EEG BASED MENTAL WORKLOAD ESTIMATION

Shabnam Samima and Monalisa Sarma*

Subir Chowdhury School of Quality and Reliability
Indian Institute of Technology Kharagpur, India, Kharagpur 721302
*Email: shabnam.samima@iitkgp.ac.in

ABSTRACT

Knowledge of the level of mental workload induced by any task is essential for optimizing load share among the operators. This helps in assessing the capability of the operators; besides, helping in allocation of tasks to the operators. Since a persistently high workload experienced by operators such as aircraft pilots and automobile drivers many times compromises their performance and safety. Despite the availability of various mental workload evaluation techniques such as heart rate variability, pupil dilation, saccades, etc., continuous assessment of mental workload is still a challenging task. In this work, we aim to evaluate the workload of the operator involved in long duration tasks. For this, experiments have been carried out in a working environment which provides tasks to be done simultaneously, tasks with a pause or break in activity and cross-functional tasks. The experiment data is recorded continuously in different modes and analyzed in segments to show the change in mental workload. The proposed approach achieves a classification accuracy of 96.6% by utilizing ANN architecture to classify the data using the defined feature set.

Index Terms— Mental workload, Electroencephalography, Power spectral density, Brain connectivity

1. INTRODUCTION

The technological revolution has significantly reduced the manual interventions of the operators. However, for effective working of the human machine system it is necessary that there is minimal error from the operators. The term "mental workload" summarizes the overall factors that are reponsible for an operator's degraded performance. The advancement of wearable and ambient sensors makes it possible to develop smart systems that can continuously monitor human activities [1]. These sensors are prominent in capturing human behaviour using the physiological parameters [2]. The signature information thus obtained can be used to establish the correlation between mental task demands, performance and human capability in environments demanding high human interventions. Electroencephalogram (EEG) being an excellent modality in capturing brain signatures can be integrated into the human-machine interfaces for measuring mental workload thereby improving human performance [3]. At critical situations and in monitoring tasks, the measurement and evaluation of mental workload can drastically reduce human errors [4]. Different methods are available for investigating mental workload [5]; however, mental workload can be easily reflected from the human brain, which makes EEG measurement as the suitable evaluation index of mental workload [6].

Estimation of mental state using EEG can be considered as a method that translates brain signal into time varying information by extracting features defining different mental states [7], [8], [9]. Considering this, the authors in [10] propose a *bag-of-words* classifier model to discriminate two task relevant states consisting of complex and concurrent tasks. Fritz *et al.* [11] use the $\beta/(\alpha+\theta)$ ratio feature for workload estimation. Roy *et al.* [12] attempted the use of ERPs for binary classification of workload based on ignored auditory probes. The authors in [13] proposed a discrete wavelet based "pattern recognition" approach to discriminate workload recorded during different cognitive conditions. Recently, the authors in [14] used a deep recurrent neural network architecture to predict the levels of cognitive load from EEG recordings. The methods discussed so far highlight that spontaneous EEG has sensitivity to change in mental workload. Thus, these methods have potential for being used as workload index in online applications. However, the generalization of the approaches to new situations and individuals still remains unsolved.

Continuous assessment of mental workload being a challenging task, the proposed work is designed to assess mental workload of a person involved in continuous monitoring task. For this, mental workload is induced by varying six different types of working memory tasks (verbal and visuo-spatial). From the recorded data, the EEG rhythms $(\theta, \alpha_1, \alpha_2, \beta_1, \beta_2, \gamma)$ are extracted to define features that indicate changes in workload. The defined feature set has been taken as input to the artificial neural network (ANN) consisting of a single hidden layer to classify them. The results demonstrates the efficacy of our proposed work which is further validated using the brain graph analysis.

2. MATERIALS

2.1. Participants

Thirty healthy volunteers (25±5 years old) were recruited on voluntary basis for the experiment. The principles of *Declaration of Helsinki* were followed to carry out the experiments. Further, the experiment protocol has been approved by the Institutional Ethical Committee (IIT/SRIC/SAO/2017). Written consent were taken from each participant prior to the experiment. All participants had normal or corrected-to-normal vision and were right-hand dominant. Participants were currently healthy and reported no history of psychiatric disorders. Besides, they were instructed to avoid alcohol, caffeine and heavy meals right before the experiments.

2.2. Equipment

The EEG data was recorded at a sampling frequency of 256 Hz having 16-bit resolution using a 64-channel scalp electrode cap (Ag/AgCl, RMS, India) following the international 10-20 system. The electrodes corresponding to the earlobes are used as reference, and the electrode at the forehead is used as ground. The impedance of the electrodes were below 10 kΩ. Nineteen electrodes: Fp1, Fp2, F3, F4, Fz, F7, F8, C3, C4, Cz, T7, T8, P7, P8, P3, P4, Pz, O1, and O2 were chosen from for the purpose of analysis.

2.3. Stimuli

The working memory stimuli [15] were displayed onto a 19.5 inch monitor (with 1600 × 900 pixels resolution) located 70 cm from the participants. The stimuli consisted of two verbal tasks (namely, operation span and reading span) and four visuo-spatial tasks (namely, Matrix span, Arrow span, Rotation span and Symmetry span).

2.4. Experiment Design

The *Working Memory Test Battery* [15] consists of complex span tasks which follow the paradigm of item storage with concurrent processing of a separate cognitive processing tasks. With these tasks, it is possible to capture the conceptual requirements of simultaneous processing and memory operations which are inherent to working memory functioning. Workload among the tasks were varied based on the difficulty level order(low, medium, high) and speed of inter element pause duration for each task.

Each task consists of nine trials with three trials for each difficulty level. The tasks were presented in the following order: Arrow span, Matrix span, Operation span, Reading span, Rotation span and Symmetry span. The inter element (trail) pause duration for Arrow span was kept at 600 ms and the duration was reduced successively by 100 ms for every tasks. The tasks follow the basic principle of *n-back* task; that is, in the first three trials the number of to-be-remembered (TBR) items are two followed by three and four respectively. In these tasks, each trial is made up of a presentation phase and a recall phase. The explanation of the various tasks is provided in [15].

3. EXPERIMENT PROTOCOL AND SIGNAL ANALYSIS

3.1. Experiment Protocol

The collected raw EEG data gives no information about mental workload changes unless some useful signatures are extracted from it. The collected EEG signals have to undergo a series of stages for extracting the meaningful information from it. The block diagram of the followed approach has been shown in Fig. 1.

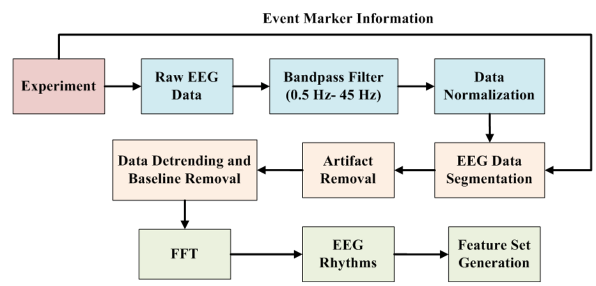

Fig. 1: Overview of the Approach

The experiment was carried out in a sound-shielded, air-conditioned laboratory at Indian Institute of Technology Kharagpur. Two dedicated computers were used for data recording and task activity. During data recording, event markers were used to mark the beginning and end of each task activity. Participants completed Questionnaire-1 and Questionnaire-2 prior to the beginning of the experiment. Questionnaire-1 gathered information about the food habits, sleep hours, vision, working hours and medical history of the participants. While Questionnaire-2 collected information about subjective alertness level. Before the experiment, each participant was given instructions of the entire experiment. Further, each participant completed sample test session for familiarizing with the tasks. At the beginning of the experiment, five minutes of eyes-opened resting state EEG were collected as baseline which is followed by the six different task activities. The entire experimental protocol is given in Fig. 2. After completion of the session, the participants again rated their perceptual alertness level in Questionnaire-2.

3.2. Signal Analysis

3.2.1. Signal Pre-processing

The raw EEG signal was first passed through a band-pass filter with a fifth-order Butterworth filter having cutoff frequen-

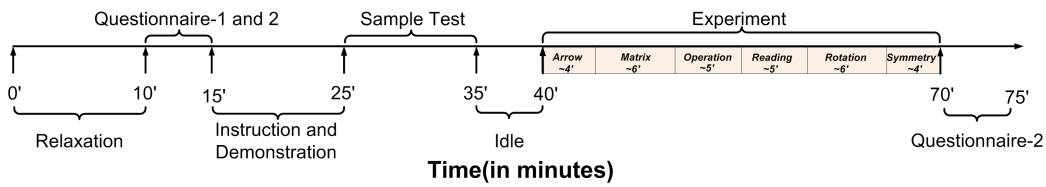

Fig. 2: Description of experimental procedure

cies of 0.5 Hz and 45 Hz. Further to reduce the power line noise, a 50 Hz notch filter has been used. Data normalization was carried out to remove any bias added during data recording. Next, the signals were segmented into epochs using the event marker information. In each EEG epoch, eye blink and movement artifact were removed using FORCe algorithm [16] followed by data detrending and baseline removal.

3.2.2. Feature Extraction

From the artifact-free EEG data, features are extracted by transforming time series data into frequency domain. For each participant, EEG rhythms $(\theta, \alpha_1, \alpha_2, \beta_1, \beta_2, \gamma)$ are extracted from each epoch of all the channels. Next, the power spectral density (PSD) is calculated for each of the extracted features. From this the feature vector is generated by concatenating all the extracted PSD rhythms of each channel. To sum up, 114 features (19 channels $\times$ 6 rhythms) are extracted from each epoch of the EEG signal. This feature vector acts as a representative set of the relevant information for detecting changes in the mental workload.

In the proposed approach, the basic idea followed for evaluating mental workload is that the recorded idle data is considered as the reference point and any changes induced during the task performance is analysed to investigate changes in the mental workload. Using this phenomenon, the feature is generated by taking ratio of EEG rhythms during task performance to the idle case. Thus, the feature set, f considered for evaluating mental workload is defined as:

$$f = \left\{ \frac{\theta^{exp}}{\theta^{idle}}, \frac{\alpha_1^{exp}}{\alpha_1^{idle}}, \frac{\alpha_2^{exp}}{\alpha_2^{idle}}, \frac{\beta_1^{exp}}{\beta_1^{idle}}, \frac{\beta_2^{exp}}{\beta_2^{idle}}, \frac{\gamma^{exp}}{\gamma^{idle}} \right\} \quad (1)$$

4. EXPERIMENTAL RESULT AND DISCUSSION

The evaluation of the proposed mental workload has been done using subjective measures and EEG data trend analysis. The brain connectivity analysis validates the proposed mental workload approach.

4.0.1. Subjective Measures

- Alertness Scale: Mental workload tasks significantly affect the state of mind of a person. With increase in mental workload, there is simultaneous reduction in arousal and alertness of an individual. For effectively analyzing the effect of mental workload, individual subjective rating has been taken before and after the experiment. The perceptual individual alertness rating has been analyzed in the likert scale of "extreme alert" to "extreme sleep" state. The t-test measure clearly indicates that the there is a remarkable difference in alertness level before and after the experiment with $p - value$ and $t - value$ 0.001121 and 4.94974 respectively, at α = 0.05.
- Task Difficulty Scale: The different tasks used in this study poses varying difficulty to each individual. To estimate the task difficulty perceived by each individual subjective rating has been taken on the scale of '0' to '10', where '0' denotes 'low difficulty' and '10' denotes 'high difficulty'. The histogram plot for the average task difficulty of the participants is shown in Fig.3.

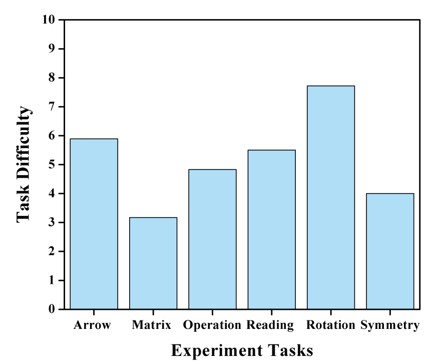

Fig. 3: Average task difficulty perceived by the participants

4.1. Workload Trend Analysis

The objective of this analysis is to identify the trend of each defined feature. The trend analysis has been done using MATLAB® version 2017a running on a PC with the following configuration: Processor: Intel(R) Core(TM) i5-6500, CPU @ 3.20GHz, RAM: 4.00 GB, System Type: 64-bit Operating System, x64-based processor. With this analysis it would be evident whether the defined features is capable of identifying the change in workload complexity with varying time. The average trend analysis for the Arrow span task (refer 3) is plotted for each of the considered features in Fig. 4.

From Fig. 4, it is can be observed that the first three features shows declining trend, while the other three shows increasing trend. The decreasing trend is due to the fact that during rest (idle) the power value of θ, α_1 and α_2 are higher than during task performance. On the other hand, while actively involved in task performance, the dominant frequency shifts to β_1, β_2 and γ rhythms. The same trend can be observed across all the other tasks considered in this experiment. However, in this paper, due to space constraint, we are unable to provide rest of the trend plots.

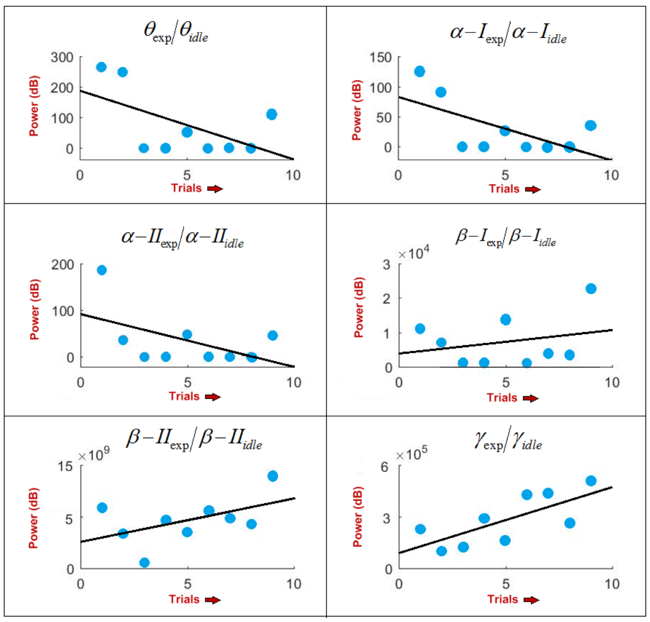

Fig. 4: Trend plot of features for Arrow span task

4.2. Classification using ANN architecture

Artificial Neural Networks (ANN) have been used to perform workload data classification. The rationale of using the ANN architecture is that mental workload computed for every person varies from one another so it is not possible to compute and classify workload using threshold limits. Thus, in our work, a multi-layer neural network is used for the classification of mental workload which uses the proposed feature vector as input. The proposed architecture comprises of a single hidden layer having default $tansig$ transfer function. In the proposed architecture, 60% of the data is used as training data and remaining 40% is used for testing. The performance of the ANN architecture is measured using the following parameters: *accuracy*, *sensitivity*, *specificity* and *precision* [17]. For the proposed approach, the overall classification accuracy is found to be 96.6%, sensitivity 97.62%, specificity 95.75% and precision 95.66%. The confusion matrix has been shown in Fig. 5.

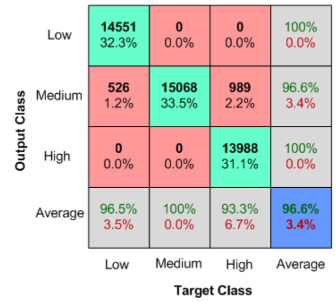

Fig. 5: Confusion matrix for classification accuracy

4.3. Brain Connectivity Analysis

The brain graph analysis [18] have been carried out to investigate the changes occurred during the different task conditions. Two parameters (that is, *strength*, *clustering coefficient*) have been evaluated for the experiment. The average strength have been calculated for the low, medium and high workload conditions. From the calculation, it can be observed that there are higher values for high workload conditions than in the low one. Further, clustering coefficient for the different task conditions have been evaluated. It has been observed that the values lowered with increase in workload. The overall findings have been tabulated in Table 1.

Table 1: Strength and clustering co-efficient values under different task conditions

Brain Connectivity Parameters	Task Complexity		
	Low	Medium	High
Average Strength	6.01	8.22	9.58
Clustering Co-efficient	0.7192	0.6205	0.4620

5. CONCLUSION

The proposed study demonstrated the evaluation of workload for the operators involved in long duration tasks. From the proposed feature vector, workload can be easily differentiated. The implementation of ANN architecture using the proposed feature vector, classified the data into the designated classes with an accuracy of 96.6%. To prove the efficacy of the proposed approach, brain connectivity analysis has been carried out. It has been observed that during more demanding task, strength value was significantly higher than the low demanding task. Besides, the clustering coefficient showed a lower value during high demanding task than in low demanding tasks. The results demonstrate the efficacy of our proposed approach. Future work will focus on using the workload metrics to design human machine interaction systems that constantly monitor workload imposed on the operators and help in task allocation based on operator's mental state.

6. REFERENCES

[1] Subhas Chandra Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE sensors journal*, vol. 15, no. 3, pp. 1321–1330, 2015.

[2] Anindya Nag, Subhas Chandra Mukhopadhyay, and Jürgen Kosel, "Wearable flexible sensors: A review," *IEEE Sensors Journal*, vol. 17, no. 13, pp. 3949–3960, 2017.

[3] Chris Berka, Daniel J Levendowski, Caitlin K Ramsey, Gene Davis, Michelle N Lumicao, Kay Stanney, Leah Reeves, Susan Harkness Regli, Patrice D Tremoulet, and Kathleen Stibler, "Evaluation of an eeg workload model in an aegis simulation environment," in *Biomonitoring for Physiological and Cognitive Performance during Military Operations*. International Society for Optics and Photonics, 2005, vol. 5797, pp. 90–100.

[4] Baljeet Singh Cheema, Shabnam Samima, Monalisa Sarma, and Debasis Samanta, "Mental workload estimation from eeg signals using machine learning algorithms," in *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer, 2018, pp. 265–284.

[5] Brad Cain, "A review of the mental workload literature," Tech. Rep., Defence Research And Development Toronto (Canada), 2007.

[6] Ranjana K Mehta and Raja Parasuraman, "Neuroergonomics: a review of applications to physical and cognitive work," *Frontiers in human neuroscience*, vol. 7, pp. 889, 2013.

[7] Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw, "Bci2000: a general-purpose brain-computer interface (bci) system," *IEEE Transactions on biomedical engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.

[8] Jens Kohlmorgen, Guido Dornhege, Mikio Braun, Benjamin Blankertz, Klaus-Robert Müller, Gabriel Curio, Konrad Hagemann, Andreas Bruns, Michael Schrauf, Wilhelm Kincses, et al., "Improving human performance in a real operating environment through real-time mental workload detection," *Toward Brain-Computer Interfacing*, pp. 409–422, 2007.

[9] Christian A Kothe and Scott Makeig, "Estimation of task workload from eeg data: new and current tools and perspectives," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 6547–6551.

[10] L. M. Merio, J. Meng, S. Gordon, B. J. Lance, T. Johnson, V. Paul, K. Robbins, J. M. Vettel, and Y. Huang, "A bag-of-words model for task-load prediction from eeg in complex environments," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 1227–1231.

[11] Thomas Fritz, Andrew Begel, Sebastian C Müller, Serap Yigit-Elliott, and Manuela Züger, "Using psycho-physiological measures to assess task difficulty in software development," in *Proceedings of the 36th International Conference on Software Engineering*. ACM, 2014, pp. 402–413.

[12] Raphaëlle N Roy, Stéphane Bonnet, Sylvie Charbonnier, and Aurélie Campagne, "Efficient workload classification based on ignored auditory probes: a proof of concept," *Frontiers in human neuroscience*, vol. 10, pp. 519, 2016.

[13] Hafeez Ullah Amin, Wajid Mumtaz, Ahmad Rauf Subhani, Mohamad Naufal Mohamad Saad, and Aamir Saeed Malik, "Classification of eeg signals based on pattern recognition approach," *Frontiers in computational neuroscience*, vol. 11, pp. 103, 2017.

[14] S. Kuanar, V. Athitsos, N. Pradhan, A. Mishra, and K. R. Rao, "Cognitive analysis of working memory load from eeg, by a deep recurrent neural network," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 2576–2580.

[15] James Stone, "Cognitive Research Tools," http://www.cognitivetools.uk/cognition, Online; Last accessed 6 August 2018.

[16] Ian Daly, Reinhold Scherer, Martin Billinger, and Gernot Müller-Putz, "Force: Fully online and automated artifact removal for brain-computer interfacing," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 23, no. 5, pp. 725–736, 2015.

[17] Abdulhamit Subasi and M Ismail Gursoy, "Eeg signal classification using pca, ica, lda and support vector machines," *Expert systems with applications*, vol. 37, no. 12, pp. 8659–8666, 2010.

[18] Nicolina Sciaraffa, Gianluca Borghini, Pietro Aricò, Gianluca Di Flumeri, Alfredo Colosimo, Anastasios Bezerianos, Nitish V Thakor, and Fabio Babiloni, "Brain interaction during cooperation: Evaluating local properties of multiple-brain network," *Brain sciences*, vol. 7, no. 7, pp. 90, 2017.