

# Anticipating human error before it happens: Towards a psychophysiological model for online prediction of mental workload

---

Jean-François Gagnon<sup>1</sup>, Gauthier Durantin<sup>2</sup>, François Vachon<sup>1</sup>,  
Mickael Causse<sup>2</sup>, Sébastien Tremblay<sup>1</sup>, & Frédéric Dehais<sup>2</sup>

<sup>1</sup>Université Laval, Québec, Canada

<sup>2</sup>Institut Supérieur de l'Aéronautique et de l'Espace, Toulouse, France

## Abstract

Mental workload is a key factor influencing the occurrence of human error; specifically in remotely-operated vehicle operations. Both low and high mental workload has been found to disrupt performance in a nonlinear fashion at a given task; however, research that has attempted to predict individual mental workload has met with little success. The objective of the present study is to investigate the potential of the dual-task paradigm and prefrontal cortex oxygenation as online measures of mental workload. Subjects performed a computerized object tracking task in which they had to follow a dynamic target with their aircraft. Task difficulty was manipulated in terms of processing load and difficulty of control: two critical sources of workload associated with remotely operating a vehicle. Mental workload was assessed by a secondary concurrent time production task and a functional near infrared spectrometer. Results show that the effects of task difficulty differ across measures of mental workload. This pattern of behavioural and neurophysiologic results suggests that the empirically-based selection of an appropriate secondary task for the measure of mental workload is critical as its sensitivity may vary considerably depending on task factors.

## Introduction

Remotely operated vehicles (ROVs) operations are becoming increasingly prevalent in a wide variety of contexts such as border security, intelligence and military operations. Undeniably, the use of ROVs in military has increased tremendously over the last decade. According to The New York Times, the U.S. Military has now over 7,000 aerial drones as compared to only about 50 a decade ago. Civilian use of ROVs is also becoming increasingly frequent as the technology is more affordable, safe, and relatively reliable.

As noted by Cooke (2006), the term “*unmanned*” that frequently qualifies such systems can be misleading. Indeed, these systems involve a strong human-in-the-loop component for which the capacity could – and should (Parasuraman & Riley, 1997) – be improved above and beyond the capacity of fully automated systems.

In D. de Waard, K. Brookhuis, F. Dehais, C. Weikert, S. Röttger, D. Manzey, S. Biede, F. Reuzeau, and P. Terrier (Eds.) (2012). *Human Factors: a view from an integrative perspective*. Proceedings HFES Europe Chapter Conference Toulouse. ISBN 978-0-945289-44-9. Available from <http://hfes-europe.org>

There is a critical need to improve human-machine interaction within ROV systems given that Human Factors issues are responsible for a large proportion of ROV accidents. For instance, a document prepared for the Office of Aerospace Medicine in the United States reports that Human Factors-related deficiencies are responsible for 21% to 67% of ROV accidents in the US Army, Navy and Air Force (Williams, 2004).

Mishaps may be attributed to the high mental demands placed on operators and the degraded environmental conditions in which ROV operations take place. Indeed, ROV operators must often deal with degraded information that decreases the quality of their situation awareness (see Chen et al., 2007, for a review). Together with these constraints, ROV operators are required to perform cognitively demanding tasks such as monitoring, target identification, and manual control. Critically, these tasks require high levels of motor control for piloting the ROV under harsh environmental conditions, which in turn, imposes high levels of cognitive processing when conducting simultaneous sub-tasks. Moreover, in an effort to reduce costs and increase efficiency, a great deal of research is concerned with increasing the ROV/operator ratio. This trend has led Cummings and Mitchell (2008) to state:

*“Because of the increased number of sensors, the volume of information, and the operational demands that will naturally occur in a multiple-vehicle control environment, excessive cognitive demands will likely be placed on operators. As a result, efficiently allocating attention between a set of dynamic tasks will be critical to both human and system performance.” - p. 451*

One could argue that adequate distribution of the operator's mental resources is important – and will become even more essential – to achieve sufficient levels of performance in the execution of ROV missions. Operators in this context must perform several tasks simultaneously, each with different priorities. It is well known, however, that humans are *cognitively bounded*, insofar as human mental capacities are fundamentally limited. Consequently, allocating more resources to a task will inevitably limit the amount of resources available for other tasks. Moreover, as these environments are highly dynamic, priorities across tasks will be expected to change as the mission develops. It is therefore important for the operator to reallocate mental resources dynamically according to changes in task priorities (Dehais Causse, Vachon, & Tremblay, 2011). This, however, is a dire challenge to human cognitive control and its limitations. Mental overload can lead to the phenomenon of *cognitive tunneling* that can be defined as the inability of the operator to reallocate his/her attention from one *task* to another. Cognitive tunneling occurs when attention is focused on specific information or areas of display while information presented outside of these areas is neglected (Thomas & Wickens, 2001).

Approaches such as adaptive automation (Sheridan, 2011) and cognitive counter-measures (Dehais, Causse, & Tremblay, 2011) attempt to solve the problem of attention allocation; however, challenges in their implementation still remain. In particular, a critical aspect of adaptive aiding system is to provide help in a timely and accurate manner (Visser & Parasuraman, 2011). Adaptive automation based on an on-line prediction of the operators' mental workload represents a promising

solution to this challenge. This study investigates how mental workload can be predicted in this context.

### *Prediction of Mental Workload*

Typically, mental workload is measured using subjective scales, psychophysiological measures, or performance at a secondary concurrent task. Assessing mental workload with subjective scales consists of asking participants to rate their perceived workload. For instance, NASA-TLX is a multi-dimensional scale that was developed to measure the workload of operators either during or directly after task performance (Hart, 2006). Although this technique is reliable, it has several limitations as it offers only a limited number of data points and represents the perceived, not the actual workload of the operator. From a practical standpoint, the use of such scales during task performance is not recommended as they are invasive and create an additional source of operator workload. However, if they are used *post hoc* the data collected are only an aggregation of the level of workload perceived across the testing session.

An alternative and promising avenue is to adopt a Neuroergonomics approach to derive operators' mental workload from brain imaging techniques and psychophysiological measurements (Parasuraman & Wilson, 2008). In the search for non-invasive and periodic measures of mental workload, recent studies investigated neurophysiological measures. For instance, functional near infrared spectrometer is an optical brain monitoring device that measures cerebral hemodynamic response within the prefrontal cortex. Using such a device, it is possible to measure mental workload across various processing load conditions. Ayaz et al. (2011) were able to associate different hemodynamic responses with a subset of task difficulty (i.e. levels of processing load) on a well-established task: the N-Back task. Although this approach yields promising results, it assumes that high task difficulty is associated with high workload. Unfortunately, workload cannot be estimated precisely with the sole properties of the task because individual factors, such as expertise, or environmental factors, such as the time of the day, impact on mental resources deployed to perform a given task. In other words, task difficulty is relatively independent from mental workload. Consequently, mental workload should be defined considering both the task and the individual performing it.

One way to assess the interaction between mental workload, the task, and the individual performing the task is the dual-task paradigm. The rationale behind the dual-task paradigm is rooted in the limited attentional capacity theory. This paradigm consists of presenting two concurrent tasks to the subjects, who are required to prioritize their cognitive resources to the primary task and perform the secondary task with the remaining resources. As the amount of cognitive resources dedicated to the execution of the primary task increases, resources available for completing the secondary task will decrease proportionally. The decrease in cognitive resources to complete the secondary task will lead to a decreased performance at the secondary task, which can then be used to infer the relative amount of resources necessary to complete the primary task. Prospective time production represents a good candidate for a secondary task as it is assumed to demand the same attentional resources that nontemporal processing requires. Indeed,

as nontemporal processing demands increase, subjectively experienced duration decreases; resulting in longer time intervals when individuals must produce a previously learnt criterion (Block et al., 2010). However, from a practical point of view, this approach is limited because the addition of a secondary task is invasive.

Within this context, the objective of the present study is to investigate the potential of the dual-task paradigm and prefrontal cortex oxygenation as *online* measures of mental workload. This will be achieved by testing if performance at a secondary task and hemodynamic of prefrontal cortex are affected by two subsets of task difficulty in the context of ROV operations – namely control difficulty and processing load.

## Method

Sixteen volunteers participated in the study (mean age = 25; SD = 4.78; 13 males). Thirteen were right-handed and six had piloting experience. Data of four participants were removed from the analyses due to problems with data collection. All subjects reported normal or corrected vision. They were all native French speakers recruited among students from ISAE campus in Toulouse, France. Subjects had two different tasks to perform concurrently: a low-fidelity flight simulator task and a time interval production task.

### *Primary Task: Low-Fidelity Flight Simulator*

The purpose of this simulation was to solicit similar cognitive functions as to those required during a real ROV flight or drone control/supervision task. This approach allows the reproduction of key features of the real-world task while keeping a high degree of experimental control. The computerized simulation involved the control of an aircraft in bird's-eye view with a joystick (see Figure 1).

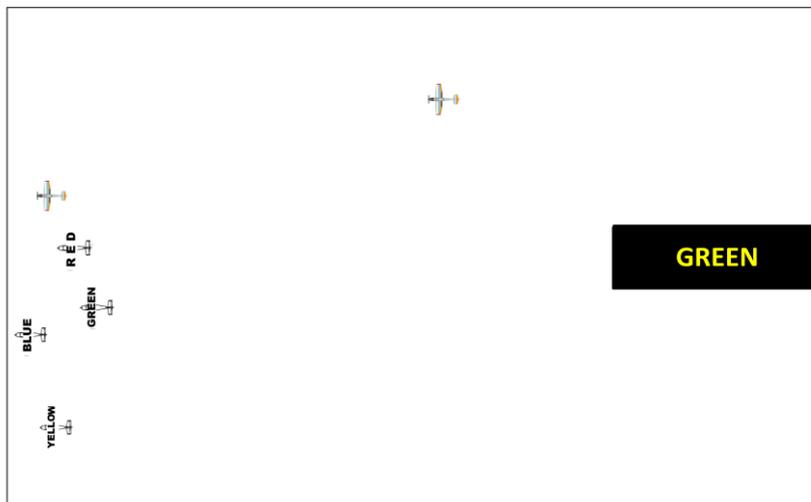


Figure 1. Low-fidelity flight simulator interface.

The subjects were instructed to minimize distance between own aircraft and a target aircraft. Own aircraft was located at about 60% from the left side of the screen. Potential target aircraft were located on the left (approximately 5% to 10% from left side). The target aircraft was specified to the subject by a visual cue presented at the right hand-side of the screen (approximately 95% from left side). A new cue was presented for 1.6 seconds every 8.6 seconds.

Task difficulty was varied in two ways: *difficulty of control* and *processing load*. There were two levels of difficulty of control (easy and hard) manipulated by varying the strength of the crosswind. The processing load was varied with an *N*-Back-like sub-task. Processing load can be varied by manipulating the number of items to be maintained and manipulated in working memory (*N*). Subjects had to target the aircraft corresponding to the last cue presented (*N*; low load condition) or the cue before (*N*-1; high load condition). The combination of the two factors yielded a  $2 \times 2$  repeated-measures design with four conditions: i) low load/easy control; ii) low load/hard control; iii) high load/easy control; and iv) high load/hard control.

#### *Secondary Task: Time-Production Task*

The secondary task was a prospective time-production task. The task involved a sound presented through loudspeakers at various times during the experimental session. The subjects had to start estimating time as soon as they heard the sound. Subjects then had to press a button on the joystick whenever they felt that the length of the sound was equal to the length of the criterion to be estimated (i.e. a previously learnt criterion of 2 s).

#### *Procedure*

Subjects were first trained at the primary and secondary tasks independently. The primary task training consisted of 10 object-tracking phases for each processing load level, and was performed at the easy level of control. Secondary task training involved a total of 130 trials. The first 110 trials provided visual feedback about the precision of time estimation. The feedback showed subjects whether their production was correct (i.e. within a 10% window around the target), too short (below the same time window) or too long (above the same time window). During the last 20 trials, subjects were not provided with any feedback. Training was necessary so that the subjects formed a good representation of the target interval to be produced during the experiment (i.e. 2 s).

After the training session, subjects achieved the four experimental sessions consecutively, each session lasting approximately six minutes. The sequence of the sessions was counterbalanced across subjects. The experimental session involved completing the primary and secondary task concurrently.

#### *Measures*

Hemodynamic of the frontal cortex was recorded with a functional near infrared spectrometer (i.e. *Biopac fNIR100*) with 16 channels. Each channel, or voxel,

records hemodynamic in terms of oxygenation level variations in comparison to a baseline. Production times at the secondary task were also recorded. Subjects received no feedback about the precision of their time estimation. Subjects filled out the NASA-TLX after each session.

Two potential online measures of mental workload were derived: performance at the secondary task and prefrontal cortex oxygenation. Performance at the secondary task was determined by the lengthening duration of the time production in comparison to the criterion. Greater production times are associated with greater levels of mental workload. Prefrontal cortex oxygenation was obtained by averaging oxygenation levels of the 16 voxels into a single measure. The overall score of the NASA-TLX was also used as an offline validation of the various levels of difficulty of the task.

## Results

Repeated-measures  $2 \times 2$  ANOVAs were carried out to test whether the effects of control difficulty and processing load were statistically significant on the three measures of workload, namely NASA-TLX, performance at the secondary task, and oxygenation level of prefrontal cortex.

### *Subjective Workload*

Figure 2 shows mean subjective workload scores (i.e. overall NASA-TLX) in each experimental condition. The ANOVA carried out of these data revealed significant main effects of processing load,  $F(1, 11) = 25.01, p < .001$ , and difficulty of control,  $F(1, 11) = 4.70, p = .053$  (trend significance), indicating higher perceived workload with high processing load and hard control. However, the two-way interaction was not significant,  $F(1, 11) < 1$ .

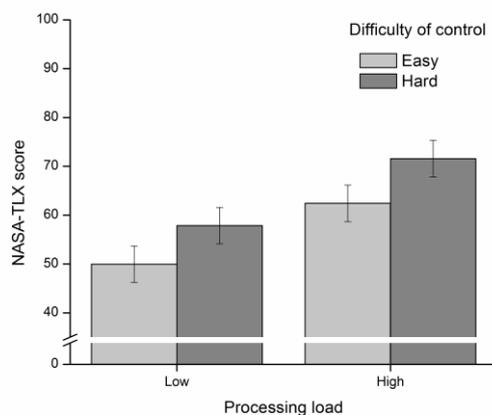


Figure 2. Mean NASA-TLX scores (+SE) by experimental conditions.

### Performance on the Secondary Task

Greater production times are associated with greater levels of mental workload. Figure 3 shows mean production time in each experimental condition. The target time was 2,000 ms. The ANOVA carried out of these data revealed significant main effects of processing load,  $F(1, 11) = 48.46$ ,  $p < .001$ , indicating greater production times in high processing load. However, both the effect of difficulty of control,  $F(1, 11) < 1$  and the two-way interaction  $F(1, 11) = 2.94$ ,  $p = .09$  were not significant.

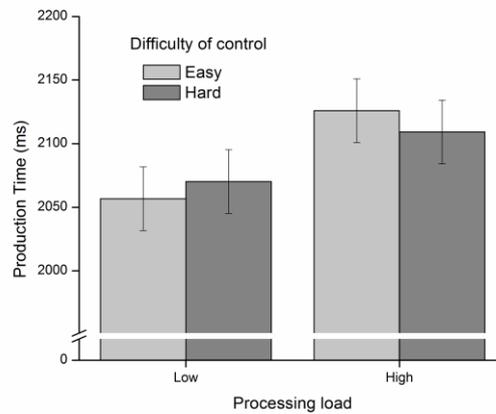


Figure 3. Mean production times in ms (+SE) by experimental conditions.

### Hemodynamics of Prefrontal Cortex

Average oxygenation levels of the 16 voxels were averaged into a single prefrontal oxygenation level (Takeushi, 2000). Figure 4 shows mean oxygenation level in each experimental condition. The ANOVA carried out of these data revealed significant main effects of processing load,  $F(1, 11) = 9.22$ ,  $p < .05$ , difficulty of control,  $F(1, 11) = 11.25$ ,  $p < .01$  (*i.e.* the mean oxygenation level increased with both processing load and control difficulty), and two-way interaction,  $F(1, 11) = 47.55$ ,  $p < .001$ . Oxygenation level increases from easy to hard control difficulty in the low processing load conditions, and it decreases in the high processing load conditions.

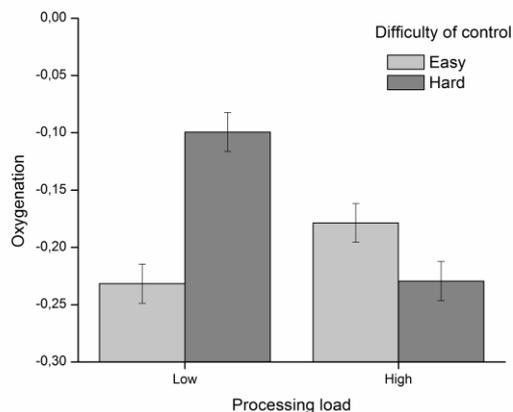


Figure 4. Mean oxygenation (+SE) by experimental conditions (normalized data).

## Discussion

Subjective workload was relatively high and varied across experimental conditions, showing that the variations were significant from the perspective of the operators performing the task. This result is an indication that the task was engaging to the subjects.

The secondary task performance was not affected by difficulty of control. This may be explained by the low sensitivity of the secondary time-production task at detecting workload associated with motor control. This finding is in line with previous research, showing that motor control may involve different resources than the ones required in timing (Robertson et al., 1999). Conversely, processing load affected secondary task performance. This result suggests that the sensitivity of this measure is adequate for detecting increased processing load during the execution of a task; a finding consistent with previous research (see Block et al., 2010). These findings do not invalidate the use of the dual-task paradigm for online measurement of workload; however, they imply that the selection of the appropriate secondary task for assessing mental workload is critical as it might not be sensitive to a wide variety of task demands. The measure of workload based on the performance at a secondary task would benefit from a characterization of the demands for which the latter is sensitive. This characterization would specify the conditions under which the measure could operate.

Mental workload as measured by oxygenation levels of the prefrontal cortex varied as a function of both processing load and control difficulty. Although this result is similar to previous findings (Ayaz et al., 2011; Takeuchi, 2000), it also shows that interactions exist between subsets of task difficulty (in this case, control difficulty and processing load). This must be taken into account in the development of a neurophysiological model of mental workload; such a model cannot be calibrated solely on the basis of processing load, for instance, as its effect on oxygenation level is modulated by difficulty of control.

From a practical standpoint, our results suggest that neurophysiological measures may exhibit complex patterns that cannot be directly associated with mental workload. Future work should further investigate how the latter issue could be resolved. For instance, one option would be to calibrate a neurophysiological model of mental workload with performance at a secondary task in a simulated ROV environment. Such a calibration could be performed by machine learning algorithms in order to best capture potential non-linear relations. If successful, this model could later be used to predict mental workload in a real ROV situation.

Overall, these findings suggest that: (1) the task used in the current study seems to be engaging for subjects; (2) the dual-task paradigm has the potential to capture some aspects of mental workload; and (3) the effect on oxygenation level is modulated by various sources of task difficulty.

## References

- Ayaz, H., Shewokis, P. a, Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2011). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage*, *59*, 36-47.
- Block, R.A., Hancock, P.A., & Zakay, D. (2010). How cognitive load affects duration judgments : A meta-analytic review. *Acta Psychologica*, *134*, 330-343.
- Cooke, N.J. (2006). Human factors of remotely operated vehicles. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*, 166–169.
- Cummings, M. L., & Mitchell, P. J. (2008). Predicting controller capacity in supervisory control of multiple UAVs. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *38*, 451–460.
- Dehais, F., Causse, M., & Tremblay, S. (2011). Mitigation of Conflicts with Automation: Use of Cognitive Countermeasures. *Human Factors*, *53*, 448-460.
- Dehais, F., Causse, M., Vachon, F., & Tremblay, S. (2011). Cognitive conflict in human-automation interactions: A psychophysiological study. *Applied Ergonomics*, *43*, 588-595.
- Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*, 904–908. SAGE Publications.
- Parasuraman, R., & Riley, V. (1997). Humans and Automation : Use, Misuse, Disuse, Abuse. *Human Factors*, *39*, 230-253.
- Parasuraman, R. & Wilson, G.F, (2008), Putting the Brain to Work: Neuroergonomics Past, Present, and Future, *Human Factors*, *50*, 468-474.
- Robertson , S. D. Zelaznik , H. N., & Lantero , D. (1999). Correlations for timing consistency among tapping and drawing tasks: evidence against a single timing process for motor control. *Journal of Experimenta Psychology. Human. Perception Performance*, *25*, 1316-1330.
- Sheridan, T. (2011). Adaptive Automation, Level of Automation, Allocation Authority, Supervisory Control, and Adaptive Control: Distinctions and Modes of Adaptation. *Systems, Man and Cybernetics, Part A: Systems*, *41*, 662-667.
- Takeuchi, Y. (2000). Change in blood volume in the brain during a simulated aircraft landing task. *Journal of occupational health*, *42*, 60-65.

- Thomas, L.C., & Wickens, C.D. (2001). Visual Displays and Cognitive Tunneling: Frames of Reference Effects on Spatial Judgments and Change Detection. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 45, 336-340.
- Visser, E. D., & Parasuraman, R. (2011). Adaptive Aiding of Human-Robot Teaming: Effects of Imperfect Automation on Performance, Trust, and Workload. *Journal of Cognitive Engineering and Decision Making*, 5, 209-231.
- Williams, K.W. (2004). *A Summary of Unmanned Aircraft Accident / Incident Data : Human Factors Implications*. Security (p. 18). Oklahoma City.