Cognitive Metrics Profiling of a Complex Task: Toward Convergent Validity with Behavioral and EEG Workload Indicators

Christopher A. Stevens (christopher.stevens.28@us.af.mil)
Air Force Research Laboratory
WPAFB, OH 45433 USA

Megan B. Morris (megan.morris.1.ctr@us.af.mil)
Ball Aerospace
Fairborn, OH 45324 USA

Christopher R. Fisher (christopher.fisher.27.ctr@us.af.mil) **Christopher Myers (christopher.myers.29@us.af.mil)**
Air Force Research Laboratory
WPAFB, OH 45433 USA

Abstract

Workload assessment remains a challenging, multi-dimensional problem. A variety of metrics are available (behavioral, physiological, subjective), but their relationships to each other and the underlying cognitive processes producing workload are not well understood. In the present paper, we extend an approach known as Cognitive Metrics Profiling to an unmanned vehicle control simulation. We show how the model predicts behavioral performance and physiological indicators of global workload in the task and produces insights about sources of workload.

Keywords: cognitive workload; ACT-R; EEG

Introduction

Cognitive workload has been the focus of empirical and theoretical investigation for many decades (Cain, 2004). Limitations in workload have captured the interest of cognitive scientists, in part, because its effects pervade a wide range of tasks, and it has direct practical implications. As an abstract concept, workload is challenging to measure directly. Several indirect measures have been used in the literature, including subjective, behavioral, and physiological measures, each with advantages and disadvantages.

Subjective or self-reported measures, are straightforward to implement, but can be obtrusive if administered during a task, and are vulnerable to biases when assessed retrospectively (Matthews et al., 2015). Behavioral measures include primary task or secondary task performance (e.g., accuracy,reaction time). Primary task performance is unobtrusive, but doesn't indicate how much capacity may remain for addiitonal tasks. Secondary tasks can indicate spare capacity, but are more obtrusive than primary tasks (Miller, 2001). Physiological measures such as electroencephalogram (EEG) and electrocardiography (ECG) have the advantage of being unobtrusive and capturing changes in workload across time; however, they are often contaminated with reactions to other factors such as the environment (Miller, 2001).

The difficulty in defining and measuring workload poses challenges for theoretical progress (Cain, 2004). Cognitive architectures have the potential to provide a much needed theoretical framework for informing existing workload metrics. Cognitive architectures have two advantages: (1) as formal models, they provide precise descriptions of cognitive mechanisms and processes that underlie workload, and (2) as integrative theories, cognitive architectures are applicable to an array of both simple and complex tasks across many cognitive domains.

Cognitive Metrics Profiling (CMP) is one of the first efforts to define and measure workload within a cognitive architecture (Gray, Schoelles, & Sims, 2005; Gray, Schoelles, & Myers, 2005). CMP uses the Adaptive Control of Thought Rational (ACT-R; Anderson et al., 2004) cognitive architecture to characterize the evolving cognitive demands of a task and link those demands to performance predictions. In CMP, workload is defined as a weighted sum of activity across multiple information processing modules (e.g., vision, motor, and declarative memory).

Initial validation of CMP has been promising. Past work has indicated a correlation between CMP and behavioral and subjective indicators of workload (Jo et al., 2012). However, CMP currently has two major limitations: 1) It has been evaluated only in small-scale tasks of short duration (e.g. paired associates) (Gray, Schoelles, & Sims, 2005; Jo et al., 2012) and 2) the relationship between CMP estimates and physiological indicators of workload has not been investigated.

In the present paper, we seek to further validate and extend CMP in two ways: first, we investigate the relationship between CMP and established EEG metrics of workload to further establish convergent validity, and second, we use CMP in an unmanned vehicle operator task to test its scaleability to more complex tasks.

Cognitive Metrics Profiling

Theory and Rationale

CMP uses the ACT-R cognitive architecture to quantify the degree to which cognitive resources (e.g., memory or vision) are taxed during a given task. A profile detailing cognitive resource usage can be analyzed to understand how task demands affect cognition. For example, if the declarative memory module is in use for 80% of the task then the memory demand of the task would be very high, making it difficult to take on additional tasks also heavy in memory demand. Alternatively, CMP can be used to measure global workload, which is defined as a weighted sum of activity across modules. Jo et al. (2012) found that global workload derived from CMP is correlated with subjective workload judgments.

Present work

In the present work, participants completed an UV task and we compared EEG and ECG metrics to workload profiles generated from CMP. We used a UV task because it induces a wide range of workload levels and taxes ACT-R modules to varying degrees. Together, these factors provide a wide range of workload conditions with which to validate CMP against physiological indicators of workload.

Predictions

Figure 3 lists the predicted relationships between physiological workload indicators and model-based workload generated from CMP. Several studies have shown EEG band frequencies correspond to manipulations of cognitive workload (Borghini et al., 2014; Lean & Shan, 2012). Alpha and theta have been shown to decrease and increase with cognitive workload, respectively. Research has also suggested that different frequencies within the alpha band capture different aspects of workload, with lower alpha (8-10 Hz) reflecting alertness and upper alpha (10-13 Hz) reflecting information-processing (Klimesch, 1999). Ratios of band frequencies have also shown some promise, with the Task Load Index (TLI), a ratio of theta and alpha, and the Engagement Index (EI), a ratio of beta to alpha and theta, increasing with increased workload and task engagement, respectively (e.g., Kamzanova et al., 2014; Freeman et al., 1999). In addition to EEG metrics, heart rate variability metrics have been shown to decrease with increased cognitive workload (Lean & Shan, 2012).

Method

Participants

Ten volunteer employees (M_{age} = 29.30; SD_{age} = 6.99; R_{age} = 19-41; $Proportion_{male}$ = 50.00%) from Wright-Patterson Air Force Base (WPAFB) who were unfamiliar with the task completed an informed consent document and participated in the study. Participants reported normal or corrected-to-normal vision, normal color vision, and normal hearing. This study was approved by the Air Force Research Laboratory (AFRL) Institutional Review Board (IRB).

Task Description

Participants completed two 60-minute missions of varying difficulty in IMPACT (Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies; Draper et al., 2017; Rowe et al., 2015)—a high fidelity UV simulator. We manipulated the task density of the missions—the number and difficulty of the tasks—to induce low vs. high levels of work load. The IMPACT environment consists of an array of monitors and two modes of communication: a microphone and headset for auditory communication, and a communication window for text communication and alerts. Alerts and instructions to complete tasks are presented primarily through the communication window and, to a lesser extent, via the headset. The primary monitor features a map of the base, a menu system for selecting and managing UVs, and a communication window. Three secondary monitors provide alternative views of the base map, UV sensor and status information, and a reference manual instructing participants how to dispatch UVs for various tasks.

During a mission, the participant must complete a variety of tasks that require scheduling and planning, resource management, multi-stage decision making, communication, and information search and acquisition. Many tasks require the participant to dispatch UVs either at predetermined times for routine surveillance or in response to sporadic security events. In order to correctly dispatch a UV, the participant must select a UV with attributes required by the task, such as the destination, UV type (e.g., aerial vs ground), optional automatic termination, and maneuver (e.g., inspection or blockade). Participants can terminate UV tasks either through scheduled automation or manually upon receiving instruction via the communication system. Other types of tasks occur periodically throughout the mission. For example, participants must reallocate resources in response to environmental or mechanical problems, and answer information queries, requiring information to be found within the interface and relayed via the communication window.

Protocol

Data collection for each participant was performed separately over the course of a single day. First, participants were trained to use the IMPACT system and to perform base defense actions. Training included participants performing a capstone mission where the experimenter revealed any behavioral errors and allowed the participant to ask questions. After training, participants were fitted with EEG and heart rate physiological sensors and participant eye gaze was calibrated to an eye tracking system. Participants then completed a low and a high task density condition. Both conditions were 60 minutes in length and were counterbalanced across participants.

Unmanned Vehicle Model

We developed a model of the UV task within the ACT-R cognitive architecture (Anderson et al., 2004). A cognitive architecture is a formal, computational framework for simulating and testing comprehensive theories of cognition (Newell, 1990). The ACT-R cognitive architecture consists of specialized information processing modules, spanning procedural and declarative memory, visual and auditory perception, speech production, and motor execution. Cognition unfolds over a series of production cycles which coordinates the flow of information among the modules. Importantly, module activity within the architecture forms the basis for workload measurement within CMP.

In the interest of brevity, we will focus on the high level strategy employed by the model, such as how it searches the interface for tasks and how it resolves conflicts between competing goals. The model's strategy is illustrated as a flow chart in Figure 1. The strategy is composed of three primary phases: an active search phase, a passive monitoring/waiting phase, and a task execution phase. During the search phase, the model inspects three locations within the interface for new tasks: (1) a message window where information queries and new task alerts appear, (2) a base map where certain problem events appear, and (3) a list of Random Anti-Terror Measures (RAMs) with target execution times and deadlines.

When a task is found, the model performs the task and rechecks the interface for new tasks that might have become available during the intervening time. When no task is found, the model proceeds to the next location. At the third location, the model compares the mission clock to the target times for the RAMs. The model will perform a RAM if the mission time is within a parameter we term leading time—a period of time preceding the deadline, during which the model will attempt to complete the task. If no RAMs can be completed, the model enters a passive monitoring phase, in which it waits for the next RAM and responds to events that pop up in the interface. Periodically, during the monitoring phase, the model will "re-calibrate" its internal clock to the mission clock in an effort to mitigate growing temporal estimation error.

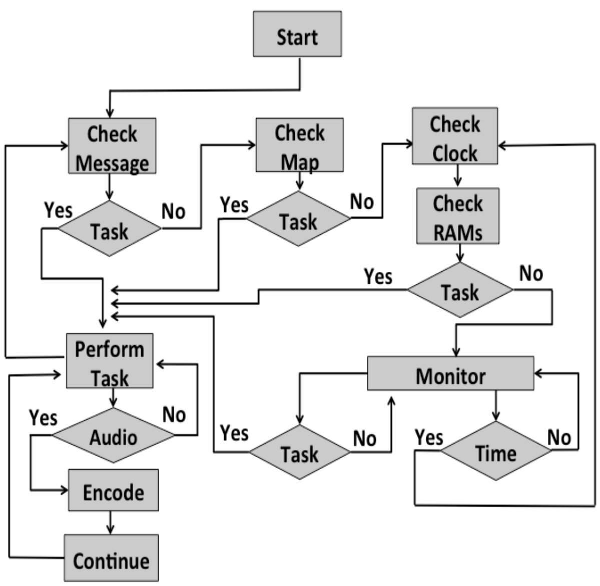

Figure 1: A flow chart of the model's high-level strategy for the UV task. Boxes represent processes and diamonds represent decision points.

The model primarily uses a first come, first served policy to manage competing task demands. This is why the model reinspects the interface for new tasks upon the completion of a current task. One minor exception to the rule occurs when an audio message is presented during an ongoing task. In this case, the model briefly suspends the ongoing task to encode the message, resumes the suspended task, and later attempts to complete the task associated with the auditory message.

Workload and Performance Measures

Physiological Workload Metrics We collected EEG, heart rate variability, and eye tracking data for our physiological workload metrics. Throughout the recordings, the eye tracking system had difficulty locating participant's eyes due to the experiment environment, producing several missing values. As a result, the ocular methodology and data are not reported. A list of workload metrics and respective calculations can be found in Table 1. EEG data was collected with a sampling rate of 500 Hz from a dry electrode Quick-20 Cognionics headset (Cognionics, CA, USA). Electrode locations followed the 10-20 system with 19 active channels (Fp1, Fp2, Fz, F3, F4, F7, F8, Cz, C3, C4, T3, T4, Pz, P3, P4, P7, P8, O1, and O2), two grounds located adjacent to Fp1 and Fp2, and a linked ear reference. Electrode contacts consisted of silver/silver chloride matrixed for conductivity. EEG data was cleaned and processed offline with an in-house script utilizing the MATLAB (The MathWorks, Inc., MA, USA) toolbox EEGLAB (Delorme & Makeig, 2004). EEG data was re-referenced to the linked ear reference and filtered using a Parks-McClellan optimal equiripple finite impulse response (FIR) band-pass (high-pass cutoff 1 Hz and low-pass cutoff 95 Hz) and notch filter (60 Hz). DC offset was removed and a recursive least squares adaptive filter was used to remove eye artifacts. To calculate EEG metrics, average band power was extracted from 10 second epochs with no overlap using a modified periodogram spectral estimator with a Hanning window. Specific metrics were then calculated from these band power values and log transformed.

Inter-beat (RR interval) data was collected with a sampling rate of 18 Hz from a Zephyr Bioharness 3.0 (Zephyr Technology Corp., MD, USA). RR interval data was cleaned offline with an in-house script by identifying outliers with a percent change strategy based on data epochs (e.g., Kemper et al., 2007; Persson et al., 2005). Outliers were removed and linear interpolation was utilized to extract values to replace the outliers (e.g., Peltola, 2012).

Performance Evaluation Due to heterogeneity of the tasks, we evaluated performance according to criteria that depended on task-specific requirements. A score of 1 was recorded if a participant satisfied a criterion and 0 otherwise. For example, events that required the deployment of a UV typically included a 3 minute deadline, correct destination, correct UV attributes (e.g., correct sensor), correct operation (e.g., aerial inspection) and a category for miscellaneous situation-dependent constraints (e.g., scheduled termination of a task). Information queries were evaluated according to a 3 minute deadline and the correctness of the response.

Model Fitting We varied two parameters that exert broad, cascading effects on the task dynamics and resource engagement: latency factor, which affects overall memory retrieval times by scaling memory activation, and leading time, which specifies how far in advance a RAM is completed relative to its deadline. In order to find the best-fitting parame-

Table 1: Physiological metrics and respective calculations.

Metric	Calculation
Alpha	Band power in range of 8 - 13 Hz located at Pz site.
Lower Alpha	Band power in range of 8 - 10 Hz located at Pz site.
Upper Alpha	Band power in range of 10 - 13 Hz located at Pz site.
Theta	Band power in range of 4 - 8 Hz located at Fz site.
Frontal Theta	Calculated as the average of theta at F3 and F4 sites (e.g., Kamzanova et al., 2014).
TLI	Calculated as theta (Fz)/alpha (Pz) (Gevins & Smith, 2003; Kamzanova et al., 2014).
EI	Calculated as beta/(alpha + theta) from averages of sites Cz, P3, Pz, and P4 (Kamzanova et al., 2014; Freeman et al., 1999).
Mean HRV	Calculated as the mean of RR intervals.
Median HRV	Calculated as the median of RR intervals.

ters, we performed a grid search in which latency factor $\in \{.5, 1.0, 1.5\}$ and leading time $\in \{2, 4, 8\}$ were varied independently. We simulated the model 20 times for each parameter set [1].

The model predicted the number of participants that satisfied each criterion for each task. The fit of the model was evaluated according to a normalized root mean squared error (NRMSE) measure based on the standard deviation of the binomial distribution. Some advantages of this approach include: (1) ease of interpretation, (2) it is more stringent at the boundaries (e.g., 90% correct) where data are less variable, and (3) it requires no pooling across heterogeneous data.

NRMSE was computed as:

$$NRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{I}\sum_{c=1}^{C_i}\left(\frac{\hat{x}_{i,c} - x_{i,c}}{\sigma_{i,c}}\right)^2}$$

where $i = [1, 2..., I]$ is an event index, $c = [1, 2..., C_i]$ is a criterion index for each event, $N = \sum_{i=1}^{I} C_i$ is the total number of criteria across all events, $x_{i,c}$ is the number of participants who satisfied criterion c for event i and $\hat{x}_{i,c}$ is the corresponding prediction. The standard deviation is computed as $\sigma_{i,c} = \sqrt{S \cdot p_{i,c} \cdot (1 - p_{i,c})}$, where S is the number of participants and $p_{i,c}$ is the proportion of participants who satisfied criterion c for event i. In cases where $p_{i,c} = 1$, we adjusted the value downward to the next possible value of $\frac{9}{10}$ to prevent division by zero.

[1] The grid search was small due to simulation times and the fact that fit was only moderately sensitive to changes in parameter values.

Results

Behavioral Performance

After excluding 8 complex tasks that were difficult to model, there were 27 and 43 tasks remaining in the low and high task density conditions, respectively. To assess our workload manipulation, we averaged across all tasks and their criteria to yield two overall accuracy scores per subject—one for low task density and one for high task density. A paired t-test ($t(9) = 5.93$, $p = .00$, $d = 1.87$) revealed an effect of task density on accuracy, high ($M = .84$) vs low ($M = .94$). As predicted, mean subjective workload, as measured by the NASA-TLX, was higher in the high ($M = 53.33$) vs low ($M = 20.58$) task density condition, (paired t-test, $t(9) = 8.75$, $p = .00$, $d = 2.77$).

Model Results

The best-fitting parameters were latency factor = 1 and leading time = 8 (NRMSE = 2.03), suggesting that subjects were proactive in setting up RAMs in advance of their deadlines. The predicted accuracy was .98 and .90 for low and high task density, respectively. Although the model tended to overestimate accuracy, it was able to capture the qualitative drop in performance.

Workload was computed according to the formulas described in (Jo et al., 2012) using consecutive time windows of 10 seconds (see Figure 2). Across the entire mission, mean workload was higher under high task density (2.30) compared to low task density (1.59), mirroring the behavioral performance results and subjective workload assessments.

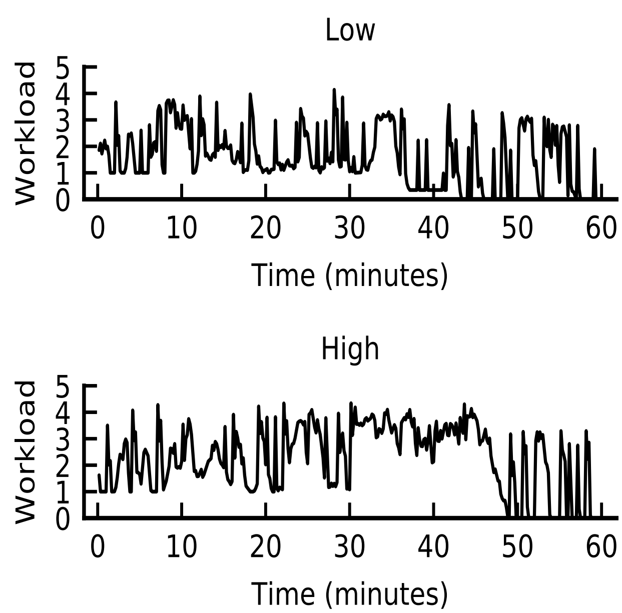

Figure 2: A comparison of global workload profiles generated by the model in the low workload and high workload conditions.

Workload Regression

We examined the association of the physiological workload metrics with workload derived from the cognitive model (values were rounded to the nearest whole number to create five workload levels, 0 - 4) and task density condition (low vs high) using linear mixed effects modeling (LMM). We performed robust linear mixed effects modeling (RLMM) from the *robustlmm* package (Koller, 2016) in *R* (R Core Team, 2017) due to violations of residual normality and homoscedasticity assumptions. Baseline models included the metric of interest and a random intercept for subjects. Augmented models included the workload predictors of interest (1. Level, 2. Level and Condition, 3. Level, Condition, and Level x Condition) and a random intercept for subjects. Robustified estimating equations from RLMM do not correspond to likelihood statistics. As a result, we could not compare the models with an ANOVA or obtain p values for the fixed effects. However, Wald confidence intervals can be used to examine the significance of the fixed effects.

RLMM analyses indicate that only EI had significant workload level and task density condition effects, suggesting increased task engagement as model workload level increased and as task density condition increased. The other EEG and heart rate metrics suggested marginal and trending effects in the expected directions, except for theta and TLI metrics in terms of task density condition (see Figure 3).

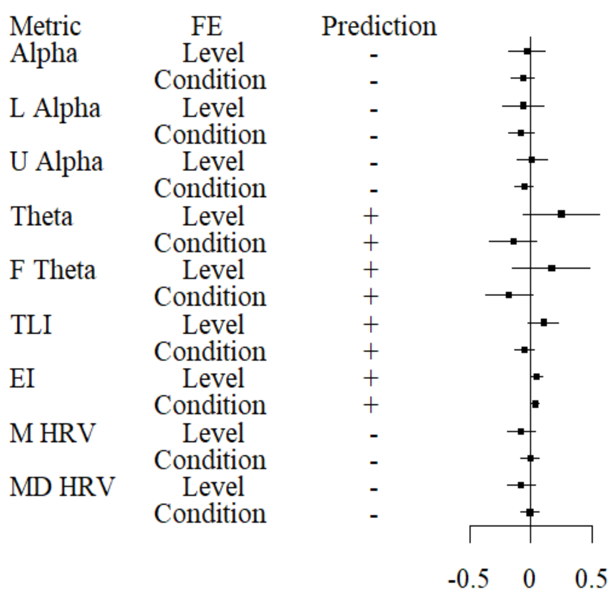

Figure 3: Predicted direction of relationship and regression coefficients between physiological workload and model-based workload. Main effects of level and condition are shown. Dots represent mean coefficient estimates and horizontal lines represent 95% confidence intervals. HRV coefficients were re-scaled by .01 for ease of presentation.

Discussion

CMP is a promising technique for characterizing workload, but it remains untested in complex environments and its convergent validity with other workload indicators, especially physiological indicators, has not been fully established. In the present study, we applied CMP to a multiple-vehicle control task that takes place over an extended time (60 minutes). Further, we examined the relationship between workload estimates generated by CMP and physiological indicators commonly associated with cognitive load or global cognitive activity. We found preliminary evidence for a relationship with one such indicator, EI, suggesting that the activity of an ACT-R model could be a valid way to characterize the cognitive resources utilized by a task. This is a potentially useful technique for predicting overall workload levels and the specific cognitive capacities affected by high workload moments.

We have provided a proof-of-concept here that CMP can discriminate between low and high workload conditions even in tasks that involve many complex interrelated subtasks over a long period of time. The CMP model predicts both an increase in subjective workload and a decrease in performance across subtasks in high complexity conditions, as was observed here. Future studies should look at CMP predictions across a wider range of task difficulties to confirm that it adequately captures the shape of the relationship between task difficulty and predicted workload.

This study adds to previous work relating ACT-R models to neural activity. It has been shown previously that buffer activity in ACT-R can be related to the BOLD signal in fMRI, suggesting that buffers may be meaningfully associated with activation of certain populations of neurons in the brain (Borst & Anderson, 2015; Qin et al., 2003). Moreover, it has been demonstrated that activity predicted by ACT-R can be recovered using Hidden Semi-Markov modeling of EEG data (Anderson et al., 2016). The present study adds to that growing body of research by suggesting that ACT-R activity may also be associated with neural indicators of cognitive activity.

We believe the tentative relationship between CMP and EI makes sense given that EI is thought to reflect multiple cognitive processes and resources. However, this is a relationship that warrants further exploration and clarification. The relationship observed here between model activity and EI is still very tentative due to noise in the indicator itself and the absence of a specific physiological model relating the two quantities. However, we propose these results justify confirmatory studies to test the hypothesis that CMP and EI may both characterize similar cognitive processes associated with workload. If this relationship is further explored, CMP may offer a potential integrative framework for behavioral, physiological, and subjective workload metrics, improving our ability to understand how they relate to each other and to the cognitive operations that they measure.

Acknowledgments

The opinions expressed herein are solely those of the authors and do not necessarily represent the opinions of the United States Government, the U.S. Department of Defense, the U.S. Air Force, or any of their subsidiaries, or employees. This research was supported by a grant from the Air Force Research Laboratory Airman Systems Directorate Chief Scientist Venture Fund. The authors wish to thank Allen Dukes and Justin Estepp for their expertise with physiological data collection and analysis. The authors further thank Sarah Spriggs and Gloria Calhoun for overseeing data collection and for their comments and suggestions on the current work. The third author was supported by the Postgraduate Research Participation Program at the Air Force Research Laboratory through the Oak Ridge Institute for Science and Education and Department of Energy.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological review*, *111*(4), 1036.

Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The discovery of processing stages: Extension of sternbergs method. *Psychological review*, *123*(5), 481.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, *44*, 58–75.

Borst, J. P., & Anderson, J. R. (2015). Using the act-r cognitive architecture in combination with fmri data. In *An introduction to model-based cognitive neuroscience* (pp. 339–352). Springer.

Cain, B. (2004). *A Review of the Mental Workload Literature* (Tech. Rep.). Defence Research and Development.

Delorme, A., & Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9–21.

Draper, M., Calhoun, G., Hansen, M., Douglass, S., Spriggs, S., Patzek, M., ... others (2017). Intelligent multi-unmanned vehicle planner with adaptive collaborative/control technologies (impact). In *19th international symposium on aviation psychology* (p. 226).

Freeman, F. G., Mikulka, P. J., Prinzel, L. J., & Scerbo, M. W. (1999). Evaluation of an adaptive automation system using three eeg indices with a visual tracking task. *Biological psychology*, *50*(1), 61–76.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 113–131.

Gray, W. D., Schoelles, M. J., & Myers, C. W. (2005). Profile before optimizing: A cognitive metrics approach to workload analysis. In *Chi 2005* (pp. 1411–1414). Portland, Oregon.

Gray, W. D., Schoelles, M. J., & Sims, C. (2005). Cognitive Metrics Profiling. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(12), 1144–1148.

Jo, S., Myung, R., & Yoon, D. (2012). Quantitative prediction of mental workload with the act-r cognitive architecture. *International Journal of Industrial Ergonomics*, *42*(4), 359–370.

Kamzanova, A. T., Kustubayeva, A. M., & Matthews, G. (2014). Use of eeg workload indices for diagnostic monitoring of vigilance decrement. *Human factors*, *56*(6), 1136–1149.

Kemper, K. J., Hamilton, C., & Atkinson, M. (2007). Heart rate variability: impact of differences in outlier identification and management strategies on common measures in three clinical populations. *Pediatric research*, *62*(3), 337.

Klimesch, W. (1999). Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, *29*(2-3), 169–195.

Koller, M. (2016). robustlmm: An R package for robust estimation of linear mixed-effects models. *Journal of Statistical Software*, *75*(6), 1–24. doi: 10.18637/jss.v075.i06

Lean, Y., & Shan, F. (2012). Brief review on physiological and biochemical evaluations of human mental workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *22*(3), 177–187.

Matthews, G., Reinerman-Jones, E., Barber, D. J., & Abich, J. I. (2015). The psychometrics of mental workload: Multiple measures are sensitive but divergent. *Human Factors*, *57*, 125-143.

Miller, S. (2001). Workload measures. *National Advanced Driving Simulator. Iowa City, United States*.

Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

Peltola, M. (2012). Role of editing of rr intervals in the analysis of heart rate variability. *Frontiers in physiology*, *3*, 148.

Persson, H., Kumlien, E., Ericson, M., & Tomson, T. (2005). Preoperative heart rate variability in relation to surgery outcome in refractory epilepsy. *Neurology*, *65*(7), 1021–1025.

Qin, Y., Sohn, M.-H., Anderson, J. R., Stenger, V. A., Fissell, K., Goode, A., & Carter, C. S. (2003). Predicting the practice effects on the blood oxygenation level-dependent (bold) function of fmri in a symbolic manipulation task. *Proceedings of the National Academy of Sciences*, *100*(8), 4951–4956.

R Core Team. (2017). *R: A language and environment fr statistical computing.* Vienna, Austria. (URL https://www.R-project.org/.)

Rowe, A. J., Spriggs, S. E., & Hooper, D. J. (2015). Fusion: a framework for human interaction with flexible-adaptive automation across multiple unmanned systems.