Neuroergonomics and Cognitive Engineering, Vol. 42, 2022, 50–56
https://doi.org/10.54941/ahfe1001818

Evaluation of Real-Time Assessment of Human Operator Workload during a Simulated Crisis Situation, Using EEG and PPG

Thomas de Groot, Johan de Heer, Rafal Hrynkiewicz, Tije Oortwijn, and Manon Tolhuisen

Thales Research & Technology NL, Human Behavior Analytics Lab, High Tech Systems Park, Haaksbergerstraat 67, 7554 PA, Hengelo (Ov), The Netherlands

ABSTRACT

We report on initial evaluation findings regarding a human cognitive state assessment tool that was tested in various operational security operations centers (SOC). This paper addresses a part of our iterative development of a human state assessment tool. First, we will introduce our motivation and context for developing such a tool. Second, we focus on the envisioned human state assessment flow. Third, we describe an instantiation of our tool in a European project named IMPETUS[1] where it was evaluated in various operational settings. Results of the evaluations are presented, and we end this paper with some preliminary conclusions.

Keywords: Human machine teaming, Human behavior analytics, Human state monitoring

INTRODUCTION

An important driver for our human state assessment research relates to collaboration or teaming between human and system entities. A Human-Machine Team has been defined (Madni and Madni, 2018) as “a purposeful combination of human and cyber-physical elements that collaboratively pursue goals that are unachievable by either individually”. In addition, Human-Machine Teaming (HMT) has been defined by the same authors as “the dynamic arrangement of humans and cyber-physical elements into a team structure that capitalizes on the respective strengths of each while circumventing their respective limitations in pursuit of shared goals”. And even one step further, Adaptive Human-Machine Teaming as “a context-aware re-organization/reconfiguration of human and cyber-physical elements into a fluid team structure that assures manageable cognitive load while exploiting the respective strengths of human and cyber-physical elements”. Definitions like these have been transformed (Laird et al., 2019) to HMT challenges such as (a) to better understand human capabilities in the context of dynamic situations, and (b) to identify what humans must know about machines

[1] https://www.impetus-project.eu

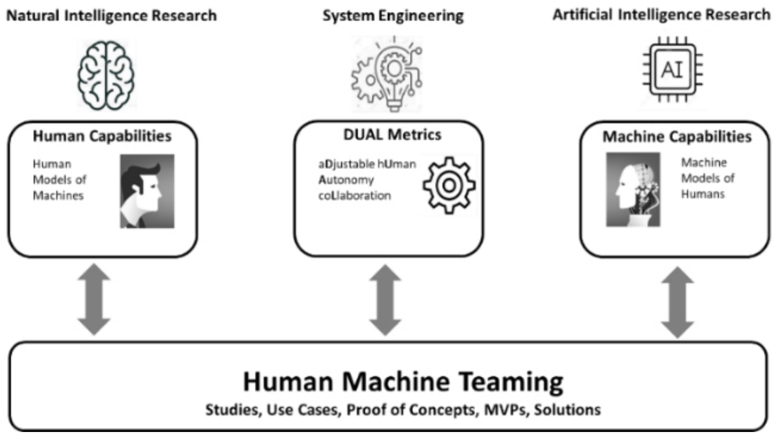

Figure 1: DUAL architecture (Taken from Alix et al., 2021).

to interact with them, including what is required in HMT to establish and maintain trust, (c) how to enhance machine capabilities to enable effective and efficient human machine teams, and (d) to model how machines reason about human teammates. In addition (Alix et al., 2021), for an adaptive human machine teaming system architecture (labeled DUAL in Figure 1) we need to evaluate HMT architectures with metrics such as mission effectiveness, autonomous platform behavior efficiency, human behavior efficiency, human behavior precursors, and collaborative metrics (Pina et al., 2008). Figure 1 illustrates these HMT challenges by bringing together research on human and machine intelligence as well as the system engineering expertise in HMT studies, use cases, proof-of-concepts, minimal viable products, and solutions.

Within the context of this broader stated HMT our focus is on assessing the human or team state as an input to align human and machine capabilities, for example, directed towards load balancing interventions or recommendations. The assessment of human and team state is also considered a necessary step in modelling how machines reason about their human teammates.

HUMAN STATE ASSESSMENT TOOL

Our research and technology roadmaps are directed on the assessment of human and team states. We focus on the development of a real-time assessment tool of human workload and team collaboration since it directly relates to human performance. Previous work suggests a tipping point between human state and - performance. In complex task environments, an optimal level of, for example, human workload correlates with a maximum level of performance. Too low or too high workload levels correlates to lower levels of performances (Yerkes and Dodson, 1908).

Our purpose is to measure biosignals of human operators, who are interacting with their equipment - and each other - while performing their specific tasks in a complex operational environment, such as a command & control room. We aim measuring biosignals continuously, in real-time and

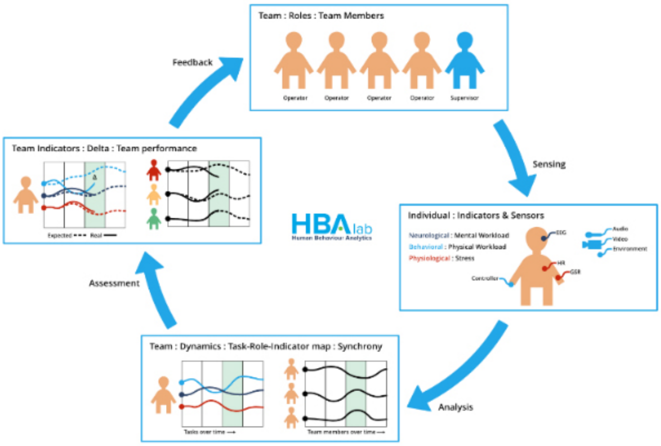

Figure 2: Schematic representation of the assessment flow: team members are sensed, neuro-physiological measurements are analyzed, workload and team collaboration is assessed, and feedback is available for interventions such as load balancing.

as unobtrusively as possible. For that, we are exploring wearable sensors detecting brainwaves (using Electroencephalograms, or EEG), sweat (using Electrodermal Activity sensors or EDA), heartbeat activity (using Photoplethysmogram, or PPG), eye and pupil activity (using eye trackers), local brain oxygenation (using Functional Near-Infrared sensors, or fNIRs) and physical activity (using software sensors capturing user input through keyboard or mouse). Additionally, environmental conditions like sound level and temperature could be recorded (using various sensors) in the room where the operators are at work. Assessments can be shown as feedback in configurable amount of detail, on individual and 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 an HMT application to close the loop and adapt their interfaces or information they provide to the operators being assessed to balance cognitive load and driving mission effectiveness.

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

EVALUATING THE ASSESSMENT TOOL IN AN OPERATIONAL ENVIRONMENT

We developed an instantiation of our assessment tool box based on the requirements derived from 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 detection of threats and help human operators dealing with threats to make 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. Threats

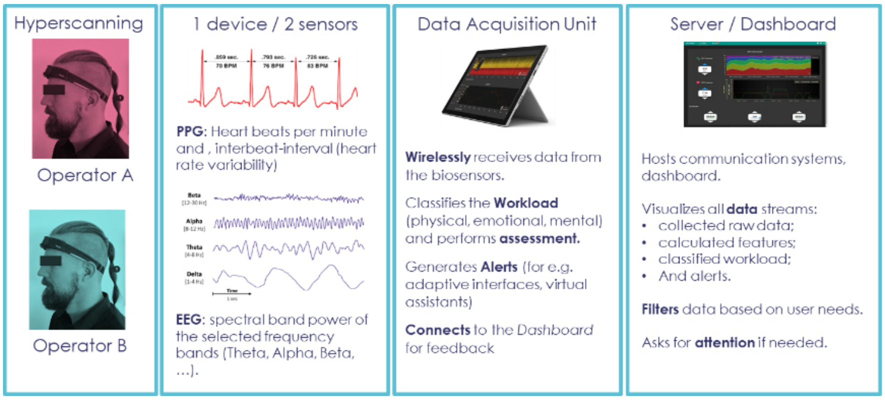

Figure 3: Workload assessment tool.

will be classified and assessed to determine an appropriate response using an approach which employs the power of AI to support human judgement. The project builds on tested technologies but enhance and combine them in a coherent and user-centered solution that goes beyond state-of-the-art in key areas such as detection, simulation & analysis, and intervention. For this project we developed a workload assessment tool for the operators of a Security Operations Center SOC. 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).

We tested our workload assessment tool (see Figure 3) in various SOCs in the cities of Oslo and Padova. We assessed the workload of SOC operators interacting in a series of simulated events. The SOC operators were wearing a Brain Computer Inter-face BCI that captured both PhotoPlethysmoGram (PPG) and ElectroEncephaloGram (EEG) signals. PPG signals were captured by a light sensor placed on the skin. The sensor records local pulse that is generated by capillary blood flow. EEG signals were captured by electrodes located on the scalp which record electrical activity of the brain. From these signals we selected features such as heart beats per minute, interbeat-interval (heart rate variability) and EEG spectral band power of the select-ed frequency bands (Theta, Alpha, Beta). The features were fed into a personalized workload assessment module consisting of a mental workload model, a stress work-load model, and a physical workload model. These models were created based on a labelled dataset that we acquired earlier using a simulation environment that represented various calibration tasks that resemble the cognitive load that human opera-tors may experience in a SOC. Data management, privacy, and ethical concerns were part of the tool design process.

Evaluation tests were performed in November 2021 in the security operation center in Oslo town hall. We extensively worked with one of the operators of that SOC to evaluate our assessment system on usability of the sensors, data collection for model building, user interface ease of use, and interpretability of the classification of work-load. In December 2021 we tested the assessment tool in the Cyber SOC as well as in the

Municipality CCTV SOC in the City of Padova. In particular, the latter environ-ment created a setting where multiple operators were present at the same time work-ing apart together on monitoring the CCTVs and take actions where appropriate. In this case the assessment tool went in hyper-scanning mode and measured and as-sessed the workload of two operators simultaneously.

Method

Before the evaluation a questionnaire was sent out to the cities and their operators to better configure the workload assessment tool according to the acceptance criteria of the operators: these were questions related to their comfort, feelings, and attitudes in using different sensors and providing personal psychological data. Based on the results we decided to use the Muse S brain computer interface that has two sensors: an EEG sensor and a PPG sensor.

In the City of Oslo, calibration tests to train personalized workload models were done 3 weeks prior to the actual validation test in Oslo: one operator participated in the (calibration) test. The calibration test took approximately 4 hours. Using these data, various workload models (physical, emotional, and mental) were trained for this specific operator. During the evaluation the same operator was using another IMPETUS tool during several roleplays just outside city hall. In parallel the workload assessment tool captured the operators' neurophysiological data using a brain computer interface, which was processed in real-time resulting in a workload classification (low, mid, high) for each workload dimension (physical, emotional, mental) and if deviations in these classifications compared to previous classifications occurred over time, then alerts were generated and visualized in the dashboard.

At the Cyber SOC in Padova one operator participated in the test. One day prior to the evaluation the operator performed the calibration task for training the operators' workload model. During the test the operator was working on a laptop (not performing any cyber risk assessments in particular). The operator was interviewed on the usability of the workload assessment tool dashboard that visualized her assessed workload in real-time while being interviewed. In addition, the Cyber SOC supervisor joined the evaluation discussion on the usability and added value of the workload assessment tool. At the CCTV SOC: two operators participated in the test. One day prior to the evaluation they performed the calibration task. Their personalized workload models were trained off-line and implemented in the workload assessment tool. Both operators were performing their normal daily activities as well as remotely viewing some of the IMPETUS tools during the test. The test included an explanation of the workload assessment tool dashboard. Both operators and their supervisor were included in the debriefing/interview afterwards. During the test, deviations in workload levels were detected and appropriate alerts were generated. These alerts were presented on the dashboard that was accessible by the supervisor of the SOC. The assessment tool enabled the supervisor to take actions when a team member was mentally and physically under or overloaded and/or stressed.

Results

Interviews and feedback from the SOC operators and supervisors illustrated that:

- Usability Sensor Set: the brain computer interface was considered comfortable and unobtrusive.
- Data collection and model training: took in total about 2-3 on average per person, which was less than expected and can be done on the same day as running the evaluation tests.
- The dashboard of the workload assessment tool is considered easy to use by supervisor and operator.
- Workload classification: personalized workload models made sense given the current task and situation at hand.
- Workload assessment tool clearly showed operational value.

CONCLUSION

We sketched out the rationale for developing a real-time human state assessment tool in the context of human machine teaming.

We evaluated an instantiation/configuration of a workload assessment tool in the IMPETUS project. We presented the findings of the evaluation, which showed the added values at the one hand side but also pointed to current limitations: new operational procedures are needed that should focus on mitigation strategies in situations where operators are suffering from workload issues affecting their level of performance during crisis management situations.

ACKNOWLEDGMENT

This paper is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No. 883286, project IMPETUS (Intelligent Management of Processes, Ethics and Technology for Urban Safety). All authors contributed equally to the paper and are listed in alphabetical order. The writers of this paper are also the makers of the workload assessment tool used in the study. This paper is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No. 883286, project IMPETUS (Intelligent Man-agement of Processes, Ethics and Technology for Urban Safety). All authors con-tributed equally to the paper and are listed in alphabetical order. The writers of this paper are also the makers of the workload assessment tool used in the study.

REFERENCES

Alix, C. Lafond, D. Mattioli, J., de Heer, J., Chattington M. and Robic, P.O. (2021) "Empowering Adaptive Human Autonomy Collaboration with Artificial Intelligence," in: 2021 16th International Conference of System of Systems Engineering (SoSE), pp. 126–131; doi: 10.1109/SOSE52739.2021.9497497.

Laird, J., Ranganath, C., and Gershmen, S. (2019) "Future Directions in Human Machine Teaming Workshop", Department of Defence USA, https://basicresearch.defense.gov/Portals/61/Future%20Directions%20in%20Hum an%20Machine%20Teaming%20Workshop%20report%20%20%28for%20public%20release%29.pdf.

Madni, A.M. and Madni, C.C. (2018) “Architectural Framework for Exploring Adaptive Human-Machine Teaming Options in Simulated Dynamic Environments”, in: Systems 2018, 6, 44; doi:10.3390/systems6040044.

Pina, P.E., Donmez, B., Cummings, M.L. (2008) “Selecting metrics to evaluate human supervisory control applications”. From Humans and Automation Laboratory (HAL), MIT.

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