Towards a multimodal bioelectrical framework for the online mental workload evaluation

Pietro Aricò- *IEEE Student Member*, Gianluca Borghini- *IEEE Student Member*, Ilenia Graziani, Fumihico Taya, Yu Sun, Anastasios Bezerianos- *IEEE Member*, Nitish V. Thakor- *IEEE Member*, Febo Cincotti- *IEEE Member*, and Fabio Babiloni- *IEEE Member*

Abstract— **In this study, a framework able to classify online different levels of mental workload induced during a simulated flight by using the combination of the Electroencephalogram (EEG) and the Heart Rate (HR) biosignals has been proposed. Ten healthy subjects were involved in the experimental protocol, performing the NASA - Multi Attribute Task Battery (MATB) over three different difficulty levels in order to simulate three classic showcases in a flight scene (cruise flight phase, flight level maintaining, and emergencies). The analyses showed that the proposed system is able to estimate online the mental workload of the subjects over the three different conditions reaching a high discriminability ($p<.05$). In addition, it has been found that the classification parameters remained stable within a week, without recalibrating the system with new parameters.**

I. Introduction

Mental workload monitoring is of particular interest in safety-critical applications where human performance is often the last controllable factor. In general as cognitive workload increases, maintaining task performance within an acceptable range becomes more difficult. Increased cognitive workload may demand more cognitive resources than that available by the operator, thus resulting in a performance degradation and an increased occurrence of errors [1]. Objective measures of mental workload based on biomarkers could be used to evaluate alternative system designs, to appropriately allocate imposed workload to minimize errors due to overloads. Such system could intervene in real-time before the operator become overloaded while performing safety-critical tasks [2]. Different systems to estimate the mental workload has been previously presented by using EEG or HR or other biometric signals [3], [4], [5]. However, all of these works presented the use of a single modality each time (e.g. only EEG, or only HR etc etc). Since it has been noted in literature as the EEG and HR are sensitive to different components of the mental workload [6], the question whether the reliability of the mental workload detection could benefit from the simultaneous use of multimodal signals (EEG, HR) arose. The purpose of the present work is to investigate the combined use of EEG and HR for the detection of the mental workload when compared to the use of the single modality alone. In this way it has been designed, implemented and evaluated a framework to quantify online the mental workload in subjects involved in managing concurrent tasks at different difficulty levels. All of these tasks are with a clear relevance for the flight control.

II. Materials and methods

A. Experimental protocol

Ten healthy voluntary male subjects (mean age = 25±3) have been involved in this study. All subjects were students and/or staff members of the National University of Singapore (NUS). The study protocol was approved by the local Ethics Committee and all subjects gave their written informed consent. In addition, all the subjects have been paid to take part at the experimental protocol. Scalp EEG has been recorded from 16 EEG electrodes (FPz, F3, Fz, F4, AF3, AF4, C3, Cz, C4, P3, Pz, P4, POz, O1, Oz, O2) referenced to the earlobes and grounded to the AFz electrode (sample rate of 256Hz). Also, the HR and the vertical EOG activity were recorded at the same time of the EEG. The task chosen to be performed by the subjects was the Multi-Attribute Task Battery (MATB, [7]) which provides a benchmark set of tasks about operator performance and workload and simulates the activities inside an aircraft's cockpit. Tasks features include an auditory communications task, a fuel resources management, an emergency lights control and a task of cursor (Figure 1).

P. Aricò is with the Dept. Physiology and Pharmacology, University “Sapienza” of Rome, Italy (corresponding author to provide phone: +39 3292973269; e-mail: arico@dis.uniroma1.it).

G. Borghini, is with IRCCS Fondazione Santa Lucia, via Ardeatina 306, 00179, Rome, Italy (e-mail: gianluca.borghini@gmail.com).

I. Graziani, is with the Dept. Physiology and Pharmacology, University “Sapienza” of Rome, Italy and BrainSigns srl (e-mail: ilenia.graziani@brainsigns.com).

F. Taya is with SINAPSE, Singapore Institute for Neurotechnology, Life Sciences Institute, National University of Singapore, Singapore (e-mail: fumihiko.taya@gmail.com).

Y. Sun is with SINAPSE, Singapore Institute for Neurotechnology, Life Sciences Institute, National University of Singapore, Singapore (e-mail: kissalladin@gmail.com).

A. Bezerianos is with SINAPSE, Singapore Institute for Neurotechnology, Life Sciences Institute, National University of Singapore, Singapore (e-mail: tassos.bezerianos@nus.edu.sg).

N. V. Thakor is with SINAPSE, Singapore Institute for Neurotechnology, Life Sciences Institute, National University of Singapore, Singapore (e-mail: sinapsedirector@gmail.com).

F. Cincotti, is with IRCCS Fondazione Santa Lucia, via Ardeatina 306, 00179, Rome, Italy (e-mail: f.cincotti@hsantalucia.it).

F. Babiloni, is with the Dept. Physiology and Pharmacology, University “Sapienza” of Rome, Italy and BrainSigns srl (e-mail: fabio.babiloni@uniroma1.it).

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

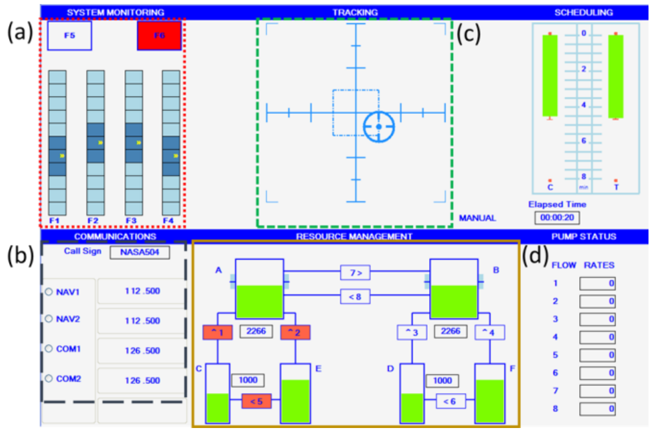

Figure 1. Screenshot of the Multi Attribute Task Battery (MATB) interface. On the top left corner (a, little dashed red box), there is the emergency lights task; on the top, in the center (b, medium dashed green box), there is the task of cursor tracking; on the left bottom corner (c, big dashed silver box), there is the radio communication task and, finally, in the center on the bottom (d, solid yellow box), there is the fuel levels managing.

In this study, they have been defined three conditions characterized by different task difficulty levels (Easy, Medium, Hard), to induce increased mental workload levels in the subject. The experimental protocol was composed by 6 recording sessions (Figure 2); the first 4 sessions were performed in two (Day 1 and Day 2) consecutive days (two per day). The last two sessions have been performed after one week from the fourth session (Day 9). A single session consisted of 7 runs. During the first 3 and the last 3 runs (offline runs), the subjects performed the three MATB difficulty levels. The fourth run (online run), consisted in a sequence of random combination of the three subtasks. This run has been used for testing online the workload evaluation system. The system has been entirely implemented in Matlab®, using the TOBI interfaces [8], which standardizes the procedures by which the different processing modules of the system exchange information.

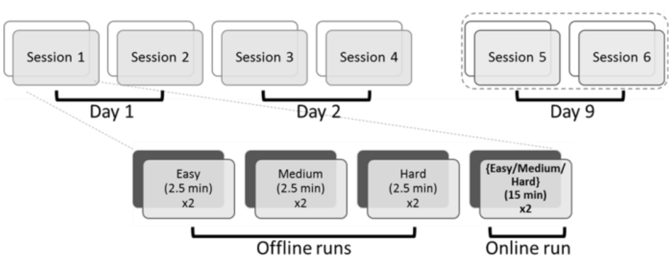

Figure 2. Experimental protocol scheme: each subject performed 6 recording sessions in three separate days; two sessions per day. The first four sessions were performed within two consecutive days, whilst the remaining two sessions were performed after about one week from the fourth session in order to test the stability of the system over time. Each session consisted of 7 runs. During the first 3 and the last 3 runs (offline runs), the subjects performed the three MATB difficulty levels (easy, medium and hard subtasks). The fourth run (online run) consisted in a sequence of random combination of the three subtasks (easy, medium, hard). Each subtask has been presented twice in the sequence, so that the total duration of the online run was 15 minutes (2.5 min each subtask)

B. *EEG, HR and fusion classifier for mental workload evaluation*

EEG: The EEG signal has been band - passed filtered (0.1-40 (Hz)) and then segmented in epochs of 2 seconds, 0.125 seconds - overlapped. The EOG signal has been used to remove the eyes-artefact contribution from each epoch of the EEG signal, by using the Gratton and Coles [9] algorithm. After that, for each epoch it has been evaluated the power spectral density (PSD) within the frequency bands involved in the mental workload estimation (theta and alpha bands [6]).

HR: As well for the EEG, the HR signal has been band - pass filtered (0.1-40 (Hz)) and then segmented in epochs of 8 seconds, 0.125 seconds – overlapped. 8 seconds of epoch length have been chosen, in order to be sure to have enough R-peaks to calculate the HR. For each epoch, only the R-peaks have been extracted and the PSD has been evaluated considering only the frequencies bins closed to the HR.

Using data from the training set (the first and the last 3 runs of the experimental session), a Stepwise Linear Discriminant Analysis (SWLDA, [9]) has been used to select the most relevant spectral features to discriminate the mental workload levels. Several moving average samples (N_{MA}) have been applied to the output of the classifiers (W_{EEG}, W_{HR}): $N_{MA}(1) = 0.125$ (s), $N_{MA}(8) = 1$ (s), $N_{MA}(16) = 2$ (s), $N_{MA}(32) = 4$ (s), $N_{MA}(64) = 8$ (s). The moving average was expected to increase the stability and the accuracy of the index with the drawback of introducing delays in the workload estimation, inducing a decrease of the workload refresh rate. Figure 3 shows the visual interface available to the operator, where the workload indexes at the different refresh rates are computed.

Fusion: A Fusion workload index has been calculated as a combination of the W_{EEG} and the W_{HR} based workload indexes. In particular, the two classifiers output have been synchronized, because their different delays, and then the new score (Fusion based workload index, W_{Fusion}) has been computed as a linear combination of the W_{EEG} and the W_{HR} score (Equation 1).

$$W_{Fusion} = aW_{EEG} + bW_{HR} \tag{1}$$

The coefficients a and b of the linear combination have been estimated for each subject by means of a simple LDA performed considering the EEG and the HR score distributions (W_{EEG} and W_{HR}) calculated over the cross validations for the three different difficulty levels. The coefficients estimated by the classifier were those who maximized the separation between the three difficulty levels.

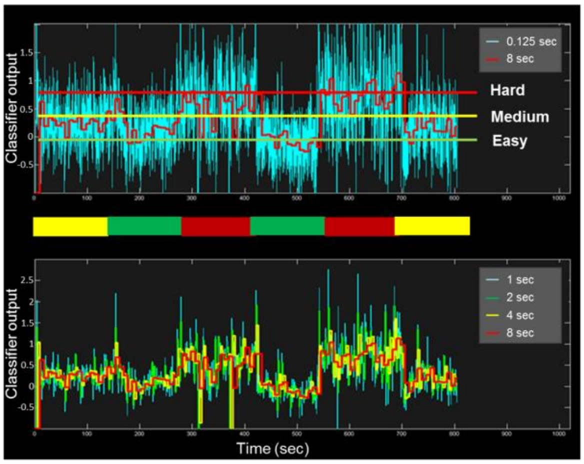

Figure 3. Screenshot of the visual interface provided to the operator that allow visualizing the fusion based workload index (W_{Fusion}) over time. In the upper side of the screen the workload index for the low and the high refresh rates are visualized. In the bottom part the $N_{MA}(x)$, x={8, 16, 32, 64} are visualized in real time. It is possible to note the variation of the index level related to the occurrence of the task difficulties

C. *Performed analyses*

System performance analyses: The dataset (*offline runs*) has been re - organized in 12 triplets (2 triplets per session) of runs (Easy, Medium and Hard subtasks). All the possible cross-validations have been considered, training the classifier with 1 triplet and testing the extracted features over the remaining triplets. The values of the Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC, [10]) describing the accuracy of the system has been calculated from the output of the classifier (for each different refresh rate). A three-way repeated measures ANOVA (CI = .95) has been performed using the classifier (EEG, HR and Fusion based), the couple of subtasks (Easy vs Hard, Easy vs Medium and Medium vs Hard), and the moving average lengths ($N_{MA}(x)$, x={1, 8, 16, 32, 64}) as factors and the related AUC values as dependent variable, for all the subjects and cross-validations. In addition, a Duncan post-hoc test has been performed in order to test the effects between all the factors.

Workload score distributions analyses: The score distributions of the single subtasks has been simulated offline within the 4th run (*online run*), by training the classifier with each triplet of runs (1st-3rd; 5th-7th) within the sessions and testing the extracted features over all the online runs (4th). Also, they have been differentiated two type of cross-validations, in order to investigate the short (INTRA) and medium (INTER) term stability of the features used for the workload classification. In particular, the INTRA type refers to the cross-validations performed considering as training sessions those related to Day 1 and Day 2 (Day 9) and as testing sessions those performed in the same days, Day 1 and Day 2 (Day 9). Contrariwise, the INTER type refers to the cross-validations performed considering as training sessions those related to Day 1 and Day 2 (Day 9) and as testing sessions those performed in the Day 9 (Day 1 and Day 2) and vice versa. Three two-way repeated measures ANOVA (CI = .95) have been performed, one for each classifier (EEG, HR and Fusion based), using subtask (Easy, Medium and Hard) and Cross-validation type (INTRA and INTER) for each subject as factors and the related workload index distributions (W_{EEG}, W_{HR} and W_{Fusion}) as dependent variables, for all the subjects.

III. RESULTS

System performance analyses: The ANOVA analyses (Figure 4) revealed no main effect of the classifiers ($F(2, 18)=.27$, $p=.76$), a main effect of conditions ($F(2, 18)=28.76$, $p=10^{-5}$) and a main effect of refresh time ($F(4, 36)=256.21$, $p=10^{-6}$). The post-hoc test showed that AUC values calculated using the EEG based classifier in the “Easy vs Medium” couple were significantly lower (all $p<10^{-6}$) than the other two ones. Also, increasing the refresh rate, the AUCs of the system significantly increase (all $p<.05$). The same behaviors have been obtained using the Fusion based classifier. For the HR based classifier, the AUC values for all the refresh time values and couples of tasks are not significantly different (all $p>.05$).

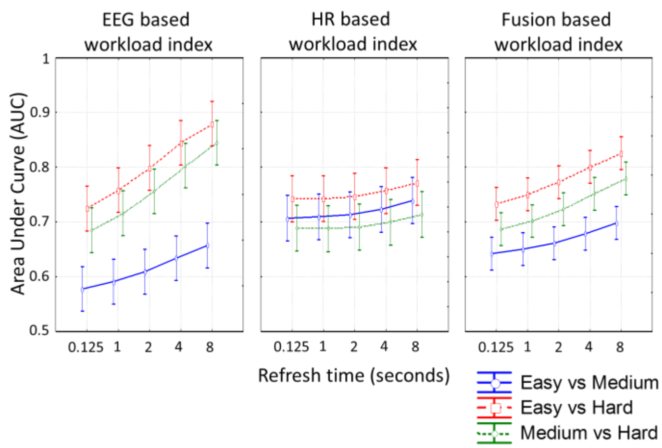

Figure 4. Mean values and related standard errors (CI = .95) of the AUC values achieved using the different classifiers (EEG, HR and Fusion-based) for each refresh time value.

Workload score distributions analyses: The ANOVA analyses (Figure 5) revealed that the score distributions related to the different subtasks (Easy, Medium and Hard) for all the three classifiers were significantly separated (EEG-based: $F(2,18)=37.84$, $p=10^{-6}$; HR-based: $F(2,18)=13.69$, $p=2.4x10^{-3}$, Fusion-based: $F(2,18)=36.52$, $p=10^{-7}$). Furthermore, no significant differences were found between the workload scores related to the INTER and the INTRA cross-validations, for each classifier (EEG-based: $F(1,9)=.20$, $p=.67$; HR-based: $F(1,9)=.85$, $p=.38$, Fusion-based: $F(1,9)=10^{-4}$, $p=.99$).

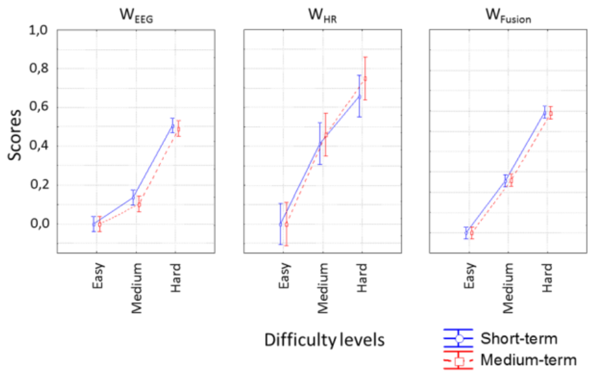

Figure 5. Mean values and related standard errors (CI = .95) of the distributions of the workload indices (W_{EEG}, W_{HR} and W_{Fusion}) evaluated by the three classifier (EEG, HR and Fusion based).

IV. Discussion

In this work, a framework to classify subject's mental workload online has been demonstrated using the brain and heart activities. The system has been tested with ten healthy subjects performing the MATB task which simulates the cockpit of an airplane. In particular, the employed tasks run over three different difficulty levels (Easy, Medium and Hard) resembling different flight conditions (cruise flight phase, flight level maintaining, and emergencies). Three different classifiers have been simulated and tested offline, by using the EEG (EEG based classifier) the HR (HR based classifier) signals alone and the combination of them (Fusion based classifier). The performance analyses as well the workload distribution analyses for all the classifiers showed a significant discriminability ($p<.05$) between the different difficulty levels when considering all the classifiers. Furthermore, the statistical analyses of the stability of the computed workload score in the short and medium terms did not show any significantly difference ($p>.05$), demonstrating that the features extracted by the classifiers are stable over the time, and that even after a week may not be necessary to recalibrate the system with new data. These aspects related to stability and accuracy are highly important for the usability point of view of the system. In fact, to use such system in real environments it could be enough to calibrate the system with the specific parameters of the operator once and then just use it without further adjustments maintaining a high reliability over at least a one-week period. The fusion-based classifier reached an AUC higher than the EEG-based classifier at fast refresh time values, and higher than the HR-based classifier at the slow refresh times. These results demonstrate that by combining information coming from different biosignals (e.g. EEG and HR), it is possible to have more reliable and faster information about the mental states of the user. This multi-modality approach can be used in real operating environments for improving the human machine interaction, not only for pilots, but also for other users, such as air traffic controllers, car drivers or more in general for all the contexts, in which the high stress conditions can cause a critical drop in performance.

V. Conclusion

In this study, a system able to estimate online the mental workload of an operator by using the combination of EEG rhythms and HR signals has been proposed. It has been demonstrated that i) the system is able to significantly differentiate three workload levels related to the three difficulty level tasks employed with a high reliability; ii) the subjective features used for the evaluation of the mental workload remain stable over one week. The innovation with respect to the current scientific literature is the possibility to predict online the mental workload of the user over three difficulty levels, obtained by using the combination of EEG and HR. Such combination of signals improved the reliability of the estimated mental states with respect to using just one information modality (e.g. only EEG, only HR). In addition, another relevant aspect is that the classification features chosen by the system are stable after a week. This aspect is strictly required in the perspective of using such system in a real environment scenario.

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. The paper reflects only the authors' views. The work is also partially supported by the Regione Lazio, through FILAS spa, in the context of the project BrainTrained, CUP: F87I12002500007.

References

[1] D. A. Norman e D. G. Bobrow, «On the analysis of performance operating characteristics», *Psychological Review*, vol. 83, n. 6, pagg. 508–510, 1976.

[2] E. A. Byrne e R. Parasuraman, «Psychophysiology and adaptive automation», *Biol Psychol*, vol. 42, n. 3, pagg. 249–268, feb. 1996.

[3] J. Kohlmorgen, G. Dornhege, M. Braun, B. Blankertz, K.-R. Müller, G. Curio, K. Hagemann, A. Bruns, M. Schrauf, e W. Kincses, «Improving human performance in a real operating environment through real-time mental workload detection», 2007.

[4] R. N. Roy, S. Charbonnier, e A. Campagne, «Probing ECG-based mental state monitoring on short time segments», *Conf Proc IEEE Eng Med Biol Soc*, vol. 2013, pagg. 6611–6614, 2013.

[5] G J. del R.Millan, J. Mourino, F. Babiloni, F. Cincotti, M. Varsta, e J. Heikkonen, «Local neural classifier for EEG-based recognition of mental tasks», in IJCNN 2000, *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, 2000, vol. 3, pagg. 632–636 vol.3.

[6] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, e 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.

[7] J. R. Comstock, «MATB - Multi-Attribute Task Battery for human operator workload and strategic behavior research», jan. 1994.

[8] P. Aricò, G. Borghini, I. Graziani, F. Bianchini, F. Cincotti, e F. Babiloni, «A brain computer interface system for the online evaluation of ATCs' workload», *Italian Journal Of Aerospace Medicine*, Rome, Italy, jun-2013.

[9] G. Gratton, M. G. Coles, e E. Donchin, «A new method for off-line removal of ocular artifact», *Electroencephalogr Clin Neurophysiol*, vol. 55, n. 4, pagg. 468–484, apr. 1983.

[10] D. Bamber, «The area above the ordinal dominance graph and the area below the receiver operating characteristic graph», *Journal of Mathematical Psychology*, vol. 12, n. 4, pagg. 387–415, nov. 1975.