On the Use of Cognitive Neurometric Indexes in Aeronautic and Air Traffic Management Environments

Gianluca Di Flumeri[1,2], Gianluca Borghini[3], Pietro Aricò[3], Alfredo Colosimo[1], Simone Pozzi[4], Stefano Bonelli[4], Alessia Golfetti[4], Wanzeng Kong[5], and Fabio Babiloni[2,6](✉)

[1] Departtment of Anatomy, Histology, Forensic Medicine and Orthopedics, Sapienza University of Rome, Rome, Italy
[2] BrainSigns srl, Rome, Italy
fabio.babiloni@uniroma1.it
[3] Departtment of Physiology and Pharmacology, Sapienza University of Rome, Rome, Italy
[4] Deep Blue srl, Rome, Italy
[5] Department of Computer Science, Hangzhou Dianzi University, Hangzhou, China
[6] Department of Molecular Medicine, University of Rome Sapienza, Rome, Italy

Abstract. In this paper the use of neurophysiological indexes for an objective evaluation of mental workload, during an ecological Air Traffic Management (ATM) task, has been proposed.

Six professional Air Traffic Controllers from the Italian ENAV (Società Nazionale per l'Assistenza al Volo) have been involved in this study. They had to perform an ecological Air Traffic Management task by using the eDEP software, a validated simulation platform developed by EUROCONTROL. In order to simulate a realistic situation, the task was developed with a continuously varying difficulty level, i.e. starting form an easy level, then increasing up to a harder one and finishing with an easy one again. During the whole task for each subject the electroencephalographic (EEG) signals were recorded in order to compute the neurophysiological workload index, and at the same time the subjective perception of the mental workload by using the Instantaneous Self-Assessment (ISA) technique. Thus, the EEG-based workload index, estimated by means of machine learning approach, by one side, and the subjective assessed workload index by the other side, have been compared in terms of correlation and difficulty levels discrimination. By the results it emerged: i) a high positive and significant correlation between the two measures, and ii) a significantly discriminability of the task different difficulty levels by using the EEG-based workload indexes, according to the ISA results.

In conclusion, this study validated the EEG-based mental workload index as an efficient objective evaluation method of the cognitive resources demand in a real operative scenario, and moreover as an index able to monitor its variations.

Keywords: EEG · EOG · Machine learning · Mental workload · Self-assessment · ATM · ATCO · eDEP

B. Blankertz et al. (Eds.): Symbiotic 2015, LNCS 9359, pp. 45–56, 2015.
DOI: 10.1007/978-3-319-24917-9_5

1 Introduction

In specific working environments where safety is paramount big issue, the human factor could be the risk reason less controllable and, at the same time, the main cause of danger. This is often because of an underestimation of the actual mental workload of the operator. In fact, as cognitive workload increases, maintaining task performance within an acceptable range becomes harder. High cognitive workload may demand more cognitive resources than those available in the human brain, resulting into performance degradation and errors commission [1]. The use of objective measures of mental workload based on biomarkers has been proposed for the evaluation of different systems design to allocate the workload, to minimize errors due to overloads or to intervene on the systems in real-time before the operators performing critical tasks become overloaded [2]. For example, few studies investigated neurophysiological indexes about the user states in safety-critical applications, such as driving, industrial environments or security surveillance. With respect to driving assistance applications, recent studies have explored the use of psychophysiological measures in a driving simulation for assessing driving performance and inattentiveness, as well as for robust detection of user intention before the braking onset [3–8].

In this regard, another example of operative environment where lack of performance or overloads may be fatal is the aviation context. Nowadays, the 80% of airplane incidents is still due to human - factors and, as the air - traffic keeps growing exponentially, the impact of new tools able to assess the interaction human – machine, in terms of cognitive resources, is becoming very important. In fact, there are evidences that the failure to perceive correctly the mental demands of a flight task, has been a causative factor in several aircraft accidents. This is true also for other operators critically involved in the air traffic managing (i.e. Air Traffic Control Officers, ATCOs). Both pilots and ATCOs categories of workers have to generate a continuous high quality performance with potential catastrophic results in occasion of error occurrence.

Focusing on the ATCOs, they have to perform a variety of tasks, including monitoring air traffic, anticipating loss of separation between aircraft, and intervening to resolve conflicts and minimize disruption to flow (for an extensive compilation of the tasks and goals of *en-route* control, see [9]). The ATCO's behavior could be measured through several human factor tools, such as the explicit measurement of errors performed during the task, or by using questionnaires related to the perception of the severity of the task executed and so forth, such as for instance the NASA-TLX or the SWAT questionnaires. Each of these methods has pros and cons, but there is not a standard one generally accepted [10], therefore the need of an objective measure becomes more important. Moreover, for their inherently subjective nature, none allows to have an objective and reliable measure of the actual cognitive demand in a real environment. Instead of only measuring secondary physiological effects, the EEG methods will offer a direct insight into the operator's state in complement to the common physiological measurements, as discussed above. There are many evidences that have underlined the correlation between the increase of the cognitive effort and the decision making in a strategy selection process during a complex task and the increase of the Electroencephalogram (EEG) *Power Spectral Density* (PSD) in the theta frequency band [4–7 Hz] over the frontal and occipital brain areas. In addition, it was also noted a corresponding decrease of the EEG PSD in the alpha frequency band [8–12 Hz] over the centro-parietal and parietal brain areas [3–6].

In a previous work [11, 12], it has been defined an algorithm able to evaluate the mental workload of novice ATCOs by using neurophysiological signals, during the execution of ATM task under different difficulty levels. Each difficulty level has been maintained constant for several minutes in order to keep the experimentation as controlled as possible. The results showed that the neurophysiological measure was able to evaluate the mental workload of the operator for each difficulty level.

On the basis of the previous results, the aim of this work was to test the reliability of the algorithm also during more ecological settings, where the difficulty of the task changes continuously. In this way, we have tested if the algorithm was able to track the fluctuation of the operators' mental workload within the operative task. In order to validate the results, the neurophysiological measure has been compared with the subjective measure of the mental workload, collected by the Instantaneous Self-Assessment (ISA) technique.

2 Methodology

2.1 Experimental Protocol and Task

Six professional ATCOs (49 ± 3.2 years) from ENAV S.p.A. (Società Nazionale per l'Assistenza al Volo, Italy) have been involved in this experimentation, in particular they have been asked to manage air-traffic under two different difficulty levels (EASY and HARD), using the ATM simulator eDEP (*Early Demonstration & Evaluation Platform*).

The eDEP software has been developed by EUROCONTROL, with the aim to produce a low-cost-lightweight, web-enabled ATM simulator platform, offering an ideal environment for research and advanced concept projects to rapid prototype applications [13]. A specific experimental protocol has been defined with the aim to highlight the investigated cognitive phenomena, that is the mental workload experienced by the subjects during the execution of the task.

In Fig. 1, a picture of the experimental setting during the task. The air-traffic task lasted about 37 minutes, during which the task difficulty varied between the two levels (EASY and HARD).

Since the eDEP software simulates a real scenario, the difficulty during the whole task varied continuously, thus, there were not constant difficulty conditions, but a

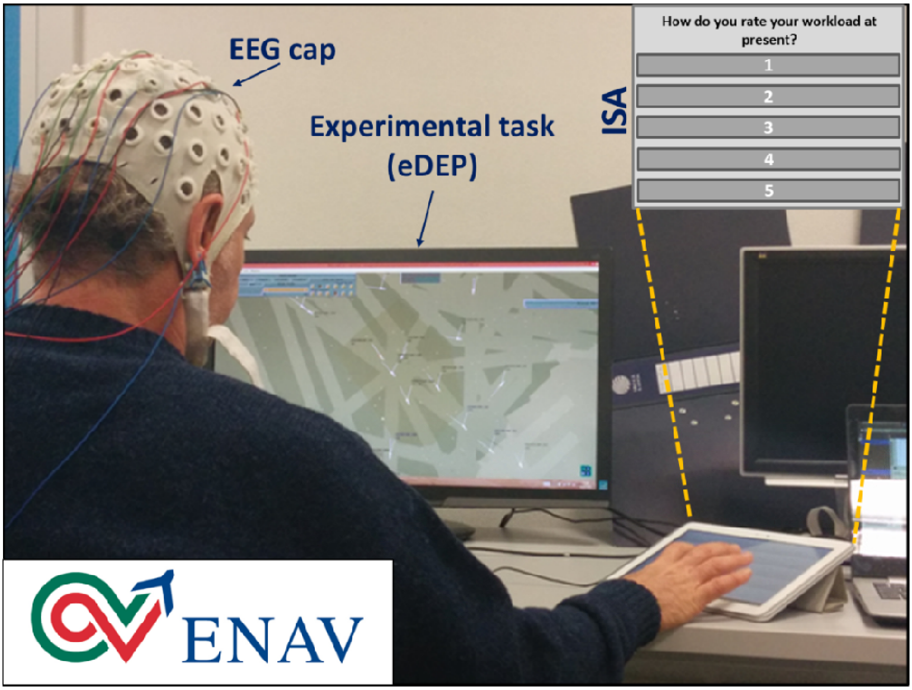

Fig. 1. Picture of the experimental setting during the task: on the right the tablet on which the ISA is presented; on the right at the top a zoom of the ISA interface.

difficulty profile designed as a "reverse – U". In other words, in the first 16 minutes ATCOs had to manage an EASY air traffic condition, in the following 16 minutes an HARD air traffic load and, in the last 5 minutes again an EASY air traffic condition (Fig. 2).

2.2 Instantaneous Self-Assessment (ISA)

Simultaneously to the execution of eDEP, ATCOs have been asked to fill the *Instantaneous Self-Assessment* (ISA).

In particular, the ISA [14] is a technique that has been developed to provide immediate subjective ratings of workload demands, from 1 (very easy) to 5 (very difficult), during the performance of a primary task, in our case, an air traffic management (ATM) task.

The ISA scale has been presented to the ATCOs every 3 minutes in the form of a colour-coded keypad on a tablet screen (10 inchs) positioned just below the main monitor (Fig 1). The keypad flashed and sounded when the workload rating was required, and the participants simply pushed the button that corresponds to their workload perception.

The ISA technique allowed a profile of the operator's workload throughout the eDEP task. The appeal of the ISA technique lies in its low resource usage and its low intrusiveness, that is, it ensured no interference with the main task.

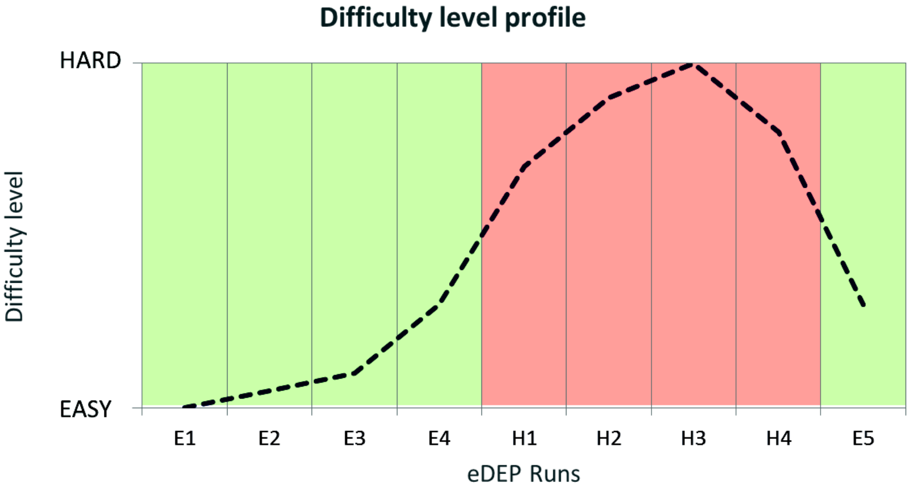

Fig. 2. Profile of the difficulty level, varying during the experimental task on the eDEP platform.

2.3 Neurophysiological Signals Acquisition and Analysis

Neurophysiological signals have been recorded by the digital monitoring *BEmicro* system (EBNeuro system). The 13 EEG channels (FPz, F3, Fz, F4, AF3, AF4, P3, Pz, P4, POz, O1, Oz and O2) and the EOG channel have been collected simultaneously with a sampling frequency of 256 (Hz). All the EEG electrodes have been referenced to both the earlobes, and the impedances of the electrodes have been kept below 10 (kΩ). The bipolar electrodes for the EOG have been positioned vertically above and below the left eye. The acquired EEG signals have been digitally band-pass filtered by a 4^{th} order Butterworth filter (low-pass filter cut-off frequency: 30 (Hz), high-pass filter cut-off frequency: 1 (Hz)). The EOG signal has been used to remove eyes-blink artifacts from the EEG data by using the *Gratton* method [15]. For other sources of artifacts, specific procedures of the *EEGLAB* toolbox, based on threshold methods have been used [16].

After the artifact rejection, the EEG signals have been segmented in epochs of 2 seconds, 0.125 (ms) shifted. The PSD has then been estimated, for each epoch and for each EEG channel, 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 [17], correlated with the mental workload variations, therefore the theta [IAF-6 ÷ IAF-2] (Hz) and alpha [IAF-2 ÷ IAF+2] (Hz) bands. Furthermore, the PSD has been calculated using a Hanning window of the same length of the considered epoch (2 seconds length, is that 0.5 (Hz) of frequency resolution. Thus, with this frequency resolution, and considering the investigated frequency range equal to [IAF-6 ÷ IAF+2] (Hz), there was 17 PSD values for each channel.

2.4 EEG-Based Workload Index

A *Stepwise Linear Discriminant Analysis* (SWLDA, [11–12]) has been used to select the most relevant spectral features, within a features domain consisted of 221 values (13 ch * 17 PSD values), to discriminate the mental workload of the subjects within the different experimental conditions (EASY and HARD). In particular, the performed SWLDA used $\alpha_{ENTER} = .05$ and $\alpha_{REMOVE} = .1$, as probabilistic criterion for including and excluding features of the SWLDA itself. Once identified such spectral features, the SWLDA assigns to each one specific weights ($w_{i\ train}$), plus a bias (b_{train}), such that the SWLDA discriminant function ($y_{train}(t)$) takes the value *1* in the hardest condition and *0* in the easiest one. This step represents the *training phase* of the classifier. Later on, the weights and the bias determined during the training phase have been used to calculate the linear discriminant function ($y_{test}(t)$) over the testing dataset (*testing phase*). Finally, a moving average of 8 seconds (8MA) has been applied to the $y_{test}(t)$ function in order to smooth it out by reducing the variance of the measures, and we defined it as *EEG-based workload index (W_{EEG}).*

Here below are reported the training SWLDA discriminant function (1, where $f_{i\ train}(t)$ represents the PSD matrix of the training dataset at the time sample *t*, and of the i^{th} feature), the testing one (2, where $f_{i\ test}(t)$ is as $f_{i\ train}(t)$ but related to the testing dataset) and the equation of the *EEG-based workload index, W_{EEG}* (3).

$$y_{train}(t) = \sum_i w_{i\ train} \cdot f_{i\ train}(t) + b_{train} \qquad (1)$$

$$y_{test}(t) = \sum_i w_{i\ train} \cdot f_{i\ test}(t) + b_{train} \qquad (2)$$

$$W_{EEG} = 8MA\ (y_{test}(t)) \qquad (3)$$

In order to have a more accurate resolution in terms of task difficulty variation, the dataset related to each subject has been segmented in 9 parts of 4 minutes each, so that we gathered 4 EASY runs (E1, E2, E3, E4), 4 HARD (H1, H2, H3 and H4) runs and another EASY run (E5). At this point, for each subject we have used the algorithm described above to train the classifier with one couple of EASY and HARD runs and to test it over the remaining eDEP difficulty conditions of the same subject. In order to appropriately choose this couple of conditions, we considered that i) the Controllers could need few minutes (i.e. E1 and E2) to become confident with the eDEP interface, so that we have used the *E3* as easy condition in the training dataset; ii) for the hard condition we have taken into account that the eDEP scenario's profile has been designed as a "reverse – U", so that we expected that the hardest condition would be in the middle. Therefore, the *H3* run has been chosen as hard condition in the training dataset.

2.5 Performed Analyses

Two kinds of statistical analysis have been performed in this study. In the first one, we estimated the Pearson's correlation coefficient between the ISA scores and the W_{EEG} measures for each run (i.e. E1, E2, E4, H1, H2, H4, E5). The E3 and H3 runs have not been considered in the analysis because they have been used for training the classifier.

In the second analysis, we tested both if the ISA and the W_{EEG} scores were statistically different over the two difficulty levels (EASY and HARD). In this way, we averaged both the EASY and HARD runs, and we performed two one-tailed student's t-tests ($\alpha = .05$) between the two classes, one for the ISA scores and one for the W_{EEG} indexes.

Before every statistical analysis, we used a *z-score* [18] correction of data for normalize the different behaviors of the subjects. In particular, we calculated this score by using the mean and the standard deviations of the related values (ISA, W_{EEG} scores) over the different runs (i.e. E1-E5, H1-H4).

3 Results

In Fig. 3, the ISA scores have been reported for each of the experimental tasks (i.e. E1-E4, H1-H4, E5). It could be appreciated as the individual perceptions of the tasks difficulties as proposed to the ATCos are in line with the intended simulations. In fact, the harder conditions (H2-H4) were perceived as more engaging cerebrally when compared to the easy ones (E1-E5). The proposed H1 task was perceived as not difficult from the ATCos involved with respect to the experimenter's judgment.

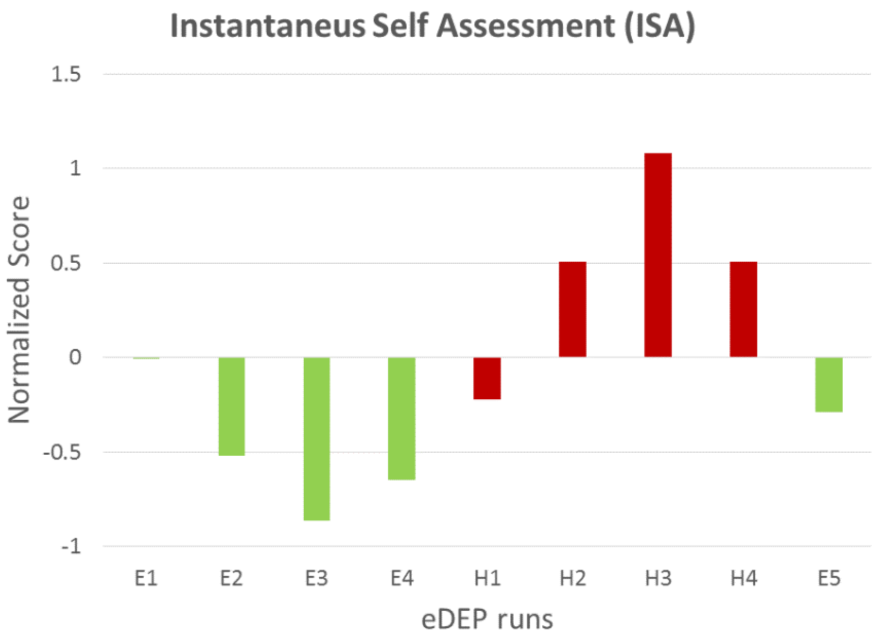

Fig. 3. ISA normalized score in the different eDEP difficulty conditions (E: EASY, H:HARD) of the ATCOs from ENAV.

Furthermore, in the Fig. 4 we showed the same representation for the EEG workload score, in which the E3 and the H3 tasks have not been reported, as stated before, since they have been used for training the classifier employed. It is possible to note that the red columns associated with the estimation of the EEG-based workload score are higher when compared to the green colums, associated to the estimation of the same score for the easier conditions. Note as the easy condition after the occurrence of a difficult sequence (E5, that arrives after the hard condition H4) was estimated of low engage from the personal judgments of the ATCOs while it was estimated still engaging above the average by using the EEG based workload index. This could suggest a major sensitivity of the EEG index to the cumulative effect of mental fatigue when compared to the personal judgment that at this moment appears to be based on external characteristic of the air traffic proposed.

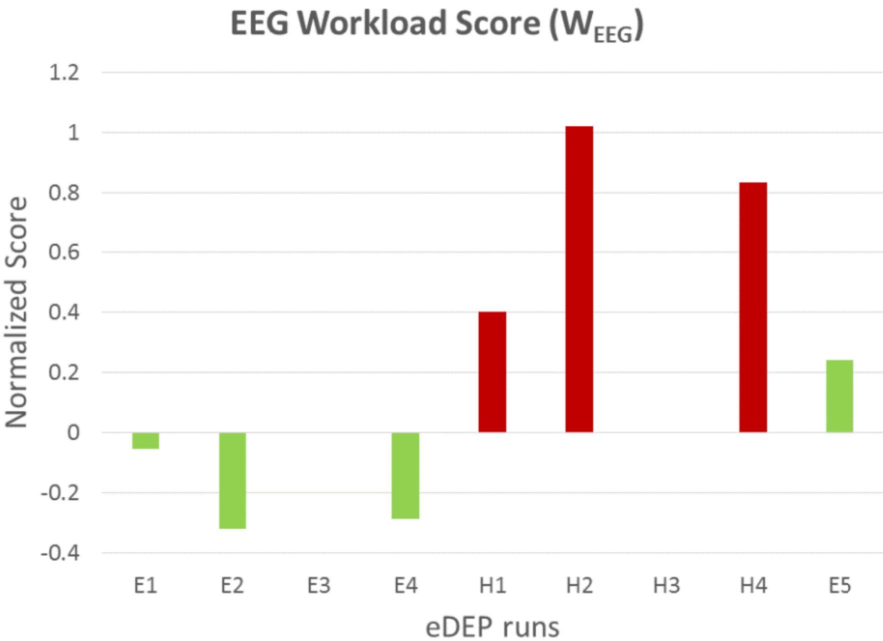

Fig. 4. Workload normalized score in the different eDEP difficulty conditions (E: EASY, H: HARD) estimated by the EEG signals of the ATCOs from ENAV.

It appears that there is a quite strong correlation with the detection of the difficult working states for EEG and ISA based indexes. In order to quantify such correlation, for the runs in which both the ISA and the W_{EEG} scores (i.e. E1 – E2 – E4 – H1 – H2 – H4 – E5) have been available, we performed a Pearson's correlation analysis. Results showed a high positive significant correlation between the two indexes trends ($R = 0.9$; $p = 0.006$), as highlighted in Fig. 5.

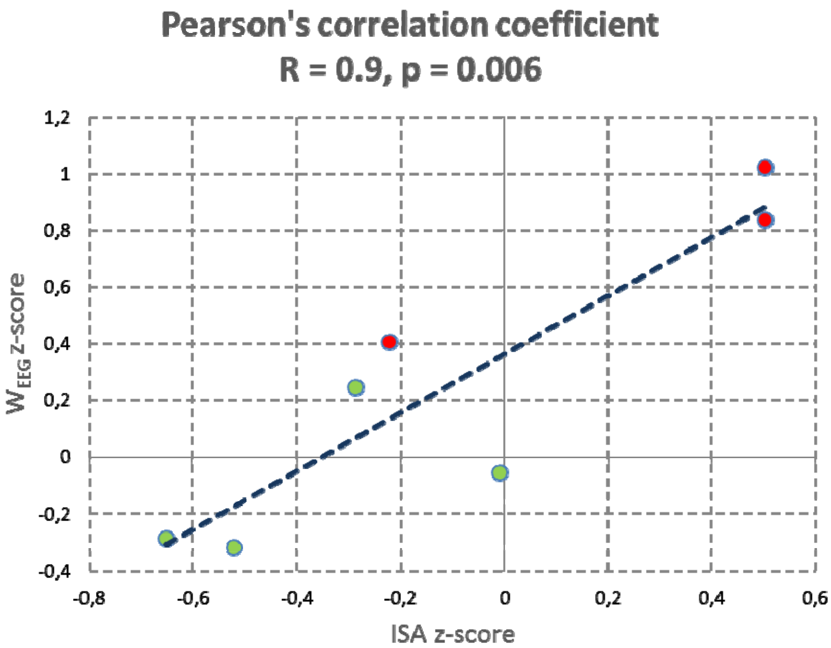

Fig. 5. Scatter plot between the normalized ISA score and the normalized W_{EEG} workload indexes. The tendency line shows the positive correlation between the brain workload perceived by the subject (ISA) and the neurophysiological measure of it (W_{EEG}).

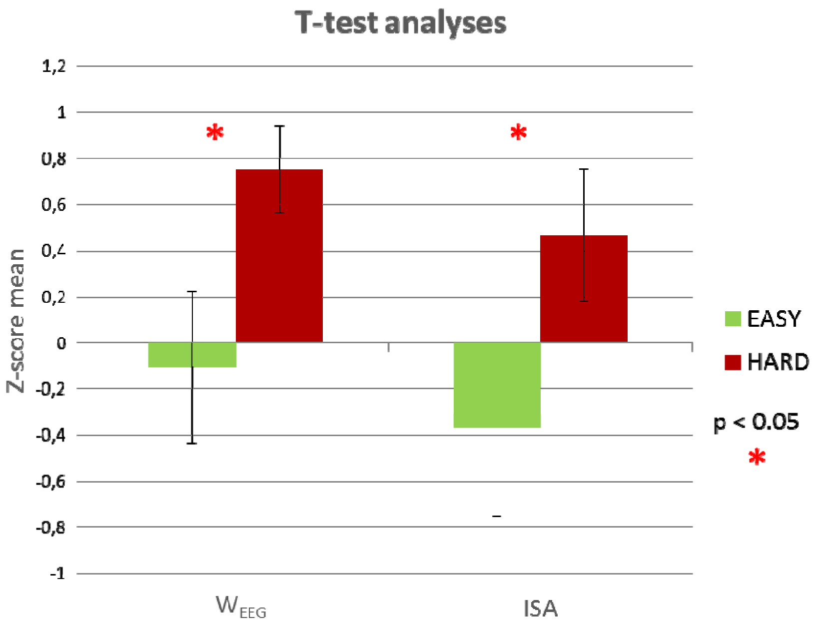

Fig. 6. The difficulty levels (EASY and HARD) of the task are significantly discriminated by both the indexes, the subjective (ISA) and the neurophysiological one (W_{EEG}).

The t-tests showed that the ISA scores related to the EASY and HARD conditions were significantly different ($p < .05$). Consistently, the overall EEG workload index calculated over the HARD conditions was significantly higher than the index calculated over the EASY tasks ($p < .05$). Figure 6 shows the results of the application of such t-tests on the different experimental conditions analyzed. The red columns are associated to the values of the analyzed indexes related to the hard working conditions while the green columns are associated to the values of the indexes related to the easy working conditions. It could be appreciated as both the use of EEG workload index as well as ISA are able to significantly distinguish the easy and the hard working conditions.

4 Discussion

Professional ATCOs have been involved in this study, where a neurophysiological workload measure (W_{EEG}) has been tested while the ATM operators performed an ecological air traffic control task. The ATCos have not been trained to use the eDEP platform before the experiments and, even if eDEP is a professional ATM simulator, during the first parts of the task (EASY1 and EASY2) they needed information and instructions to learn how to use correctly its interface. This aspect has been confirmed both by the ISA (Fig. 3) and by the EEG workload index (Fig. 4). In fact, during the E1 and E2 runs subjects showed higher workload perception (ISA) and physiological increment (W_{EEG}) of the workload than during the next EASY runs. Furthermore, as the eDEP scenario's profile has been designed as a "reverse – U", both the ISA score and the mental workload (W_{EEG}) index showed the same shape, confirmed by a high and significant correlation index ($R = 0.9$; $p = 0.006$).

In conclusion, both the workload perception (ISA) and the neurophysiological (W_{EEG}) measures showed a significant discriminability ($p<.05$) between the difficulty levels (EASY and HARD).

Our previous studies [5, 6, 11, 12, 19] showed the possibility to track the mental workload of the user even online, during simulated tasks in laboratory settings. In those studies, the difficulty of the task has been maintained constant for each experimental condition.

The results of the actual study confirmed that the neurophysiological workload measure can be used as a reliable index of the mental workload experienced by an operator also in ecological working scenario, where the difficulty of the task has not a discrete, but a continuous, profile. With the aim of confirming these results, further experiments will be performed over a bigger experimental sample size of ATCos, and probably with a greater resolution in terms of difficulty levels, in order to ensure that the estimated index is actually related to the experienced brain workload.

5 Conclusions

An algorithm able to track the mental workload of the user by using its brain activity, while performing an ecological operative task has been proposed in this study.

Results showed that the neurophysiological workload index (W_{EEG}) i) showed a high significant correlation with the perceived workload (ISA) and ii) was able to discriminate significantly two different difficulty levels, according to the ATCOs self-assessment.

We can then conclude that neurophysiological measures could provide objective evaluation of cognitive phenomena, e.g. the mental workload, both in real-time (on-line) and in ecological environments. In fact, questionnaires or rating scales might not fit real operative settings, where the operators (e.g. ATCOs) have to be focused exclusively on the task and they could not pay attention to secondary task(s), with the aim to provide data about their cognitive state, probably increasing the final task demand and operating in dangerous condition (under or over-load zone).

Acknowledgments. The present work was supported by the European Commission by FP7 project "ALICIA" and co-financed by Euro Control acting on behalf of the SESAR joint undertaking as part of work package E in the SESAR program and NINA project. The grants provided to FB by the Italian Minister of University and Education under the PRIN 2012 and that provided by the Minister of Foreign Affairs for the bilateral relation between Italy and China for the project "Neuropredictor" are gratefully acknowledged.

References

1. Norman, D.A., Bobrow, D.G.: On the analysis of performance operating characteristics. Psychological Review **83**(6), 508–510 (1976)
2. Byrne, E.A., Parasuraman, R.: Psychophysiology and adaptive automation. Biol Psychol. **42**(3), 249–268 (1996)
3. Welke, S., Jurgensohn, T., Roetting, M.: Single-trial detection of cognitive processes for increasing traffic safety. In: Proceedings of the 21st (Esv) International Technical Conference on the Enhanced Safety of Vehicles, Held June 2009, Stuttgart, Germany (2009). http://trid.trb.org/view.aspx?id=1100165. Accessed November 10, 2011
4. Venthur, B., Blankertz, B., Gugler, M.F., Curio, G.: Novel applications of BCI technology: psychophysiological optimization of working conditions in industry. In: 2010 IEEE International Conference on Systems Man and Cybernetics (SMC), pp. 417–421. IEEE (2010)
5. Borghini, G., Arico, P., Astolfi, L., Toppi, J., Cincotti, F., Mattia, D., et al.: Frontal EEG theta changes assess the training improvements of novices in flight simulation tasks. In: Conf Proc IEEE Eng Med Biol Soc. 2013, pp. 6619–6622 (2013)
6. Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F.: Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci Biobehav Rev., October 30, 2012
7. Haufe, S., Kim, J., Kim, I., Sonnleitner, A., Schrauf, M., Curio, G., Blankertz, B.: Electrophysiology-based detection of emergency braking intention in real-world driving. J Neural Eng **11**(5), 056011 (2014)
8. Haufe, S., Treder, M.S., Gugler, M.F., Sagebaum, M., Curio, G., Blankertz, B.: EEG potentials predict upcoming emergency brakings during simulated driving. J Neural Eng **8**, 056001 (2011)
9. Rodgers, M.D., Drechsler, G.K.: Conversion of the CTA, Inc., en route operations concept database into a formal sentence outline job task taxonomy. Washington, DC: Federal Aviation Administration Office of Aviation Medicine (1993)

10. Rubio, S., Díaz, E., Martín, J., Puente, J.M.: Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. Applied Psychology. **53**(1), 61–86 (2004)
11. Aricò, P., Borghini, G., Graziani, I., Imbert, J.-P., Granger, G., Benhacene, R., et al.: Air-Traffic-Controllers (ATCO) - Neurophysiological analysis of training and workload ATCO. IJASM **12**, 18–35 (2015)
12. Arico, P., Borghini, G., Graziani, I., Taya, F., Sun, Y., Bezerianos, A., et al.: Towards a multimodal bioelectrical framework for the online mental workload evaluation. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3001–3004 (2014)
13. Owen, S., Stainton, C.: eDEP Technical Document – CWP Performance report. 2006 Jan. Report No.: Reference GL/eDEP/EEC/TRSC60-2004/WP6.4
14. Kirwan, B., Scaife, R., Kennedy, R.: Investigating complexity factors in UK air traffic management. Human Factors and Aerospace Safety 1(2) (2001). http://trid.trb.org/view.aspx?id=717648. Accessed April 28, 2014
15. Gratton, G., Coles, M.G., Donchin, E.: A new method for off-line removal of ocular artifact. Electroencephalogr Clin Neurophysiol. **55**(4), 468–484 (1983)
16. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. **134**(1), 9–21 (2004)
17. Klimesch, W.: EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res Brain Res Rev. **29**(2–3), 169–195 (1999)
18. Zhang, J.-H., Chung, T.D.Y., Oldenburg, K.R.: A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J Biomol Screen. **4**(2), 67–73 (1999)
19. Borghini, G., Aricò, P., Graziani, I., Salinari, S., Sun, Y., Taya, F., et al.: Quantitative Assessment of the Training Improvement in a Motor-Cognitive Task by Using EEG, ECG and EOG Signals. Brain Topogr. January 22, 2015