Scientific Research Publishing

Journal of Behavioral and Brain Science, 2022, 12, 474-484
https://www.scirp.org/journal/jbbs
ISSN Online: 2160-5874
ISSN Print: 2160-5866

# Cognitive Workload Assessment of Aircraft Pilots

**Maxime Antoine, Hamdi Ben Abdessalem, Claude Frasson***

Departement d'Informatique et de Recherche Operationnelle, Universite de Montreal, Montreal, Canada
Email: maxime.antoine@umontreal.ca, hamdi.ben.abdessalem@umontreal.ca, *frasson@iro.umontreal.ca

**How to cite this paper:** Antoine, M., Ben Abdessalem, H. and Frasson, C. (2022) Cognitive Workload Assessment of Aircraft Pilots. *Journal of Behavioral and Brain Science*, **12**, 474-484.
https://doi.org/10.4236/jbbs.2022.1210027

**Received:** August 25, 2022
**Accepted:** October 7, 2022
**Published:** October 10, 2022

## Abstract

In this research, we study the cognitive workload of aircraft pilots during a simulated takeoff procedure. We propose a proof-of-concept setup environment to gather heart rate, pupil dilation, and brain cognitive workload data during an A320 takeoff within a simulator. Experiments were performed during which we collected 136 takeoffs across 13 pilots for more than 9 hours of time-series data. Moreover, this paper investigates the correlations between heart rate, pupil dilation, and cognitive workload during such exercise and found that a spike in cognitive load during a critical moment, such as an engine failure, augments a pilot's heart rate and pupil dilation. Results show that a critical moment within a takeoff procedure increases a pilot's cognitive load. Next, we used a stacked-LSTM model to predict cognitive workload 5 seconds into the future. The model was able to produce accurate predictions.

## Keywords

Cognitive Workload, Aviation, Heart Rate, Pupil Dilation, Machine Learning, Deep Learning, EEG

## 1. Introduction

Hart and Staveland from NASA describe cognitive workload (CW) as the user's perceived level of mental effort influenced by task load and task design [1]. Because of this, CW has been a hot topic in the research community for the last decades. While many studies exist around measuring CW, few studies are done on how to measure CW in real-time. This method of measuring CW is relevant for different industries where measuring the CW of a person could help predict future behavior and avoid poor decisions. This method is particularly relevant in aviation, where passengers' security is the top priority. CW within the aviation industry, which englobes the mental capacity of a pilot to perform the maneuvers of an aircraft, varies according to the piloting stage and the number of tasks converging at the same time on the pilot. This load can lead to pilot errors with serious consequences. While the aviation industry has a variety of procedures to minimize human errors, they still happen.

The National Transportation Safety Board (NTSB), the organism that tracks aviation accidents in the United States, reports that 80% of all aviation accidents are due to human errors, with sometimes dramatic consequences [2]. Most errors occur during the takeoff or landing procedure. Most errors occur during the takeoff or landing procedure. Because of this, measuring their CW could help the aviation industry better understand a pilot's mental state when errors happen and help pilots during flights by recommendation. To understand and measure the CW of a pilot in real-time, different cognitive and physical measurements need to be combined and analyzed to understand a pilot's mental state at a given time and predict it accordingly. For example, physical measurements such as the airplane metrics combined with other measurements related to CW, such as the heart rate (HR) or the pupil dilation (PD) of a pilot, could be used to predict the pilot's future CW better.

Considering previous problem and investigations, we formulate the following hypotheses:

**Hypothesis 1:** Is it possible to measure the CW of pilots in real-time during a takeoff experience?

**Hypothesis 2:** Is it possible to establish a correlation between the measured CW and the measured PD and HR during a critical event?

**Hypothesis 3:** Is it possible to predict the CW of a pilot based on his previous behavior?

To answer these hypotheses, this paper will be organized as follows: First, a related work section will give an overview of the current work done around the subject of cognitive workload in aviation and research linking CW with other body parts such as HR and PD. Next, a software solution created to measure, monitor, and trigger different events and data from a single interface will be described in the Section 3. The experiment setup and progress will be described in Section 4. Section 5 will describe the results from the experiments and answer the three hypotheses. Finally, Section 6 will describe the future work regarding this paper.

## 2. Related Work

Different measurement methods exist to measure CW, subjective, performance, physiological, and behavioral measures [3]. This research used a combination of physiological and behavioral measures as they offer the ability to measure a person's behavior in real-time. Moreover, different studies found that somebody regions directly relate to a person's cognitive load. These regions include the heart and pupils of a person.

### 2.1. Heart Rate

Sosnowski *et al.* demonstrated that an increase in the task's difficulty increased the HR of the learners [4]. Another study from Jerčić *et al.* used the HR in addition to the PD of each learner to measure its CW and attention [5]. Lang *et al.* showed that recall of pleasant and unpleasant memories prompts HR acceleration, showing that arousal determines the HR of a person [6].

### 2.2. Pupil Dilation

Kahneman *et al.* used PD as part of his empirical foundation for his attention theory. Since then, several studies have been done using pupillary dilation as a proxy for evaluating CW. The pupil dilates when cognitive load increases until task demands exceed the available cognitive resources [7]. Different studies used PD as a proxy for the cognitive load. Zekveld *et al.* used PD to observe the correlation between hearing loss, age, and cognitive ability. PD was used to measure the CW from different learners [7]. Their study showed that the pupil response systematically increased with decreasing speech intelligibility. Lastly, Palinko *et al.* used PD to estimate the CW of drivers that were speaking and driving simultaneously [8].

### 2.3. Cognitive Workload

While there is no agreement on its definition, cognitive workload can be seen in terms of resources or mental energy expended, including memory effort, decision making or alertness [9]. It indicates the amount of effort invested as well as users' involvement level. To extract the CW from the EEG headset, this research uses a third-party software called Mentor [10]. This software is a module from the NCO software, a proprietary software of the BMU lab that represents the convergence of multiple years of research into one extensive program called NCO [11]. The program uses machine learning models to extract the EEG signals, interpret them and transform them into a readable cognitive workload score ranging from 0 to 100 [10]. This paper uses this score to represent the current CW of a pilot at a particular point in time.

## 3. Methodology

One problem with modern simulators is that they often lack a way of triggering or monitoring specific events. To solve this, this research created an Event Generator software solution to start, capture, and log different aircraft failures to measure the variation of a pilot's CW, HR, and PD in real-time. An overview of the Architecture is shown in **Figure 1**.

Moreover, the proposed solution is simulator agnostic, meaning it can be plugged into different simulators if desired. While many commercial flight simulators exist, this research focused on the X-Plane 11 simulator, offering realistic aircraft with different programmable variables and solutions compared to other simulators. The Event generator is a multithreaded Node.js and Python

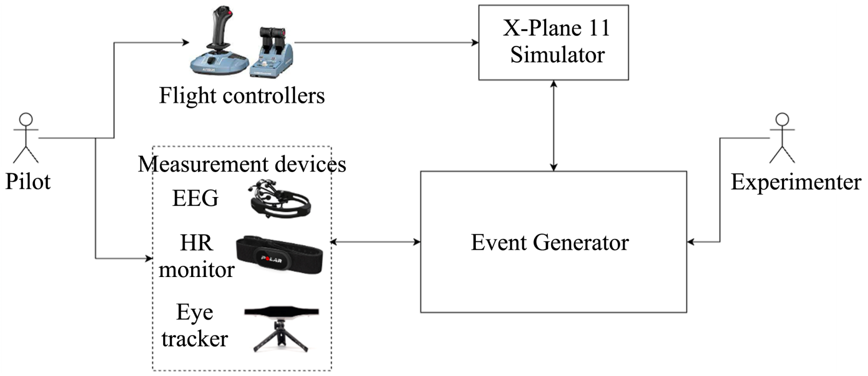

**Figure 1.** An overview of the event generator architecture.

server running concurrently with a React.js frontend architecture. This research made it possible to run multiple server instances on different machines in real-time over a local network to avoid CPU overload. It communicates via HTTP and WebSockets to transfer real-time information to the server or external programs if necessary. Here is an overview of the functionalities of the Event Generator:

- Monitor measurement tools during the experiments.
- Monitor, log, and trigger events related to the aviation industry.
- See real-time data being saved.
- Create, update, and delete plug-and-play scenarios for an experiment.

Moreover, the Event Generator communicates with every data gathering device and safely saves the current extracted data in a specific file in a scalable way. Here is an overview of all the modules that are present within the program:

- Heart rate module: Measures at a frequency of 1 Hz the heart rate of a participant (in bpm) using a heart rate monitor.
- Eye-tracking module: Measures at a frequency of 60 Hz the pupil dilation of a participant using an eye tracker.
- EEG module: Using an EEG headset logically measures and saves the raw, modified, and multiple EEG data points.
- Simulator executor module: Separate module triggering certain events sent to the simulator via UDP.
- Logging module: Other modules can use this module to save data in a particular format.
- Screen recording module: Starts and stops the screen recording using OBS and WebSockets.

## 4. Experiments

This research did an experiment using the previously mentioned methodology to measure the real-time cognitive activity of pilots. The experiment aims to measure pilots' CW, HR, and PD during a takeoff procedure in an Airbus A320 within the X-Plane 11 simulator. The participants had to release the parking brake, do a takeoff procedure, and climb until 3000 ft without using autopilot.

Six different scenarios were created, as described in Table 1. Moreover, participants were divided into two groups, one debuting with failure scenarios and the other with standard scenarios, to cancel the learning effect of doing multiple takeoffs. CW, HR, and PD were measured using an EEG headset from OpenBCI running the proprietary NCO software from BMU [10], a Polar H10 heart rate monitor strap, and the Gazepoint GP3 eye tracker.

## 4.1. Participants

In total, 13 participants with seven pilots, including five A320 pilots, completed the experiment for a total of 136 takeoffs. The participants in this study were required to work from Bombardier or CAE and work in a field closely related to the aviation industry. The six participants with no piloting license were engineers working on a specific aircraft at Bombardier and CAE, knowing most aircraft maneuvers. Participants were all males with an average age of 36 years (±8.8 years), 604 flight hours, and 8.5 years of piloting experience.

## 4.2. Procedure

Figure 2 shows the experiment environment with the participant (left) and the pilot monitor (middle).

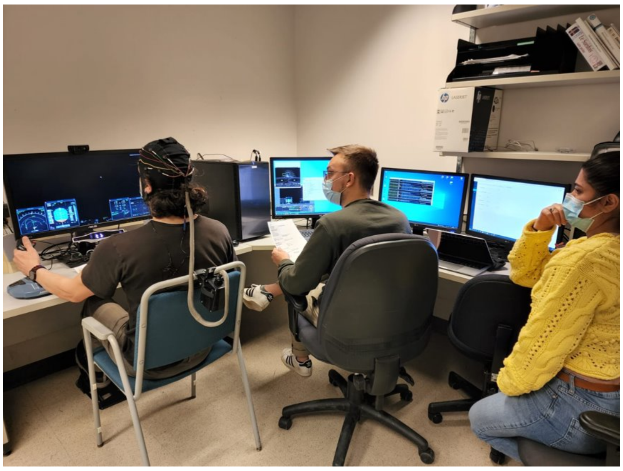

Figure 2. The experiment with the participant (left) and the pilot monitor (middle).

Table 1. An overview of the different scenarios.

| Scenario | Time | Weather | Failure |
|---|---|---|---|
| 1 | 1:45 pm | No wind, no clouds | No |
| 2 | 6:00 am | Clouds at 2700 ft, rain | No |
| 3 | 9:00 pm | No wind, no clouds | No |
| 4 | 5:30 am | No wind, no clouds | Yes, EF at 80 knots |
| 5 | 6:00 am | 15 knots crosswind | Yes, EF at 140 knots |
| 6 | 6:00 am | Low visibility, rain | Yes, EF at 80 knots |

The experiment followed a strict procedure that the partners and the ethics committee approved. Before the experiment, participants received a detailed document of the A320 takeoff procedure to familiarize themselves with the simulator's handling characteristics. On the day of the experiment, participants were given a 30-minute familiarization flight to learn to take off an A320 aircraft with the simulator setup. Moreover, participants were familiarized with the different failure procedures they had to follow during the familiarization. A pilot monitor (experimenter) helped the participant takeoff conform to an Airbus A320 takeoff. After 30 minutes, the measurement tools were installed, and the actual experiment started. The participants did not know the scenarios in advance. This was purposely done to generate more cognitive workload for every scenario. The experiment was divided into two 20-minute sessions, one with failures and one without failures. During the session with failures, the participant knew there could be failures during each takeoff but did not know which type of failure or when a failure could happen. Moreover, the participant did not know the weather conditions in advance of each scenario.

## 5. Results and Discussion

This study's first hypothesis is to demonstrate if it is possible to measure the CW of a pilot in real-time. To investigate this, we started by analyzing if a cognitive load-intensive event, such as an engine failure, triggers a significant change in CW. To this end, this study analyzed the CW of the participants during a 4 second time window around an engine failure as shown on Figure 3. The results show a mean maximum CW of 0.39 and a mean min CW of 0.21. Using F-tests, this study found an F-value of 5.424 and a p-value of 0.029, rejecting the null hypothesis. This result shows the matching correlation between the measured data and the events from the simulator, proving the hypothesis.

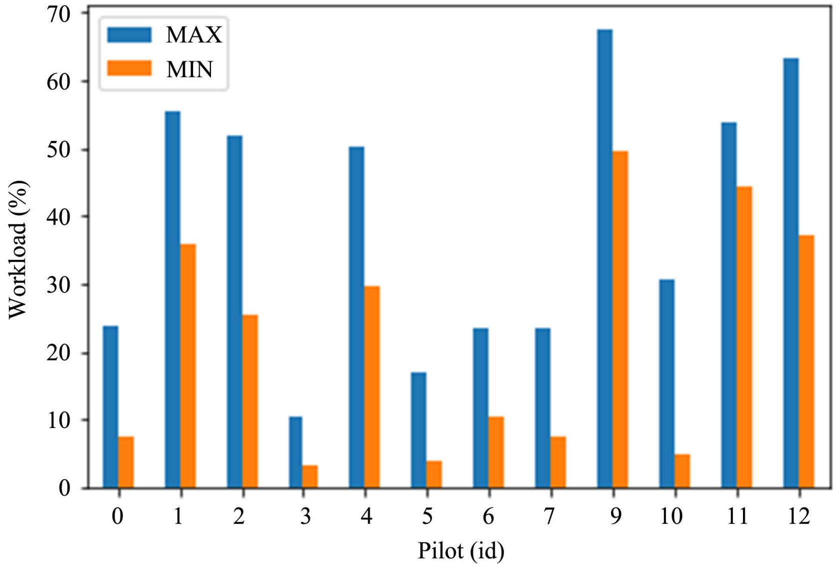

Figure 3. Min-max value of CW within a 4-second timeframe.

The second hypothesis was: Is it possible to establish a correlation between the measured CW and the measured PD and HR during a critical event? Using previous hypothesis' results, this study analyzed the consequences of a critical event on HR and PD before and after the event. Figure 4 shows each pilot's average HR and PD before and after an engine failure. For HR, over a 20-second timeframe, the mean HR before the event was 71.75 bpm and 80.41 bpm after the event. This research used an F-test resulting in an F-value of 10.75, giving a p-value of 0.00317, proving that a CW increase during a cognitive-intensive event also increases the HR. Using the same procedure for PD, over a 4-second timeframe gave a mean PD before the event of 15.23 pixels in contrast to 17.79 pixels after the event. This results in an F-value of 6.35, giving a p-value of 0.0187, rejecting the null hypothesis.

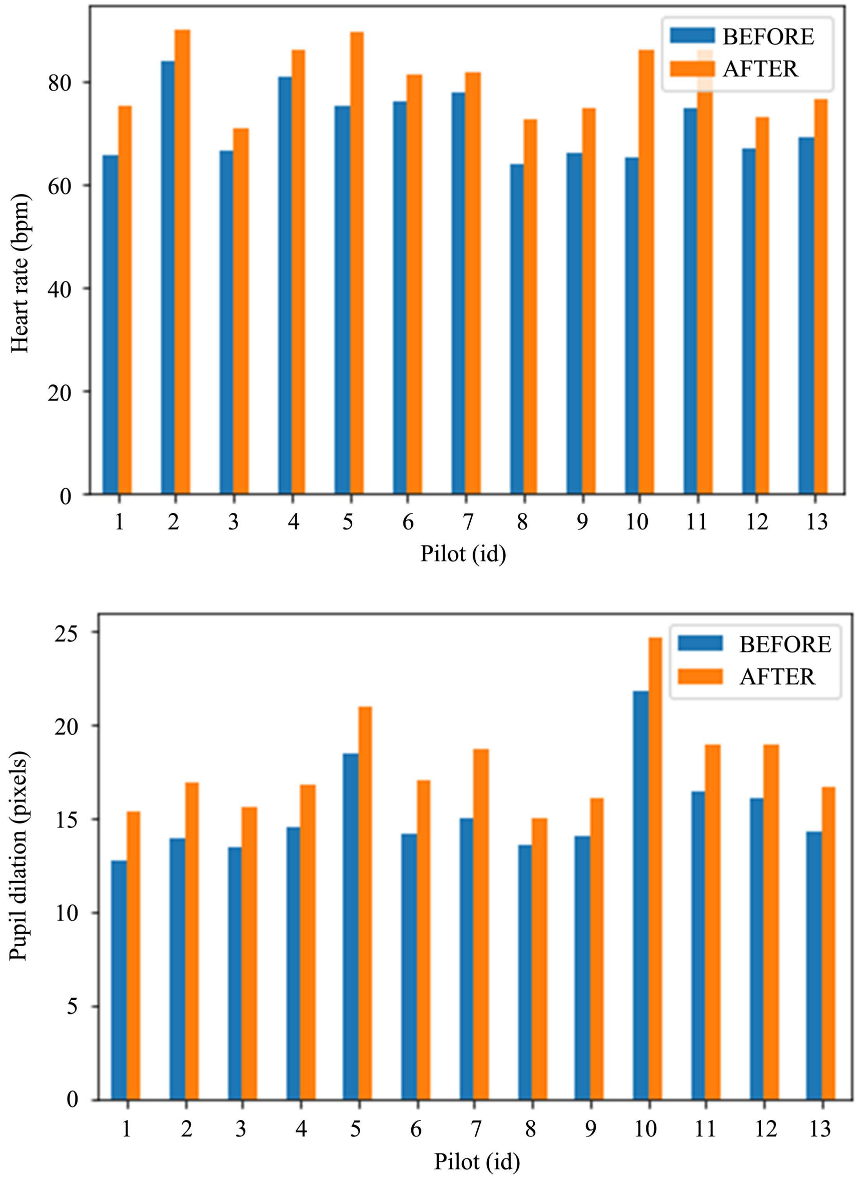

**Figure 4.** Average HR (Upper) and PD (Bottom) before and after engine failure.

The last hypothesis of this study is: Is it possible to predict the CW of a pilot based on his previous behavior? To solve this, different machine learning and deep learning models were trained and tested using the data from the experiment. Table 2 shows the results of the different models. The data was preprocessed by removing unnecessary columns and rows with empty values. Moreover, the HR, PD, and CW columns were smoothed using the Butterworth filter to remove noise. The data was also standard scaled and one-hot encoded, and the top ten features were selected using their correlation with CW using a Pearson correlation matrix.

This study found that a stacked-LSTM model, consisting of two LSTM models using the ADAM optimizer and dropout layers at 0.5, gave the best overall results, as shown in Table 2. Using HR, PD, flight logs, and previous CW, this model could predict future CW values 5 seconds in the future. Figure 5 shows the predictions of the model on unseen data.

Table 2. Models performances.

| Model | MSE | RMSE | MAE |
|---|---|---|---|
| Ridge Regression | 474.70 | 21.79 | 19.67 |
| SVR | 626.62 | 25.03 | 22.31 |
| MLP | 537.11 | 23.18 | 23.18 |
| CNN | 497.73 | 22.31 | 18.52 |
| BI-LSTM | 99.6 | 9.98 | 7.81 |
| **Stacked-LSTM** | **44.09** | **6.64** | **5.28** |

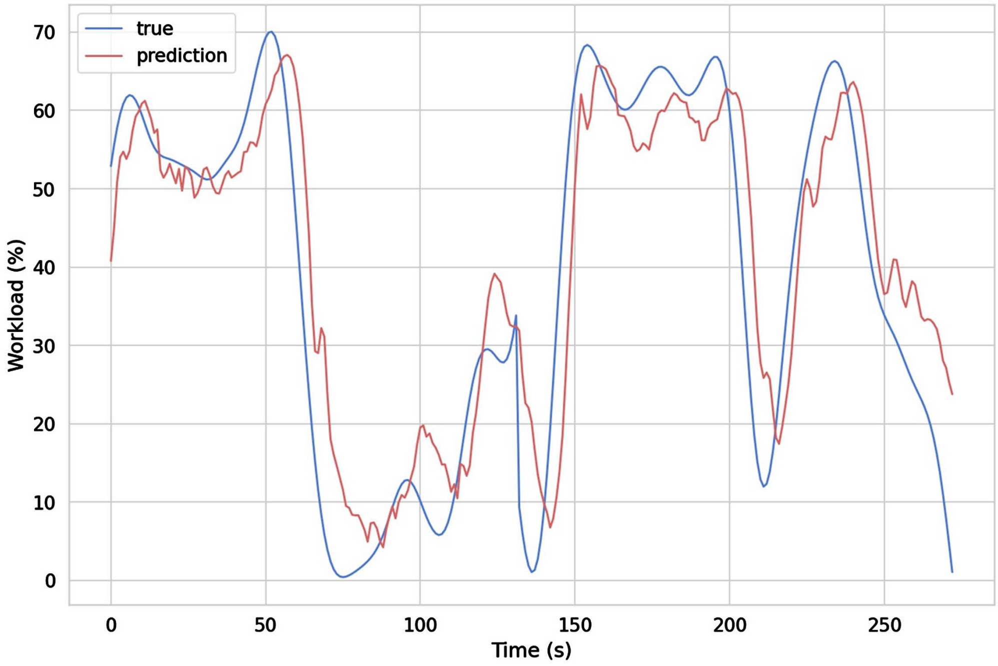

Figure 5. Stacked LSTM model predictions (red) vs true labels (blue).

### 5.1. Discussion

Based on the results in the previous section, we conclude it is possible to predict future cognitive workload based on a pilot's past behavior using data from the HR monitor, eye tracker, EEG, aircraft events and logs combined with a stacked-LSTM model. Different approaches were used to try to predict the CW of a pilot. One idea was not to use any previous CW to predict future CW. This method was used for the SVR, Ridge regression, MLP, and CNN models. As shown in Table 2, this method did not result in reliable predictions. This could be because predicting CW only by looking at the current timestamp does not give enough information for the models to predict anything confidently. Moreover, LSTM models were trained without using previous CW as input and did not yield reliable results either. We also tried to add lag in the CW data. This method shifted all CW x seconds into the future (shifted x rows), so the model could analyze the change in PD or HR to predict the current CW. Nevertheless, even when using lag, we could not predict the next CW. Currently, the model is only capable to predict CW 5 seconds into the future. We tried shorter and longer timestamps to predict the CW of a pilot using different LSTM models. The 5-second future time prediction was chosen based on trial and error. It can be noted that the further the prediction in time, the worse the model's performance. The model fails to predict CW over 15 seconds accurately. Furthermore, predicting the CW of an aircraft pilot more than 10 seconds into the future during a critical event would not make sense as every second counts during such event.

### 5.2. Limitations

The model is limited to making CW predictions of a pilot during a takeoff procedure. Moreover, it can, now, only make predictions during engine failures before or after V1 and during standard takeoffs. Another limitation is that the relationship found between HR, PD, and CW is only during an engine-failure critical event.

## 6. Conclusion

In this research, we investigated the real-time cognitive workload of airline pilots during a takeoff procedure. We tried to investigate the possibility of measuring CW during takeoff using an EEG headset, a heart rate monitor, and an eye tracker. To achieve this, we created a software solution that can trigger failure events and measure the CW, HR, PD, and other simulator events in real-time during a simulated takeoff. It was used during experiments that gathered 13 participants, including seven pilots and six aircraft engineers. Out of those seven pilots, five were A320 pilots. In total, the experiment gathered 136 takeoffs combined across six different scenarios, proving that it is possible to measure the CW of a pilot in real-time. Using this data, we determined that during a critical event, the CW, HR, and PD of a pilot increased, proving the second hypothesis of this paper. Lastly, to prove the third hypothesis, we compared and analyzed the effectiveness of using machine learning and deep learning models to predict the cognitive load of a pilot in real-time using his past behavior. We found that a stacked-LSTM resulted in the best predictions with an MSE of 44.09, an RMSE of 6.64, and an MAE of 5.28. This model was able to predict the CW of a pilot 5 seconds into the future. This study is part of the Pilot AI project and acts as a starting point for future research. The gathered data and the experimental setup will be used for future research and models regarding the Pilot AI project.

## Acknowledgements

We acknowledge the support from NSERC, Alliance, CRIAQ, CAE, Bombardier, and BMU for funding this work.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] Hart, S.G. and Staveland, L.E. (1988) Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.A. and Meshkati, N., Eds., *Advances in Psychology*, Vol. 52, North-Holland, 139-183. https://doi.org/10.1016/S0166-4115(08)62386-9

[2] Panish, Boyle, Ravipudi and Shea. Aviation and Plane Crash Statistics. https://www.psbr.law/aviation_accident_statistics.html

[3] Sevcenko, N., Ninaus, M., Wortha, F., Moeller, K. and Gerjets, P. (2021) Measuring Cognitive Load Using In-Game Metrics of a Serious Simulation Game. *Frontiers in Psychology*, **12**. https://doi.org/10.3389/fpsyg.2021.572437

[4] Sosnowski, T., Krzywosz-Rynkiewicz, B. and Roguska, J. (2004) Program Running versus Problem Solving: Mental Task Effect on Tonic Heart Rate. *Psychophysiology*, **41**, 467-475. https://doi.org/10.1111/j.1469-8986.2004.00171.x

[5] Jerčić, P., Sennersten, C. and Lindley, C. (2020) Modeling Cognitive Load and Physiological Arousal through Pupil Diameter and Heart Rate. *Multimedia Tools and Applications*, **79**, 3145-3159. https://doi.org/10.1007/s11042-018-6518-z

[6] Lang, P.J., Greenwald, M.K., Bradley, M.M. and Hamm, A.O. (1993) Looking at Pictures: Affective, Facial, Visceral, and Behavioral Reactions. *Psychophysiology*, **30**, 261-273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

[7] Zekveld, A.A., Kramer, S.E. and Festen, J.M. (2011) Cognitive Load during Speech Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil Response. *Ear and Hearing*, **32**, 498-510. https://doi.org/10.1097/AUD.0b013e31820512bb

[8] Palinko, O., Kun, A.L., Shyrokov, A. and Heeman, P. (2010) Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator. *Proceedings of the* 2010 *Symposium on Eye-Tracking Research & Applications*, 141-144. https://doi.org/10.1145/1743666.1743701

[9] Chaouachi, M., Jraidi, I. and Frasson, C. (2011) Modeling Mental Workload Using EEG Features for Intelligent Systems. In: Konstan, J., Conejo, R., Marzo, J.L. and Oliver, N., Eds., UMAP, Springer, Berlin Heidelberg, Vol. 6787, 50-61.

https://doi.org/10.1007/978-3-642-22362-4_5

[10] Chaouachi, M., Jraidi, I. and Frasson, C. (2015) MENTOR: A Physiologically Controlled Tutoring System. In: Ricci, F., Bontcheva, K., Conlan, O. and Lawless, S., Eds., *User Modeling, Adaptation and Personalization*, Springer International Publishing, 56-67. https://doi.org/10.1007/978-3-319-20267-9_5

[11] Benlamine, M.S., Chaouachi, M., Frasson, C. and Dufresne, A. (2016) Physiology-Based Recognition of Facial Micro-Expressions Using EEG and Identification of the Relevant Sensors by Emotion. *Proceedings of the 3rd International Conference on Physiological Computing Systems*, 130-137. https://doi.org/10.5220/0006002701300137