CONTINUOUS DETECTION OF WORKLOAD OVERLOAD: AN FNIRS APPROACH

Horia A. Maior, Matthew Pike, Max L. Wilson, Sarah Sharples

Horizon, Nottingham University
Mixed Reality Lab, Nottingham University
Mixed Reality Lab, Nottingham University
Human Factors, Nottingham University

Functional Near-Infrared Spectroscopy (fNIRS) is a brain imaging technique that offers the potential to provide continuous, detailed insight into human mental workload, enabling an objective means of detecting overload conditions during complex tasks. When compared to other brain imaging techniques, fNIRS provides a non-invasive, portable and reliable measure that lends itself towards more ecologically valid settings. Our findings confirm a correlation between fNIRS and NASA-TLX subjective workload questionnaire. Our results provide novel insights into fNIRS and its relation to mental workload, and we propose the use of fNIRS as a continuous objective tool for detecting task overload situations.

Introduction

Humans have limited capabilities for processing information (memorising, making decisions and performing tasks), therefore designing tasks for users requires careful consideration. Understanding and identifying users' limitations has always been a challenge within work contexts, our aim is to advance the current understanding in one particular respect - workload overload.

Mental workload (MWL) is described by Hart and Staveland (1988) as a relationship between the mental processing capabilities and the demands imposed by a task. Non-optimal MWL level will result in human performance issues such as slower task performance and errors such as slips, lapses or mistakes.

Measurements such as primary task performance, secondary task performance, and subjective ratings are commonly used methods of measuring MWL. These surveys are usually obtained after the task has been completed, commonly missing essential information about user's experiences during the task.

To address this issue, we are using a non-invasive, real time brain monitoring technique called functional Near Infrared Spectroscopy (fNIRS) to objectively measure participants' physiological changes (indicative of brain activity and MWL) during tasks. fNIRS is an emerging neuroimaging technique used for monitoring brain activation (Villringer, 1993). It has the properties for being non-invasive, portable, inexpensive and suitable for periods of extended monitoring relative to other neuroimaging techniques. fNIRS measures the hemodynamic response - the delivery of blood to active neuronal tissues and it is designed to be placed directly upon a participants scalp, typically targeting the prefrontal cortex (PFC). While some other brain sensing techniques like functional Magnetic Resonance Imaging (fMRI) require minimal or no movement from users, fNIRS can be successfully used while seated naturally at a computer (Solovey, 2009). Further, because fNIRS is an optical based technology rather than electrical (such as Electroencephalography (EEG)), it permits more natural movements such as those associated with using a computer without introducing significant artefacts to the data.

Measuring Mental Workload

To use fNIRS as a tool for measuring MWL, we first need to understand what cognitive processes fNIRS is detecting. Peck (2013) identified three reasons that support fNIRS being an appropriate tool for measuring MWL:

1. fMRI studies (D'Esposito, 1999) have confirmed that a decrease in deoxygenated hemoglobin indicates an increase in brain activity. When a brain region becomes active, it requires more oxygen. To meet these demands, there is an increase in oxygenated hemoglobin, resulting in a decrease of deoxygenated hemoglobin.
2. Peck (2013) found a correlation between the NASA-TLX (Hart and Staveland, 1988) subjective questionnaire and deoxygenated hemoglobin levels in fNIRS data during a visual task.
3. Peck (2013) successfully managed to distinguish between various levels of n-back visual tasks with fNIRS, suggesting that the different levels of n-back tasks induce different levels of MWL. Confirming with the fMRI findings, fNIRS deoxygenated hemoglobin levels had lower values during 3-back task compared with 1-back task.

We believe that the inclusion of this novel measurement complements existing task evaluation measures such as NASA-TLX. We must note the potential negatives associated with this type of technology. fNIRS is an emerging technology therefore does not have the associated supporting research proving its correctness. Studies have correlated the measurements to those observed with fMRI (Toronov, 2001), specifically the BOLD signal. Additionally, in the current state of technology, fNIRS can only be used to detect a level of workload (high or low), leaving a distinct lack of mapping between the readings recorded with fNIRS and the actual cognitive or emotional states.

Detecting Overload with fNIRS

MWL can be described as the amount of resources an operator uses when performing a specific task. These resources are limited; therefore, a problem arises when a task requires the operator to use more resources than are maximally available. This state is known as a human operator overload, and normally results in a significant drop in performance. Therefore, it is important to consider the operators optimum level of workload (not overloaded) throughout the task.

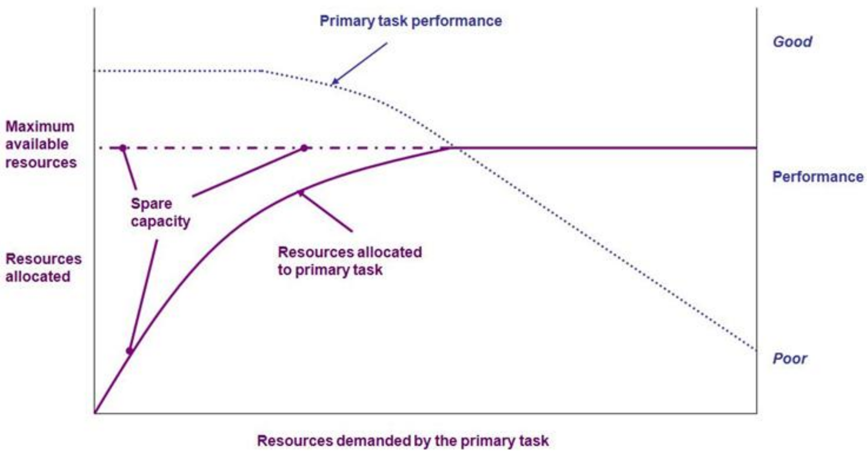

Figure 1: Limited Resource Model

One theoretical model of workload presented by Megaw (2005) is the Limited Resource Model (adapted from Kahnemanns (1973) and Wickens (Wickens and Hollands, 1999)). One of the model's assumptions is that the "limited processing capacity" (which Kahenemanns refers to as "attentional resources") has to be shared between a number of physiological processes such as perception, cognition and response. The Limited Resource Model graph (see Figure 1) describes the relationship between the task demands (Resources demanded by the primary task, the x-axis on the graph), the resources allocated to the task (the left y-axis on the graph), and the impact on task performance (the second y-axis on the graph). When task demands increase, more resources need to be allocated (therefore the spare capacity decreases). When allocated resources reach a point near the maximum available resources, a drop in performance is expected as the operator cannot cope with the task demands.

We propose using an fNIRS device as a continuous measure for detecting Resources allocated in accordance with the Limited Resource Model. Specifically we are interested in using fNIRS for detecting overloaded situations (the states where resources allocated are near the maximum available resources and the spare capacity is minimal).

Methodology

The aim of this study is to identify how fNIRS is suitable in the application of detecting MWL (specifically MWL overload). We aim to validate the measures obtained with fNIRS by correlating it against measures recorded via the subjective MWL questionnaire NASA-TLX. With these aims in mind, we formulate the following research questions:

1. How can we identify MWL using fNIRS?
2. Is there a correlation between fNIRS and NASA-TLX?

To answer these research questions we devised a task that tested participants at solving mathematical problems of varying difficulty under two separate conditions, baseline (C1) and verbal (C2). The baseline required participants to simply solve the mathematical problem, whereas the verbal condition introduced a nonsense verbal utterance ("Blah") that participants were required to verbalise whilst solving the mathematical problem.

From the research questions above, we propose the following hypotheses:

A. There will be a significant difference in performance between conditions C1 and C2.
B. There will be a relationship between the NASA-TLX ratings and fNIRS brain data.
C. There will be a difference in MWL between conditions C1 and C2.

Task

A mathematical problem was chosen as it allowed us to vary difficulty (and therefore resource requirements). This problem is a variation on what is commonly known as the countdown problem. Participants have sixty seconds to get as close to a target number as possible, using six given numbers. Each number may only be used once and participants can use addition, subtraction, multiplication, and division to reach the target number.

Sixteen versions of the task were generated at varying difficulties across the two conditions. Difficulty was classified by one researcher and two independent judges. Difficulty was judged in four categories: easy, quite easy, quite hard, and hard. Inter-rater agreeability was confirmed with a Cohen's Kappa test, where the researcher achieved scores of 0.6419 (substantial agreement) with the first independent judge, and 0.8571 (almost perfect agreement) with the second.

Participants

Twenty participants (14 Males, 6 Females) with an average age of 28.55 years were recruited to take part in the study. All participants had normal or corrected vision and reported no history of head trauma or brain damage. Participants provided informed consent, and were compensated with £15 Amazon vouchers.

Procedure

Participants were fitted with the fNIRS brain imaging device, which was placed upon their forehead targeting the PFC. Participants performed two practice runs of the task (under baseline condition) to familiarise themselves and reduce the impact of learning under the first condition.

Conditions were counterbalanced and each condition included eight of the experimental tasks described previously. For each of the eight tasks in each condition, participants were given sixty seconds to attempt the problem. All calculations were performed mentally and no pen and paper was provided. After the sixty seconds had elapsed, participants were prompted to enter the number they had achieved during the calculation period.

To avoid participants simply entering the target number, they were prompted to recall their solution. The solutions provided by participants were recorded by the researcher on pen and paper. After each condition, participants completed a standard NASA TLX form to subjectively rate their MWL during the task.

We collected a variety of data during the study:

1. Performance - We measured primary task performance according to distance from the target answer as a percentage, where 100% was the perfect answer.
2. NASA-TLX - We used a weighted NASA-TLX questionnaire, based on the weighted average ratings of six subscales including, in order: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration.
3. fNIRS Data - fNIRS data was recorded using a fNIRS300 setup and the associated COBI Studio recording software provided by Biopac Systems inc. 16 channels of data were recorded, providing oxygenated (HbO) and deoxygenated (Hb) hemoglobin readings in each of the channels. Oxygenation values were calculated using the Modified Beer-Lambert Law (MBLL) (Villringer and Chance, 1997) and the data was pre-processed using the NIRS-SPM software package (Ye, 2009).

Results

The aim of this study is to identify whether fNIRS is suitable in the detection of MWL, with the eventual aim of using the imaging technique to detect workload overload. Our hypotheses state that MWL detection in fNIRS should correlate with the measurements observed in the NASA-TLX scale, and thus show that fNIRS is indeed capable of identifying MWL. We report below results that agree with these hypotheses.

The aim of hypothesis A was to identify if there was a significant difference in performance between conditions C1 and C2. In other words, we investigate whether or not participants felt overloaded during condition C2 (as it places additional demands on users through the continuous verbalisation of nonsense words). We did not identify any significant differences in performance between the two conditions (Pair 8 in Table 1). We can attribute this to the findings of Geddie (2001), who in Chapter 3, states: "two systems with the same level of overall performance may impose quite different levels of workload on operators". In relation to MWL, and specifically overload situations, it is not necessarily the case that our task was not demanding enough to elicit an overload state. Rather, since the performance measure used here is an average across all problems in each condition, some of the overload situations may be hidden through averaging.

Hypothesis B states that there is a statistical relationship between NASA-TLX and the measures obtained from the fNIRS device. Correlations were found to support this hypothesis:

- A Pearson correlation ($r = -0.340$, $p = 0.03$) exists between overall deoxygenated hemoglobin and the mental effort subscale measure of NASA-TLX. This finding agrees with (Peck, 2013) who found that decreases in deoxygenated hemoglobin correlated with increased mental effort in NASA-TLX.
- A Spearman test ($r=-0.352$, $p=0.02$) identified a negative correlation between the Total oxygenation (HbO + Hb) and Mental Demand subscale from NASA-TLX questionnaire.

Hypothesis C stated that a difference in MWL would be observed between conditions C1 and C2 in the study. We found a significant amount of evidence to support this hypothesis. A t-test on the mental demands subscale from NASA-TLX reveals a significant difference between C1 and C2 ($p = 0.025$, Pair 2 in Table 1). Similarly, for the physical demands and mental effort scales from NASA-TLX, C2 was significantly more demanding ($p = 0.012$) and required more effort ($p = 0.04$) than C1 (Pair 3 & 6 in Table 1). This supports Geddies' (2001) findings, showing that despite there being no impact in performance, there was a difference in participants' MWL.

Table 1: Study Findings

Paired Samples Test		t	df	Sig. (2-tailed)
Pair 1	Weighted Nasa Score C1 - WeightedMWL2 C2	-1.042	19	.310
Pair 2	Nasa-TLX Mental Demands C1 - Nasa-TLX Mental Demands C2	-2.437	19	.025
Pair 3	Nasa-TLX Physical Demands C1 - Nasa-TLX Physical Demands C2	-2.785	19	.012
Pair 4	Nasa-TLX Temporal Demands C1 - Nasa-TLX Temporal Demands C2	.980	19	.339
Pair 5	Nasa-TLX Performance C1 - Nasa-TLX Performance C2	-1.045	19	.309
Pair 6	Nasa-TLX Mental Effort C1 - Nasa-TLX Mental Effort C2	-2.204	19	.040
Pair 7	Nasa-TLX Frustration C1 - Nasa-TLX Frustration C2	-.071	19	.944
Pair 8	Average Distance From Target C1 - Average Distance From Target C2	.005	19	.996
Pair 9	Average Time Spent C1 - Average Time Spent C2	1.547	19	.138
Pair 10	OverallOxy C1 - OverallOxy C2	-1.643	19	.117
Pair 11	OverallDeOxy C1 - OverallDeOxy C2	-.940	19	.359
Pair 12	OverallTotal C1 - OverallTotal C2	-1.468	19	.158
Pair 13	OxyL C1 - OxyL C2	-1.163	19	.259
Pair 14	DeOxyL C1 - DeOxyL C2	-1.021	19	.320
Pair 15	TotalL C1 - TotalL C2	-1.730	19	.100
Pair 16	OxyR C1 - OxyR C2	-.586	19	.565
Pair 17	DeOxyR C1 - DeOxyR C2	.114	19	.910
Pair 18	TotalR C1 - TotalR C2	-.348	19	.731

Despite not finding any significant differences with fNIRS data alone, we found fNIRS to be complementary to existing measures such as NASA-TLX. We believe that increasing the number of participants would increase power, reduce type II error and positively impact our findings. Also we note that overload situations might be hidden through averaging the fNIRS data over conditions. Data that cannot be identified with the NASA-TLX questionnaire might be detected using fNIRS.

Conclusion

The aims of this research were to investigate whether fNIRS is a suitable technique for detecting MWL. To do this we devised a mathematical task with varying difficulties to elicit different workload requirements from participants. Additionally, we introduced another condition which included a nonsense utterance, requiring additional, non-complementary resources.

The correlation of fNIRS and NASA-TLX provides an insight into fNIRS ability to detect MWL. Coinciding with the findings from (Peck, 2013) and (D'Esposito, 1999), we believe that fNIRS is in fact capable of detecting MWL. This demonstrates the need to continue researching how to detect specific workload states such as MWL overload. We believe that the combination of complementary measures (NASA-TLX, fNIRS and other measures of MWL) will provide greater insight into MWL.

Our future work will look at expanding on these findings with the aim of being able to detect varying degrees of MWL in accordance with the Limited Resource Model. We have proposed to look at problems on an individual basis. fNIRS property of being a continuous measure enables the detection of MWL states that are not observable in NASA-TLX data alone.

References

D'Esposito, M., Zarahn, E., & Aguirre, G. K. (1999). *Event-related functional MRI: implications for cognitive psychology*. Psychological bulletin, 125(1), 155.

Geddie, J. C., Boer, L. C., Edwards, R. J., Enderwick, T. P., & Graff, N. (2001).*NATO Guidelines on Human Engineering Testing and Evaluation* (No. RTO-TR-021). NATO RESEARCH AND TECHNOLOGY ORGANIZATION NEUILLY-SUR-SEINE (FRANCE).

Hart, S. G., & Staveland, L. E. (1988). *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research.* Human mental workload,1(3), 139-183.

Hollands, J. G., & Wickens, C. D. (1999). *Engineering psychology and human performance*. New Jersey: Prentice Hall.

Kahneman, D.(1973) *Attention and effort*,. Prenctice Hall, Englewood Cliffs, NJ.

Megaw, T. (2005) *The definition and measurement of mental workload*. Evaluation of human work, 525-551.

Peck, E. M. M., Yuksel, B. F., Ottley, A., Jacob, R. J., & Chang, R. (2013). *Using fNIRS brain sensing to evaluate information visualization interfaces*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 473-482). ACM.

Solovey, E. T., Girouard, A., Chauncey, K., Hirshfield, L. M., Sassaroli, A., Zheng, F., & Jacob, R. J. (2009). *Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines*. In Proceedings of the 22nd annual ACM symposium on User interface software and technology (pp. 157-166). ACM.

Toronov, V., Webb, A., Choi, J. H., Wolf, M., Michalos, A., Gratton, E., & Hueber, D. (2001). *Investigation of human brain hemodynamics by simultaneous near-infrared spectroscopy and functional magnetic resonance imaging*. Medical physics, 28, 521.

Villringer, A., Planck, J., Hock, C., Schleinkofer, L., & Dirnagl, U. (1993). *Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults*. Neuroscience letters, 154(1), 101-104.

Villringer, A., & Chance, B. (1997). *Non-invasive optical spectroscopy and imaging of human brain function*. Trends in neurosciences, 20(10), 435-442.

Ye, J. C., Tak, S., Jang, K. E., Jung, J., & Jang, J. (2009). *NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy*. Neuroimage,44(2), 428-447.