Development of an AI Pipeline for Real Time Assessment of Fighter Pilots' Mental State Based on Hybrid Stream Processing

Magnus Bång[1], Jens Alfredson[2], Bence Nagy[1]
Anna-Marija Trinkune[1] and Björn J E Johansson[1]

[1]Department of Computer and Information Science, Linkoping University, SWEDEN
[2]Saab AB, Aeronautics, SWEDEN

magnus.bang@liu.se jens.alfredson@saabgroup.com

ABSTRACT

Future fighter pilots and remote operators will need advanced decision and attention support to cope with increased complexity, uncertain information and multi agent coordination. Monitoring the pilot's mental state and awareness and making it available to the system could enable better human-system collaboration and increased joint performance. Research has shown that different psycho-physiological measurements techniques can be used to assess multiple cognitive and affective states such as mental workload, attention, fatigue as well as job-related variables such as task difficulty and task completion. However, little work has been devoted to real-time assessment of multiple sensing techniques and the temporal sequence of signs. We have developed an experimental AI pipeline for investigating fighter pilot mental states in real time using eye tracking (saccades, fixation times etc), electrodermal activity (EDA) and heart rate variables (e.g., HR and HRV). The system utilizes a hybrid analytics approach comprising data stream processing and machine learning (ML) which enable real time analysis and time-based inference of the different signal events. We report on the advantages and disadvantages of the approach, present results from ongoing empirical experiments with the system and discuss possible applications for advanced attention guidance.

***Keywords**: AI, Fighter pilot, Mental state, Psychophysiology, Fighter aircraft.*

1.INTRODUCTION

Adaptive Automation [1,2] is a promising approach to support operators and keep their workload at appropriate levels. Today, there are a number of sensor technologies that can be worn or embedded in our physical work environments such as eye tracking glasses and smart watches. These developments make it possible to create advanced applications that track pilots' health and cognitive states in relation to the operational tasks and that can provide support - when needed [3]. Hence, future work environments will likely measure the performance, stress and level of attention of individuals and groups with the goal of optimizing and balancing tasks individually and among groups. However, this approach requires methods and algorithms to adequately classify and assess the cognitive state of the operator in terms of workload, stress and attention levels in real time [4,5,6] which can only be achieved through the use of psychophysiological sensors. Future applications include pilot environment for future military concepts, manned as well as unmanned, with suitable levels of autonomy to assist the pilot and decision support to cope with effects of information overload. Further, a separation between various types of automation for specific tasks with known user cognitive demands could be beneficial [7]. Our research aims to explore the potential of pattern identification of various psychophysiological responses to workload induced tasks and attention. The goal is to understand the relationship among these responses and signals to be used for future adaptive automation technologies, with the aim of reducing operators' mental workload, improving attention, hence, securing performance levels.

In this paper, we present explorative and ongoing work on a sequential time-based analysis engine used to classify and validate multiple sensor data and psychophysiological phenomena in real time. Moreover, we present the data collection approach and experimental setup to validate hypotheses on the patterns of the eye, heart and electrodermal response to external stimuli in a pilot environment. First, we provide the background with regards to the psychometric measurements for assessing cognitive states that we use in this research. Second, we discuss the AI-pipeline for processing data streams in real time. Third, we present our experimental setup using eye tracking, heartrate and electrodermal response in a virtual reality environment. In the remainder of the paper, we present pros and cons of the approach.

2. BACKGROUND

Psychophysiological measures such as heart rate (HR), heart rate variability (HRV), electrodermal activity (EDA) as well as eye movements are suitable candidates to investigate subject's stress, workload and attention levels [8,9]. Furthermore, measurements of psychophysiological response "have the advantages of objectivity and are well suited to continual monitoring, supporting augmented cognition for the stressed operator" [10, p 811]. In high complexity environments such as airplane piloting it might even be necessary to simultaneously measure several attributes in order to capture the state of operators' cognitive processes [8]. However, most studies have looked at single physiological parameters and their correlation to the human cognitive states rather than combining multiple variables to strengthen the analytic approach.

2.1 Psychophysiological Responses

Heart Rate and Heart Rate Variability

Fluctuations of the heart are caused by feedback from the central nervous system (CNS) to the peripheral autonomic receptors. HR and HRV are common measures used to assess mental workload levels in individuals. HRV tracks fluctuations in heart rate beat-to-beat and has proven to be indicative of elevated workload levels [11]. Studies comparing differences in HRV measures across different levels of workload have observed significant differences in HRV between low and high workload [c.f. 12]. Part of the evidence suggests that an increased workload presents itself as an elevated HR and decreased HRV [13, 14]. Even though HR and HRV appear to present as inverted to each other, researchers have argued that they display different bodily mechanics and are thus useful to measure in parallel [15].

In order to record HR activity, it is necessary to collect data of the inter-beat interval (IBI) which can further be converted to the well-known measure of beats per minute (bpm). The HRV can be acquired by similar methods, however this measure is highly sensitive to noisy data. Skipped or extra beats of the heart affect the HRV data heavily and can lead to serious errors if not corrected [4]. Other drawbacks of using data obtained from the heart include the sensitivity of HR towards emotional stress as well as physical effort [6].

Electrodermal Activity

EDA is a psychophysiological measure which records changes in skin conductance caused by fluctuations in perspiration level. Perspiration is another physiological attribute controlled by the ANS and can thus be connected to mental state fluctuations. EDA is widely used in research in order to study both physiological and psychological constructs.

In order to record EDA activity, two techniques have been proposed and are currently in use - exosomatic and endosomatic measurement. Exosomatic measurement is the most commonly used of the two and functions by applying a small direct current generated by an external resistor to the skin [14, 16]. When exposed to an external stimulus, the electrodermal level (EDL) either increases (if data is presented as skin conduction) or decreases (if presented as skin resistance). These fluctuations can then be classified as the phasic electrodermal responses (EDRs) [14, 16]. In some cases, EDRs are elicited without the presence of known external stimuli and are then referred to as nonspecific EDRs (NS. EDRs). The Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures [16] recommends a latency window of 1-3 sec for determining if an EDR has been caused by the chosen stimulus, however latencies of 4 sec and longer can occur. EDRs appearing quicker than 1 sec are less likely to be accurate due to processing speed, ANS response speed and the time it takes for sweat to penetrate the skin.

When recording EDA data, it is important to take note of the individual variability in EDR that is likely to occur given demographic differences such as age, gender and culture. Evidence shows that older adults (over 60) experience smaller EDRs than do younger adults. This can likely be attributed to changes in the peripheral and central nervous system caused by age. The amount of active sweat glands also decreases with age and could be a reason for the decrease in EDR levels [16].

Eye Movement

Eye movement measure is most frequently used in research on cognitive processes and mental states as it provides an insight on operator attention allocation. Since visual attention is the primary step in cognitive processing, using eye movement as a psychophysiological measure could benefit in tracking the operators' situational awareness (SA) and mental state [17, 18]. Saccade rate, fixation frequency and duration are some of the measures potentially reflective of cognitive workload levels [19]. Blink rate also serves as a measure for cognitive workload as it correlates with other eye movement measures [4].

Eye tracking and eye movement measures prove a good alternative to measuring operator workload and SA compared to other commonly used measures such as subjective assessments due to eye tracking technology allowing uninterrupted data collection and objective observation. Out of all psychophysiological measures, eye movement and eye tracking data are also considered the most unobtrusive as it usually does not require the operator to be connected to external devices compared to e.g., EDA sensors [17]. A drawback when using eye tracking measures for research purposes stems from the eye being sensitive to light. If not controlled for, changes in e.g., ambient lightning can disturb the data [17]. If using equipment such as virtual reality (VR), however, lighting is always kept at a constant or can be manipulated by the researchers in accordance with study requirements.

The eye movements tracked within this research study are blink rate and saccade rate. Blink rate records the amount of blinks the subject performs during a specified time frame (usually, seconds). Previous studies have observed that increase in workload and especially visual load tends to correlate with lower blink rates [20].

The hypotheses based on these observations is that when operators experience higher visual load, they tend to avoid blinking as much so not to miss important visual cues during task performance. However, results in the literature vary (e.g. [21] where blink rate increased during higher task difficulty) and it has been proposed that differences in blink rate could be indicative of the kind of task subjects are expected to perform [22].

Saccades are the movement of the eye between two fixations and are known to be the fastest movement of the human body. An interesting feature of saccadic movement is that the subject is virtually blind for the duration of the saccade [22]. Different saccadic features are used in a variety of research studies, e.g., whether saccadic rate can be indicative of mental workload. Previous research has proposed that saccadic rate increases when the subject experiences higher workload (e.g. [21]).

3. APPROACH

3.1 System Implementation

The pipeline for analyzing eye tracking, EDA and HR was developed using Apache Flink [23]. Flink is a programming framework and distributed processing engine for computations over unbounded and bounded data streams. Flink is particularly suited for the task of identifying events in continuous data streams and compute the temporal relationship among them. Figure 1 shows the overall architecture of the pipeline.

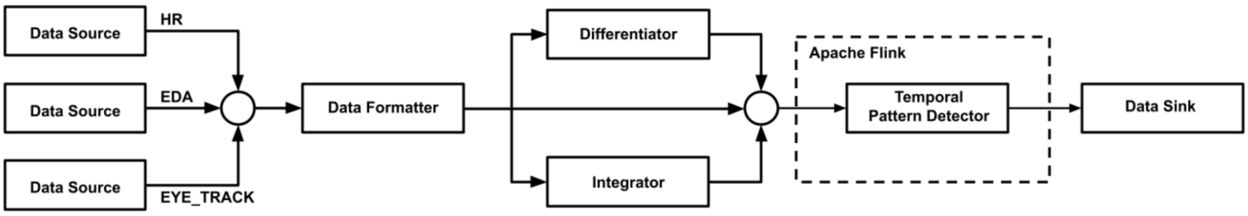

Figure 1: Architecture of the pipeline for analyzing psychophysiological data in real time.

Temporal Pattern Detection Algorithm

A general temporal pattern detector was implemented in Flink. The goal was to be able to detect certain patterns expressed by sets of psychophysiological variables as they change over time. The sought-after pattern was defined using a set of temporal conditions. A temporal condition is bound to a variable, and defines a spatial range and a temporal range for the variable. The spatial range is defined by a lower and an upper bound for the variable. The temporal range is defined by a time offset from the start of the matched pattern and a tolerance. The tolerance sets the range of time around the offset that the condition needs to be fulfilled in. For the condition to be fulfilled, the value of the variable needs to be between the defined lower and upper bound and the distance in time from offset needs to be at most tolerance. Once a condition is fulfilled, it is marked as fulfilled until the detector is reset. Once all conditions in the set are fulfilled, the detector signals that the pattern has been detected using a predefined output packet. If a condition is failed, meaning that the specified temporal range has passed without the condition being fulfilled, the pattern detector is reset and tries to match the pattern from the beginning. Figure 2 shows the principle of the temporal pattern detector.

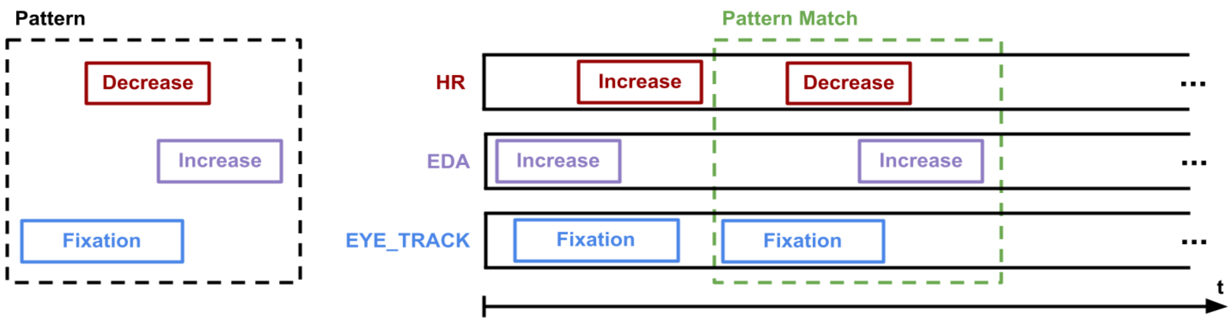

Figure 2: The temporal pattern detection engine classifies events on singular channels and looks for overarching temporal patterns across the channels in real time. In this case the system looks for a pattern and sequence beginning with a pupil fixation (possible object detection), followed by a small HR disturbance and a subsequent rise in EDA (within 1500 ms).

3.2 Experimental Setup

The current experimental setup consists of a HTC Vive Pro Eye Virtual Reality system [24] with integrated eye tracking, Shimmer3 for EDA and HR measurements. A virtual work environment was developed for VIVE in Unity, displaying an operator task similar to NASA's MATB-II [25] placed in 3D space to simulate a cockpit environment. MATB-II is a software that simulates cockpit tasks to evaluate operator performance and workload. This was developed to create a task suitable also for non-pilots. At present, we are investigating if it is possible – using the detector to identify and quantify - the attention level of the pilot. To do this, we have designed a task where the research participant is to maintain the main task (MATB-II, balancing tank level task) as well as identifying and classifying incoming spheres of different types (colors) (see Figure 3).

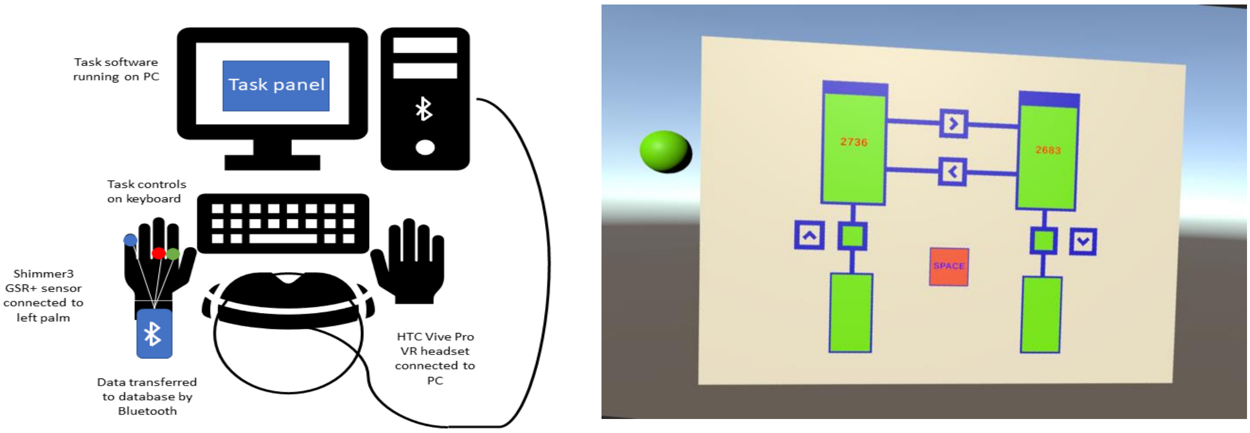

Figure 3: Left: The setup to investigate temporal patterns among EDA, HR, and eye tracking consisting of VIVE VR goggles and Shimmer3 data collection system for capturing EDA and HR. Right: The user task implemented in VIVE consists of a standard modified NASA MATB-II task augmented with an additional attention task consistsing of identifying and classifying different types of spheres that show up intermittently during the test session.

In order to interact with the experimental tasks, participants use a standard type swedish keyboard. The keys used for task interaction are the four arrow keys, space bar and the "Enter" key. The hypothesis is if it is possible for our algorithm to identify if the user can identify a "new" uncommon type of sphere, replicating a new unidentified object in the pilots' 3D view.

4. RESULTS AND DISCUSSION

In this paper, we have presented a first take on a framework for analysing multiple real time psychophysiological signals for future fighter pilot applications. Our framework shows potential, and the system allows for more advanced real time processing such as analysing the significance and timing of fast paced saccade and fixation sequences in relation to the HR and EDA. This would allow a system to know when the operator has observed a high significance object and, potentially, assessed the situation. Moreover, using high frequency eye tracking it should be possible to identify and analyse pupil micropatterns and its potential relation to other physiological signs. Further developments could be to include frontal lobe EEG and FNIR.

Machine Learning methods are promising in this context since new possibilities of identifying hidden patterns and relationships among eye, HR and EDA which are difficult to identify using ordinary research tools and methods (i.e., clustering methods etc). For example, it should be possible to apply standard LSTM [c.f., 26] processing to predict signs and physiological responses of a stimuli. Ready-trained Machine

Learning models can be integrated into our framework and, hence used in our system for real time test.

A challenge with our approach is that the sensors used must provide data with millisecond accuracy which is challenging due to internal signal processing and filtering. Calibration and synchronisation are required, however, lag in signal processing naturally diverges from the main idea of real time assessment of the pilots cognitive state. Nevertheless, and to conclude, the approach proposed provide a basis for assessing and understanding eye tracking data more deeply, by also combining HR and EDA signals in the analysis. This approach is still in its infancy and more research needs to be done to validate it. Future work includes more applied settings for fighter pilot decision making, as well as further refinement of data collection and analysis.

ACKNOWLEDGEMENTS

This research was funded by Swedish Defence Material Administration (HMI HuFLaB) and NFFP (National Aviation Research Programme, NFFP7- 2019-02773) funded by VINNOVA (Swedish Governmental Agency for Innovation Systems), the Swedish Armed Forces and the Swedish Defence Material Administration. The research was also funded by the Swedish Transport Administration (grant TRV 2020/138317).

REFERENCES

[1] Parasuraman, R., Bahri, T., Deaton, J.E., Morrison, J.G., & Barnes, M. (1992). Theory and design of adaptive automation in aviation systems (Technical Report No. NAWCADWAR- 92033-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.

[2] Scerbo, M. W., Freeman, F. G., & Mikulka, P. J. (2003). A brain-based system for adaptive automation. Theoretical Issues in Ergonomic Science, 4, 200-219.

[3] Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neuroscience and Biobehavioral Reviews, 44, 58–75.

[4] Magnusson, S. (2002). Similarities and Differences in Psychophysiological Reactions Between Simulated and Real Air-to-Ground Missions. The International Journal of Aviation Psychology, 12(1), 49–61.

[5] Teo, G., Matthews, G., Reinerman-Jones, L., & Barber, D. (2020). Adaptive aiding with an individualized workload model based on psychophysiological measures. Human-Intelligent Systems Integration, 2.

[6] Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T.-P. (2001). Eye Activity Correlates of Workload during a Visuospatial Memory Task. Human Factors, 43(1), 111–121.

[7] Taylor, G. S., Reinerman-Jones, L. E., Szalma, J. L., Mouloua, M., & Hancock, P. A. (2013). What to Automate: Addressing the Multidimensionality of Cognitive Resources Through System Design. Journal of Cognitive Engineering and Decision Making, 7(4), 311–329.

[8] Lee, K. (2010). Effects Of Flight Factors On Pilot Performance, Workload, And Stress At Final Approach To Landing Phase Of Flight. 123.

[9] Daviaux, Y., Bey, C., Arsac, L., Morellec, O., & Lini, S. (2019). Feedback on the Use of Matb-Ii Task

for Modeling of Cognitive Control Levels Through Psycho-Physiological Biosignals. 7.

[10] Matthews, G. (2016). Multidimensional Profiling of Task Stress States for Human Factors: A Brief Review. Human Factors, 58(6), 801–813.

[11] Muñoz, J., Pereira, F., & Karapanos, E. (2016, September 1). Workload Management Through Glanceable Feedback: The Role of Heart Rate Variability.

[12] Nagasawa, T., & Hagiwara, H. (2016). Workload Induces Changes in Hemodynamics, Respiratory Rate and Heart Rate Variability. 2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE), 176–181.

[13] Shakouri, M., Ikuma, L. H., Aghazadeh, F., & Nahmens, I. (2018). Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: The case of highway work zones. International Journal of Industrial Ergonomics, 66, 136–145.

[14] Waard, D. de. (1996). The measurement of drivers' mental workload. Rijksuniv.

[15] Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator Effort and the Measurement of Heart-Rate Variability. Human Factors, 29(2), 161–170.

[16] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures. (2012). Publication recommendations for electrodermal measurements: Publication standards for EDA. Psychophysiology, 49(8), 1017–1034.

[17] Moacdieh, N. M., Devlin, S. P., Jundi, H., & Riggs, S. L. (2020). Effects of Workload and Workload Transitions on Attention Allocation in a Dual-Task Environment: Evidence From Eye Tracking Metrics. Journal of Cognitive Engineering and Decision Making, 14(2), 132–151.

[18] Yu, C.-S., Wang, E. M.-Y., Li, W.-C., & Braithwaite, G. (2014). Pilots' visual scan patterns and situation awareness in flight operations. Aviation, Space, and Environmental Medicine, 85(7), 708–714.

[19] Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. Biological Psychology, 42(3), 361–377.

[20] Niezgoda, M., Tarnowski, A., Kruszewski, M., & Kamiński, T. (2015). Towards testing auditory–vocal interfaces and detecting distraction while driving: A comparison of eye-movement measures in the assessment of cognitive workload. Transportation Research Part F: Traffic Psychology and Behaviour, 32, 23–34.

[21] Takahashi, K., Nakayama, M., & Shimizu, Y. (2000). The response of eye-movement and pupil size to audio instruction while viewing a moving target. Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, 131–138.

[22] Holmqvist, K. (Ed.). (2011). Eye tracking: A comprehensive guide to methods and measures. Oxford University Press.

[23] Apache Flink: Stateful Computations over Data Streams. (n.d.). Retrieved September 30, 2021, from https://flink.apache.org/

[24] "VIVE Pro Eye Specs & User Guide - Developer Resources." Accessed September 30, 2021.

https://developer.vive.com/resources/vive-sense/hardware-guide/vive-pro-eye-specs-user-guide/.

[25] Comstock, J. R., & Arnegard, R. J. (1992). The multi-attribute task battery for human operator workload and strategic behavior research (NAS 1.15:104174). https://ntrs.nasa.gov/citations/19920007912

[26] Fagerström, J., Bång, M., Wilhelms, D. et al. LiSep LSTM: A Machine Learning Algorithm for Early Detection of Septic Shock. Sci Rep **9,** 15132 (2019). https://doi.org/10.1038/s41598-019-51219-4