Measuring Mental Workload: Assessing the FaceLAB4 Eye Tracking Device

Dion Parera

School of Psychology, University of Western Australia

CEED Partner: Defence Science and Technology Organisation (DSTO)

Abstract

Understanding mental workload is a key point in human factors research. One of the more recent developments in the measurement of mental workload is the remote eye tracking FaceLAB4 instrument and software developed by 'Seeing Machines' (2004). This project will assess the reliability and validity of the FaceLAB4 device and software as a measure of mental workload. The results of this study are expected to further knowledge on alternative methods of measuring mental workload as well as the relationship that several psychophysiological measures have with subjective measures of mental workload. The accurate measurement of mental workload is advantageous to the DSTO in that it allows the organization to better design computer interfaces to reduce the mental workload of submariners, thereby increasing their performance and combat efficiency

1.0 Introduction

As modern technology has made it increasingly possible for humans to reduce the amount of physical work they have to perform, it is easy to presume that we are now at a stage where our overall workload has been significantly reduced. However, though our *physical* workload has to an extent been reduced by technological advancements, the supervision and management of these technological systems has now imposed a *mental* form of workload on us. The concept of mental workload is a key component in the study of human factors, and has lead to much research and analysis over recent years.

Mental workload can be defined as being the difference between the cognitive demands of a particular job or task, and the available attentional capacity of the operator (Wickens, 1992). It is the task demands that are placed upon an individual, and the individual's ability to cope with those demands. The mental workload ratio is such that mental workload is high when the difference between the cognitive demands of a task and the operator's attentional resources are low, or indeed, when the cognitive demands outstrip available attentional resources; conversely, when more mental resources are available than the task demands, then mental workload is said to be low (Kantowitz, 1988).

When a person is overloaded (high mental workload), then there is an associated increase in stress, fatigue errors and accident rates. Alternatively, too low a workload (underloading) can lead to boredom, complacency and "mental lapses" (Rubio et al., 2004). Knowledge regarding a person's available mental capacity when performing a task is beneficial since it may have implications for training and/or task and equipment design.

2.0 Project Background

The Defence Science and Technology Organisation (DSTO) is interested in employing objective metrics to support development of their measurement techniques of mental workload, and has commissioned this CEED project to assess a relatively new measure of mental workload, as well as determining the relationship between several mental workload metrics. Eye tracking and physiological measurements will be taken using the FaceLAB4 eye tracker and BIOPAC, respectively – new instruments that have yet to be used by the DSTO experimentally on a large scale and whose reliability and validity are still to be determined.

With advanced knowledge on the instruments' capabilities, the DSTO would be better able to use the instruments in designing submarine interfaces to reduce the mental workload of personnel. The effect of this reduced workload is two fold. Firstly, it will serve to reduce the stress levels of submariners, thus increasing their performance and efficiency, which is especially crucial in combat situations. Since the first priority of a submarine crew is to ensure the safety of the submarine and its personnel, a well-designed and intuitive interface would make it easier and less stressful for the submariners to maintain that level of safety.

Secondly, if the mental workload associated with each task can be reduced, then several tasks may be combined to be performed by one person. This has the effect of reducing the number of personnel required, which is of value during the current shortage of skilled applicants. A lesser number of personnel would also reduce training costs. Training costs would also be reduced if the computer interfaces were more intuitive and easier to use, since less time would have to be taken to train personnel on how to use them.

3.0 The Measurement of Mental Workload

Techniques for measuring mental workload fall into three general paradigms: (i) performance-based measures, (ii) subjective procedures, and (iii) physiological indices (Veltman & Gaillard, 1996). Given the aims of the project, and since performance measures have been shown to be a relatively poor measure of mental workload (Veltman & Gaillard, 1996), the current project will focus upon physiological indices and subjective procedures as measures of mental workload.

3.1 Physiological Measures

It is well known that physiological measures are indicative of physical activity levels. For example, extent of physical exertion and heart rate are positively correlated. These physiological measures (e.g. heart rate, respiration rate, skin conductance) can also be used to reveal aspects of human mental behaviour, performance and/or capacity which are otherwise unobservable (Bentley, Kieboom & Morris, 2007), such as mental activity.

Prior to technological advancements within the past decade or so, the hardware used to take psycho-physiological measures were large and impractical to move and the collection and analysis of the resultant data was tedious and prone to error (Bentley, Kieboom & Morris, 2007). As such, the technology was limited to laboratory use and was not appropriate for mental workload assessments in the field. Present-day computer equipment such as the BIOPAC is far more powerful as well as portable. Developments in software usability have made the collection and analysis of the data much more accessible and user-friendly.

Using the BIOPAC, the current study will use three psychophysiological measures to assess mental workload: Heart Rate, Respiration Rate, and Skin Conductance. Each of these metrics have been previously validated as accurate measures of mental workload.

3.1.1 Heart Rate

Heart rate is a well-established measure of mental workload and is typically measured using an electrocardiogram (ECG). At rest, the normal human heart rate is approximately 72 beats per minute (BPM) (Andreassi, 2000). The general application in psychophysiology is that increases in heart rate indicate increases in mental demands and workload (Wilson, 2002).

3.1.2 Respiration Rate

Respiration (the supplying of oxygen to the body's cells and the removal of carbon dioxide) is also a common biometric associated with the level of mental workload. The most commonly measured aspect of respiration by psycho-physiologists is respiration rate. The normal resting respiration rate for humans is approximately 12-16 breaths per minute (Stern, Ray & Quigley, 2001), and this changes with increases in mental strain.

3.1.3 Skin Conductance

Skin conductance (or galvanic skin response, GSR) is a measure of the electrical conductivity of the skin and is reliant upon eccrine sweat glands on the hands and feet which respond primarily to psychological stimulation, whereas other sweat glands respond mainly to temperature changes (Stern, Ray & Quigley, 2001). The higher the psychological stimulation (brought about by stress or emotion), the higher the sweat output of these glands, and therefore the higher the skin conductance. Therefore, it follows that in circumstances of higher mental workload (and consequently higher stress), skin conductance increases.

3.2 Ocular Measures

To assess the FaceLAB4 eye tracker, eye movements will also be focused upon as a measure of mental workload.

The FaceLAB4 system purchased by the DSTO works by using remote cameras to track the eyes' movements; this is achieved through focusing the camera(s) on such ocular features as the corneal reflection of a light source, or the pupil centre (Duchowski, 2003). To differentiate between head and eye movements, the FaceLAB4 eye tracking cameras focus on the features of a person's face (such as the corners of their eyes, eyebrows, mouth, etc.) to measure both eye and head movement concurrently.

3.2.1 Blink Frequency and Duration

An eye blink occurs when the upper and lower eyelids touch momentarily, thereby temporarily hiding the eye (Andreassi, 2000). In a relaxed state, humans generally blink approximately 15 to 20 times per minute, with the majority of these being *involuntary blinks*, that is, blinks occurring spontaneously to ensure the surface of the eye remains adequately moist (Andreassi, 2000).

In tasks requiring a close monitoring of events, blink frequency has been shown to decrease in order to obtain as much visual information as possible, and to decrease the time during which no information can be obtained (Wilson, 2002).

Conversely, several studies have indicated that blink frequency increases when an individual is stressed and under an increased mental workload (e.g. Poole & Ball, 2004). An increase in blink frequency can also be taken as an indicator of general negative mood states such as nervousness or fatigue (Tecce, 1992).

3.2.2 PERCLOS

PERCLOS (*Per*centage of Eye *Clos*ure) is a recently developed parameter that has been touted as the most reliable and valid ocular measure of fatigue (Mallis, 1999). Research has shown that mental workload and fatigue tend to co-vary with increased task demands (Kakizaki, Oka & Kurimori, 1992).

PERCLOS is defined as the percentage of time that a person's eyes are more than a given percentage closed, usually between 70% and 80% (FaceLAB 4.2 User Manual, 2004). It has often been used to assess drivers' fatigue and drowsiness. PERCLOS then is a useful measure to determine whether a task is inducing a physical "tiredness" on an individual and how this may impact upon their perception of their mental workload.

3.3 Subjective Measures

The most common and easily administered method used to measure mental workload however, is still through the use of subjective questionnaires. The subjective assessment of workload is based on the presumption that an increase in mental workload is linked to perceived effort, and that this can be adequately reported via questionnaires.

As well as the aforementioned physiological and ocular measurements, this project will also use the NASA-TLX (NASA Task Load Index) rating scale to offer a subjective measure of participants' workload, with the aim of determining whether changes in a person's psychophysiological state are accompanied by a related change in the person's subjective experience of their workload.

4.0 Current Project

Given the measurement of several biometrics, it is hypothesized that if ocular movement reflects mental workload then – (i) they will correlate with other measures of workload, and (ii) they will show criterion related validity.

Additionally, a factor analysis will indicate whether the various metrics are measuring the same underlying latent construct. If all the measures relate to mental workload, then high inter-measure correlations and a single factor are expected to emerge.

5.0 Experimental Method

A total of 43 students from the University of Western Australia (UWA) have taken part in the study thus far, consisting of 15 males and 28 females, with a mean age of 22 years (*SD* = 3.75).

5.1 Experimental Task: Unmanned Underwater Vehicle (UUV) Simulation

Physiological and ocular measures were obtained while participants performed a piloting task in a computer-based unmanned underwater vehicle (UUV) simulation. Mental workload was manipulated by having the participants perform two separate scenarios differing in difficulty, which was manipulated by altering the display style and underwater terrain. Since the complexity of the simulation requires the participant to use numerous cognitive processes, it is an ideal tool for the assessment of multiple measures of mental workload.

The simulation was capable of presenting three different forms of display: Baseline, Inside-Out and Outside-In. A screenshot of the Baseline display is shown in Figure 1:

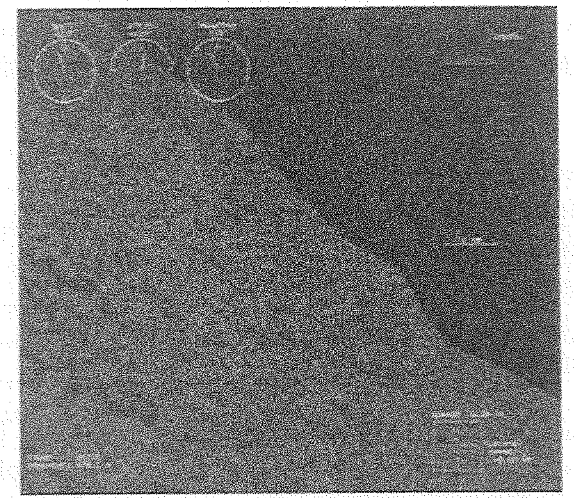

Figure 1. Baseline display design.

The Inside-Out and Outside-In displays differ in whether the vehicle or the earth's coordinates are used as the frame of reference. The Inside-Out display was characterized by a display in which a stationary symbol representing the vehicle is placed in the middle of the display and is used as the point of reference (Donovan & Triggs, 2006). In contrast, an Outside-In display design uses the symbol for the horizon as the stationary point of reference whilst the vehicle symbol moves to indicate the vehicle's orientation (Donovan & Triggs, 2006). These display designs are illustrated in Figure 2.

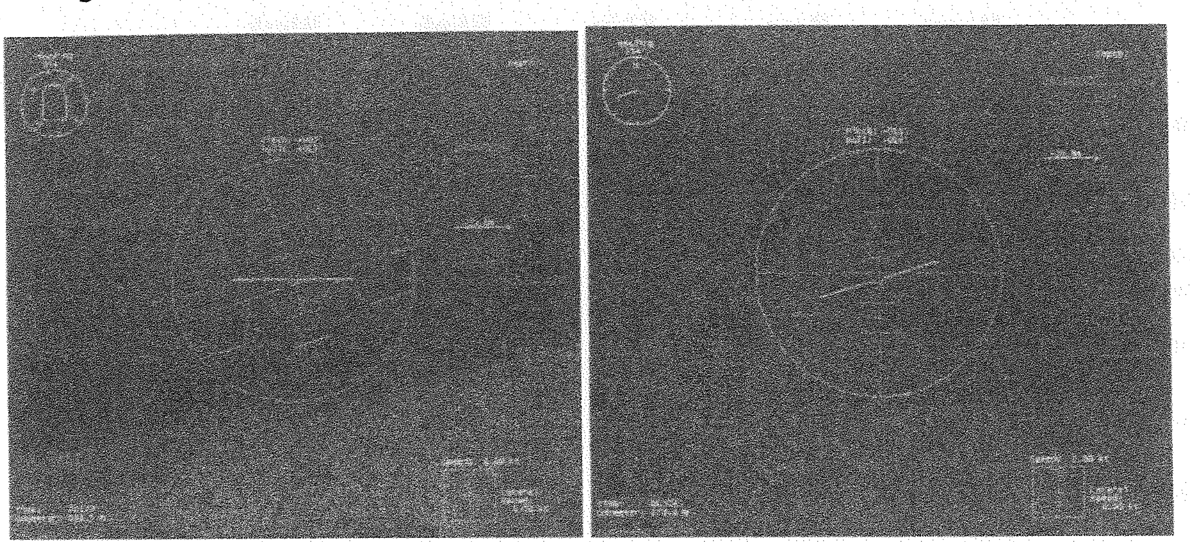

Figure 2. Outside-In (left) and Inside-Out (right) Display Designs

Past research using the UUV simulation has shown that the Outside-In display is generally regarded as more intuitive and easier to use, with overall performance significantly improved when using it (Donovan & Triggs, 2006). This provides the rationale for using the Outside-In display as the "Easy" difficulty condition and the Inside-Out display as the "Hard" condition in the current study. Additionally, the underwater terrain in the "Easy" condition was flat and level, whereas the terrain in the "Hard" condition was made hilly and undulating.

Participants were required to pilot their UUV in order to find and reach several waypoints, which appeared on the display as orange diamonds floating in the water column or on the sea floor. Additionally, in order to further increase the participant's mental workload, the simulation would pause and black out at random intervals during each scenario and present a question to the participant designed to assess their knowledge of the vehicle's navigational status at that moment. Each question was asked twice, thus pausing the simulation a total of ten times; participants were required to verbalise their answers. The questions asked are presented in Table 1 below.

Table 1. Navigational questions for the UUV simulation

1. Estimate your current pitch angle (in degrees).
2. Estimate your current roll angle (in degrees).
3. Estimate your current heading (in degrees).
4. Estimate your current depth (in metres).
5. Estimate your current speed (in knots).

After each scenario, the participant was given a copy of the NASA Task Load Index (NASA-TLX) questionnaire to complete. When the participant had completed both scenarios and completed the two NASA-TLX questionnaires, they were asked which display style of the two they preferred, and their role in the experiment ended.

6.0 Preliminary Results

As data is still being collected, only preliminary analyses have been conducted. A correlational analysis was performed on the data acquired thus far, the results of which are presented in Table 2 below.

Table 2. Pearson Correlations between the Mental Workload Metrics

	Heart Rate	Resp. Rate	Skin Conduct.	Blink Freq.	Blink Dura.	PERCLOS	NASA-TLX
Heart Rate	1.00*	0.04	-0.05	-0.01	0.04	0.02	0.18
Resp. Rate	---	1.00*	0.02	-0.10	-0.17	-0.04	-0.25*
Skin Conduct.	---	---	1.00*	-0.11	-0.08	0.14	-0.13
Blink Freq.	---	---	---	1.00*	0.20*	0.20*	0.09
Blink Dura.	---	---	---	---	1.00*	0.25*	-0.04
PERCLOS	---	---	---	---	---	1.00*	0.12
NASA-TLX	---	---	---	---	---	---	1.00*

* $p < .05$

Thus far, only few correlations are significant, and all are weak correlations at best. Of note is that Blink Frequency is significantly and positively correlated with Blink Duration ($r = 0.20, p < .05$) and PERCLOS ($r = 0.20, p < .05$). Additionally, PERCLOS is significantly and positively correlated with Blink Duration ($r_{(84)} = 0.25, p < .05$).

One-tailed paired samples t-tests revealed a significant difference in NASA-TLX scores, $t_{(42)} = -3.52$, $p < .05$, with the Hard condition being consistently rated as being more workload intensive than the Easy condition.

There was also a significant effect for Heart Rate, $t_{(42)} = -1.86$, $p < .05$, with a higher rate in the Hard condition than the Easy. PERCLOS significantly decreased in the Hard condition, $t_{(42)} = 2.18$, $p < .05$. There were no other significant differences between the conditions for any other measures.

A possible explanation for the lack of significant results is that since the physiological and ocular measures are tracking variables, any effects present may be getting "washed out" over the length of the task. Therefore, it was deemed more effective to conduct additional analyses specifically comparing the physiological and ocular measures at the points when the navigational questions were asked, and comparing them to the participant's baseline physiological and ocular states at the beginning and end of each scenario. From the twelve data sets that were analysed as such thus far, it was found that Respiration Rate significantly decreased when a question was asked, $t_{(119)} = 2.25$, $p < .05$. Additionally, Skin Conductance significantly increased when a question was asked, $t_{(119)} = -9.75$, $p < .05$.

No other metrics reached significance, however since the bulk of the data is still to be analysed in this way, it is expected that the significance of the results would change as more data is analysed.

7.0 Discussion

The significant difference in NASA-TLX scores between the conditions indicates that at a subjective level at least, the Hard condition was perceived to be more difficult than the Easy condition. The significant increase in Heart Rate for the Hard condition is to be expected given past research linking increases in Heart Rate with increases in mental strain (Wilson, 2002). The significant decrease in PERCLOS for the Hard condition can be explained by the fact that since the Easy scenario was flat, level, and overall easier to navigate, it may have been more conducive to making the participant more bored and drowsy than in the Hard scenario which required the participant to be more mentally active.

The significant decrease in Respiration Rate during the periods when a navigational question was asked can be explained by the fact that the participant was verbalising their answer to he navigational question, thus causing the decrease in Respiration Rate. The increase in Skin Conductance is congruent with what is known as the *startle response*, referring to a group of physiological responses caused by a sudden and unexpected stimulus (Andreassi, 2000). The physiological responses elicited by such a stimulus include an increase in blink frequency and an increase in heart rate, as well as an increase in skin conductance (Andreassi, 2000). With further analysis the data, it is expected that Blink Frequency and Heart Rate would also reach significance.

The significant correlations between the ocular measures illustrate the interlinked nature of the ocular measures – that as Blink Frequency and Duration increase, so does PERCLOS, since it is a measure of the amount of time that the eyes are closed. This demonstrates the criterion-related validity of the measures.

Given that the current project is constrained by the limited time available for testing, further research may benefit from a longer experimental task which allows tracking of the ocular metrics across a longer period of time. Alternatively, an experiment of the same length or even shorter could be conducted with a task with an increased effect size, that is, an increased difference between the "Easy" and "Hard" conditions.

8.0 References

Andreassi, J. L. (2000). *Psychophysiology: Human Behaviour and Physiological Response*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Bentley, T., Kieboom, H., & Morris, A. (2007). Psycho-Physiological Measurement in Design and Evaluation: What It's Good For. *SimTecT 2007 Simulation Conference: Simulation - Improving Capability and Competitiveness*. Brisbane, Australia.

Donovan, S-L., & Triggs, T. (2006). *Investigating the Effects of Display Design on Unmanned Underwater Vehicle Pilot Simulation*. Fishermans Bend, Victoria: Maritime Platforms Division, Defence Science and Technology Organisation.

Kakizaki, T., Oka, T., & Kurimori, S. (1992). Changes in Mental Workload and Fatigue During Performance of a Mental Task. *Sangyo Igaku*, vol. 34(6), pp. 565-573.

Kantowitz, B. H. (1988). Mental Workload. In P. A. Hancock (Ed.) *Human factors psychology*. Amsterdam: Elsevier.

Mallis, M. M. (1999) *Evaluation of Techniques for Drowsiness Detection: Experiment on Performance-based Validation of Fatigue-tracking Technologies*. (Doctoral dissertation, Drexel University, 1990).

Poole, A., & Ball, L. (2004). Eye tracking in human-computer interaction and usability research: current status and future prospects. *Encyclopedia of Human Computer Interaction*. Idea Group.

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX and Workload Profile methods. *Applied Psychology: An International Review*, vol. 53(1), pp. 61-86.

Stern, R.M., Ray, W.J., & Quigley, K.S. (2001). *Psychophysiological Recording*. New York: Oxford University Press.

Tecce, J. J. (1992). Psychology, Physiology, and Experimental Psychology. In *McGraw-Hill Yearbook of Science and Technology*. New York: McGraw-Hill.

Veltman, J. A. & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, vol. 42, pp. 323-342.

Wickens, C. D. (1992). *Engineering psychology and human performance*. New York: HarperCollins.

Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology*, vol. 12(1), pp. 3-18.