Advantages of objective-analytical measures include: (1) consistency in the workload values produced, (2) the ability to include workload values differentiated for resources or channels, and (3) the ability to calculate workload values on any time scale. However, it should be noted that constructing accurate models is both a science and an art; the quality of the workload estimates rely heavily upon the completeness of the task analysis as well as the analyst's understanding of the system, operator, task, context, and the modeling methodology.

*Objective-empirical* measures feature direct measurement of either task performance or physiological states. Task performance measures focus on error frequency, number of errors, response time, and response accuracy to determine the level of cognitive workload. Direct measurement of task performance can be conducted in a way that is transparent to the user, and it can be conducted continuously throughout the performance of the task. However, the relationship between performance and workload is non-linear, with both high and low workload associated with poor performance, with higher task performance between these workload extremes (Teigen 1994). Furthermore, despite task performance measures ability to monitor real-time changes in performance, performance changes likely lag causal changes in workload. That is, an operator may experience excessively high workload for a period of time before the change becomes apparent in their performance. Lag is especially detrimental to adaptive systems: if the system only alters workload after a performance change is detected, it is too late to *prevent* that performance degradation. These problems with using performance as an indicator of workload: non-specificity, non-linearity and lag, have encouraged the expedited search for other objective-empirical measures of workload.

Technological advances in sensor development and computer processing have enabled the use of physiological measures for assessing operator workload. Objective-empirical physiological measures use biological feedback to estimate cognitive workload. Common physiological measures include heart rate, heart rate variability, respiration, skin response, pupil dilation, eye movement/fixation, blink rate, and brain activity. Physiological measures are particularly well suited for adaptive automation purposes because they provide immediate feedback, are highly sensitive to change, and can be designed for minimal-to-no intrusiveness in task performance. In order for objective-empirical physiological measures to be effective, researchers need to be able to distinguish useful, workload-relevant information from physiological fluctuations caused by environment and irrelevant biological processes. Just as workload does not have a linear correlation with performance, it is likely that the relationship between physiological data and workload will be highly complex.

## 1.2 Workload Measurement Application

While subjective-empirical, or self-report measures, are relatively easy to use and widely accepted, the measurements are typically taken after task completion and often consist of a single, cumulative rating. These features make self-reported workload unrealistic for incorporation into adaptive systems. Thus, while workload measurement has historically defaulted to subjective-empirical measures, more recently interest has grown in using physiological measures to estimate workload (Parasuraman and Wilson 2008; Warm and Parasuraman 2007). Physiological measures have the advantage of being specific to the individual, relatively un-intrusive to the task, continuously measurable, and available in real-time. However, using physiological measures for workload estimation has challenges, one being that physiological state is affected by numerous factors besides workload. Separating the workload “signal” from the biological and environmental “noise” is a daunting task, especially when the relationship between the physiological measure and workload is not well-established.

In order to define this relationship between workload and physiological measures, researchers have begun to use a dual collection approach, collecting both self-reported measures and physiological measures (e.g. Wilson and Russell 2004; Taylor et al. 2013). This dual collection provides an opportunity to interpret and validate objective-empirical physiological data using the subjective-empirical self-reports. However, this validation is difficult because—in order to reduce task interruptions—subjective-empirical measures are sampled infrequently, sometimes only at the completion of the task. In contrast, objective-empirical measurements are recorded continuously, and provide frequently-updated information over the course of a trial. Thus, linking the low-sample-rate subjective-empirical measurement to the high-sample-rate objective-empirical measurements poses a significant challenge. While the series of objective-empirical measurements could be down-sampled or averaged over a longer time period to match the subjective-empirical sample rate, this process discards potentially relevant information, and may produce meaningless values for certain types of physiological data which could vary rapidly in the time between the less-frequently measured samples.

Objective-analytical measures possess the unique potential of being able to bridge the gap between subjective-empirical and objective-empirical measurements. Objective-analytical measures are produced from detailed task analyses of the individual's behavior when performing the task. The task analysis assigns workload estimates to each low-level activity the subject performs during the task. Regardless of task duration, workload can be estimated by aggregating workload values from all the activities the subject was performing in an arbitrarily-sized time interval. As a result, workload values are generated bottom-up from the lowest level of activity. While objective-analytical values are scale-compatible with the subjective workload estimates subjects normally provide at the end of a trial, objective-analytical measures avoid being influenced by estimation biases occurring when subjects have to recall their experiences and declare their cumulative workload over long time intervals.

Objective-analytical measures can be calculated continuously over the course of a task, allowing for comparison on the natural (continuous) time-scale of the objective-empirical measurements. Moreover, objective-analytical measurements can be consolidated in meaningful ways (time-weighted average, peak value, and sustained peak-value) to validate against well-established subjective-empirical measures. This bridge between self-reported subjective-empirical measures and objective-empirical measures will enable a time-series characterization of workload based on physiological measures in real time, a necessary feature of workload-based dynamic task allocation, such as adaptive automation.

## 2 Purpose

This paper argues that objective-analytical measures of workload possess numerous benefits that can significantly enhance current empirical workload measurement and analysis techniques. Specifically, objective-analytical methods have the unique ability to be able to connect subjective-empirical workload measures, such as NASA-TLX to objective-empirical measures, such as physiological readings. This connection enables the creation of predictive machine learning algorithms, which paves the way to use physiological measures in real-time augmentation (such as dynamic task allocation) to improve operator performance.

## 3 Method

Our study begins by collecting subjective- and objective-empirical data from a human-in-the-loop experiment. Next, objective-analytical workload measurements are generated using the improved performance research integration tool (IMPRINT). IMPRINT simulation models are created for each subject for each trial in order to calculate continuous workload values that correspond directly to the physiological data collected for each trial. These workload values are established using the visual, auditory, cognitive, psychomotor (VACP) method (Bierbaum et al. 1989). Next, the objective-analytical models are validated using the subjective-empirical data. The continuous workload profiles generated through IMPRINT allow for the training of a model tree to

discover the relationship between physiological data and VACP workload values. To demonstrate efficacy of our method, the model tree algorithms are evaluated on their ability to infer objective-analytic VACP workload from objective-empirical physiological measurements. Algorithm performance is reported using root mean squared error (RMSE).

### 3.1 Human-in-the-Loop Study

The human-in-the-loop study was conducted by the HUMAN Lab, Air Force Research Laboratory, at Wright-Patterson AFB. The study included 12 participants (8 male, 4 female; ranging from 18–46 years of age, with a mean of 25.66), performing remotely piloted aircraft tasks using a synthetic task environment. The participants performed two tasks, a surveillance task and a tracking task. In the surveillance task, the participant operated a simulated drone-mounted video camera to search through a desert marketplace for a high-value target designated as the figure holding a rifle. Upon finding the target, the participant used the camera to track the target (who was on foot) until the target left the observation area. The task difficulty was affected by manipulating the number of distracter figures (12 or 48) and the video image quality (high or low noise), creating a  $2 \times 2$  factorial design. Each of the 4 surveillance conditions was completed 4 times by each participant.

The second task was the tracking task, which consisted of operating the video camera to follow a high-value target whose position is already known. In the tracking task, the target is on a motorcycle, and moves at considerably faster speed than the surveillance targets, making it more difficult to keep the target in the camera image and requiring faster reaction time from the participants. The task difficulty was altered by manipulating the number of targets to follow (either one or two) and the terrain type (urban or rural), creating a  $2 \times 2$  factorial design. Each of the 4 tracking conditions was completed 4 times by each participant.

NASA-TLX scores were collected at the end of each trial: each participant provided 16 surveillance scores and 16 tracking scores. Physiological data collected included 49 electroencephalography (EEG) measurements (7 cranial node sites, 7 brainwave frequency bands), 4 pupillometry measurements ([diameter, quality]  $\times$  [filtered, raw]), 2 electrooculography (EOG) measurements (blink rate, fixation), 2 electrocardiography (ECG) measurements (heart rate and heart rate variability) and 2 respiration measurements (amplitude, frequency).

### 3.2 IMPRINT Models and Model Validation

A task analysis was conducted on the surveillance and tracking tasks in order to develop task networks that captured the task flows, decision logic, and user interactions. Participant performance and response times were used to calculate timing for each of the lowest level tasks. Workload demand values were assigned to each of the lowest level tasks using the VACP method. VACP builds upon multiple resource theory (Wickens 2002) to capture workload demand across 7 resource channels: visual,

auditory, cognitive, fine psychomotor, gross psychomotor, speech, and tactile. Each channel has a 0–7 demand scale, with specific values tied to descriptive anchors (e.g. a visual reading task has a value of 5.9). Each of the lowest level tasks in the network is assigned a demand value for each channel. Demand values are then summed across channels and tasks at each point in time to generate an overall workload score. This workload score for each point in time enables the generation of a workload profile (Fig. 2).

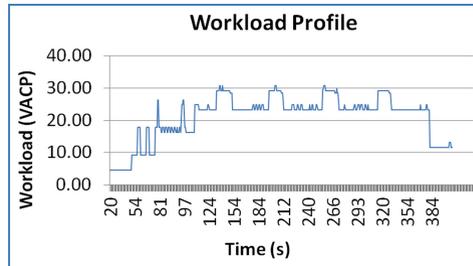


Fig. 2. Example VACP workload profile

Workload is validated using a correlation analysis that pairs model VACP predicted workload per trial with self-reported NASA-TLX values for each participant for each task (tracking and surveillance). The subjective interpretation of the NASA-TLX scale, makes this dimension unique to each individual, thus it is necessary to perform separate correlations for each participant. This validation produces a set of 24 correlations (12 subjects  $\times$  2 tasks). The tracking task correlations ranged from 0.31–0.87 with a mean of 0.61. Correlations values above 0.60 are considered as having a “marked degree of correlation” (Franzblau 1958), and thus the correlations from the tracking task meet the criteria for satisfactory validation. The surveillance task correlations ranged from 0.1–0.63 with a mean of 0.37. Lower correlations for the surveillance task are largely due to the lack of differentiation in the NASA-TLX self-reported scores; an ANOVA for the NASA TLX by condition does not find any statistical difference between the four surveillance conditions. Thus, the inability to validate these models is attributed to the design of the human-in-the-loop experiment rather than an issue with the models.

### 3.3 Inferring Workload with Machine Learning Algorithms

Validated IMPRINT models generate profiles which characterize operator workload throughout the course of a task. As workload varies, operator physiological response is affected. By determining the relationship between workload and physiological response, physiological state can be used to infer the workload the operator is experiencing. This section describes machine learning methods for inferring operator workload.

Our proposed method estimates operator workload in real time from physiological measurements using supervised machine learning algorithms to train predictive models.

We evaluated this method using algorithms trained on a subset of the objective-empirical physiological measurements and their corresponding objective-analytic workload values computed using IMPRINT. Once trained, the ability of the models to infer workload from physiological data was assessed on hold-out data that was not used for training.

IMPRINT produces a log of discrete workload changes. The log contains a series of timestamps and new workload values which were captured as each workload-change event in the simulation of the task occurred. Between any two events in the log, workload is constant. These discrete-event workload values were sampled at 1 Hz to create a time-series interpretation of operator workload which could be associated with corresponding physiological data. Time series data for all seven VACP channels of cognitive workload were captured and summed to produce an overall workload value at 1 Hz for each of the 32 scenarios the twelve subjects accomplished during the experiment.

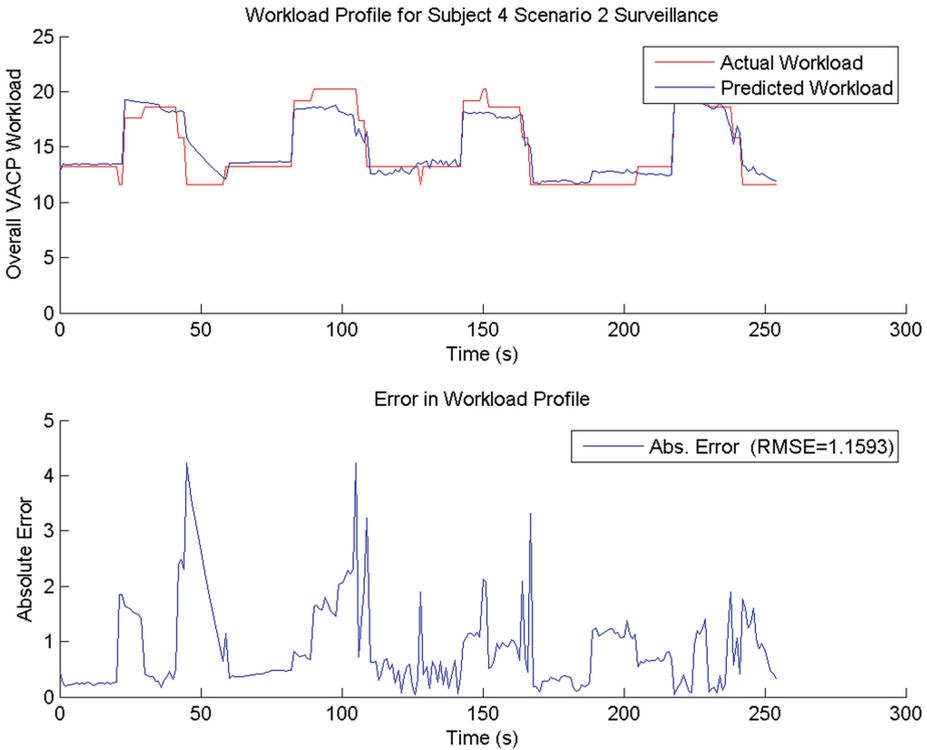
During the execution of the sixteen surveillance and sixteen tracking tasks, operator physiological data was recorded for each of the twelve subjects. The researchers who conducted the study performed additional processing on the sensor data to generate physiological data in commonly-used formats, then re-sampled at 1 Hz. Sixty brain, heart, respiratory and eye physiological data features were collected.

Inferring real-valued workload from real-valued physiological data is a specific example of a more general activity known as regression. Three supervised-learning regression algorithms were evaluated for their ability to infer workload from physiological data: linear regression, model trees, and the multi-layer perceptron (also known as an artificial neural network). Due to space limitations, and because it had the best performance of the algorithms evaluated, only the model tree algorithm (Quinlan 1992, Wang and Witten 1997, Fong 2010) are discussed further in this section, although the performance of all three algorithms is presented in Sect. 4. Before a model can be used for prediction, it must be trained. We discuss the training and testing process for model trees next.

During training, a model tree uses an iterative hierarchal process to split a dataset according to a particular feature value. Each branch in the tree captures the value of a single feature which best splits the remainder of the data into evenly-sized subgroups. Once the training is complete, the tree can be used to infer workload from the feature values observed in the physiological data. To infer workload value from an observation, the algorithm evaluates the observation in the context of the first node in the tree, and decides which branch to follow based on the value of the feature being evaluated in that branch. This process is repeated until the observation is evaluated at a leaf node. At a leaf node of the tree, an inference is made on the value of workload.

## 4 Results

To be useful, our method must demonstrate that objective-empirical physiological measurements can be used to infer objective-analytic workload. This section examines the efficacy of the supervised machine learning algorithms to infer workload from physiological data. Figure 3 depicts an example of overall actual and inferred



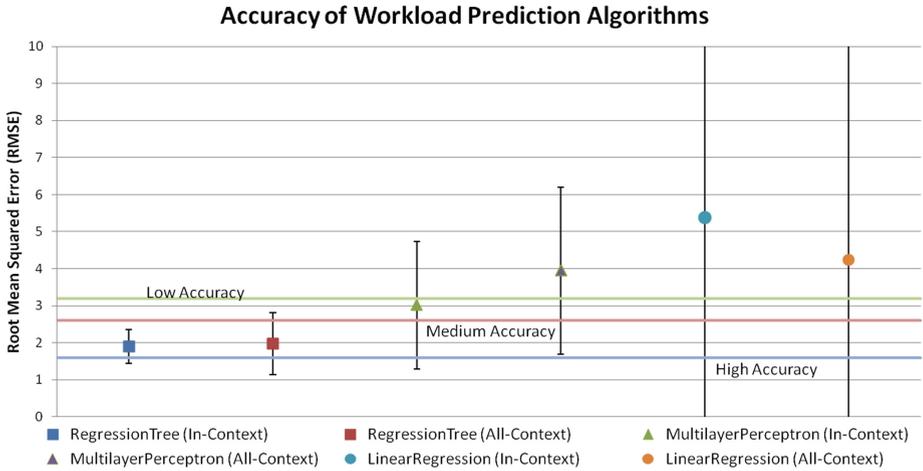
**Fig. 3.** Inferring operator workload. (Top) Workload profile and model tree prediction for surveillance task. (Bottom) Absolute prediction error of tracking task.

(predicted by model tree) VACP workload for one subject's surveillance task. The distance between the inferred workload and the actual workload is the absolute prediction error.

Efficacy was evaluated for three supervised learning algorithms: model tree, multilayer perceptron (60-30-1), and linear regression. For training, each model used 75 % of the data and tested on the remaining unseen 25 % of the data. The workload prediction results are presented in Fig. 4. Accuracy of three prediction algorithms with error bars representing the standard deviation of root mean squared errors (RMSE) and 95 % confidence intervals. Root mean squared error (RMSE) is used as the figure of merit for model performance. Low, medium, and high accuracy thresholds are depicted for reference. These accuracy thresholds have meaning in the workload prediction task:

Low (RMSE < 3.2): At low accuracy in a two-target scenario the model could differentiate between the situation where the operator is tracking both targets or tracking no targets.

Med (RMSE < 2.6): At medium accuracy the algorithm could determine if the secondary task (question and answer) was being performed or not.



**Fig. 4.** Accuracy of three prediction algorithms with error bars representing the standard deviation of root mean squared errors (RMSE) and 95 % confidence intervals. “In-context” workload prediction performance is measured using a model trained on the same person in the same conditions as the testing scenario. “All-context” workload prediction performance is measured using a model trained using more general data from all scenarios.

High (RMSE < 1.6): At high accuracy in the one-target scenario, the algorithm could differentiate between an operator who was successfully tracking the target and one who was attempting to reacquire a lost target.

Our evaluation suggests model trees are the most accurate supervised learning algorithm for inferring workload from physiological data. In most cases, this predictor can differentiate between tasks the operator is performing, and in some cases the algorithm is able to indicate whether or not the operator is successfully at performing the task. This finding suggests the algorithm’s ability to infer workload from physiological sensor data may be accurate enough to facilitate dynamic task allocation in real-time.

## 5 Discussion and Future Directions

Subjective-empirical workload measures such as NASA-TLX provide quick, well-accepted measurements of workload, but are not practical for use in adaptive systems, which require continuous, real-time measurements of operator workload. Objective-empirical workload measures, especially physiological measures, possess the ability to fill this gap, by enabling an adaptive system to directly monitor operators and predict their cognitive states. However, predicting cognitive workload requires a means of interpreting or connecting physiological readings to levels of workload.

Objective-analytical measures, such as IMPRINT’s VACP, can fill this gap, by providing workload values that correspond to the change in demand on an individual’s cognitive resource during task performance. These objective-analytical outputs can be

validated using subjective-empirical measures. Once validated, workload outputs from objective-analytical models and physiological data can be used to train machine-learning algorithms. The trained algorithms can be used to predict workload from physiological data, facilitating real-time dynamic task allocation.

This study demonstrated this technique using IMPRINT VACP workload profiles, validated using NASA-TLX scores. These workload profiles are trained using a model tree to predict operator workload from physiological sensor data—a task that would not have been possible with NASA-TLX alone. In this preliminary work, the model tree was able to successfully differentiate between the tasks the operator was performing and in some conditions, whether the operator was successfully accomplishing the task or not.

The next step for this research is to establish an experiment protocol which collects physiological data on the full space of workload values (underload, peak performance, and overload) for each operator. This data is needed to train machine learning algorithms so they can generate real-time workload predictions. Further research also includes extension of the physiological-IMPRINT model analysis to specific resource channels in an effort to train machine learning algorithms to predict workload on those specific resource channels. This channel-specific analysis will enable identification of tasks that are appropriate for dynamic task allocation or for adaptive presentation of information via alternate modalities (e.g. moving information from visual to audio modality).

## References

- Aldrich, T.B., Szabo, S.M.: A methodology for predicting crew workload in new weapon systems. In: Proceedings of the Human Factors Society 30th Annual Meeting (1986)
- Bailey, N.R., Scerbo, M.W., Freeman, F.G., Mikulka, P.J., Scott, L.A.: Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Hum. Factors* **48**(4), 693–709 (2006)
- Bierbaum, C.R., Szabo, S.M., Aldrich, T.B.: Task analysis of the UH-60 mission and decision rules for developing a UH-60 workload prediction model No. AD-A210763. U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria (1989)
- Boles, D.B., Adair, L.P.: The multiple resources questionnaire (MRQ). In: Proceedings of the Human Factors & Ergonomics Society Annual Meeting, pp. 1790–1794 (2001)
- Cooper, G.E., Harper, R.P.: The use of pilot ratings in the evaluation of aircraft handling qualities No. NASA Ames Technical Report NASA TN-D-5153). NASA Ames Research Center, Moffett Field (1969)
- De Visser, E.J., Parasuraman, R.: Adaptive aiding of human-robot teaming: effects of imperfect automation on performance, trust, and workload. *J. Cogn. Eng. Decis. Making* **5**(2), 209–231 (2011)
- Fong, A., Sibley, C., Cole, A., Baldwin, C., Coyne, J.: A comparison of artificial neural networks, logistic regressions, and classification trees for modeling mental workload in real-time. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 54, pp. 1709–1712 (2010)
- Franzblau, Abraham N.: A Primer of Statistics for Non-Statisticians. Harcourt, Brace & World, New York (1958)

- Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*, pp. 139–183. Elsevier, Amsterdam (1988)
- Jung, H.S.: Establishment of overall workload assessment technique for various tasks and workplaces. *Int. J. Ind. Ergon.* **28**(6), 341–353 (2001)
- Kaber, D.B., Kim, S.H.: Understanding cognitive strategy with adaptive automation in dual-task performance using computational cognitive models. *J. Cogn. Eng. Decis. Making* **5**(3), 309–331 (2011)
- Parasuraman, R., Barnes, M., Cosenzo, K.: Adaptive automation for human-robot teaming in future command and control systems. *Int. C2 J.* **1**(2), 43–68 (2007)
- Parasuraman, R., Cosenzo, K.A., De Visser, E.J.: Adaptive automation for human supervision of multiple uninhabited vehicles: effects on change detection, situation awareness, and mental workload. *Mil. Psychol.* **21**(2), 270–297 (2009)
- Parasuraman, R., Wilson, G.F.: Putting the brain to work: neuroergonomics past, present, and future. *Hum. Factors* **50**(3), 468–474 (2008)
- Parks, D.L., Boucek Jr., G.P.: Workload prediction, diagnosis, and continuing challenges. In: McMillan, G.R., Beevis, D., Salas, E., Strub, M.H., Sutton, R., Van Breda, L. (eds.) *Applications of Human Performance Models to System Design*, pp. 47–64. Plenum, New York (1989)
- North, R.A., Riley, V.A.: W/Index: a predictive model of operator workload. In: Mc-Millan, G., Beevis, D., Salas, E., Strub, M., Sutton, R., Van Breda's, L. (eds.) *Applications of Human Performance Models to System Design*, pp. 81–90. Springer, USA (1989)
- Quinlan, R.J.: Learning with continuous classes. In: 5th Australian Joint Conference on Artificial Intelligence, pp. 343–348 (1992)
- Reid, G.B., Nygren, T.E.: The subjective workload assessment technique: a scaling procedure for measuring mental workload. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*, pp. 185–218. Elsevier, Amsterdam (1988)
- Scerbo, M.W.: Adaptive automation. In: Parasuraman, R., Rizzo, M. (eds.) *Neuroergonomics: The Brain at Work*, pp. 238–252. Oxford University Press, New York (2007)
- Sheridan, T.B.: Adaptive automation, level of automation, allocation authority, supervisory control, and adaptive control: distinctions and modes of adaptation. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **41**(4), 662–667 (2011)
- Taylor, G., Reinerman-Jones, L., Szalma, J., Mouloua, M., Hancock, P.: What to automate: addressing the multidimensionality of cognitive resources through system design. *J. Cogn. Eng. Decis. Making* **7**(4), 311–329 (2013)
- Teigen, K.H.: Yerkes-Dodson: a Law for all seasons. *Theory Psychol.* **4**(4), 525–547 (1994)
- Tsang, P.S., Velazquez, V.L.: Diagnosticity and multidimensional subjective workload ratings. *Ergonomics* **39**(3), 358–381 (1996)
- Wang, Y., Witten, I.H.: Induction of model trees for predicting continuous classes. In: *Poster Papers of the 9th European Conference on Machine Learning* (1997)
- Warm, J.S., Parasuraman, R.: Cerebral hemodynamics and vigilance. In: Parasuraman, R., Rizzo, M. (eds.) *Neuroergonomics: The Brain at Work*, pp. 146–158. Oxford University Press, New York (2007)
- Wickens, C.D.: Multiple resources and performance prediction. *Theor. Issues Ergon. Sci.* **3**(2), 159–177 (2002)
- Wilson, G.F., Russell, C.A.: Psychophysiological determined adaptive aiding in a simulated UCAV task. *Human Performance, Situation Awareness, and Automation: Current Research and Trends*, pp. 200–204. Erlbaum Associates, Mahwah (2004)