

The Realtime Assessment of Mental Workload by Means of Multiple Bio-Signals

Masterthesis Report

Methodology and Statistics for the Behavioural,
Biomedical and Social Sciences

Utrecht University

Bart-Jan Boverhof, 6000142

Thesis Supervisor

Prof.dr.ir. B.P. Veldkamp

Date

October 14, 2020

Word count

747

1 Introduction

The topic of mental workload (MWL) has received widespread attention across a variety of different fields, amongst others the field of ergonomics (Young, Brookhuis, Wickens, & Hancock, 2015), human factors (Pretorius & Cilliers, 2007) and neurosciences (Shuggi, Oh, Shewokis, & Gentili, 2017). A simple definition of MWL is the demand placed upon humans whilst conducting a specific task. As pointed out by de Waard (1996), such a definition is too shallow, for it defines workload solely externally. It is of importance to acknowledge that MWL is person-specific, for the amount of experienced MWL ushered by a given task differs across people (De Waard & te Groningen, 1996).

A commonly employed measure of workload is the NASA-Task Load Index questionnaire, operationalizing workload in clusters of six different dimensions (Hart, 2006). A potential problem arising with such measurements, is that they are conducted post-experiment, posing the risk of jeopardizing the quality of the data by introducing subjectivity. An example of such a danger resides in the observer bias, proving that actors participating in an experiment tend to over-exaggerate the treatment effect when having to report it themselves post-experiment (Mahtani, Spencer, Brassey, & Heneghan, 2018). An approach inherently less prone towards such biases is the employment of bio-signals during an experiment.

The main objective of the current research is to provide a flexible framework with which researchers can measure MWL using the specific bio-signals (hereafter modalities) that are most feasible within their context. The latter framework will aim to classify MWL in real-time, i.e. whilst the experiment takes place. Ultimately, this line of research pursues the ability to conduct a dynamic experiment, the development of which can be altered by in real-time. The following research question is delineated:

Which type of framework is most feasible for analyzing MWL in real-time by means of various bio-signals?

2 Data and Methods

In order to answer the aforementioned question, data will be collected utilizing three different modalities, surrounding which the framework will be build. In order to attempt this in real-time, a technique called Lab Streaming Layer (LSL) is utilized during the data collection process. This enables data streams from all modalities to be properly synchronized, whilst satisfying the principle of modularity (i.e. the possibility to add various modalities) (Kothe, Medine, & Grivich, 2018). More elaborate explorations of the appropriate architecture of this real-time framework are currently being endeavoured, but fall outside the scope (and word limit) of this proposal. The framework will be created in Python, by utilizing the machine learning toolbox Pytorch (Paszke et al., 2017)

2.1 Data

The experimental setting for data collection is the spaceship bridge simulator videogame Empty Epsilon, in which respondents are required to carry out tasks on a virtual spaceship bridge (Daid & Nallath, 2016). This experiment is instituted by the Brain Computer Interfaces (BCI) testbed lab hosted by the University of Twente (UT) and carried out in cooperation with Thales. Ethical approval has been attained at the UT, and will be requested at the UU in the near future.

2.2 Methods

2.2.1 Electroencephalogram (EEG)

The first modality is a technique called electroencephalogram (EEG), which detects electrical activity in the brain using electrodes. Research by (Schirrneister et al., 2017) contrasted several differently designed convolutional neural networks (ConvNets) against the baseline method for EEG data (FBCSP). Altogether, a deep ConvNet with four convolutional-max-pooling blocks was found to perform well, and will be utilized consequently (Schirrneister et al., 2017).

2.2.2 Galvanic Skin Response (GSR)

The second modality is the technique called Galvanic Skin Response (GSR), measuring sweat gland on the hands and hereby inferring arousal. This technique was found to be an objective predictor of MWL (Shi, Ruiz, Taib, Choi, & Chen, 2007). For analyzing GSR data, a hybrid model combining a convolutional neural network (CNN) with a long-short-term memory (LSTM) model is proposed (Sun, Hong, Li, & Ren, 2019). Such a model hybrid model was contrasted with a regular CNN (2 layers), after which the hybrid model was found to perform best (Dolmans, Poel, van 't Klooster, & Veldkamp, in press). Consequently, a model stacking 2 convolutional layers with 2 LSTM layers will be utilized.

2.2.3 Photoplethysmography (PPG)

The third modality constitutes Photoplethysmography (PPG), which is a technique utilized to measure heart rate, and is also proven to be a suitable indicator of MWL (Zhang et al., 2018). A hybrid model, stacking two convolutional layers and two LSTM was found to perform best, for which this model will be utilized (Biswas et al., 2019).

2.3 Model Evaluation

The performance of the real-time framework will be assessed by contrasting it with a non-real-time framework. Specifically, the quality of performance for each modality individually will be assessed. This validation will be endeavored by means of six widely utilized performance measures within the fields of deep and machine learning: accuracy, sensitivity, specificity, PPV, NPV and F1. The framework that performs best across these measures will be considered the optimal framework.

References

- Biswas, D., Everson, L., Liu, M., Panwar, M., Verhoef, B.-E., Patki, S., ... others (2019). Cornet: Deep learning framework for ppg-based heart rate estimation and biometric identification in ambulant environment. *IEEE transactions on biomedical circuits and systems*, 13(2), 282–291.
- Daid, & Nallath. (2016). *Empty epsilon multiplayer spaceship bridge simulation*. URL: <https://github.com/daid/EmptyEpsilon>. GitHub.
- De Waard, D., & te Groningen, R. (1996). *The measurement of drivers' mental workload*. Groningen University, Traffic Research Center Netherlands.
- Dolmans, T., Poel, M., van 't Klooster, J.-W., & Veldkamp, B. (in press). Perceived mental workload detection using intermediate fusion multi-modal networks.
- Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, pp. 904–908).
- Kothe, C., Medine, D., & Grivich, M. (2018). Lab streaming layer (2014). URL: <https://github.com/scn/labstreaminglayer> (visited on 02/01/2019).
- Mahtani, K., Spencer, E. A., Brassey, J., & Heneghan, C. (2018). Catalogue of bias: observer bias. *BMJ Evidence-Based Medicine*, 23(1), 23.
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... Lerer, A. (2017). Automatic differentiation in pytorch.
- Pretorius, A., & Cilliers, P. (2007). Development of a mental workload index: A systems approach. *Ergonomics*, 50(9), 1503–1515.
- Schirrmester, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggenesperger, K., Tangermann, M., ... Ball, T. (2017). Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11), 5391–5420.
- Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). Galvanic skin response (gsr) as an index of cognitive load. In *Chi'07 extended abstracts on human factors in computing systems* (pp. 2651–2656).
- Shuggi, I. M., Oh, H., Shewokis, P. A., & Gentili, R. J. (2017). Mental workload and motor performance dynamics during practice of reaching movements under various levels of task difficulty. *Neuroscience*, 360, 166–179.
- Sun, X., Hong, T., Li, C., & Ren, F. (2019). Hybrid spatiotemporal models for sentiment classification via galvanic skin response. *Neurocomputing*, 358, 385–400.
- Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics*, 58(1), 1–17.
- Zhang, X., Lyu, Y., Hu, X., Hu, Z., Shi, Y., & Yin, H. (2018). Evaluating photoplethysmogram as a real-time cognitive load assessment during game playing. *International Journal of Human-Computer Interaction*, 34(8), 695–706.