Skill, Rule and Knowledge-based Behaviors Detection during Realistic ATM Simulations by Means of ATCOs' Brain Activity

Gianluca Borghini, Pietro Aricò, Gianluca Di Flumeri, Ilenia Graziani, Alfredo Colosimo, Serenella Salinari, Fabio Babiloni
"Sapienza" University of Rome
Rome, Italy
borghini@dis.uniroma1.it;
ilenia.graziani@gmail.com;
[pietro.arico; gianluca.diflumeri; alfredo.colosimo; serenella.salinari; fabio.babiloni]@uniroma1.it

Jean-Paul Imbert, Géraud Granger, Railane Benhacene
Ecole Nationale de l'Aviation Civile
Toulouse, France
[jean-paul.imbert; geraud.granger; railane.benhacene]@recherche.enac.fr

Alessia Golfetti, Stefano Bonelli, Simone Pozzi
Deep Blue Research and Consulting
Rome, Italy
[alessia.golfetti; stefano.bonelli; simone.pozzi]@dblue.it

***Abstract*—The aim of this work was to test a neuro-physiological methodology able to discriminate the *Skill* (S), *Rule* (R) and *Knowledge* (K) based cognitive control levels of Air-Traffic-Controllers' performing realistic traffic management tasks . The three categories of human behaviours have been associated to specific cognitive functions (e.g. attention, memory, decision making) already investigated with Electroencephalography (EEG) measurements. A link between S-R-K behaviours and expected frequency bands configurations has been hypothesized. Eventually, specific events have been designed to trigger S, R and K like behaviours and then integrated into realistic Air Traffic Management (ATM) simulations. A machine-learning algorithm has been used to differentiate the three different levels of cognitive control by using brain features extracted from the EEG rhythms of different brain areas, that is, the frontal theta and the parietal alpha activities. Twelve professional Air-Traffic-Controllers (ATCOs) from the *École Nationale de l'Aviation Civile* (ENAC) of Toulouse (France) have been involved in the study. The results showed that the algorithm was able to differentiate with high discrimination accuracy (AUC > 0.7) the three S-R-K cognitive behaviours during simulated air-traffic scenarios in an ecological ATM environment**.

Keywords - EEG; Skill, Rule, Knowledge, ATM, ATCO.

I. INTRODUCTION

According to the framework proposed by Rasmussen [1], the terms *Skill*, *Rule* and *Knowledge*-based (S-R-K) refer to "the degree of conscious control exercised by the individual over his or her activities, depending on the degree of familiarity with the task and the environment". The S-R-K behaviours represent three dynamic and parallel cognitive levels of expertise, where the control of behaviour continuously shifts from a level to another one.

At the skill-based level, the behaviour is regulated by the lowest level of conscious involvement and is characterized by highly routinized and automated activities. In fact, skill-based mode refers to "the smooth execution of highly practiced, largely physical actions in which there is virtually no conscious monitoring". In the ATM environment, a large portion of the expert controller's observable behaviour is skill-based: cursor positioning, command entry, use of phraseology.

Rule-based behaviour is also activated in familiar work situations, but it is distinguished from skill-based behaviour, as "it requires some degrees of conscious involvement and attention. Situation assessment leads to recognition of which procedures apply to particular familiar situations". Events implemented at this level are typical control task, such as re-routing, conflicts detection and management, and coordination.

When faced with unfamiliar situations, where no solutions are already available, it is necessary to move to the knowledge-based level of behaviour. At this level, the User "carries out a task in an almost completely conscious manner. This would occur in a situation where a beginner is performing the task (e.g. a trainee at the beginning of its training) or where an expert is facing with a completely novel situation. In either such cases, the User would have to exert considerable mental effort to assess the situation, and his or her responses are likely to be slow. Also, after each control action, the User would need to review its effect before taking further action, which would probably further slow-down the responses to the situation" [2].

This framework of human performance is a useful means to figure out how humans can deal with ambiguous situations, solve familiar or unfamiliar situations, quickly react to the environmental requests, and set new problems in an efficient and flexible way. It is also a powerful framework to orient design and evaluation of new interface system.

The end research goal of this study is to define quantitative, reliable and valid neuro-physiological indicators for the S-R-K levels. This paper presents the testing and validation of such indicators, using data collected from Air Traffic Controllers' brain activity. The development of the actual SRK mental classifier, to identify S-R-K behaviours in real time would be the next step of the research.

II. MATERIALS AND METHODS

A. *Literature review*

Two steps were performed to select which brain features to analyse:

- First, S-R-K levels of control were associated to specific cognitive functions,
- Then a literature review was performed on neuroscience research, in order to identify which brain activity frequency bands were linked to the identified cognitive functions.

SRK	Cognitive processes	Bands	Location	Channels[1]
Skill	High Automated processes and long term memory (procedural) Low Executive control (attention and working memory) Low Attention No Decision-making (resolution of conflicts and error detection) No Problem solving	Low theta	Frontal	Fc3-Fc4 F5-Fc6 F8
		High alpha	Posterior	C3-C4 C5-C6 C1-Cp2 P1-P2 P5-P6 T7-T8 O1-O2
Rule	Less automated processes and long term memory (procedural) than Skill level More executive control (attention and working memory) than Skill level More Attention than Skill level No Decision-making (resolution of conflicts and error detection) No Problem solving	Increased theta respect to Skill level	Frontal	F1-F2 Fc3-Fc4 F5-Fc6 F7-F8 Af7-F6
		Decreased alpha respect to Skill level	Parietal	P1-P2 P5-P6 Cp3-Cp4
Knowledge	No automated processes and long term memory (procedural) Executive control (attention and working memory) High Attention Decision-making (resolution of conflicts and error detection) Problem solving	High theta	Frontal	Af3-Af4 Fp1-Fp2 Af7-Fpz F1-F2 F6-F7-F8
		Low alpha	Parietal	-[2]
		High gamma	Parieto-occipital	-[3]

Table 1: Association between levels of performance and EEG bands.

Literature evidences show that an increase of electroencephalographic (EEG) power spectral density (PSD), especially over the frontal cortex, in the theta band (4 - 7 Hz), and an EEG PSD decrease in the alpha band (8-12 Hz), over the parietal cortex, have been observed when:

- increase of required automated processes [3][4],
- increase of demands on executive control (attention and working memory) [5]–[8],
- activation of decision-making processes, like resolution of conflicts and error detection [9], or problem solving [10],
- increase of mental workload [11] and task complexity [12].

Based on such evidences, the following hypotheses have been formulated:

- *Skill-based behaviour*: high parietal alpha activity increment with respect to the Rule and Knowledge behaviours.
- *Rule-based behaviour*: frontal theta activity increment and a lower parietal alpha rhythm increase than in the Skill-based condition.
- *Knowledge-based behaviour*: higher frontal theta activation than in the Rule and Skill conditions, and parietal alpha activity increment.

[1] For the mapping between EEG channels and Brodmann areas, see http://www.brainm.com/software/pubs/dg/BA_10-20_ROI_Talairach/nearesteeg.htm

[2] No specific channels available from the literature review for this specific level of performance.

[3] No specific channels available, just generic identification of involved brain areas.

B. *Experimental subjects*

Twelve professional (age: 40.41±5.54) ATCOs from the *École Nationale de l'Aviation Civile* (ENAC) of Toulouse (France) have been involved in this study. They were selected in order to have a homogeneous experimental group in terms of age and expertise. They all participated to 5 sessions of one-hour on the experimental platform, like the following: Training 1, Experiment 1 (with EEG recording), Training 2, Experiment 2 (with EEG recording), Training 3. These sessions were followed by a final experiment session, named Experiment 3, where data were collected. The experiments have also been attended by two External Expert ATCOs, and two Pseudo-Pilots, who have interacted with the ATCOs with the aim to simulate real-flight communications and to modulate specific S-R-K events. During the sessions, the Experts sat behind the controllers, listening to R/T communications, observing the radar display, monitoring and triggering SRK events and taking note of anything considered relevant. They gathered data both on the performance of the ATCOs, in terms of air-traffic management, and on how the Controllers reacted in the different S-R-K events.

C. *Experimental ATM scenario*

ATCOs have been asked to perform an ATM simulation using the research simulator hosted at ENAC (Figure 1). The ATM scenario included three levels of difficulty, easy, medium, hard, and lasted 45 minutes. The same experiment was also used to validate a EEG based workload index [13]. The traffic complexity has been modulated by the number of aircraft in the controlled sector and the geometry of conflicts. Six S-R-K events (two for each type, S1, R1, K1, S2, R2, K2) have been inserted into the ATM scenario within coherent difficulty conditions (Figure 2). The S-R-K events have been designed to maximise the realism of ATC tasks (see the following section for more details). The system was calibrated recording ATCOs' brain activity in a *Baseline condition* (rest conditions, with closed and open eyes) and in a *Reference condition* (ATCOs looked at the radar screen without reacting, where two conflict-free flights were being presented). The calibration took place for each ATCO at the beginning of the simulation session.

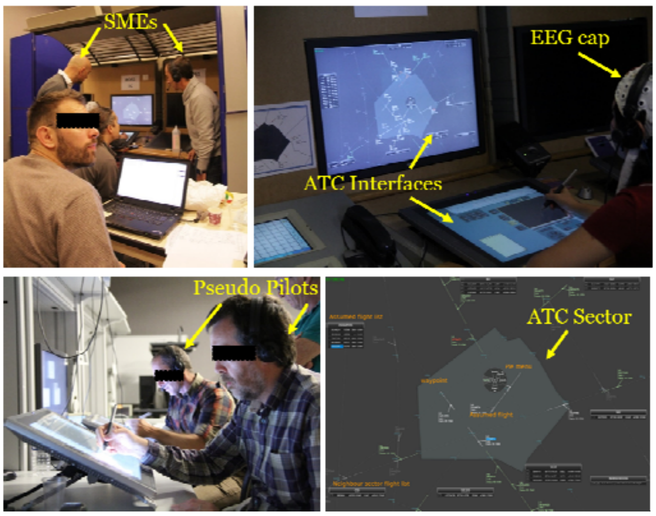

Figure 1. Experimental setup: prototypal ATCO working positions developed by ENAC (Toulouse, France) for a research simulator. The ATCO's brain activity has been recorded continuosly and S, R and K events have been marked in order to recognize them within the entire EEG recording.

D. *S-R-K events*

A Subject Matter Expert (SME) controller from the *Ente Nazionale di Assistenza al Volo* (ENAV, Rome, Italy) was involved in order to create realistic and not disruptive SRK events during the simulation. The events represented an attempt to induce ATCO behaviours associated with S-R-K levels during usual normal air traffic conditions. The following considerations apply to the design of the S-R-K events.

The *S* events were basic interactions with the interface, during the task execution. As the ATCOs participated to previous sessions, we were able to track their progress and make sure they had acquired a high level of proficiency and expertise in the use of the platform interface. This check was needed as not all interface interactions can be classified as Skill-based, only those that are actually carried out at that level. Controllers were asked to visualize the distance between two aircraft (Distance event) and to display the Flight Plan (FPL) trajectory of each aircraft present in the controlled sector (Display FPs).

The *R* events were mainly control-tasks and conflicts-resolutions, during which controllers were also performing skill-events (interaction with the interface). In the two "conflict event" presented, Controllers had to detect and solve a conflict by using the menu of the interface and assigning new altitudes and headings. The hypothesis was that routine conflict detection task represents a familiar situation for ATCOs. Therefore, Controllers would recognize the correct procedures and familiar solutions and then to apply them to solve the conflict.

The *K* events integrated in the scenario were unusual situations. This uncertainty led the Controllers to analyse the situation and to find out the right procedure to cope with the unexpected event. In other words, the ATCOs initially had to analyse the unusual situation (problem setting at the Knowledge-based level) and then came back to the Rule-based level to select the right procedures (problem solving at the Rule-based level, without the need of developing a new solution). In the first Knowledge-based event, "deviation event", Controllers were expected to detect and understand that an aircraft was not following the route filled in the flight plan (FPL). Once contacted, the pilot would state that he was following the right FPL. Controllers needed to understand what was going on.

In the second Knowledge-based event "Unidentified Flying Object (UFO)", the Pseudo-Pilot reported an unknown-traffic detected by the Traffic Collision Avoidance System (TCAS) and a TCAS resolution advisory to avoid a mid-air collision. This unknown aircraft was not displayed on the Controller's radar image. The ATCO was expected to ask additional information to the Pseudo-Pilot. After the avoidance manoeuvre (descent), the Pseudo-Pilot would ask for his previous flight-level, which would display as not changed on the ATCO's HMI. The Controller would see neither the aircraft responsible of the TCAS advisory nor the implementation of the avoidance manoeuvre.

Each event lasted for about 30 seconds. Due to the need of inserting SRK events into realistic ATM tasks, it was not possible to create "pure" SRK behaviours. While Skill events were basic interactions with the interface and the ATCO could be almost entirely focused on them, the same could not be done for the Rule and Knowledge levels. The Rule events were control tasks and conflicts resolutions, during which controllers were also performing at a skill level (having to interact with the interface to handle them). The Knowledge events involved the three levels: Skill + Rule + Knowledge. Considering the need of building realistic situations and taking into consideration the limitations of a simulation, it was possible to prepare events triggering a uncertainty state in controllers, in other words situations that were peculiar enough to make controllers focus on them to try and recognise a familiar situation. After this initial "what is going on?" state, controllers usually came back to the rule level, finding a procedure to solve the problem or deciding to ignore it. In the latter case, their interpretation was that it was pilot’s responsibility or simply not impacting traffic safety.

E. Physiological signals recording and pre-processing

The neurophysiological signals have been recorded by the digital monitoring *BEmicro* system (EBNeuro system). The thirteen EEG channels (FPz, F3, Fz, F4, AF3, AF4, P3, Pz, P4, POz, O1, Oz, O2) and the EOG channel have been collected with a sampling frequency of 256 (Hz). All the EEG electrodes have been referenced to both the earlobes, grounded to the left mastoid, and the impedances of the electrodes were kept below 10 (kΩ). The bipolar electrodes for the EOG have been positioned vertically above the left eye. The acquired EEG signals have been digitally band-pass filtered by a 4th order Butterworth filter (low-pass filter cut-off frequency: 30 (Hz), high-pass filter cut-off frequency: 1 (Hz)) and the EOG signal has been used to artefacts remove eyes-blink from the EEG data by using the [14]. Although EEG is designed to record cerebral activity, it also records electrical activities arising from sites other than the brain. The recorded activity that is not of cerebral origin is termed “*artifact*” and can be divided into physiologic and extraphysiologic artifacts. While physiologic artifacts are generated from the patient, they arise from sources other than the brain (i.e. body). Extra-physiologic artifacts arise from outside the body (i.e. equipment, environment). For other sources of artefacts on the EEG signal, specific procedures of the EEGLAB toolbox, based on threshold methods have been used [15]. In particular, three methods have been used for the artefacts rejection: the threshold criteria, the trend estimation and the sample-to-sample difference. In the threshold criteria the EEG epoch has been marked as “artefact” if the EEG amplitude was higher than ±100 (μV). In the trend estimation, the EEG epoch has been interpolated in order to check the slope of the trend within the considered epoch. If such slope was higher than 3 (no-physiological variation), the considered epoch has been marked as “artefact”. The last check calculated the difference between consecutive EEG samples. If such difference, in terms of amplitude, was higher than 25 (μV), it meant that an abrupt variation (no-physiological) happened, thus it was marked as “artefact”. At the end, the EEG epoch marked as “artefact” have been removed from the EEG recording with the aim to have a clean EEG signals from which estimate the brain parameters for the different analyses. The Power Spectral Density (PSD) has then been estimated by using the Fast Fourier Transform (FFT) in the EEG frequency bands defined for each subject by the estimation of the Individual Alpha Frequency (IAF) value [6].

F. S-R-K estimation

The classification algorithm *automatic stop Stepwise Linear Discriminant Analysis* (asSWLDA, patent pending [16]), developed by the “Sapienza” University of Rome’s spin-off, BrainSigns, has been used to select the most relevant brain spectral features to discriminate the three S-R-K cognitive levels. In particular, the algorithm has been trained by using the brain areas and rhythms found in the scientific literature and described previously (frontal theta and parietal alpha bands). In this way, the algorithm has been trained with brain features extracted from one triplet of S-R-K events (S1, R1, K1) and then tested on the remaining triplet (S2, R2, K2) and vice-versa. For each testing triplet, we calculated the Area Under Curve (AUC) values of the Receiver Operating Characteristic (ROC, [17]) by considering couples between S-R-K distributions. The AUC values related to the discrimination accuracy between the three couples of conditions (S vs R, S vs K, R vs K) have been calculated and analysed for each Controller. It has to be underlined that an

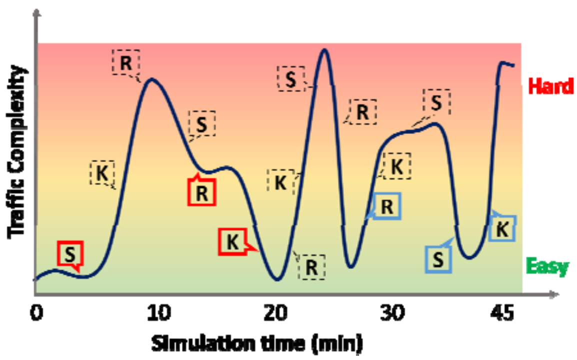

Figure 2. The ATM scenario has been designed with different difficulty conditions (levels of complexity).. The S-R-K events have been selected within coherent difficulty conditions (in red and light-blue squares).

AUC of 1 means a perfect discrimination between the considered classes (S vs R, S vs K, R vs K). On the contrary, if the AUC is equal to 0.5, the algorithm is not able to discriminate the classes. More in general, if the AUC is higher than 0.5 and lower than 0.7, the classification is good, if it is higher than 0.7, the classification is optimum, in other word the classes can be discriminated.

G. Performed analyses

In order to test the effectiveness of the algorithm, for each couple of conditions (S vs R, R vs K, S vs K), we have compared the AUC distributions obtained from the experimental data of all the ATCOs (*Measured AUC*), with the same distributions centred on 0.5 (*Random AUC*), situation corresponding to the *chance level*. As stated before, an AUC of 0.5 means that the algorithm is not able to discriminate the two conditions. We compared the *Random AUC* distributions with the *Measured AUC*, by using three two tailed student t-tests (α=0.05), in order to demonstrate the reliability of the algorithm.

III. Results

The area under curve (AUC) related to the three couples of conditions (S vs R, S vs K, R vs K) have been calculated and reported in figure 4, together with the AUC distribution centred in 0.5 (*Random* AUC). In particular, the results of the statistical tests highlighted that the *Measured AUC* distributions were significantly higher than the *Random AUC* distributions ($p<0.001$). In other words, the machine-learning algorithm was able to discriminate the S-R-K conditions high reliability (AUC > 0.7), thus providing information about the level of cognitive control of the ATCOs.

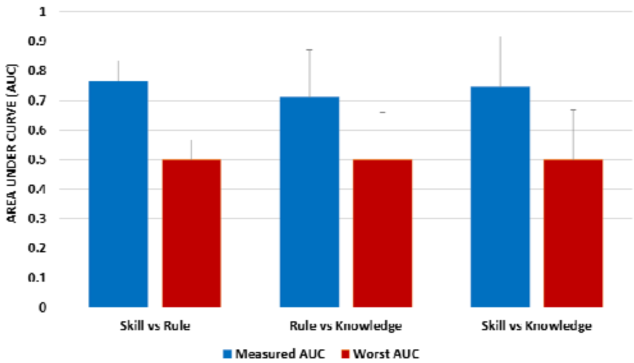

Figure 3. Error bars (CI=.95) related to the *Measured AUC* and the Random *AUC* distributions, related to the three couples of conditions (S vs R, S vs K and R vs K).

IV. Discussion

The research hypothesis of this study was tested with positive results. The algorithm was able to differentiate significantly the ATCOs' cognitive control behaviours (induced by the ad-hoc designed scenario events) with an accuracy higher than 70%. This result should be taken with due consideration of the limitations of this study.

In the design of the ATM scenarios, we introduced S-R-K events compatible with the air traffic situation of that moment, keeping the ATM simulation as realistic as possible. For this reason, the S-R-K events did not fit perfectly with the Rasmussen's model. However, they fit our purpose of triggering skill-based, rule-based and a "what is going on" responses (the closest we could get to the Knowledge-based level with experts).

The aim of the analysis was not to identify moments or "pure Skill", or "pure Rule, "or pure Knowledge" within specific events, nor to develop a real-time SRK classifier. Our goal was to test the possibility of using brain activity to discriminate these cognitive control behaviours. In other words, the S, R and K events have been designed with the aim to investigate the possibility to define a metric, based on specific brain features, by which discriminating and quantitatively estimating the cognitive control behaviour (S, R or K) during the execution of an ATM task. The capability of distinguishing these levels is a pre-condition for the actual development of a real-time SRK classifier.

V. Conclusions

This study demonstrated that it was possible to assess with a high reliability the ATCO's cognitive control level (S-R-K) by monitoring her/his brain activity. To our knowledge, there are no corresponding studies in the existing literature.

Several studies mention the different information processing levels of Skill, Rule, and Knowledge, but we found no mention in these studies of the associated neurophysiological indicators. The aim of this study was to address this gap, by identifying neuro-physiological indicators that could potentially be used to discriminate the S, R and K levels.

The results represent a promising step further in the analysis of human behaviour and demonstrate the possibility of developing new HF tools able to discriminate, also in real-time, the level of operators' cognitive control during ecological tasks. Another possible use might be an online tool for triggering *Adaptive Automations* (AA, [18]-[21]), in which the system behaves depending on the Operator's current level of cognitive control. The authors already implemented a similar solution, but only based on the real-time monitoring of the workload level. Finally, as the level of cognitive control during a task is related to the level of User's expertise, this tool can also be used to track the level of training reached by the User.

Acknowledgment

This work is co-financed by EUROCONTROL on behalf of the SESAR Joint Undertaking in the context of SESAR Work Package E - NINA research project (www.nina-wpe.eu).

References

[1] J. Rasmussen, *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York, NY, USA: Elsevier Science Inc., 1986.
[2] D. Embrey, Understanding human behaviour and error. *Human Reliability Associates*, *1*, 1-10, 2005.
[3] E. Kandel, J. Schwartz, and T. Jessell, *Principles of Neural Science, Fourth Edition*. McGraw-Hill Companies,Incorporated, 2000.
[4] G. Borghini, P. Aricò, I. Graziani, S. Salinari, Y. Sun, F. Taya, A. Bezerianos, N. V. Thakor, and F. Babiloni, "Quantitative Assessment of the Training Improvement in a Motor-Cognitive Task by Using EEG, ECG and EOG Signals," *Brain Topogr.*, Jan. 2015.

[5] Gevins, A and M. E. Smith, "Neurophysiological measures of cognitive workload during human-computer interaction," *Theor. Issues Ergon. Sci.*, vol. 4, no. 1–2, 2003.

[6] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Res. Brain Res. Rev.*, vol. 29, no. 2–3, pp. 169–195, Apr. 1999.

[7] M. B. Sterman, C. A. Mann, D. A. Kaiser, and B. Y. Suyenobu, "Multiband topographic EEG analysis of a simulated visuomotor aviation task," *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.*, vol. 16, no. 1, pp. 49–56, Feb. 1994.

[8] O. Jensen, J. Kaiser, and J.-P. Lachaux, "Human gamma-frequency oscillations associated with attention and memory," *Trends Neurosci.*, vol. 30, no. 7, pp. 317–324, Jul. 2007.

[9] M. Beulen. "The role of theta oscillations in memory and decision making", *Neuroscience and Cognition*. Utrecht University, 2011.

[10] S. Sandkühler and J. Bhattacharya, "Deconstructing insight: EEG correlates of insightful problem solving," *PloS One*, vol. 3, no. 1, p. e1459, 2008.

[11] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neurosci. Biobehav. Rev.*, Oct. 2012.

[12] A. Gevins, M. E. Smith, H. Leong, L. McEvoy, S. Whitfield, R. Du, and G. Rush, "Monitoring working memory load during computer-based tasks with EEG pattern recognition methods," *Hum. Factors*, vol. 40, no. 1, pp. 79–91, Mar. 1998.

[13] Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Graziani, I., Imbert, J. P., ... & Babiloni, F. (2015, August). Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks. In Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE (pp. 7242-7245). IEEE.

[14] G. Gratton, M. G. Coles, and E. Donchin, "A new method for off-line removal of ocular artifact," *Electroencephalogr. Clin. Neurophysiol.*, vol. 55, no. 4, pp. 468–484, Apr. 1983.

[15] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[16] P. Aricò, G. Borghini, G. Di Flumeri and F. Babiloni, "A Method for the estimation of mental state, in particular of the mental workload and its device" ("Metodo per stimare uno stato mentale, in particolare un carico di lavoro, e relativo apparato"). Patent pending. Number P1108IT00, 2015.

[17] D. Bamber, "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph," *J. Math. Psychol.*, vol. 12, no. 4, pp. 387–415, Nov. 1975.

[18] E. A. Byme, R. Parasuraman, "Psychophysiology and Adaptive Automation", Bilogical psychology, Vol. 42, No. 3, 1996, pp 249-268.

[19] Parasuraman R., Mouloua M., Molloy R., Effects of Adaptive task Allocation on Monitoring of Automated Systems, in Human Factors 38(4), 665-679, 1996.

[20] M. Uebbing-Rumke, H. Gürlük, D. Schulze-Kissing, "Adaptive Automation Support for Time-Based Operations in ATC", Third International Air Transport & Operations Symposium ATOS, Delft, Nederland, 2012.

[21] G.F. Wilson, J.D. Lambet, C.A. Russell, "Performance enhancement with Real-Time Physiologically controlled Adaptive Aiding", Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 44, No. 13, pp. 61-64, 2000.