Update on Our Ongoing Evaluation of Our Workload Monitoring System During a Simulated Event

Thomas de Groot, Manon L. Tolhuisen, Rafal Hrynkiewicz, Tije Oortwijn and Johan de Heer
Thales - Human Behavior Analytics Lab, Zuidelijke Havenweg 40, 7554 RR Hengelo, The Netherlands

Keywords: Human Behaviour Analytics, Human Cognitive State Monitoring.

Abstract: In complex task environments, especially during crisis management scenarios, optimal performance is essential. Workload levels are associated with levels of performance. We aim to develop a human cognitive monitoring tool that aids in reaching optimal performance levels by providing insight into the experienced workload of one or multiple operators. Here, we give an update on our ongoing evaluation findings regarding our human workload monitoring tool that was tested in operational security operations centers in the context of the IMPETUS project.

1 INTRODUCTION

This paper addresses our workload assessment tool that was evaluated in a series of simulated crisis management scenarios in two smart cities. The human workload monitoring system (WMS) was developed to monitor the workload of human operators in real time during crisis management operations and provide feedback when workload levels are suboptimal. Operators were working in a Security Operations Centre (SOC) and the workload assessment tool sent out alerts when observed levels of workload were different than expected. This work was done on a European project named IMPETUS (Gorman et al., 2023). We report an update on our evaluation findings.

2 WORKLOAD ASSESSMENT TOOL

Our tool focuses on the monitoring of human workload and team collaboration since both constructs directly impact human performance. The inverted U-shape relation between the human state and performance suggests a tipping point indicating that at an appropriate level of state maximum performance can be expected. Note that this relation has been associated with complex task environments such as command and control settings. In addition, there is no absolute value attached to the human state level, and the definition of what is the appropriate state is likely to vary across operators. But the general notion is clear, too low, or too high workload levels reduce the level of performance (Yerkes, 1908).

During a crisis management scenario, SOC operators are interacting with their equipment and with each other, while performing their specific tasks. Mostly, operators must process information coming from multiple input channels, both visual and auditory, and subsequently act by communicating with the system or their team members.

The number and complexity of the tasks will affect the experienced workload and alter both the operator's behavior and biosignals. From these two, biosignals are a more suitable information source for workload assessment since the alteration of the biosignals in response to workload is involuntary, while behavior can intentionally be altered (Giannakakis, 2022). We measure biosignals continuously, in real time, and as unobtrusively as possible.

The end goal of the monitoring tool is to provide timely feedback and assure that operators can perform their tasks without being overloaded or overstressed which might impede their work and introduce unwanted reduced effectiveness of the operators.

Feedback provided by the WMS is based on a rule system that is configurable. For example, the rule could be to generate an alert when workload levels exceed a certain threshold for a longer period, say 3 minutes. Assessments are shown as feedback in a configurable amount of detail, on individual and

de Groot, T., Tolhuisen, M., Hrynkiewicz, R., Oortwijn, T. and de Heer, J.
Update on Our Ongoing Evaluation of Our Workload Monitoring System During a Simulated Event.
DOI: 10.5220/0011962100003622
In *Proceedings of the 1st International Conference on Cognitive Aircraft Systems (ICCAS 2022)*, pages 73-76
ISBN: 978-989-758-657-6

aggregated (team) levels, to person or persons of choosing, in the form of a (digital) dashboard. This feedback can also be used in the context of a Human Machine Teaming application to close the loop (Fig. 1) and adapt the human-machine interface to the operators being assessed to balance the cognitive load and drive mission effectiveness.

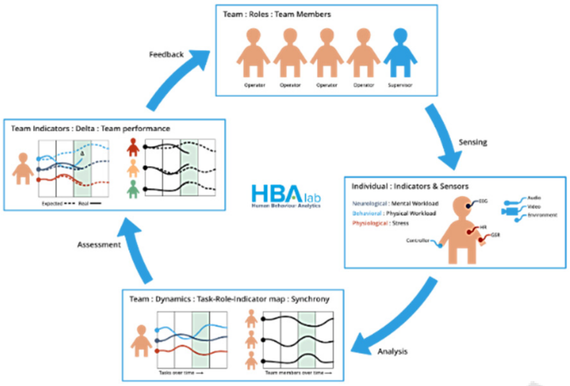

Figure 1: Schematic representation of the assessment flow: team members are sensed, neuro-physiological measurements are analyzed, workload and team collaboration are assessed, and feedback is available for interventions such as load balancing.

3 EVALUATING THE HUMAN WORKLOAD MONITORING SYSTEM IN AN OPERATIONAL ENVIRONMENT

We tested the WMS in the IMPETUS project. The goal of IMPETUS is to provide city authorities with new means to address security issues in public spaces. Using data gathered from multiple sources, the project aims to facilitate the detection of threats and help human operators to deal with threats by making better-informed decisions. IMPETUS will detect potential threats by using AI techniques to search social media and the deep/dark web for unusual and suspicious activities and to analyze available smart city data complying with ethical, legal, and societal issues (ELSI). Threats will be classified and assessed to determine an appropriate response using an approach that employs the power of AI to support situational awareness, human judgment, problem-solving, sense-making, and decision-making. The project builds on tested technologies but enhances and combines them in a coherent and user-centered solution that goes beyond the state-of-the-art in key areas such as detection, simulation & analysis, and intervention. For IMPETUS we configured our workload assessment tool to the requirements of the project Part of the research in the IMPETUS project is to evaluate all tools in an operational environment provided by two partner cities Oslo (Norway) and Padova (Italy).

3.1 Method

We tested our WMS (Fig 2) in various SOCs in the cities of Oslo and Padova. We monitored the workload of SOC operators interacting in a series of simulated events. The SOC operators were wearing a Muse S, which captures both PhotoPlethysmo-Gram (PPG) and ElectroEncephaloGram (EEG) signals. We collected data from a single PPG sensor that was located on the skin. The PPG sensor records the capillary blood flow that can be translated to the local pulse. Four EEG electrodes located at the scalp recorded the electrical activity of the brain, i.e., EEG signals. From these signals, we computed features, including multiple heart rate variability features and the EEG spectral band power of the Theta, Alpha, Beta, and Gamma frequency bands.

Before the simulated events, we collected data from the operator while performing a calibration task. These data were used to train personalized models for the mental, emotional, and physical workload. The calibration task included a controlled environment in which the operator had to perform multiple tasks with varying difficulty to simulate the range in the cognitive load that the operators may experience in a SOC. Data management, privacy, and ethical concerns were part of the tool design process.

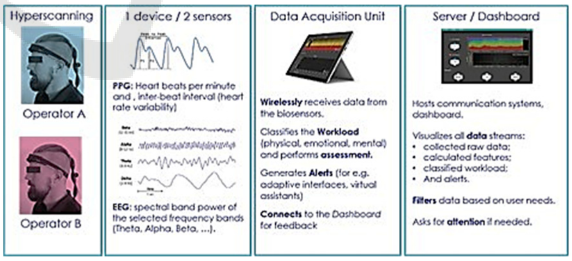

Figure 2: Human workload monitoring tool.

The initial evaluation sessions were performed in November and December 2021 in the SOC in Oslo town hall and the Cyber SOC and Municipality CCTV SOC in the City of Padova, respectively. As reported earlier (De Groot, 2022), during these sessions we evaluated the usability and interpretability of the HMT in collaboration with multiple SOC operators in a single-operator and multiple-operator setting. Subsequent evaluation sessions were organized in August and September 2022. During these sessions, the usability of the WMS was further evaluated in a scripted scenario that simulated a realistic event. Additionally, we evaluated the experienced and hypothesized impact of the WMS on real-life operations, now and in the future.

From the first event, in Oslo, we learned that the collection of the calibration data for the models the day before the actual evaluation was suboptimal since the operators were distracted by the organization of the actual event. Therefore, for the second event in Padova, we collected the calibration data a few weeks upfront. During the evaluation, the operators used several other IMPETUS tools during several roleplays just outside the city hall. In Oslo, a single operator participated in the simulated event. The Oslo city hall was closed to the public, but the SOC was still operational. At the Cyber SOC in Padova, the operator was able to fully focus on the evaluation scenario. In parallel, the workload assessment tool captured the operators' neuro-physiological data using the Muse S, which was processed in real time resulting in a workload classification (low, medium, high) for each workload dimension (physical, emotional, mental). If the workload classifications remained high for over three minutes an alert was generated and visualized in the dashboard.

The test included an explanation of the workload assessment tool dashboard. The alerts were presented on the IMPETUS dashboard which was accessible by the supervisor of the SOC. The supervisor also had access to the WMS dashboard.

The assessment tool enabled the supervisor to act when a team member was mentally and physically under or overloaded and/or stressed. Both operators and their supervisors were included in the debriefing/interview afterward. During the debrief we asked the operators and supervisors about their experience with the tool, specifically focusing on:

- the time needed for the calibration and training of the tool,
- the impact of the tool on their normal activities,
- the impact of wearing the sensors,
- the influence of the HMT on the experienced workload, and potential cyber security issues.

3.2 Results and Discussion

From the evaluation, we have learned that:

- The calibration task, including the setup, takes around 2 to 3 hours. The design choice is based on personalized workload models given the variability in perceived workload between subjects. However, the enrolment procedure could be optimized with an online learning procedure where we start off with a generic workload model that is periodically or continuously adapted over time.
- The collection of calibration data is preferably collected at a moment when the operator is not distracted by other activities. This points to a potential bias or skewness in the dataset used for model training that may impact our workload prediction. However, it is challenging to design a calibration task that results in a calibration dataset with evenly distributed workload labels since the experienced complexity of the calibration task varies between subjects.
- The dashboard of the workload assessment tool is considered informative, and easy to use by both supervisor and operator. Alerts and feedback are preferably shown to the supervisors instead of the operators because the operators experienced increased workload due to the visibility of the HMT results. The operators and supervisors saw the potential of monitoring the workload levels during daily life. However, further exploration is needed to determine the actions required after an alert is generated. These issues reflect the operational embedding of human state assessment tools in general. An objective standardized human assessment tool is not part of current procedures and mitigating strategies relating to human error.
- During the simulated scenario, like the previous evaluation, the MUSE S headband was considered comfortable, unobtrusive, and easy to wear. Also, here the design choice is characterized by the trade-off between a number of channels to measure EEG and therefore potentially an increase in model accuracy versus usability requirements related to unobtrusive measurements.

4 CONCLUSIONS

We reported an update on our findings from the evaluation of our WMS during a simulated event that was organized in the context of the IMPETUS project. The WMS was intuitive, and the sensors were not impacting their daily activities. Future work should focus on validating the models and exploring strategies for situations where operators deviate from workload levels that affect their level of performance during crisis management situations.

ACKNOWLEDGEMENTS

This study is part of IMPETUS a research and innovation project on the Intelligent Management of Processes, Ethics, and Technology for Urban Safety. This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 883286. All authors contributed equally to the paper and are listed in alphabetical order. The writers of this paper are also the makers of the WL assessment tool used in the study.

REFERENCES

Gorman (2023) *IMPETUS* https://www.impetus-project.eu

Yerkes, R.M., Dodson, J.D. (1908) *The relation of strength of stimulus to rapidity of habit-formation.* Journal of Comparative Neurology and Psychology. 18 (5): 459–482. doi:10.1002/cne.920180503.

Giannakakis, Grigoriadis, Giannakakie *et al.* (2022) *Review on psychological stress detection using biosignals* IEEE Transactions on Affective Computing, 13 (1): 440-460, doi: 10.1109/TAFFC.2019.2927337

De Groot, T., Heer, J., Hrynkiewicz, R., Tolhuisen, M., Oortwijn, T. (2022). *Evaluation of Real-time Assessment of Human Operator Workload during a Simulated Crisis Situation, Using EEG and PPG.* In: Hasan Ayaz (eds) Neuroergonomics and Cognitive Engineering. AHFE International Conference. AHFE Open Access, vol 42. AHFE International, USA. doi: 10.54941/ahfe1001818