AerospaceEurope Conference2021
WARSAW, POLAND | 23 -26 NOVEMBER 2021
PSAA
CEAS
Council of European Aerospace Societies

APPLICATION OF BIOFEEDBACK AND ADAPTIVE AUTOMATION FOR UAV OPERATOR PERFORMANCE ENHANCEMENT

Antoni Kopyt, Marcin Adamczuk & Antonina Waszkiewicz

Institute of Aeronautics and Applied Mechanics, Warsaw University of Technology, Warsaw, Poland

Abstract

The paper presents a research on biofeedback analysis in the context of UAV operators and development of the adaptive system for operators, so their performance is maximum. In the first, theoretical part, methods of measuring human psycho-physical state are considered. For each measurement method, a review of commercially available measurement equipment - especially in terms of data availability - has been performed. Taking into account the most useful and affordable devices, concepts of measurement station was developed. On the basis of these concepts an actual stand was built. It enables measurements of specific biofeedback parameters i.e. bioelectrical activity of brain, heart rate and skin conductivity. The research part concerns the measurement of selected parameters, their analysis and further use of the obtained results. Following research focuses on the integration of the results of individual data in order to determine the operator's stress/strain level . Base on those data an engagement index (EI) will be determined so that it can be used in further research as input to the adaptive system. Having the engagement index calculated during the tasks the further, more advanced research is delivered. The development of an adaptive system capable of adjusting the level of task difficulty accordingly to the estimation of functional state of the human operator. Depending on the circumstances both, mental underload and overload, can lead to decreased performance. Performance decrements resulting from the state of mental underload can be associated with loss of situational awareness, insufficient attentional resources and deskilling. The role of the designed system is to assist the operator to maintain optimum engagement, increase his ability to cope with the tasks and as a result maximize his performance. Adaptive automation invocation processes are based on real-time operator performance and physiological assessment, along with subjective self-reported workload provided by the operator himself by filling NASA TLX questionnaire. Considering individual differences and ambiguous assessment of the operator's mental state human behavior is difficult to define with typical logic approach. Therefore, the assessment of human workload is done using fuzzy modeling approach taking performance index, engagement index (based on physiological measurements as described in the first part) and NASA TLX score as input parameters. The effectiveness of the developed adaptive automation system was verified in real-time by conducting human-in-the-loop experiments during which the operators were performing the MATB-II tasks.

Keywords: human operator, UAV training, biofeedback, workload assessment, adaptive automation

1. Introduction

Several experiments have demonstrated that there is an inverted U-shape relationship between workload and performance. As presented in the figure below, the operator's ability to cope with task demands is maximized when he experiences moderate levels of workload. Several studies have indicated that both extreme conditions, mental overload and underload, negatively affect task performance as well as the operator's welfare, and need to be mitigated [1]. Excessive mental workload can lead to stress, frustration, confusion, fatigue and delayed information processing. Excessive levels of acute stress have detrimental effect on cognitive processes such as attention, working memory, memory retrieval and decision making. Highly demanding tasks, time pressure and stressful circumstances make human operator prone to committing errors [1]. On the other hand, at a low level of mental workload human operator can experience boredom, annoyance and mind wandering that leads to vigilance decrements [2].

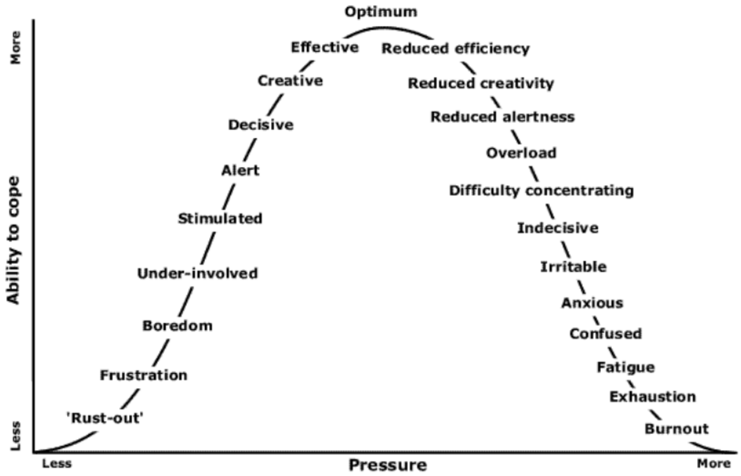


Figure 1– Pressure/efficiency curve.

The switch from active processing to passive monitoring might be a result of implementing higher levels of automation that in general are expected to improve the efficiency and capacity of the system but on the other hand introduce some difficulties for human operators. One of the major consequences of automation is out-of-the-loop performance problem when as a result of allocation of system functions to an automated controller, the operator is removed from a control loop. It decreases their ability to observe system parameter changes and take over manual operations in case of automation failure. The operator has reduced situation awareness and cannot monitor the system efficiently. It is associated with decreased alertness, complacency – human over-trust in computer controllers, increased reaction times and manual control skills decay. These consequences can impact human performance under normal operating conditions and in case of system failure [3].

2. Biofeedback loop implementation for adaptive systems

Physiological variations reflect fluctuations in operator’s functional state and can be used to evaluate the mental workload experienced while executing a task. Going further, results of real-time analysis of biofeedback measurements can provide an adaptive control input to biocybernetic systems, capable of dynamically adjusting their mode of operation to the human mental state [4].

Adaptive automation systems have proved effective in mitigating decreased performance caused by either under- or overload. These systems are able to dynamically implement automated aids in response to situational changes e.g. operator’s workload [5].

Changes in mode of operation of adaptive system can be initiated based on e.g. the operators performance, physiological assessment or subjective workload rating. In this study, these three measures are integrated in order to evaluate operator’s mental workload and trigger changes in adaptive automation if necessary.

Possible types of system adaptation include changes in functions allocation, task scheduling and levels of difficulty. Additionally, some systems can adjust the GUI layout, enhance the guidance of attention by providing visual or vocal cues and features or modify the amount of information that is presented to the operator. In this study the adaptive system adjust its mode of operation to the operator’s functional state by changing the level of task difficulty.

It has been demonstrated that adaptive automation is superior to static automation as it preserves the operator's skill level, ensures that the operator's workload is maintained within the optimum range, guarantees continuous task involvement and improved situation awareness, and as a result improves the operator's performance. It also leads to improved management of automation failures and degraded conditions, including detection of the failure and management of recovery, since the skills needed to performed the task are maintained [6].

3. Methods of measuring psychophysical parameters

When analyzing different types of measurement methods, the following were decided upon due to the usefulness of the measurement data, accuracy, ease of use, price, among others.

Table 1

Measuring method	Pros	Cons
EEG	data usability, accuracy, non-invasiveness, does not restrict mobility	price, general data (unable to distinguish between emotions, e.g. fear – excitement)
HR	price, data usability, easy analysis	unable to distinguish between emotions, does not need to respond with increased psychological stress
GSR	price, ease of implementation, data usability	accuracy

On the basis of [7], a table was created in which the relationship between the results of measurements measured by different methods and the load of the tested person is presented.

Table 2

Measuring method	Lack of involvement	Excessive involvement
EEG	increase in α wave activity, decrease in β and θ wave activity	increase in α wave activity
HR	HR decrease	HR increase
GSR	decrease in skin conductance values	increase in skin conductance values

3.1 Electroencephalography (EEG)

Electroencephalography is the measurement of the bioelectrical activity of the brain [8][9]. EEG measures the electrical potentials produced by the cells and dendrites of pyramidal neurons. The signals are measured in the μV range (0.5 to 100 μV) at low frequency (0.5 to about 40 Hz). They are usually referred to as rhythms and are classified into five frequency bands (Table 3).

Table 3

No.	Brain waves	Frequency [Hz]
1	Delta (δ)	0,5 – 4
2	Theta (θ)	4 – 8
3	Alpha (α)	8 – 13
4	Beta (β)	13 – 30
5	Gamma (γ)	36 – 44

On the basis of observation and various studies, it has been established which waves dominate in certain human mental states. They can be defined as follows:

- alpha waves are dominant during the relaxed state,
- beta waves occur during the state of readiness; they are divided into low, middle and beta 2 waves. Low beta dominates in the process of recalling information, it accompanies learning processes.

Middle waves occur during rapid brain work. The least desired are beta 2 waves, which appear during excessive excitement, nervousness.

- gamma waves are responsible for the process of intensive thinking,
- delta waves occur mainly during deep sleep,
- theta waves dominate during meditation, hypnosis; during their occurrence stress is reduced and creativity is increased.

To determine the engagement index for an EEG study, the relationship [10] can be used:

$$E = \frac{\beta}{(\alpha+\theta)} \tag{1}$$

The index was established taking into account that an increase in beta waves is associated with an increase in brain activity during mental effort, while an increase in alpha and theta waves is associated with reduced mental performance and alertness.

3.2 Heart rate (HR)

Heart rate (HR) can be defined as the number of heartbeats per unit time. This value is usually given as beats per minute (BPM). As stress increases, a person's HR increases. The normal pulse rate is between 60 and 100 BPM [11] and depends on many factors, including age, health, physical activity, therefore the limit above which the stress level increases is determined individually. To determine the limit above which the pulse value indicates cognitive load of the examined person, the following rule has been adopted: 10% of the average pulse value is added to the resting pulse value, e.g. in the case of resting pulse of 70 BPM, values above 77 BPM may indicate stress of the examined person. Such a rule was established taking into consideration the nature of the test - lack of physical activity, therefore heart rate changes are not so big and dynamic. The longer the pulse stays above the limit, the more demanding the task is. The final coefficient can be defined as:

$$E_{HR} = 1 - \frac{stress\_points\_measurement}{all\_measurement} \tag{2}$$

3.3 Galvanic Skin Response (GSR)

The galvanic skin response is measured by recording the electrical resistance of the skin [12][13]. This resistance depends on skin hydration, which changes during secretion of sweat by human sweat glands. In this situation, sweat becomes a good conductor between the skin and blood vessels. The conductivity increases with increasing humidity. Sweat production increases under human load, physical work and thus the value of human body resistance decreases. An increase in the GSR level corresponds with an increase in the load of the subject. The main advantages of GSR measurement are its simplicity - two electrodes placed e.g. on the toes of the tested person are enough for measurement, non-invasiveness and easiness of results analysis.

4. Measurement Equipment

The following table summarizes the equipment for each measurement method used in the study.

Figure 2 – From the left: Neurosky Mindwave, Polar H10 (heart rate sensor), Neurobit Opitma+ 4

Table 4

Measurement	Name	Output data	Unit	Remarks
EEG	Neurosky Mindwave	Alpha1, Alpha2, Beta1, Beta2, Gamma1, Gamma2, Delta, Theta, Attention, Meditation	-	Export to .csv file
HR	Polar H10	Heart Rate	BPM	Export to .csv file or connection to Matlab
GSR	Neurobit Optima+ 4	Conductivity	S (Siemens)	Export to .edf file

4.1 Measurement station

As part of the work, a measuring station was prepared.
It consists of the equipment listed above:

- EEG – Neurosky Mindwave Mobile 2
- GSR – Neurobit Optima+ 4
- HR – Polar H10

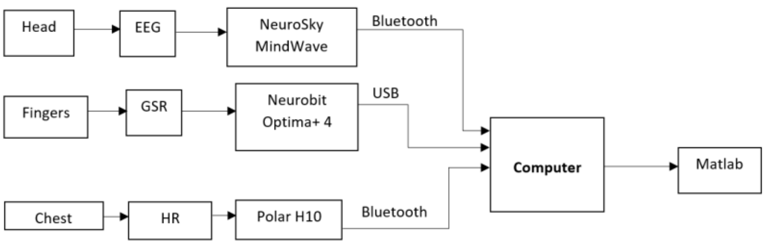


Figure 3 – Measurement station scheme

The data collected from the EEG measuring device do not have units, so they should only be compared with each other. It is possible to measure all types of brain waves and export the results to a ".csv" file. The GSR sensor can be connected to the program cooperating with Neurobit Optima device. Output data is saved in ".edf" file. The Polar H10 sensor can be connected to the "Polar Flow" application available for phones and tablets or directly to the Matlab environment. The target is measured human pulse in BPM.

5. Preliminary tests

After the review of the measurement methods a series of tests were prepared to validate each sensor, to test the code written and to check that the psychophysical parameters do indeed change with the change in cognitive load. Free online tests from *www.humanbenchmark.com* were used.
They were divided into three categories, depending on the issue they addressed, in order to see how exercise in different areas affects people. Two tests were conducted in each category:
Reaction time category:

- Task 1: test reaction time by clicking the mouse button when the screen changed from red to green
- Task 2: click as quickly as possible on the successively appearing targets.

Figure patterns category:

- Task 1: memorise and associate words; the participants had to decide whether the displayed word had already appeared before or whether it was a new word
- Task 2: memorise the displayed sequence of squares and then mark it correctly.

Number patterns category:

- Task 1: memorise the displayed number, with each level the number of digits increased
- Task 2: memorise the order of the digits on the squares, and then to mark each of the covered squares in the right order.

5.1 Results

The tests were conducted for four participants. Based on the results of person A, the measurement analysis, data processing and final result will be presented.

5.1.1 EEG Data

The following graphs show the raw data of the measurements made with the EEG and the processed data. The data has been passed through a low-pass filter, smoothed using Matlab's *smooth* function and outliers removed.

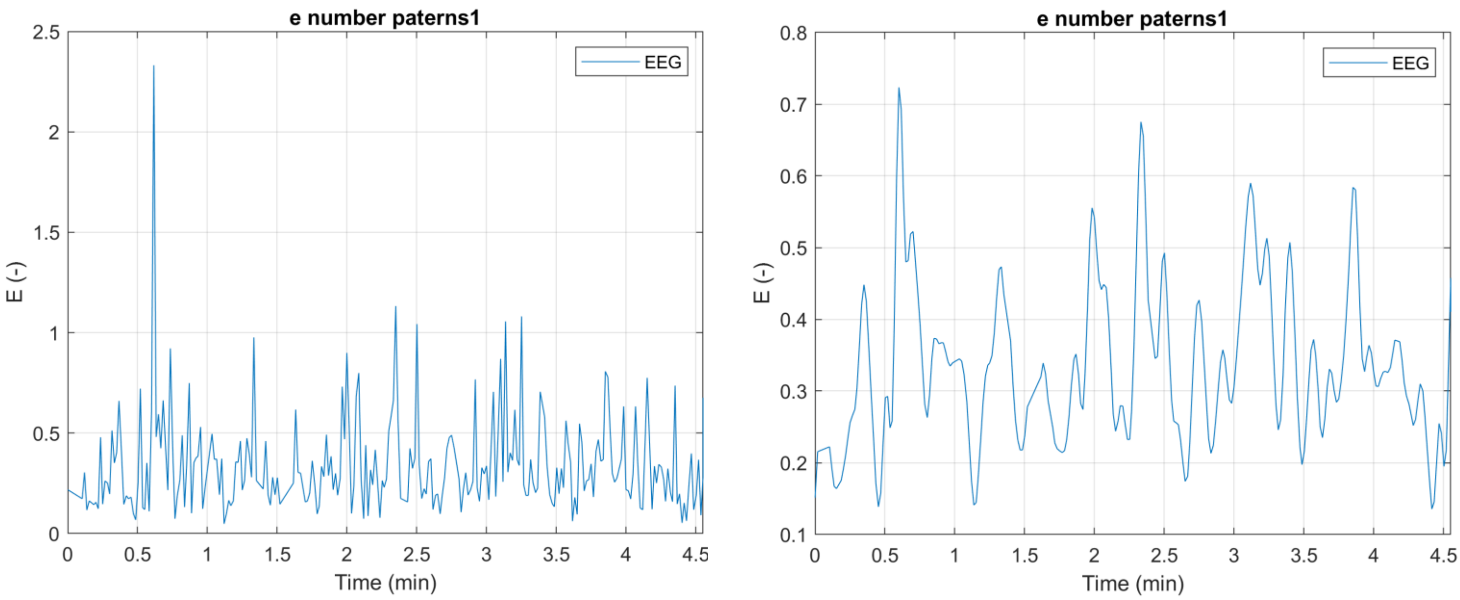


Figure 4 – On the left raw data of EEG engagement index, on the right processed data

5.1.2 HR Data

For heart rate measurements, only the removal of outliers in the data was applied.

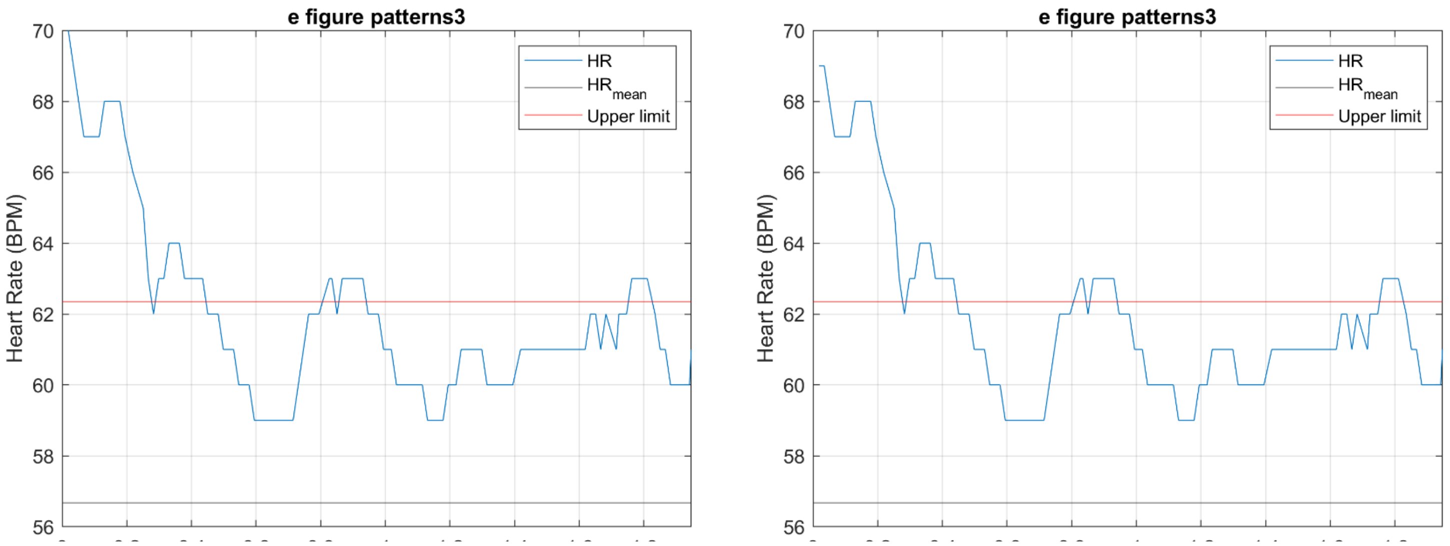


Figure 5 – On the left raw data of EEG engagement index, on the right processed data

5.1.3 GSR Data

For data collected with GSR, smoothing was used and outliers were removed.

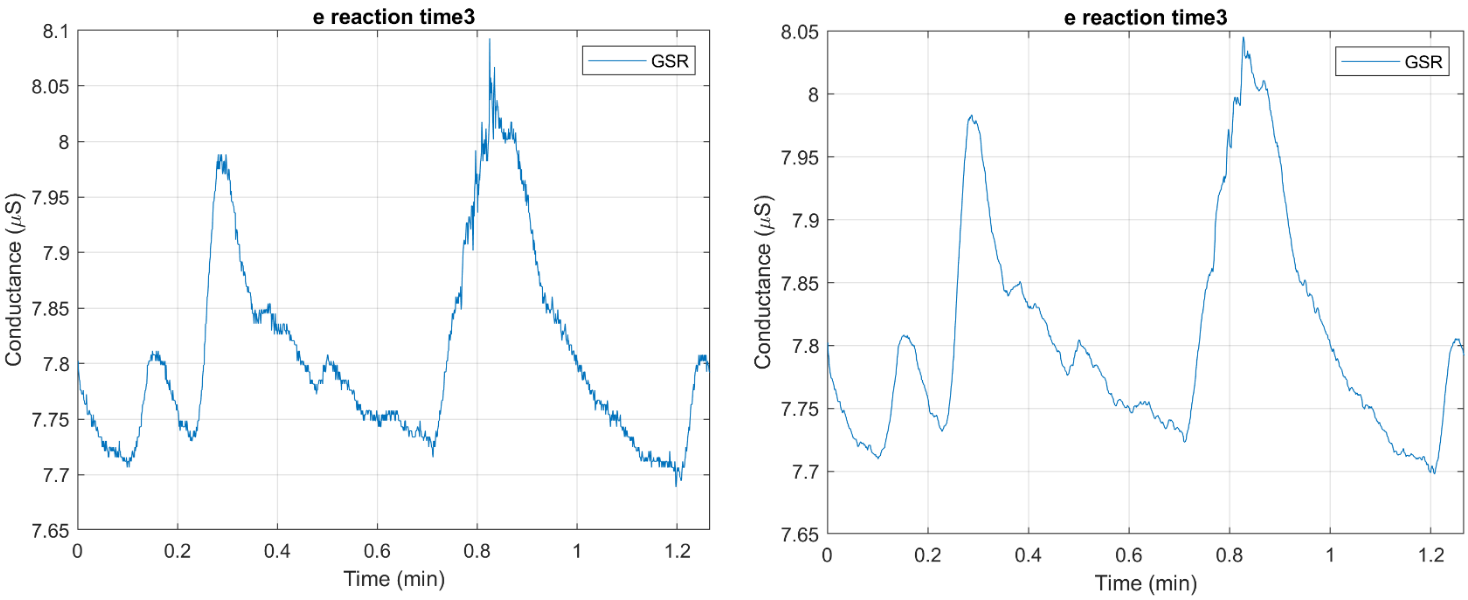


Figure 6 – On the left raw data of EEG engagement index, on the right processed data

5.1.4 Test results

The table below shows the results and the calculated engagement index of person A. The coefficients obtained with biofeedback mean:

- in the case of GSR measurement the baseline value was subtracted from the mean value of each attempt, the disadvantage of this measurement is that the average value increases with the duration of the test,
- in the case of EEG measurement, the higher the value of the index, the greater the involvement/effort of the subject,
- in case of HR the closer the value is to 1, the more relaxed the subject was - the better he/she coped; the closer the value is to 0, the more stressed the subject was.

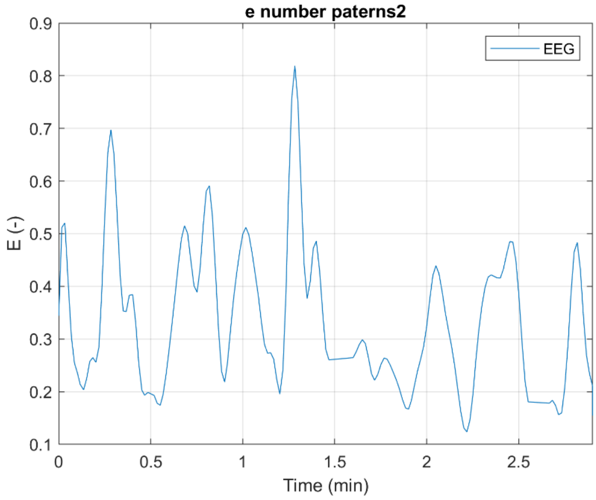
Analyzing the following test categories:

- Reaction time: worst results on first attempt, lowest hr index indicating stress; best results on second attempt, medium level of stress from hr, lower gsr value than on third attempt, low index eeg indicating less involvement
- Figure patterns: the best result was obtained on the second attempt, high eeg index indicates high involvement, low hr index indicates increased stress; on the first attempt slightly worse results were obtained, but the person was less stressed and less involved; on the third attempt coefficients indicate low stress and involvement, which may be due to fatigue during the test
- Number patterns: as in the previous cases, the best result was obtained on the second attempt, the EEG indicates quite high involvement and the hr index indicates medium relaxation.

Table 5

		A			
Relax	e_eeg_mean	0,335			
	e_hr_mean	1			
	gsr_mean	5,393			
		I attempt	II attempt	III attempt	**Mean**
Reaction time	Reaction time	413	387	358	**386,0**
	Aim trainer	582	488	541	**537,0**
	e_eeg_mean	0,312	0,301	0,327	**0,313**
	e_hr_mean	0,506	0,548	0,651	**0,568**
	gsr_mean	1,402	2,156	2,414	**1,991**
Figure patterns	Verbal memory	29	52	20	**33,7**
	Visual memory	9	10	9	**9,3**
	e_eeg_mean	0,316	0,338	0,276	**0,310**
	e_hr_mean	0,986	0,429	0,692	**0,702**
	gsr_mean	1,286	2,128	2,481	**1,965**
Number patterns	Number memory	7	10	8	**8,3**
	Chimp test	12	8	9	**9,7**
	e_eeg_mean	0,271	0,322	0,329	**0,307**
	e_hr_mean	0,415	0,511	0,662	**0,529**
	gsr_mean	1,755	2,481	2,470	**2,235**

In addition, sample graphs for person A during the second attempt at the number patterns task are presented.



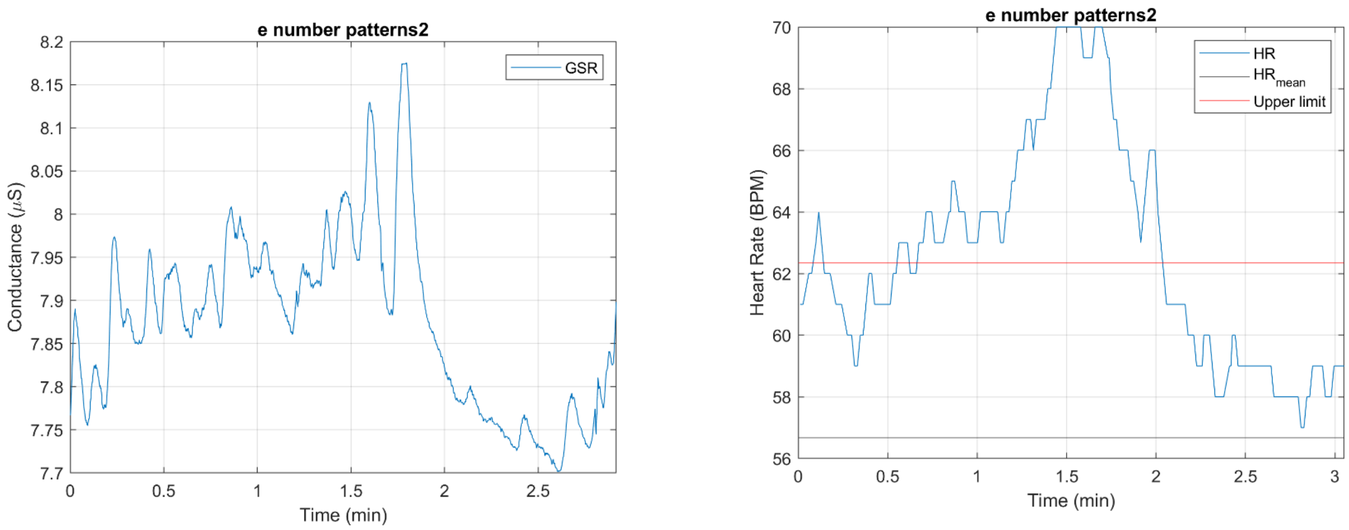


Figure 7 - EEG, HR and GSR results for person A during second attempt of number patterns task

In all the graphs you can see an increase in cognitive load from around minute 1 and a decrease around minute 2. On the EEG chart this occurs a little earlier before 1 and before 2 minutes.

Table 6

		A				B				C				D			
Relax	e_eeg_mean	0,335				0,331				0,260				0,275			
	e_hr_mean	1				0,977				0,909				1			
	gsr_mean	5,393				1,820				3,054				3,453			
		I attempt	II attempt	III attempt	**Mean**	I attempt	II attempt	III attempt	**Mean**	I attempt	II attempt	III attempt	**Mean**	I attempt	II attempt	III attempt	**Mean**
Reaction time	Reaction time	413	387	358	**386,0**	471	392	393	**418,7**	338	337	376	**350,3**	350	362	499	**403,7**
	Aim trainer	582	488	541	**537,0**	713	668	611	**664,0**	538	466	495	**499,7**	572	572	524	**556,0**
	e_eeg_mean	0,312	0,301	0,327	**0,313**	0,351	0,393	0,286	**0,343**	0,400	0,399	0,324	**0,375**	0,327	0,271	0,310	**0,303**
	e_hr_mean	0,506	0,548	0,651	**0,568**	0,632	0,657	0,677	**0,655**	1,000	1,000	1,000	**1,000**	0,663	1	1	**0,888**
	gsr_mean	1,402	2,156	2,414	**1,991**	-0,060	1,226	1,523	**0,897**	2,065	2,959	3,537	**2,853**	4,604	14,814	12,925	**10,781**
Figure patterns	Verbal memory	29	52	20	**33,7**	62	42	45	**49,7**	27	29	75	**43,7**	22	29	35	**28,7**
	Visual memory	9	10	9	**9,3**	9	10	8	**9,0**	10	11	11	**10,7**	11	11	10	**10,7**
	e_eeg_mean	0,316	0,338	0,276	**0,310**	0,497	0,319	0,354	**0,390**	0,420	0,440	0,428	**0,429**	0,312	0,324	0,338	**0,324**
	e_hr_mean	0,986	0,429	0,692	**0,702**	0,099	0,088	0,260	**0,149**	1,000	1,000	0,945	**0,982**	0,327	0,865	0,608	**0,600**
	gsr_mean	1,286	2,128	2,481	**1,965**	0,464	1,383	1,709	**1,185**	2,461	3,128	3,689	**3,093**	10,397	14,772	13,201	**12,790**
Number patterns	Number memory	7	10	8	**8,3**	10	11	10	**10,3**	11	9	7	**9,0**	-	9	11	**10,0**
	Chimp test	12	8	9	**9,7**	12	10	9	**10,3**	10	10	12	**10,7**	-	11	10	**10,5**
	e_eeg_mean	0,271	0,322	0,329	**0,307**	0,328	0,306	0,338	**0,324**	0,370	0,358	0,392	**0,373**	-	0,322	0,334	**0,328**
	e_hr_mean	0,415	0,511	0,662	**0,529**	0,065	0	0,012	**0,026**	0,875	0,843	0,875	**0,864**	-	0,559	0,543	**0,551**
	gsr_mean	1,755	2,481	2,470	**2,235**	0,622	2,041	2,522	**1,728**	3,053	3,548	3,757	**3,453**	-	14,540	14,009	**14,274**

The results of the individual tasks correspond to the results obtained with the sensors. Person C achieved the best results, the engagement index for EEG was the highest (also comparing to baseline values) - highest engagement, the engagement index for HR was the highest - lowest stress.
Comparing person A and D, who received similar results, it can be observed:

- Reaction time: A performed better than D, the biofeedback indexes indicate this (higher E_{eeg} so more engaged and higher E_{HR} so less stressed)
- Figure patterns: both subjects had similar test results, taking into account the biofeedback indexes A did better (lower E_{eeg} or not so much effort and higher E_{HR} so less stressed)
- Number pattern: better results were obtained by D, the biofeedback results also show this (higher E_{eeg} and higher E_{HR})

Person B performed quite well but was also very engaged and stressed - high E_{eeg} values and very low E_{HR} values.
Finally, based on the tests performed, it can be concluded that biophysical parameters change with cognitive load and engagement.

6. Designed system

6.1 Input data for adaptive system

The changes in mode of operation in the developed system are based on the assessment of operator's functional state. It is evaluated using performance measurement (Performance Index) and engagement assessment (Engagement Index). Two different engagement measurement techniques can be applied, objective and subjective. The former one is based on biofeedback analysis as described in the first part of the article. Subjective workload can be measured using several types of questionnaires enabling the operator to rate the level of workload that he experiences. In this study a simplified version of NASA Task Load Index was used. It enables the operator to subjectively rate the workload along six different categories: mental demand, physical demand, temporal demand, performance, frustration and effort.
The final goal is use the three parameters, performance and both subjective and objective workload measurement, to initiate changes in task difficulty. However, in case when some data are not available or it's not possible to ask the operator for subjective assessment due to e.g. time constraints, the system can operate using only one or two of these parameters. Since the research on adaptive automation was conducted in parallel with the one exploring the effects of workload on biofeedback, the method of evaluating workload based on physiological measurements for this particular study was not developed yet. In case of the experiments described below system operation was driven by either only Performance Index or Performance Index and NASA TLX results.

6.2 Fuzzy logic approach for workload assessment

Due to individual differences, uncertainty of human behaviour and the criteria for the operator's mental state evaluation are ambiguous and difficult to define with typical logic approach. It was decided to apply Fuzzy Logic for human functional state assessment. Considering its high ability to cope with non-crisp processes it is assumed to be efficient in analyzing human non-linear behavior.
It was necessary to define membership functions for all inputs (Performance and Engagement – objective and subjective) and for the output (mental workload) of the fuzzy model. The next step was to define rules that describe the relations between inputs and the output.

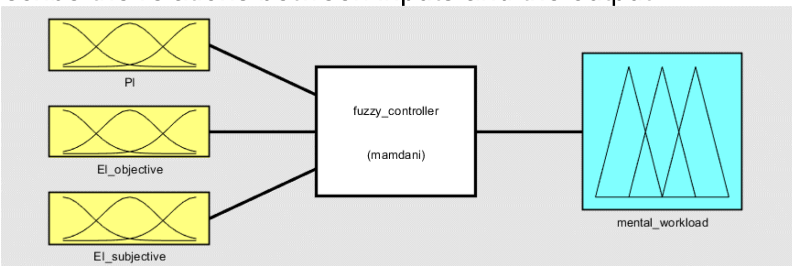


Figure 8 - FIS (Fuzzy Inference System) Process Chart

6.2.1 *Membership functions*

The values and ranges of the membership functions were specified following the analysis of baseline experiment described below. Each membership function can be defined with two parameters: levels of signal and curve of the signal. In this case the membership functions of all inputs and output are triangular or trapezoidal.
Performance Index was divided into three levels: low, average and high.
Engagement Indices, both subjective and objective, also have three levels:

- Low – the operator has little situation awareness and might be not able to cope with an emergency situation
- Optimum – the operator is engaged in the task but is not overloaded and can react to the emergency
- High – the operator is stressed, frustrated, and is not able to cope with current task demands

Five levels of fuzzy logic output, mental workload, were defined: very low, low, optimum, high, very high.

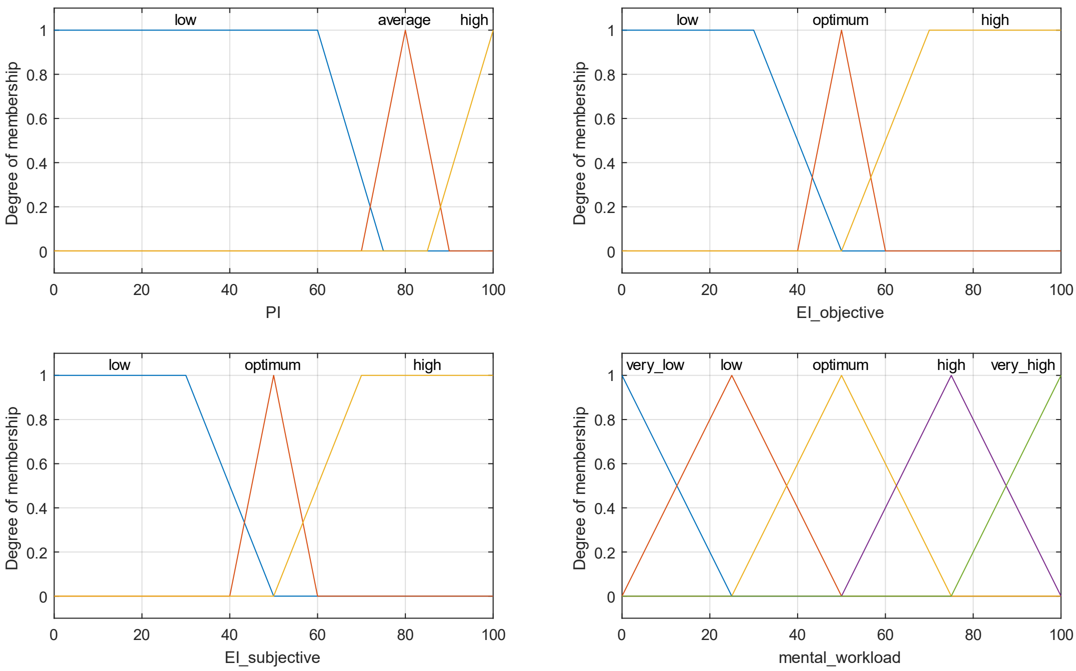


Figure 9 – Membership Functions

6.2.2 Fuzzy Rules

After specifying the membership functions, the rules specifying the logic used to map input data to the relevant output were created. The rules were initially based on the assumptions described earlier illustrated on Figure 1 and were later confirmed during human-in-the-loop experiments. Fuzzy rules are written in the following form:

If (**PI** *low*) and (**EI_objective** *low*) and (**EI_subjective** *low*) than **mental_workload** *very low*

6.3 Simulink model

The adaptive system design was implemented in MATLAB/Simulink.

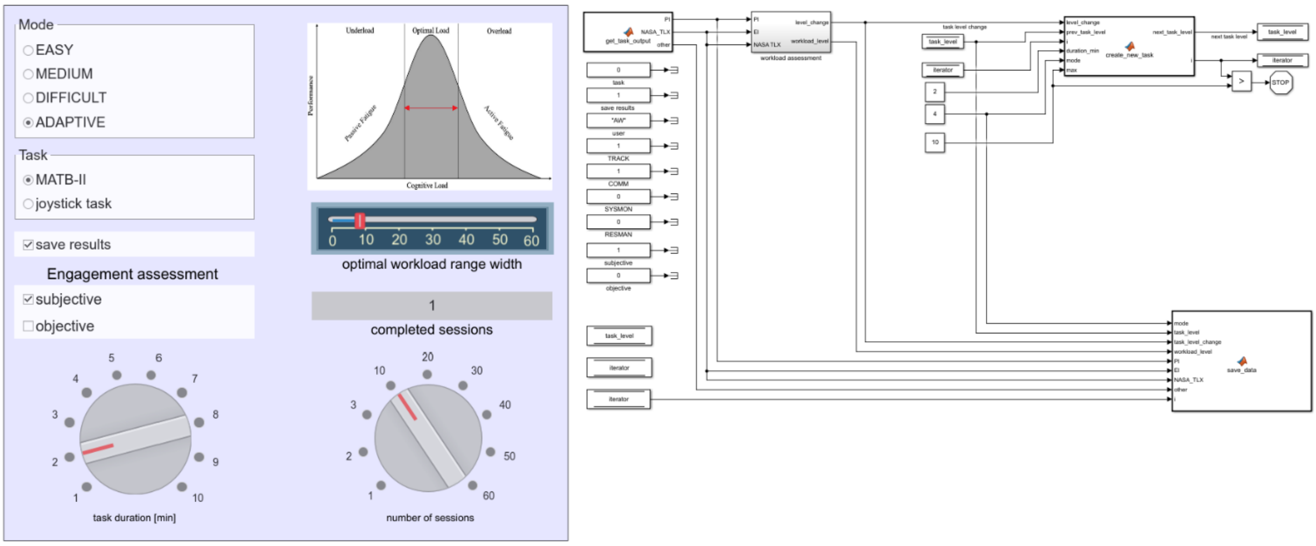


Figure 10 - Adaptive system Simulink model

The system provides a user with the possibility to choose between four modes of operation: three difficulty levels and adaptive mode where task difficulty is dynamically adjusted by the system.
Apart from that, a user can specify the width of optimal workload range within which it is assumed that the operator's engagement and performance are optimal and task difficulty does not need to be changed. Additionally, one can select the task to be executed, engagement assessment methods (objective/subjective) as well as task durations and number of task sessions.
The program launches the task application for the amount of time defined as task duration, then collects performance and engagement data that serve as inputs to fuzzy controller that evaluates the operator's mental workload. If the value of mental workload fits within the optimal range, task difficulty remains the same. However, if the state of underload or overload is detected, task level is increased or decreased respectively.

7. Methodology

The adaptive system was tested during human-in-the-loop experiments in order to determine if it is effective in balancing workload and augmenting task performance. It was expected that implementing the system would help to neutralize workload by increasing it in the underload condition and lowering in the overload condition. It was further assumed that neutralizing workload would positively affect task performance.

7.1 Tasks and experimental conditions

The system was initially designed to be applied to one of two different tasks, Multi-Attribute Task Battery II or joystick task, both described below. With minor modifications in the Simulink model it can work with any other task set. The only requirement is the possibility to change task difficulty and measure performance.

7.1.1 Multi-Attribute Task Battery II

Multi-Attribute Task Battery II is a computer based multitasking environment designed by the NASA to evaluate operator's mental workload and performance during the execution of a benchmark set of tasks. The tasks are supposed to be similar to activities that crew members need to perform in flight.

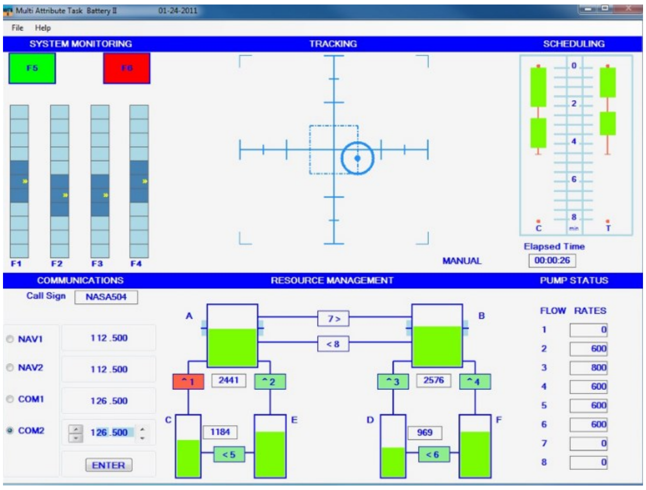


Figure 11 - Multi-Attribute Task Battery II [14]

System Monitoring (SYSMON) task is divided into two sub tasks: scales and warning lights. Each scale has an indicator light that fluctuates in the middle. When it gets near the top or the bottom, the user should respond through the keyboard or with a mouse. In normal conditions the lights' colors are green and white. When the green turns off or the white turns red, it indicates system failure that the operator should react to by pressing the affected rectangle.
Communications (COMM) task simulates pilot interaction with air traffic controller. Operator hears messages asking him to change frequency of a specific radio but should respond only to the requests using the aircraft call sign of *NASA504*.
While executing Tracking (TRACK) task the operator uses joystick to keep the target at the grid center. The task operates in either automatic or manual mode. The mode of operation can be switched several times during task execution indicating automation failure or fix.
The goal of the Resource Management (RESMAN) task is to maintain the amount of fuel in tanks A and B at the optimal level of 2500 units by transferring fuel from lower supply tanks through the use of pumps. When a pump is red, it is not operational and cannot be turned on.
The table below presents task parameters assigned to different levels of difficulty. The values represent the number of events per each 60 seconds of tasks execution.
For SYSMON tasks, the difference of difficulty was mainly manifested on different number of warnings. Difficulty of COMM tasks was manipulated by changing the amount of requests that the operator should respond to (OWN) or ignore (OTHER). In case of TRACK tack the difficulty was determined by the proportion of manual versus automatic states. For RESMAN tasks, difficulty was distinguished by changing the number of pump failures.
Additionally, in difficult and medium task conditions the operator is supposed to execute all MATB tasks while in easy task conditions only 2 tasks are active: TRACK and another one randomly selected from remaining: SYSMON, COMM and RESMAN.

Table 7 - MATB II task difficulty levels

difficulty level	**SYSMON**		**TRACK**	**COMM**		**RESMAN**
	lights	scales	time in MANUAL mode	OWN	OTHER	pump failures
easy	0	1	25 s	1	0	1
medium	1	1	35 s	1	1	3
difficult	3	4	40 s	2	2	6

Performance Index is obtained as average score across all executed tasks. The formulas used to calculate individual task scores are presented below.
In case of system monitoring and communications tasks performance is determined by calculating the percentage of time the task was in an incorrect state in relation to the total time it could be in a correct state. The formula used to compute the percentage of incorrect system time X_i is:

$$X_i = 1 - \frac{C \cdot RT + N \cdot T}{(C+N) \cdot T} \tag{1}$$

RT is the average response time, T is timeout value after which the task is reset if there is no operator response, C is the number of correct responses and N represents the number of times the system timed out without user response. Then the subtask's score is determined as:

$$SYSMON\ Task\ Index = 100 \cdot (1 - X_{SYSMON}) \tag{2}$$

$$COMM\ Task\ Index = 100 \cdot (1 - X_{COMM}) \tag{3}$$

The tracking task performance is calculated using the RMSE (root mean square error) in pixels between the target and the grid center.

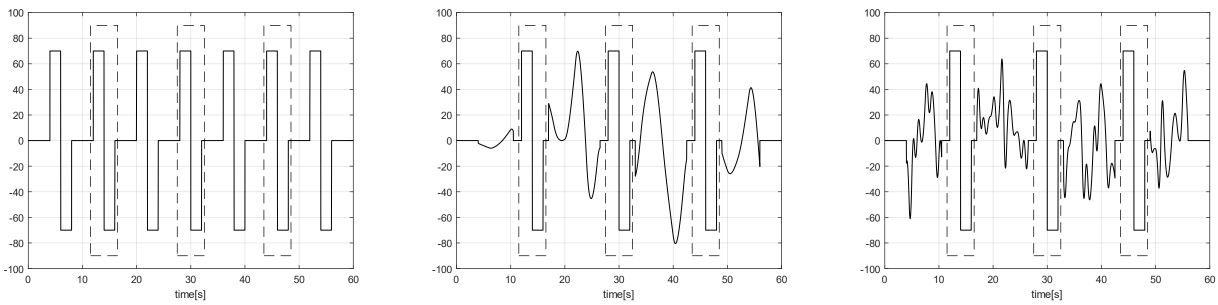
$$TRACK\ Task\ Index = 100 \cdot \left(1 - \frac{RMSE}{Max\ Distance}\right) \tag{4}$$

The resource management performance was evaluated by calculating mean fuel level error in tanks A and B in relation to desired fuel level L_F.

$$RESMAN\ Task\ Index = 100 \cdot \left(1 - \frac{E_{A,B}}{L_F}\right) \tag{5}$$

7.1.2 Joystick task

The second task was designed similarly to the tracking task implemented in [15]. The operator is asked to follow the signal appearing on the screen using a joystick. Three different task conditions can be applied.



(a) (b) (c)

Figure 12 – joystick task difficulty levels (a – easy, b – medium, c - difficult)

Following the approach implemented in [15], in each task condition the reference signal is supplemented with the same pulse signal. Human performance evaluation is based on the analysis of the effectiveness in these selected time intervals (marked by dashed rectangles) when reference signal was identical regardless of current task difficulty level. Task score is based on the characteristics of human model which is identified by using the trajectory of the signal that the subject had to follow as an input and the response of the operator as an output. For this research human dynamic model was adapted from [16]:

$$H(s) = \frac{K(1+T_z s)e^{-T_d s})}{(T_w^2 s^2 + 2\zeta T_w s + 1)} \tag{6}$$

In case of this task Performance Index is calculated using step-response characteristics for the identified human dynamic model, rise time T_r and peak overshoot PO.

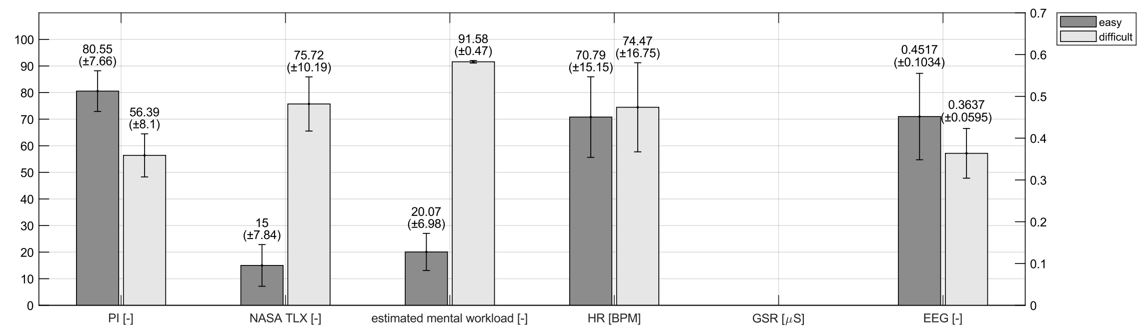
$$PI = 100 \cdot \left(\frac{0{,}5}{1+T_r} + \frac{0{,}5}{1+PO}\right) \tag{7}$$

7.2 Baseline experiments

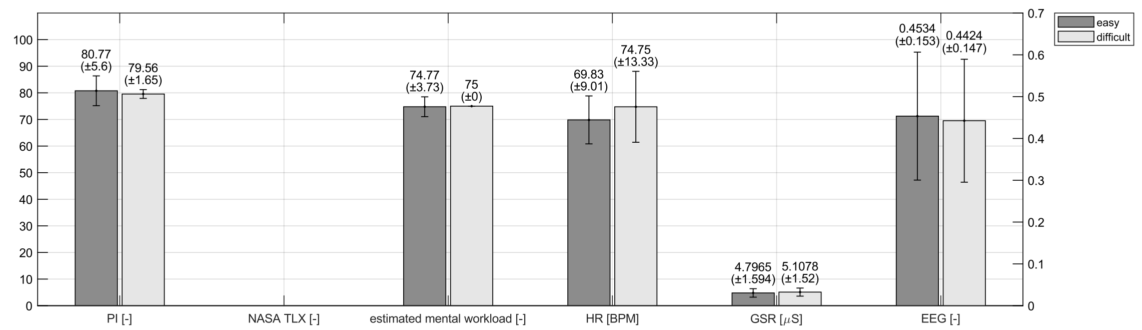
In order to confirm initial assumptions and determine threshold values and ranges for membership functions baseline experiments were conducted first. Three participants were performing first 3 sessions (60 seconds each) of joystick task and then 1 session (120 seconds) of MATB II and at low and high level of difficulty. During task execution both, their performance as well physiological reactions, were registered. While participants were performing joystick task, their heart rate (HR), electroencephalography (EEG) and galvanic skin response (GRS) were measured. Since during MATB II tasks the subjects needed to use both hands, GSR data were not collected. Participants were asked to fill NASA TLX questionnaire only when executing MATB II tasks. In case of joystick task, workload assessment was based only on performance data.

Table 8 - Results of baseline experiments (mean and standard deviation)

task	difficulty level	Performance Index	NASA TLX	heart rate [BPM]	EEG index [-]	GSR [μS]	mental workload
MATB II	easy	80,55 (7,66)	15 (7,84)	70,79 (15,15)	0,4517 (0,1034)	-	20,07 (6,98)
	difficult	56,39 (8,10)	75,72 (10,19)	74,47 (16,75)	0,3637 (0,0595)	-	91,58 (0,47)
joystick task	easy	80,77 (5,60)	-	69,83 (9,01)	0,4534 (0,1532)	4,7965 (1,5943)	74,77 (3,73)
	difficult	79,56 (1,65)	-	74,75 (13,33)	0,4424 (0,1467)	5,1078 (1,5198)	75 (0)



(a)



(b)

Figure 13 - Results of baseline experiments (a - MATB II, b - joystick task)

Considering the results of mental workload assessment during joystick tasks, the system might not be able to correctly identify whether performance decrement is due to overload or underload if evaluation is made using solely performance data. In case of MATB II tasks, where NASA TLX results were also available, the results of human workload assessment properly reflect experimental task conditions.

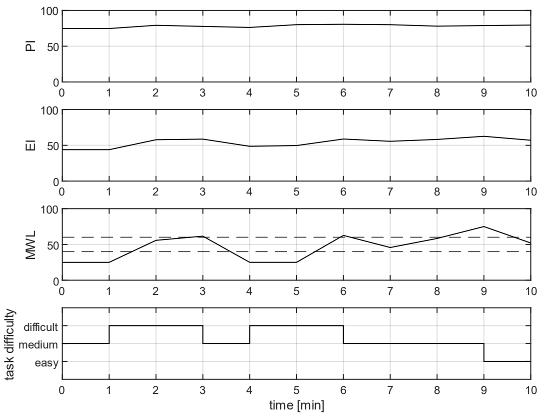
7.3 Adaptive automation system validation

In order to evaluate the impact of using the developed system, three subjects were asked to complete the same tasks but this time task difficulty was adjusted in response to variations in estimated workload. They were performing the tasks in 10 minute sessions, after each 60 seconds the system could change task difficulty accordingly to the operator's functional state. Analysis of the results demonstrates that the system was efficient in maintaining the perceived workload within reasonable range. The overall performance did not significantly improve comparing to the baseline experiments, however during the whole experiment it remained relatively high. The table below presents the summary of task results performed by humans assisted by the adaptive system developed for this research.

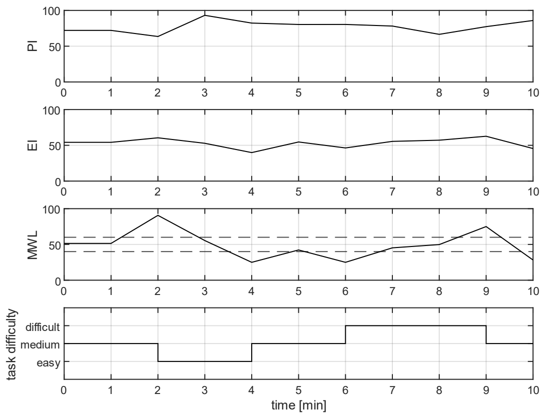
Table 9 - Results of adaptive system experiments (mean and standard deviation)

task	Performance Index	NASA TLX	mental workload
MATB II	75,55 (8,25)	53,18 (6,83)	48,08 (20,08)
joystick task	76,42 (6,06)	53,28 (12,06)	48,75 (22,51)

Based on the experiment data of one of the subjects, the figures below demonstrate how the system operates in time and how it reacts to changes in human performance and self-reported workload.



(a)



(b)

Figure 14 - Experiment data (a – joystick task, b – MATB II)

The main variables logged during the experiment are shown graphically. PI demonstrates variations in task performance, EI presents fluctuations of self-reported engagement and MWL shows the results of mental workload assessment obtained using fuzzy logic approach. The dashed lines represent threshold values for optimal workload. The bottom trace shows how task difficulty was adjusted throughout the experiment. As expected, the system detects changes in human performance and engagement and reacts accordingly by increasing or decreasing task difficulty.

8. Potential applications

The idea presented in the paper shows, that the application of an adaptive system might be a useful tool in various complex-system domains. The authors of [5] have demonstrated that such system can be successfully applied to reduce workload and increase performance in Air Traffic Management context. They could also be applied in videogames domain where physiological signals can be used to infer relevant mental states in order to adapt game difficulty to a more desirable level and keep the gamer sufficiently challenged and engaged [17]. This methodology can also be applied to other, safety-critical, domains such as in aviation and driving e.g. to adaptively switch control of the plane or car depending on the needs of the operator [18]. The research described in [19] proves that adaptive automation can be implemented to create an enhanced command and control infrastructure that enabled more effective operation of unmanned vehicles by reducing workload and improving situation awareness, coordination and performance of SUAS (Small Unmanned Aircraft System) crews. This solution could also be beneficial in the process of UAV operator's training for both the candidate and the instructor. Adjusting task demands to individual needs would adapt the system to the candidate and keep him challenged and engaged without imposing excessive stress and overload. By monitoring candidate's mental state the system could also be used to detect how the potential candidate reacts to the growing stress. Providing the instructor with both, task score and candidate's mental state, would give him wider field of view of the real state of the pilot to adequately assess his abilities.

In general, the application of this kind of technology could enhance the cooperation between humans and machines, improve the overall performance and as a result contribute to higher safety standards [20].

9. Copyright Statement

References

[1] Longo L., A defeasible reasoning framework for human mental workload representation and assessment. *Behaviour & Information Technology*, Vol. 34, No. 8, pp 758–786, 2015

[2] Gouraud J., Delorme A., Berberian B. Influence of automation on mind wandering frequency in sustained attention. *Consciousness and Cognition,* Vol. 66, No. 1, pp 54-64, 2018

[3] Kaber DB, Endsley MR (1997) Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. Process Saf Prog 16(3):126–131

[4] Fairclough S.H. and Venables L. Prediction of subjective states from psychophysiology: A multivariate approach. *Biological Psychology* , Vol. 71, No. 1, pp 100–110, 2006.

[5] Aricò P., Borghini G., Di Flumeri G., Colosimo A., Bonelli S., Golfetti A., Pozzi S., Imbert J.-P., Granger G., Benhacene R. and Babiloni F. Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Frontiers in Human Neuroscience*, Vol. 10, No. 1, pp 539, 2016

[6] O'Hara J. and Higgins J. Adaptive Automation. Current Status and Challenges, 2020.

[7] Dehais F., Lafont A., Roy R. and Fairclough S. A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance. *Frontiers in Neuroscience*, Vol. 14, pp 268, 2020

[8] Teplan M. Fundamentals of EEG Measurement. *Measurement Science Review*, Vol. 2, Section 2, 2002

[9] Hasan M. K., Rusho R. Z., Hossain T. M., Ghosh T. K. and Ahmad M. Design and simulation of cost effective wireless EEG acquisition system for patient monitoring. *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*, pp. 1-5, 2014

[10] Kosmyna N., Maes P. An EEG-Based Closed-Loop Biofeedback System for Real-Time Monitoring and Improvement of Engagement for Personalized Learning. *Sensors*, 2019

[11] British Heart Foundation. *„Your heart rate".* www.bhf.rg.uk, 2014

[12] Maurya A.K. Impact of node density on the performance of GSR & TORA routing protocols. *5th International Conference on Computer and Communication Technology*, 2014.

[13] Shi, Y., Ruiz, N., Taib, R., Choi, E.H., and Chen, F. Galvanic skin response (GSR) as an index of cognitive load. *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, pp 2651-2656, 2007

[14] MATB-II: Revised Multi-Attribute Task Battery, https://matb.larc.nasa.gov/

[15] Kopyt, A., Zużewicz, K. and Bartuzi, P. Experimental identification of a mathematical model of human operator working under mental stress. *Acta of Bioengineering and Biomechanics,* Vol. 19, No. 3, pp 177-185, 2017

[16] Gittleman B., Dwan W.E. and Smiley C.S., System Identification Human Tracking Response. *IEEE Transactions On Education*, Vol. 35, No. 1, pp 31–37, 1992

[17] Pope, A.T., Stephens, C.L., and Gilleade, K. Biocybernetic Adaptation as Biofeedback Training Method, *Advances in Physiological Computing*, 2014

[18] de Visser, Ewart J. et al. Designing an Adaptive Automation System for Human Supervision of Unmanned Vehicles: A Bridge from Theory to Practice. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* Vol. 52, pp 221 – 225, 2008

[19] E. J. de Visser *et al*., TECRA: C2 application of adaptive automation theory. *2010 IEEE Aerospace Conference*, pp. 1-12, 2010

[20] Di Flumeri, G., De Crescenzio, F., Berberian, B., Ohneiser, O., Kramer, J., Aricò, P., Borghini, G., Babiloni, F., Bagassi, S., and Piastra, S. Brain-Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems. *Frontiers in human neuroscience*, Vol. 13, pp 296, 2019