# Modeling the Operator Functional State for Emergency Response Management

**Mark Parent**
Co-DOT Laboratory, Université Laval
mark.parent@psy.ulaval.ca

**Jean-François Gagnon**
Thales Research and Technology, Canada
jean-francois.gagnon@ca.thalesgroup.com

**Tiago H. Falk**
MuSAE Lab
falk@emt.inrs.ca

**Sébastien Tremblay**
Co-DOT Laboratory, Université Laval
sebastien.tremblay@psy.ulaval.ca

**ABSTRACT**

New technologies are available for emergency management experts to help them cope with challenges such as information overload, multitasking and fatigue. Among these technologies, a wide variety of physiological sensors can now be deployed to measure the Operator Functional State (OFS). To be truly useful, such measures should not only characterize the overall OFS, but also the specific dimensions such as stress or mental workload. This experiment aimed to (1) design a multi-dimensional model of OFS, and (2) test its application to an emergency management situation. First, physiological data of participants were collected during controlled experimental tasks. Then, a support vector classifier of mental workload and stress was trained. Finally, the resulting model was tested during an emergency management simulation. Results suggest that the model could be applied to emergency management situations, and leave the door open for its application to emergency response on the field.

**Keywords**

Emergency Management, Simulation, Operator Functional State, Mental workload, Stress,

## INTRODUCTION

As cities grow, the need for urban security is steadily rising since city authorities need to ensure the safety and well-being of their citizens, protect public assets, and make optimum use of city resources. Moreover, authorities need to coordinate different security agencies to prevent crime, reduce incident rates and response times, and manage emergencies and unpredictable crisis situations, such as industrial accidents, terrorist alerts or natural disasters. With the rise of smart cities around the world and the so-called 'internet-of-things' revolution, smart urban security has emerged where sensors and systems are deployed across the city to enable security forces to handle incidents and coordinate emergency response by police, fire crews, first responders and private operators. Ideally, these systems are connected to an integrated operations control center, thus providing real-time decision support and managing resources for emergency response. With such technological advances, control center operators are now faced with cognitive challenges such as information overload, multitasking, interruptions, and fatigue, all of which can reduce the efficiency of emergency response. Moreover, first responders (e.g., policemen, firefighters, paramedics) are also exposed to a combination of physical and mental factors that contribute to high levels of workload, stress, and fatigue, thus increasing the likelihood of errors, which in many cases can be life threatening.

Today, several tools, such as intelligent video surveillance systems (Dufour, 2012), have become available to assist security experts in reliably conducting their work. As such, it is imperative that these tools provide relevant and on-time assistance to human operators. The real-time monitoring of the Operator Functional State (OFS) could potentially contribute to augmenting human performance and responsiveness (e.g., Galster & Johnson, 2013). OFS is defined as “the multidimensional pattern of human psychophysiological condition that mediates performance in relation to physiological and psychological costs” (Carter, Cheuvront & Sawka, 2004). According to OFS theory, maintaining one’s performance in a primary task requires a psychological and physiological cost that can negatively affect performance in other tasks (Hockey, 2005). OFS can help enhance human-computer interaction by providing information about the current human status (e.g., under-loaded/over-loaded) and vulnerabilities (e.g., attentive/fatigued). As such, OFS is a multi-dimensional concept. Measures associated with OFS can benefit human-human interaction in team management, or could prove useful in training situations. For example, stress level of trainees could be monitored in order to keep trainees in their zone of maximal adaptation (i.e., neither too easy nor too hard) (Warm, Parasuraman, Matthews, 2008). With the burgeoning of bio-behavioural monitoring wearable technologies and the so called internet of things, deploying sensors in the field to assess, in real-time, several OFS dimensions has become a reality.

### Bio-Behavioural monitoring

Several studies have demonstrated the feasibility of physiological state measurement with emerging wearable bio-behavioural monitoring devices. For example, the classification of mental workload using a combination of brain, heart or respiratory measures has been shown possible (Casson, 2014; Wilson & Russell, 1999). Broadly speaking, physiological measurement can be divided into two categories: central and peripheral measurement. Central measures of the nervous systems have demonstrated to be highly beneficial to the assessment of concepts associated with OFS (e.g., Hogervorst, Brouwer & van Erp, 2014), with electroencephalography (EEG) emerging as a prime technology. In emergency response situations, however, ambulatory constraints might prevent the use of central measures sensors, particularly EEG, which is extremely sensitive to movement artifacts. For this reason, this study focuses on the use of peripheral sensors. Peripheral measures include, but are not limited to: 1) cardiac measures, 2) electrodermal response, and 3) ocular metrics such as the pupil diameter, blink frequency and fixation/saccade patterns.

Cardiac measures are possibly one of the most accessible physiological measures available. Heart beat detection, for example, is readily available across numerous devices, including watches, chest straps, and smart shirts. Consequently, heart rate, and its derivative metrics such as heart rate variability (HRV, see Vuksanović & Gal, 2007), have been investigated in many domains such as sport (Aubert, Seps & Beckers, 2003), health (Kleiger, Miller, Bigger & Moss 1987) and cognition (Nickel & Nachreiner, 2003). Mental load and stress are also known to influence electrodermal activity (EDA) (i.e.: variations of skin conductance caused by sweat). Research has suggested that EDA can help discriminate between mental load and stress in office-like settings (Setz et al., 2010). Eye metrics, in turn, can include fixations/saccade metrics, blink patterns and pupillary measurement. Eye movement metrics are numerous, but these metrics are often task-bound and cannot be extrapolated to other tasks. Notwithstanding, research has shown that ocular metrics can be used to measure OFS determinants such as workload (Benedetto et al., 2011) and fatigue (Cazzoli et al., 2014).

### Current Challenges

Despite advances in bio-monitoring devices, measuring determinants of OFS is problematic as the task is often ill-defined. For example, fatigue can be conceptualized as a physical exhaustion, but is often also used in the scope of mental strain after long working periods (see: Balkin, Horrey, Graeber, Czeisler & Dinges, 2011; Borghini, Astolfi, Vecchiato, Mattia & Babiloni, 2012). Workload also has several conceptualizations. As noted by Hart and Staveland (1988), an individual might evaluate his workload by his evaluation of the task difficulty while another might consider his effort input. The relative subjectivity of the concepts associated with OFS is detrimental to the development of psychophysiological models because feedback used for calibrating such models is incoherent across individuals. Therefore, some authors have called for a better conceptualization and manipulation of the states of interests (Brouwer, Zander, Van Erp, Korteling & Bronkhorst, 2015). In addition, physiological patterns of states of interests (such as fatigue or workload) are known to vary considerably across individuals, tasks, and time of the day (Wang, Hope, Wang, Ji & Gray,2012). In a recent article, Matthews et al. (2015) have called for algorithms that can adapt to the multidimensional construct of workload. This aspect is specifically relevant in the context of emergency response because it implies that the development of a model specific to this context will be required for achieving high level of classification accuracy, mandatory for safety-critical applications.

## OBJECTIVES

The objectives of the study are: 1) address the issues of using reliable physiological metrics to design a multidimensional OFS model, and 2) assess the behavior of this OFS model within the context of emergency management. To achieve these goals, first we describe the tasks used to calibrate the models and therefore the conceptualization of the states of interest: namely mental workload and stress. Second, we present the results with a strong emphasis on the behaviour of the models in an emergency management task. We then discuss implications for real-time monitoring of operators involved in emergency management.

## METHOD

### Participants

Twenty-two volunteers (12 males, aged 20 to 34 years) were recruited on the university campus and received financial compensation for their participation. Inclusion criteria were having normal or corrected vision and no known health issues. Participants were required to wear a Zephyr BioHarness 3 chest strap to measure cardiac activity. Electrodermal activity was measured using a Biopac MP100. Electrodes were placed on the medial phalanx of the non-dominant hand. Finally, an ASL Mobile Eye tracking glasses was used to measure ocular activity. Participants wore headphone during the experiment. The ethics committee of Université Laval granted its approval for this study.

### Tasks

We operationalized mental workload and stress in two tasks, namely the N-back and visual search, and then used a third one - emergency management simulation - to validate the model in a new realistic context.

The N-Back is a computerized task that requires participants to identify a target letter among a series of distractors presented sequentially in time (one every 2 seconds). The participants are presented a series of letters and must tell whether the actual letter is the same (target) or a different one (distractor) from the $N^{th}$ previous letter (Figure 1, left). Workload was manipulated by varying 'N'; here N=1 and N=2 was used to elicit low and high mental load, respectively. Visual search, in turn, is a computerized task that requires the participants to identify a target letter among a series of distractors (Figure 1, right). The task requires participants to visually scan the screen to search for the target letter. Workload was manipulated by varying the difficulty of the task (i.e., either finding an “E” amongst distractors or finding a non-inclined consonant amongst a greater number of distractors. In low stress conditions, these two tasks were performed undisturbed. In the high stress condition, on the other hand, participants were presented with a loud aversive sound through their headphones. The aversive sound was played randomly during the task, but the probability of occurring was higher if the participants’ performance was poor.

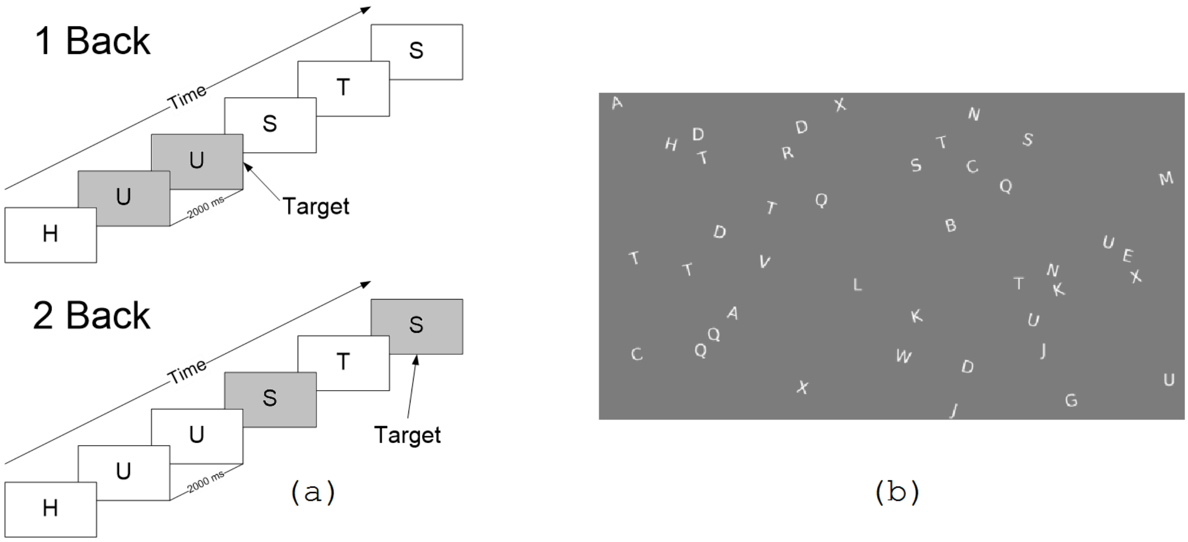


**Figure 1. Representation of the N-Back task (left) and the Visual Search task (right)**

Lastly, the emergency management platform used was SYnRGY (see Figure 2). To ensure realism, the platform was validated by collaborating and performing a task analysis with an emergency management team (Gagnon, Couderc, Rivest, Banbury & Tremblay, 2013). In the context of physiological research, such a simulation allows fitting operators with various sensors. In this experiment, we used the custom scenario editor to recreate a brief episode of a police service dispatch task. Participants had to use various tools in order to 1) assess their workforce; 2) assess the properties of the incidents; 3) the properties of the requests for support; and, 4) dispatch units, follow their progress and reallocate units to optimize response. In addition to physiological measures, subjective evaluations were collected using the NASA-TLX (Hart & Staveland, 1988). In order to evaluate stress, we added a custom stress question (i.e.: "*How stressful was the task?").*

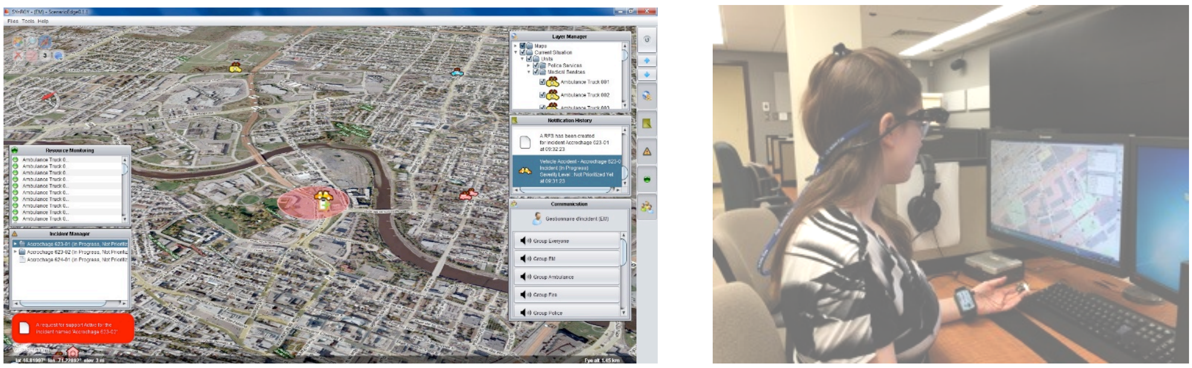

**Figure 2: Screenshot of the SYnRGY simulation (left) and the experimental setup (right)**

### Experiment Design

The two first tasks (N-Back and visual search) were each repeated four times. Workload and stress were manipulated: easy/calm, easy/stressful, hard/calm, or hard/stressful. A 5-minute break was inserted between conditions to avoid carry over effects of physiological response. NASA-TLX was administered after each condition. Conditions were counterbalanced across participants. Lastly, all participants completed the SYnRGY emergency management simulation task. Physiological data was collected during all experimental sessions. The aforementioned metrics were extracted from the heart rate, EDA and ocular signals, and used to train a support vector classifier (SVC) to discriminate the different levels of workload and stress. To avoid overtraining, data from the first two tasks was used to train, validate and test the classifiers and the unseen data from the third task (simulated emergency management) was used to assess if the predicted levels of workload and stress showed sensitivity to the simulation. The SVC implementation from Matlab Statistics and Machine Learning Toolbox was used with the following parameters: (autoscale: true; kernel function: Gaussian radial basis; method: least square). To optimize results, a grid search was performed on the box constrain and the Gaussian radial basis sigma.

## RESULTS

Results showed that the SVC was able to reach accuracies of approximately 70% for the two first tasks (i.e.: N-Back and visual search). Detailed results are shown in table 1.

| | Workload accuracy | Stress accuracy |
|---|---|---|
| Training | 74.52 % | 69.81 % |
| Validation | 77.14 % | 77.14 % |
| Test | 68.57 % | 71.42 % |

**Table 1. Classification accuracy of the training, validation and test sets for the N-Back and visual search (VS) tasks.**

Using the resulting SVC model, we classified levels of workload and stress of participants during the emergency management simulation. In order to verify the progression of workload and stress during the simulation, we plotted the percentage of participants classified in the high workload or high stress category at five points during the simulation (see figure 3).

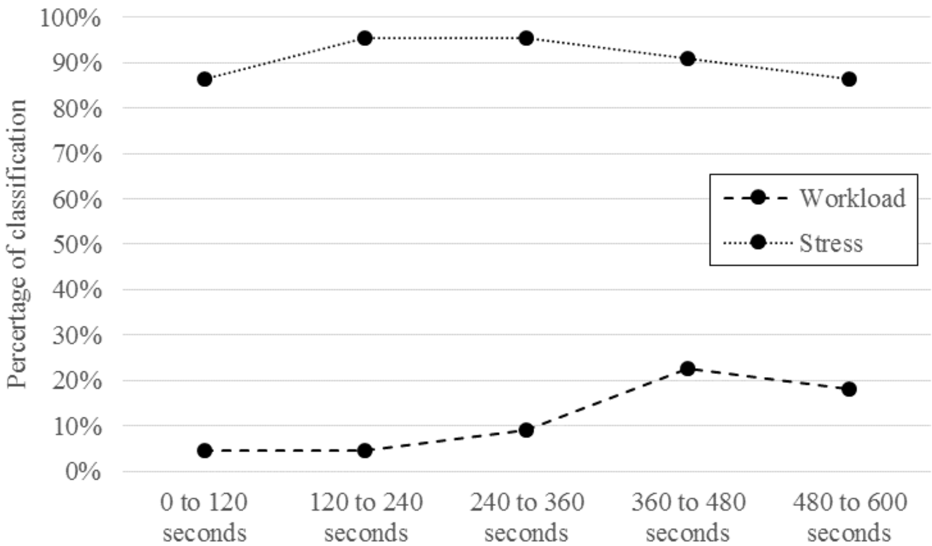


**Figure 3: Percentage of participants classified in the high workload/stress condition during 5 time periods**

Analysis of variance (ANOVA) and subsequent multiple comparisons on the subjective workload scores show that there was a significant difference between the two workload conditions (2 first tasks) and the SYnRGY task $F(2,62) = 43.46$, $p < 0.001$. Similarly, ANOVA performed on the subjective stress score showed that there was a significant difference between the stress manipulation on the two first tasks and the SYnRGY task $F(2,62) = 40.33$, $p < 0.001$. Results are shown in figure 4. Multiple comparisons were corrected with the Bonferroni procedure.

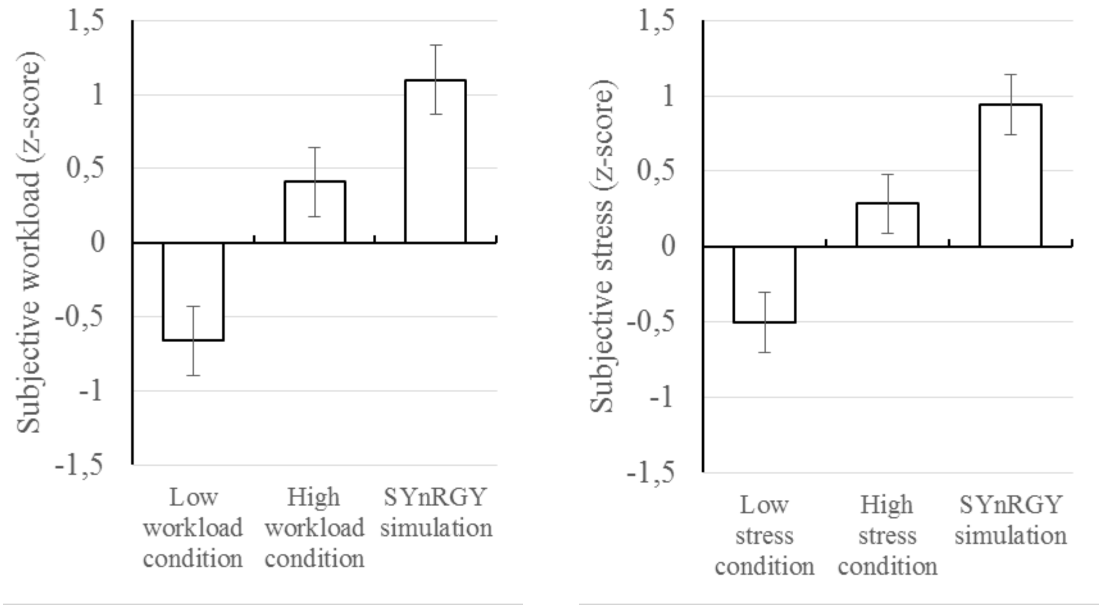


**Figure 4: Subjective evaluation of workload (z-score) during the low and high workload condition (N-Back, visual search) and the emergency management simulation (left). Subjective evaluation of stress (z-score) during the low and high stress condition (N-Back, visual search) and the emergency management simulation (right)**

## DISCUSSION

Results suggest that the support vector classifiers were able to reach reasonable accuracy when classifying the two OFS dimensions when using the N-Back and visual search tasks. These are promising results considering the absence of central nervous measures, such as EEG. While efforts must still be made to achieve physiological measurement to real-work settings (Balkin et al., 2011), such results suggest that it would be possible to implement OFS measurement in first responders or emergency management personnel.

Results also show that nearly all participants were classified as highly stressed during the emergency management simulation. This could be explained by the fact that the emergency management simulation was probably much more stressful than the two first tasks because it involved a much more realistic cover story. Moreover, the emergency management simulation required a long and elaborate tutorial which might have contributed to building up expectations. In addition, as opposed to N-Back and visual search tasks, the emergency management simulation task involved “accidents”, “victims” and “bombs”, which might have further contributed to participant stress. The increase between the first and the second period could be attributed to the fact that there were no incidents during the first minute. The decrease of stress, starting after the third period (240 to 360 seconds), could also be attributed to the participants obtaining a better grasp of the task and, therefore, calming slightly. A larger number of participants would however be required to see if this effect is significant.

Conversely, nearly all participants were classified as having a low-workload during the emergency management simulation. This pattern does not fit with the subjective ratings that show an increase in perceived workload during the simulation when compared to the two workload conditions of the N-Back and visual search tasks. This effect could be attributed to a sub-optimal SVC. However, there seems to be a surge in the classification of workload around the fourth period (360 to 480 seconds). It was during this period that the most intense incident (i.e., the bomb threat) occurred, suggesting that the classifier might be able to detect short-term variations of workload. An alternative hypothesis could be that the high difficulty of the emergency simulation task combined with the high stress prevented participants from investing optimal mental effort or concentration in the task. While the two first tasks were selected to include both mental/memory (i.e., N-Back) and visual/attentive (i.e., visual search) components, it is possible that the emergency management simulation task elicited a different type of mental load on the participants. Such results could be explained by recent work showing the multidimensional construct of workload, as well as the poor correlation between physiological and subjective evaluation of workload (Matthews et al., 2015).

Our empirical work suggests that measurement of the OFS is possible even when relying only on peripheral sensors such as a chest strap, skin conductance electrodes and portable eye tracking glasses. We believe this is an advantage over visual inspection (performed by teammates for instance) because it automates the assessment and allows for complex non-linear combination of cues for a more accurate assessment. As for the psychological dimensions that can be monitored from the analysis of the physiological signal, although we investigated workload and its relation to stress, other OFS dimensions such as fatigue fluctuate during emergency response operations and mediate performance as well. Developers of intelligent systems with the capability to assist emergency response personnel should therefore not base their system on a sole dimension.

## REFENCES

1. Aubert, A. E., Seps, B., & Beckers, F. (2003). Heart rate variability in athletes. *Sports medicine*, *33*(12), 889-919.
2. Balkin, T. J., Horrey, W. J., Graeber, R. C., Czeisler, C. A., & Dinges, D. F. (2011). The challenges and opportunities of technological approaches to fatigue management. *Accident Analysis & Prevention*, *43*(2), 565-572.
3. Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011). Driver workload and eye blink duration. *Transportation research part F: traffic psychology and behaviour*, *14*(3), 199-208.
4. 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.
5. Brouwer, A. M., Zander, T. O., van Erp, J. B., Korteling, J. E., & Bronkhorst, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in neuroscience*,*9*.
6. Carter, R., Cheuvront, S. N., & Sawka, M. N. (2004). *Operator Functional State Assessment (l'évaluation de l'aptitude opérationnelle de l'opérateur humain)*. Army research institute.
7. Casson, A. J. (2014). Artificial Neural Network classification of operator workload with an assessment of time variation and noise-enhancement to increase performance. *Frontiers in Neuroscience*, 8(December), 372. doi:10.3389/fnins.2014.00372
8. Cazzoli, D., Antoniades, C. A., Kennard, C., Nyffeler, T., Bassetti, C. L., & Müri, R. M. (2014). Eye Movements Discriminate Fatigue Due to Chronotypical Factors and Time Spent on Task–A Double Dissociation. *PloS one*, *9*(1).
9. Dufour, J. Y. (Ed.). (2012). *Intelligent video surveillance systems*. John Wiley & Sons.
10. Gagnon, J. F., Couderc, F., Rivest, M., Banbury, S., & Tremblay, S. (2013, May). Using SYnRGY to Support Design and Validation Studies of Emergency Management Solutions. In *Proceedings of the 10th International ISCRAM Conference*.
11. Galster, S. M., & Johnson, E. M. (2013). *Sense-assess-augment: A taxonomy for human effectiveness* (No. AFRL-RH-WP-TM-2013-0002). Air Force Research Lab Wright-Patterson Afb Oh Human Effectiveness Directorate.
12. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*,*52*, 139-183.
13. Hockey, G. R. J. (2005). *Operator functional state: The prediction of breakdown in human performance* (pp. 373-394). Oxford, UK: OUP.
14. Hogervorst, M. A., Brouwer, A. M., & van Erp, J. B. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience*, *8*.
15. Kleiger, R. E., Miller, J. P., Bigger, J. T., & Moss, A. J. (1987). Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *The American journal of cardiology*, 59(4), 256-262.

16. Matthews, G., Reinerman-Jones, L., Wohleber, R., Lin, J., Mercado, J., & Abich IV, J. (2015). Workload Is Multidimensional, Not Unitary: What Now?. In *Foundations of Augmented Cognition* (pp. 44-55). Springer International Publishing.
17. Nickel, P., & Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(4), 575-590.
18. Setz, C., Arnrich, B., Schumm, J., La Marca, R., Troster, G., & Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable EDA device. *Information Technology in Biomedicine, IEEE Transactions on*, *14*(2), 410-417.
19. Vuksanović, V., & Gal, V. (2007). Heart rate variability in mental stress aloud. *Medical engineering & physics*, *29*(3), 344-349.
20. Wang, Z., Hope, R. M., Wang, Z., Ji, Q., & Gray, W. D. (2012). Cross-subject workload classification with a hierarchical Bayes model. *NeuroImage*, *59*(1), 64-69.
21. Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 433-441.