Human Factors in Transportation, Vol. 95, 2023, 539–548
https://doi.org/10.54941/ahfe1003839

Student Air Traffic Controllers' Performance Under Conditions of Increased Workload

Terézia Pilmannová, Lenka Hanáková, Vladimír Socha, Michal Freigang, and Stanislav Schmidt

Department of Air Transport, Faculty of Transportation Sciences, Czech Technical University in Prague, 110 00, Czech Republic

ABSTRACT

The work of an Air Traffic Controller (ATCO) involves a constantly changing mental workload and stressful situations to which one must react adequately. Candidates for this profession are expected to demonstrate sufficient resilience and adaptability to such situations already in the selection procedure and during the training process. However, the complexity of Air Traffic Controllers' tasks has increased dramatically in the last decades. This change has mainly become apparent in the number of additional skills a trainee has to possess even before initiating the On the Job (OJT) Training. These include, for example, the ability to obtain optimal flight efficiency or flexibility in planning aircraft sequences. To address the additional complexity, Training Organizations are constantly optimizing training processes to prepare a candidate able to withstand challenging situations with varying workloads. To investigate trainees' reactions to increased workload in the context of conducted errors, an experiment was organized examining 14 subjects that participated in a simulator exercise with increasing intensity. The experiment aimed to assess the impact of increased workload, utilizing ECG measurement, on the trainees' performance and subsequent error analyses. To analyze the ECG measurement, data in the form of RR intervals were filtered and further processed by heart rate variability analysis (HRVAS) software by time-frequency analysis. A subject's individual psychophysiological condition could be further obtained based on these data. Next, an error analysis was conducted, observing 10 categories of errors, including wrong phraseology, application of wrong ATC procedures, separation infringement, and loss of situational awareness. The experiment has shown that the students are prone to conduct mainly errors associated with basic procedures while with time and increasing workload, also errors connected to aircraft separation, situational awareness, and flight efficiency. This paper aims to assess student Air Traffic Controllers' psychophysiological state and performance during times of increased workload and provide an overview of physiological reactions and subsequent mistakes conducted by the trainees. Also, it should serve as an outlook of possible drawbacks that have a negative impact on the initial training progress and should therefore be addressed with special care. Therefore, the experiment's results can be used to adjust basic training where the most common errors occurring during increased workload can be addressed and trained within the enhanced practical part of the basic training.

Keywords: Human factors, Air traffic control, Situational awareness, Workload, Training, Errors

INTRODUCTION

The complexity of Air Traffic Controllers (ATCO) tasks has in the last decades dramatically increased. Therefore, a transformation of the ATCO training syllabus was inevitable to ensure the ability of future ATCOs to cope with this changing environment. This change has mainly become apparent in the number of practical skills a trainee has to possess even before initiating the On the Job (OJT) training. Further, the emphasis on practical competencies has consequently increased demand for synthetic training throughout the licensing process (Eurocontrol, 2010). However, a crucial indicator of the training success is not only a well-prepared candidate for the future ATCO role but also the time aspect that forces the training organizations to compress the training syllabus to a minimum feasible number of days.

Even though Air Traffic Controller basic training is mainly theoretical, it is also composed of practical skills assessment based on the formulated performance objectives per Commission Regulation (EU) 340/2015, PART D.030 (European Commission, 2015). This assessment, conducted on an ATC simulator, expects the candidate to be proficient in basics like phraseology or familiarity with a professional environment but also proper ATC procedures application and situational awareness, which are complex tasks and require sufficient time the trainee has to spend on a simulator. To be able to focus on the content of the practical part so that the quality of trainees is acceptable and the aspect of cost-efficiency of the basic training is met, it is essential to target the possible problems, and deficiencies trainees may experience already from the early stages of the initial training. Even though the initial training syllabus is set as per the Commission Regulation (EU) 340/2015, there is no notion of how the actual training process shall be organized. Therefore, there might be substantial differences among training organizations in terms of the length of the training or design of the practical training part (Juričić et al., 2020).

No studies could be found assessing student ATCOs' workload during the initial training and its effect on learning progress. In this paper, the workload is understood as a combination of the ATCO mental demand a particular task or task sequence requires and the physiological response. In particular, none of the research deals with trainees' psychophysiological reaction to increased workload generated from increasing task load and an effect on actual performance regarding the accuracy, situational awareness, and flight efficiency. Identification of errors is a complex problem in the case of air traffic control. It is not always possible to unambiguously determine whether an error has been conducted, as it often happens in the course of a particular situation or as a consequence of more actions. Several methods have been developed to identify and classify the error and its source to objectivize the error analysis of aviation operational personnel. The US Federal Aviation Administration has, in the study on Examining ATC Operational Errors Using the Human Factors Analysis and Classification System, proposed an error classification system to recognize the source of error and its consequences easily. The assessment matrix also involves managerial actions to ensure corrective actions. The classification includes (Scarborough et al., 2005):

- Data Posting actions (incorrect data input, incorrect or late data update),
- Radar Display (misidentification, inappropriate use of displayed data),
- Communication Error (Phraseology, Misunderstanding, Readback),
- Coordination (Violation of separation standards, Incorrect altitude, flight level, route of flight, speed).

On the other hand, Donmez et al. apply in the research on *Service Quality and Efficiency-Oriented Training Design in Air Traffic Management* a method of decreasing the percentage score every time a student ATCO conducts any of the previously specified mistakes. The gravity of the mistake is represented in the percentage value. Therefore, in case of a loss of separation, the student will lose 30% of the initial 100%. Mistakes in vectoring, radiocommunication, and flight strip updates are graded by a loss of 5 to 3% (Donmez et al., 2022).

This paper aims to connect the two aspects, evaluate the physiological response toward increased workload, and assess mistakes the trainees are prone to make while solving complex situations during simulation exercises. Two approaches were applied to assess trainees' actual performance and workload; an error assessment analyzing the type of errors and their frequency with increasing workload and ECG measurements to determine the participant's physiological condition.

The results may help understand the physiological response of a trainee during an increased workload and its effect on the quality of air traffic control. The results are also valuable information on what processes should be especially emphasized during the initial training to prevent the most common types of errors and increase the training successes.

MATERIALS AND METHODS

Participants

The experiment aimed to put subjects in a situation with a high workload so that their performance and physiological response could be tested. Fourteen subjects aged 18 to 28 years participated in the measurement. The experiment subjects were Department of Air Transport students with no particular experience in ATC but considerable knowledge of standards and procedures in air traffic control and familiarity with the simulation environment. It is therefore supposed that the level of skills and knowledge of the subjects are equivalent to student Air Traffic Controllers conducting the basic training. The baseline level of knowledge included basic procedures in air traffic control, familiarity with the performance characteristics of the most recognizable aircraft, radar vectoring, and a basic understanding of aviation phraseology. All subjects were informed about the process and risks of the experiment, with the possibility to withdraw at any stage, and subjects participated voluntarily. The experiment was conducted following ethical principles for research involving human subjects (Rickmann, 1964).

Experimental Setup

For the purposes of this measurement, all subjects had equal access and conditions. Initially, they had a briefing on the experiment's rules, which included familiarization with the area of responsibility, standard procedures for departing and arriving traffic, completing paper flight strips, letters of agreement with other air traffic control units, and simulator controls. The briefing also included a practical introduction.

The entire area of responsibility is fictional to obtain consistency. Each subject completed a 90-minute exercise without interruption. The scenario consisted of 75 IFR traffic and a few aircraft with a change of flight rules. The first 60 minutes was the control under standard operating conditions with an increasing workload in the form of increased traffic and potential conflicts. At approximately the 60th minute, the controller should have been at the edge of his mental capacity when the airport traffic control tower performed verbal coordination, informing that the airport was suspending operations and not accepting any flights due to the failure of lighting and electrical equipment. The subject had to immediately abort all approaches and send the aircraft to holding patterns. He was also supposed to inform everyone about the situation and the unpredictably long delay. The untangling and sequencing of traffic after lasted until the end of the exercise, approximately 20 to 30 minutes. Due to the short-term overload the situation caused, the fatigue of subjects became apparent.

For assessment purposes, the exercise scenario described above has been evaluated in accordance with the Eurocontrol methodology for developing ATCO synthetic training. While each activity an ATCO has to conduct is awarded a particular point score. The methodology states that an ATCO can handle maximum workload points of 300 (Eurocontrol, 2010). The scenario used for the experiment has a workload course displayed in Figure 1.

The experimental scenario displayed in Figure 1 is based on the four most significant factors affecting the ATCO workload:

- 1st peak (212 Workload points) – increase is caused by the rise in the number of aircraft on frequency, from 8 up to 14 in several seconds. ATCO must form a plan.
- 2nd peak (191 Workload points) – 8 aircraft were vectored, 6 altitude changes, 4 speed control acts.

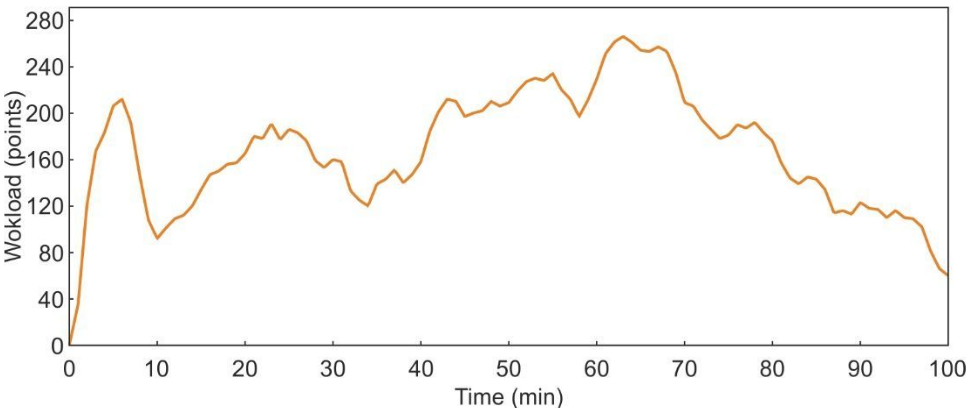

Figure 1: Scenario workload score.

- 3rd peak (234 Workload points) – 12 aircraft on frequency, 9 altitude changes, 5 aircraft were vectored, and 5 acts of verbal coordination were required.
- 4th peak (266 Workload points) – 15 aircraft on frequency, 11 aircraft were vectored (holding patterns included), 7 altitude changes.

Based on the Eurocontrol methodology, the exercise described can be evaluated as Intermediate. Eurocontrol recommends three peaks with ca. 4–5-minute-long valleys in between these peaks. Peaks are recommended to have around a 30% increase in the maximum Workload value, with the described exercise having only around 15–20%. Because the graph produced is the result of a purely mathematical process, it can lead to misconceptions about the workload, especially if there are too many sudden changes happening in a short period of time that cannot be reflected by natural mental activity. Also, to obtain knowledge about participants' performance, ten types of errors were assessed during the experiment that can be further divided into three categories; critical errors, serious errors but not time-critical, and errors connected to the experience and proficiency of subjects, see Table 1.

Errors were assessed as cumulative values, and no further analyses of possible error subgroups was conducted. Situational Awareness in the context of conducted experiment refers to an ability to maintain an overview of all movements and their intended flight operations within the area of ATCOs responsibility in a 3D airspace. Further, the assessment of Flight Efficiency aims to evaluate the ATCO's ability to provide the aircraft with the most optimal flight path.

A heart rate variability (HRV) analysis was used to assess the subjects' physiological state for objective workload monitoring. Cardiac activity was measured by a single-lead electrocardiograph using the VLV-Lab (Albertov Research Center, Prague, CZ), providing an electrocardiogram (ECG) with a sampling frequency of 1000 Hz (Hon et al., 2015).

The system provides an unfiltered ECG curve, so the data must be pre-processed first. For this purpose, the Pan-Tompkins algorithm (Pan and Tompkins, 1985) was used to detect individual R-peaks (Sedghamiz, 2014). The RR-intervals were calculated to obtain RR-interval vectors based on

Table 1. Error assessment scheme.

Error type	Severity
Separation minima infringement	Critical error
Airspace violation	
Unsafe clearance	
Wrong readback	Serious error
Late/incorrect target identification	
Failure to confirm ATIS/ QNH/ Mode C information	
Incorrect phraseology	
Situational Awareness	Error caused by lack of experience and proficiency
Flight Efficiency	
Submitting aircraft to a wrong frequency	

the detected R-peaks. These so-called tachograms were further used for HRV analysis in the frequency domain. The ratio of power spectral density (PSD) in a low-frequency band (LF, 0.04–0.15 Hz) and high-frequency band (HF, 0.15–0.4 Hz) calculated from tachograms was selected as the relevant parameters depicting the current load status from the ECG measurements as it was proven as useful for such applications (Hsu et al., 2015). This parameter reflects the so-called sympathovagal balance, corresponding to the degree of the current activity of the sympathetic and parasympathetic divisions of the autonomic nervous system. This parameter is believed to be a reliable indicator of the current psychophysiological state, which is affected by workload variations (Hsu et al., 2015). Also, PSD in separated bands (LF, HF) were used as it puts the LF/HF ratio into context.

The HRV analysis in frequency analysis involves the calculation of PSD, but this loses the time information. Therefore, given the paper's objectives, it was necessary to perform a time-frequency analysis via Short Time Fourier Transform (STFT). The above analysis was performed using Matlab software implantation HRVAS (Ramshur, 2010).

After obtaining the LF and HF waveforms, these were exported to be further processed in the Matlab 2022a software environment (MathWorks, Natick, MA, USA). The LF/HF ratio was then calculated from the LF and HF vectors. In order to allow for inter-individual comparisons, inter-, and intra-individual variability had to be suppressed as much as possible. The above was done by using standardization, specifically z-score, which converts the original dataset to a zero-mean dataset, with the distance of each data point from the mean measured in standard deviations (Carey and Delaney, 2010). Standardization was performed separately for each subject, and the resulting z-score datasets were further averaged to create a standardized LF/HF curve and LF and HF curves for follow-up purposes.For averaging purposes, it was first necessary to unify the timeline of the datasets due to the difference in the experiment duration between various subjects, as some of them managed the traffic more efficiently and finished the assignment in a shorter time. Thus, all assessed time series were relativized to the planned exercise duration (for comparison with the workload curve), i.e., to the range of 0–100 minutes. Standardizing all data series,' and sample sizes were necessary for further evaluation. For this purpose, given the nature of the data, linear interpolation was used.

The trend extraction from the noisy LF/HF, LF, and HF mean curves was performed by polynomial fitting (a polynomial of degree 20 was used for the purpose of sufficient smoothing while preserving the character of the curve). The curve was fitted with a 95% confidence interval. To allow visual comparison of the obtained mean curves with the calculated workload score curve, the latter was normalized to the same range as the LF/HF curve. Statistical comparison of the "similarity" of the two curves was performed using cross-correlation (Buck et al., 1997).

RESULTS

The results of physiological conditions during the experiment are presented in Figure 2 A–C. These have shown that the physiological reaction of

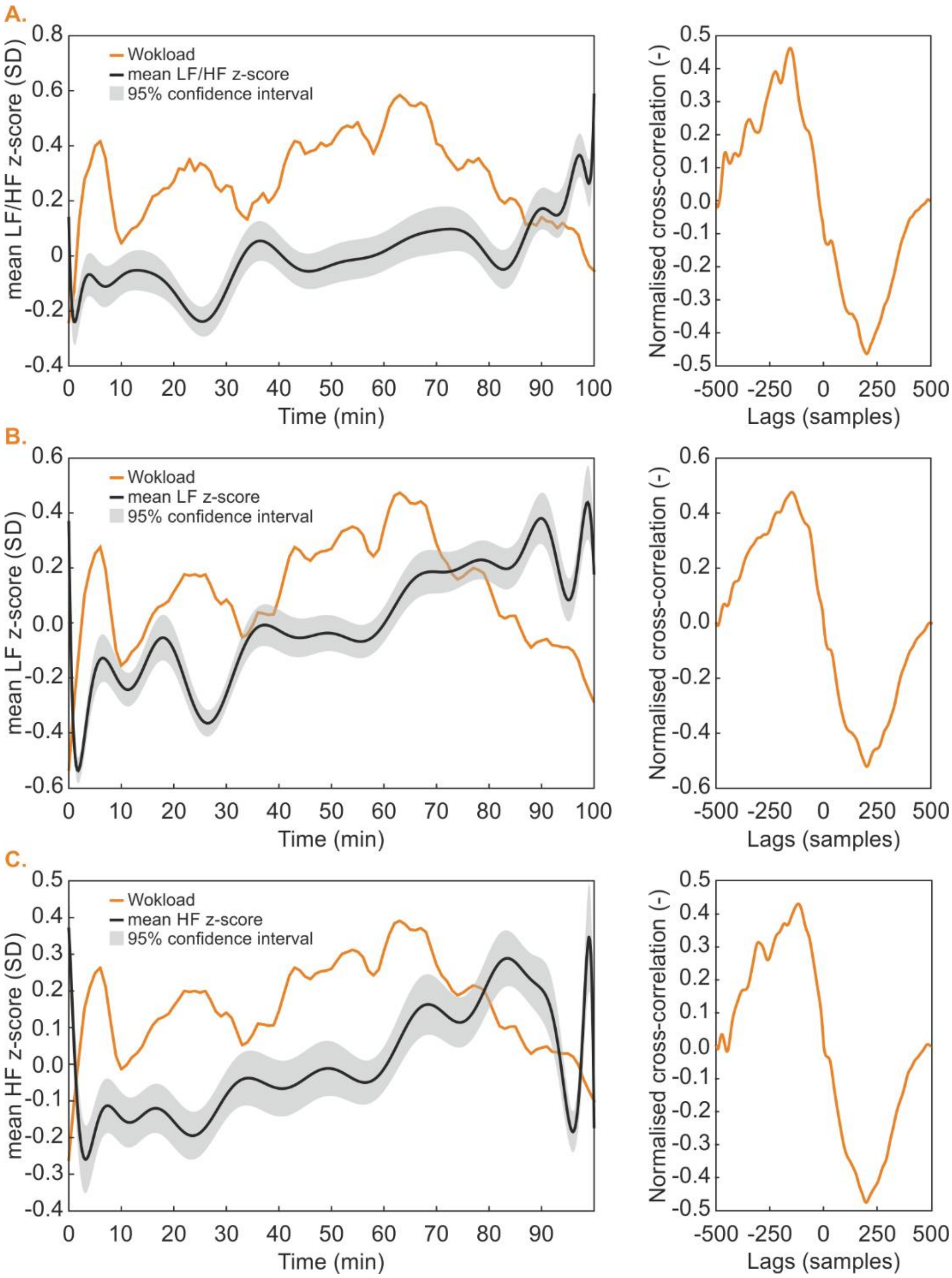

Figure 2: Comparison of mean standardized heart rate variability parameters and workload curve and normalized cross-corelation between those two curves for LF/HF (A), LF (B), and HF (C).

a student ATCO correlates to the anticipated workload to some extent. It is clear from the results that there is a gradual increase in all three parameters during the measurement, with several significant fluctuations (decreases), especially at about the 25th minute and before the end of the measurement. The HF and LF/HF parameters follow an inverse trend to the load curve, i.e., as the exercise load increases, these parameters decrease and vice versa.

Although the LF parameter trend is similar to the other two, there is a smoother increase in the parameter except for the aforementioned 25th minute. Also, there is a slight reactionary shift apparent that is caused by a normal physiological reaction to an increased workload. The cross-correlation results are similar for all three monitored parameters, with the highest correlation occurring at a shift of approximately 150 samples, corresponding to about 35 minutes. The zero-shift results indicate a weak, negative correlation (see Figure 2).

Another aspect of measurement investigated is the classification of the errors made by the subjects during the experiment. However, in this type of experiment, error assessment is often challenging as it is not possible to clearly determine what is considered an error and what is not, given that the subjects were not actual air traffic controllers-in-training. Nevertheless, the error assessment has shown that while facing lower workloads, the subjects are prone to conduct errors associated with basic procedures (Identification, QNH, or submitting the aircraft to the wrong frequency). With time and increasing workload and fatigue, errors connected to aircraft separation, situational awareness, and flight efficiency were more frequent. Specific overall occurrences of all errors based on their type and severity among all subjects are presented in Figure 3.

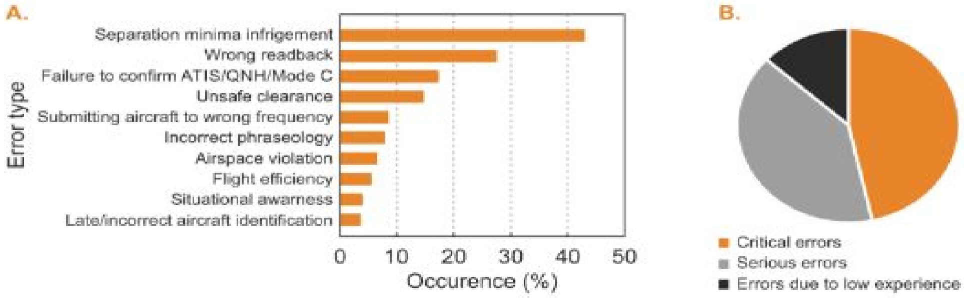

Figure 3: Error occurrence by type (A) and severity (B).

DISCUSSION

The results of HRV, to some extent, do not fulfill the expectations before the experiment, as it was supposed that a non-trained individual would react to complex situations simulated in the scenario by an increase in LF/HF. On the other hand, this ratio reacts very well inversely to the load, i.e., when it increases, it decreases, and vice versa. The individual behavior of each division of the autonomic nervous system can explain the above. Although it is generally understood that the LF/HF ratio reflects increased load by its increase, some authors contradict this fact. For example, Kim et al. find that LF/HF increases with load in individuals who prefer passive stress coping strategies. At the same time, it decreases in individuals who prefer active stress coping strategies (Kim et al., 2014). Other authors then contradict the simplification of the sympathetic/parasympathetic relationship to a linear dependence represented by the LF/HF ratio (Billman, 2013; von Rosenberg et al., 2017). In any case, the results presented in Figure 2 show that cardiac activity reflects the expected stress, albeit opposite to how LF/HF is generally understood. The low correlation between physiological parameters and the assumed workload is then likely related to the different magnitude of the physiological response and the slight shift between the physiological data and the workload curve (measurement difficulty) that is assumed, given the inertia of the physiological response and the inter- and intra-individual variability of the subjects. At the same time, this shift is not and cannot be constant over the entire measurement period, also taking into account the fact that some tasks may be easier for a certain part of the subjects than for another, i.e., the perceived load may be different and at the same time the time required to perform the task may vary slightly. These facts are then likely to negatively impact the statistical comparison of the individual curves, as evident from the graphs in Figure 2.

Regarding the error analysis, the results do not, of course, include the individual psychological state of the subjects, their level of fatigue or wakefulness, or their current emotional state, which may have influenced the results of the measurements. The results, however, indicate that errors categorized as serious (Wrong readback, Late target identification, failure to confirm ATIS/QNH, or Wrong phraseology) are failures that should be specially addressed already in the initial phase of the ATCO training. These can be obtained by incorporating practical exercises already in the early stages of the initial training so that the student ATCOs get enough time to develop proper practical habits in line with the theoretical education they receive in the first part of the Air Traffic Controller Training.

CONCLUSION

The number of skills that the student ATCO has to possess even before initiating the practical part of the training has increased in the last decade due to the complexity of air traffic. The results of the above-described experiment show that the coordination of theoretical education and practical exercises already from the beginning of the ATCO training might positively affect the ability to apply procedures correctly as the recently explained theoretical concepts are immediately applied in praxis. The early introduction of practical training may also influence the trainee's ability to respond to an increased workload, thus improving overall performance. This might affect the training duration and successes kept at a minimum feasible time for economic and operational reasons.

REFERENCES

Buck, J. R., Daniel, M. M., & Singer, A. C. (1997). Computer explorations in signals and systems using MATLAB. Prentice-Hall, Inc.

Carey, J. J., & Delaney, M. F. (2010). T-scores and Z-scores. Clinical reviews in bone and mineral metabolism, 8, 113–121.

Donmez, Kadir., Kaplan. Zekeriya, Aybek Cetek, Fulya. Rogošić, Tea. Juričić, Biljana.

Eurocontrol. (2010). Hum-Sim Principles for the design of radar simulation exercises and course development.

European Commission (2022). Commission Regulation (EU) 340/2015. Web: https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX%3A32015R0340.

Francetić, Ivana. (2022). Service Quality and Efficiency-Oriented Training Design in Air Traffic Management. Transportation Research Procedia, Volume 64, Pages 345–355.

Hon, Z., Smrcka, P., Hana, K., Kaspar, J et al. (2015). A surveillance system for enhancing the safety of rescue teams. Communications - Scientific Letters of the University of Zilina, vol. 17, no. 1, pp. 81–86.

Hsu, B. W., Wang, M. J. J., Chen, C. Y., & Chen, F. (2015). Effective indices for monitoring mental workload while performing multiple tasks. Perceptual and motor skills, 121(1), 94–117.

Juričić, Biljana. Antulov-Fantulin, Bruno, Rogošić, Tea. (2020) Project ATCOSIMA – Air Traffic Control Simulations at the Faculty of Transport and Traffic Sciences. Website: https://hrcak.srce.hr/file/35581

Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. IEEE transactions on biomedical engineering, (3), 230–236.

Ramshur Jr, J. T. (2010). Design, evaluation, and application of heart rate variability analysis software (HRVAS).

Rickham (1964). Human experimentation. code of ethics of the world medical association. declaration of helsinki. *Br. Med. J.* 2, 177–177. doi: 10.1136/bmj.2.5402.177.

Scarborough, Alfretia, Bailey, Lerry. Pounds Julia. (2005). Examining ATC Operational Errors Using the Human Factors Analysis and Classification System, Oklahoma City, OK.

Sedghamiz, H. (2014). Matlab Implementation of Pan Tompkins ECG QRS detector. [Data set]. Unpublished. https://doi.org/10.13140/RG.2.2.14202.59841.