Unsupervised Approach for Measurement of Cognitive Load using EEG Signals

[1]Diptesh Das, *Member, IEEE,* Debatri Chatterjee and Aniruddha Sinha, *Senior Member, IEEE*

***Abstract*— Individuals exhibit different levels of cognitive load for a given mental task. Measurement of cognitive load can enable real-time personalized content generation for distant learning, usability testing of applications on mobile devices and other areas related to human interactions. Electroencephalogram (EEG) signals can be used to analyze the brain-signals and measure the cognitive load. We have used a low cost and commercially available neuro-headset as the EEG device. A universal model, generated by supervised learning algorithms, for different levels of cognitive load cannot work for all individuals due to the issue of normalization. In this paper, we propose an unsupervised approach for measuring the level of cognitive load on an individual for a given stimulus. Results indicate that the unsupervised approach is comparable and sometimes better than supervised (e.g. support vector machine) method. Further, in the unsupervised domain, the Component based Fuzzy c-Means (CFCM) outperforms the traditional Fuzzy c-Means (FCM) in terms of the measurement accuracy of the cognitive load.**

I. Introduction

THE definition of cognitive load, according to Wikipedia is the "load related to the executive control of working memory" [1]. In simpler words, it is the amount of load experienced, by the working memory, to perform a task. The cognitive load mostly depends on working memory capacity [2]. An optimum level [3] of cognitive load for an individual is desirable, in order to perform a task in a satisfactory manner. Various application use cases ranging from monitoring pilots [4] to normal truck drivers [5], adaptive learning systems [6] and practical training sessions [7] demand for real-time monitoring of individuals' cognitive loads, user-interface (UI) evaluation and testing [8] [9]. Apart from the mission critical activities, the need for constant monitoring of cognitive load can be realistic only if the EEG measuring device is of low cost and easy to wear. Recently, a couple of such devices are available (neuro-headset from Emotive [10] and Neurosky [11]) which have reinforced the research in such application areas.

The EEG signals have been successfully used in various motor control applications [12] [13] [14]. Most of the cognitive load related works use functional magnetic resonance imaging (fMRI) [15], functional near infrared spectroscope (fNIRs) [16], positron emission tomography (PET) [17] etc. Alternately pupil dilation [18] and galvanic skin responses [19] have been used to detect the stress, anxiety and cognitive load. Statistical approaches are used to measure the cognitive load for students [20] without using any EEG device. There have been limited amount of work done in measurement of cognitive loads with low cost EEG devices.

In the present work, we have used the 14 leads Emotive EEG device which is wearable on the head. Though it is low cost and easily available, a practical problem lies in the repeatability of the EEG signals. In this regard there are two issues – intensity and signal range variation among different individuals and the variation for same individual for different trials. The reason for the first one is mainly due to the difference in the shape of the head or scalp for different individuals. The second variation is due to spatial shift in the positioning of the leads from one trial to another for the same individual. This is usually termed as normalization issue. Models generated using supervised learning cannot address this unless the EEG signals are normalized before training. Hence in this paper, we have proposed an unsupervised approach for measurement of cognitive load.

The objective is to identify the stimulus that imparts higher cognitive load to an individual than another stimulus using the EEG signals. In order to achieve the same, initially the EEG signals are split into sliding windows of few seconds (~10 sec.). These raw EEG signal are fed into the input of CSP (Common Spatial Pattern) [21] to filter out the two most widely separated signals based on their variance. These CSP filtered signals are then used to extract the meaningful features to be used for determining the level of cognitive load imparted by a stimulus (dataset). In the present paper we focus on two levels of the cognitive load – low and high. Standard Fuzzy c-Means (FCM) and its extension Component-wise Fuzzy c-Means (CFCM) algorithms have been used as the unsupervised approach. A comparison is performed with the supervised approach using a standard SVM (Support Vector Machine) classifier to generate the training model and classify the test dataset using the feature set. One example of classifying EEG signals using SVM classifier is explained in [22]. The use of unsupervised approach to measure different levels of cognitive loads overcomes the issues related to normalization and relaxes the need for initial training of individual users before the measurements. Therefore the unsupervised approach enhances its wide spread applicability for any real-time measurement of cognitive

[1]Diptesh Das, Debatri Chatterjee and Aniruddha Sinha are with Innovation Labs, Tata Consultancy Services Ltd., India. {e-mail: diptesh.das, debatri.chatterjee, aniruddha.s}@tcs.com

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

load in any educational institute as well as professional pedagogical (or training) institute.

The paper is organized as follows. Section II provides the details of Test Stimuli used for the experiment. In section III we have described the data acquisition tool and the details of proposed methodology. Section IV gives the details of the experiments conducted, followed by the results and performance analysis in section V. Finally the paper is concluded in section VI.

II. DESIGN OF TEST STIMULI

The cognitive load can be broadly classified into three types – *Intrinsic*, *Extraneous* and *Germane* [23]. Each of these types effect the learning and decision making process of an individual. Intrinsic load refers to the complexity of the task. The additional load due to the method of presentation of the task refers to Extraneous load. The load due to learning a new technology or schema refers to the Germane load. In the current work, we have used three different types of stimulus namely – (i) Stroop Test [24] – this is mostly Extraneous load (ii) Generic UI usability Test – Extraneous load and (iii) Logical reasoning – mostly Germane load. For generic GUI usability test, we have used two different onscreen keyboard layouts (Fig. 5 and Fig. 6) which can be operated by a remote control. These two layouts were compared on the basis of user-friendliness. All of the above stimuli are having varying Intrinsic load.

Each of these stimuli is constituted for two levels of complexity – low and high as detailed below.

A. *Stimuli for Stroop test*

In typical Stroop test the users are asked to read out the colours of each text in each slide while wearing the EEG headset. Typical example of Stroop slides used for two levels of cognitive loads are shown in Fig. 1 and Fig. 2.

In Fig. 1 (for low cognitive load) the name of the colour was printed in a colour denoted by the name while in Fig. 2 (for high cognitive load) the name was printed in a colour not denoted by the name. In both the cases the users were asked to read the colour inferred by the text instead of name of the colour.

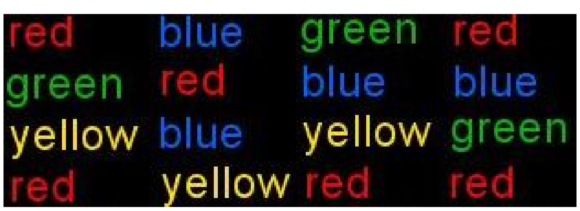

Fig. 1. Stroop slide for low cognitive load.

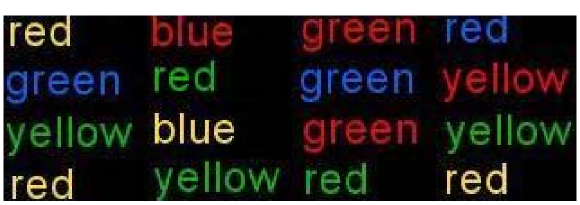

Fig. 2. Stroop slide for high cognitive load.

B. *Logical reasoning test*

In this test a set of logical reasoning questions having different levels of cognitive loads were designed. The questions are prepared based on the laws of propositional logic as described in [25]. Questions containing single logical error were designed to impart low cognitive load whereas questions containing multiple logical error imparted high cognitive load. The experimental test set consisted of three slides for low load and three slides for high load. Typical examples of such slides are shown in Fig. 3 and Fig. 4 for low and high cognitive load respectively.

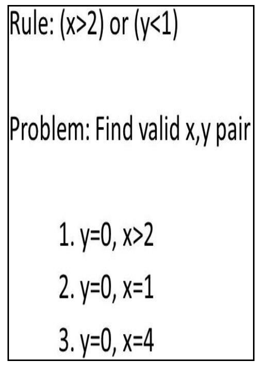

Fig. 3. Question for low load.

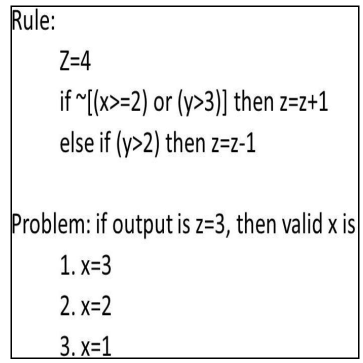

Fig. 4. Question for high load.

C. *Typing using onscreen-keyboards*

In this experiment, the users are asked to type some phrases using two different onscreen layouts shown in Fig. 5 and Fig. 6. Theoretical model (KLM-GOMS) as well as extensive user study establishes the fact that the Hierarchical layout imparts lower cognitive load compared to QWERTY layout [26] [27]. The test phrases are taken from Mackenzie's standard phrase sets [28]. The average length of the phrases selected is 20 characters.

Fig. 5. Hierarchical keyboard layout.

Fig. 6. QWERTY keyboard layout.

III. PROPOSED METHODOLOGY

The objective of this research is to compare the performance of supervised and unsupervised classification methodologies while classifying the cognitive load. While performing the tasks, the EEG signals are acquired from the users. As the time intervals for executing tasks for different stimuli are different, we apply dynamic time wrapping algorithm [29] to align them in the same time-scale.

A. EEG data acquisition and analysis tool

For recording the EEG signals, we have used 14–channel Emotiv neuro-headset. The device uses a standard 10-20 electrode configuration for channel fixing. The sampling frequency of the device was set to 128Hz. The raw EEG data acquired was pre-filtered using a notch filter at power line frequency. The device also provides two reference channels (DRL, CMS) at P3/P4 locations offering good spatial resolution. The headset is shown in Fig. 7.

The acquired EEG data are processed in a window by window basis. Window size is defined based on the maximum discrimination of Cognitive Score (CS) between high and low cognitive loads. A typical window size of 10 sec has been experimentally chosen based on the Stroop test dataset [25].

Fig. 7. Emotive neuro-headset.

The raw EEG signal (in a particular window) suffers from poor spatial resolution. Hence we used CSP filters to maximize the variance of band-pass filtered EEG signals from one class while minimizing that from other class [21]. From these filtered signals, EEG features are extracted. The features used in the present work are log variance [30], Hjorth parameters [31] [32] and different band power estimates [33] [34].

The feature vectors thus obtained are then fed to both supervised and unsupervised learning algorithms to finally measure the level of the cognitive load on the users.

B. Classification of cognitive load using supervised learning

A standard Support Vector Machine (SVM) classifier is used to classify the cognitive load of users. For the purpose of training, EEG data from standard Stroop test undertaken by a group of users have been used. This trained model is further used to classify the EEG data obtained from the same set of users exposed to two different sets of stimuli. Here in our experiments, different on-screen keyboard layouts and logical reasoning problems are used as two different testing stimuli as explained in [22], [25].

C. Classification of cognitive load using unsupervised learning

EEG signals generated from the same individual at different instances for the same stimulus do not remain the same because of the normalization issues as discussed before. Thus it becomes absolutely necessary to train the classifier for individual users every time we need to measure the cognitive load of that user to achieve best performance of the classifier. This may not be feasible under all circumstances. To overcome this difficulty we have opted for unsupervised learning algorithms for differentiating the levels of cognitive loads. In an unsupervised learning algorithm, data points are grouped into two or more clusters based on the data similarity. In the present work we have used traditional FCM algorithm and an extension of the traditional Fuzzy c-Means algorithm called Component-wise Fuzzy c-Means (CFCM) algorithm [35] and have compared their performances while measuring the cognitive load for two levels.

Traditional FCM is an unsupervised algorithm, where each data points are induced by a membership value which is a function of the distance measure of the data point from the cluster centers. It is an iterative process of finding the cluster centers based on these membership values. At the end of each iteration, the membership values are used for hard assignment of the data points to one of the clusters.

But in CFCM, instead of giving an overall membership for a data point, each individual dimensions of the data-point are assigned a membership of belongingness to a cluster. This enhances the performance of clustering where odd parametric values of one or fewer dimensions lead to a bad clustering of data points. However instead of giving a complete freedom, these individual memberships of belongingness at component level is further aggregated while calculating the cluster centroid. Another important aspect of CFCM is that it automatically updates the value of fuzziness control parameter m based on the stability criterion of cluster centers. Therefore whatever be the initial value of m, the algorithm automatically tunes m in the stability region of cluster centers.

The present approach has been evaluated on two levels of clustering, but it can further be employed to multi-level cognitive load measurement in a seamless manner. The overall methodology using both supervised and unsupervised learning algorithms has been depicted in a flowchart given in Fig. 8.

IV. EXPERIMENTS

Three sets of experiment are conducted to generate the training and testing database for the proposed approach.

A. Experiment 1

A group of 15 users (male) were selected in the age group of 25-30 years. Female users were avoided because of the problems in placement of EEG headset due to their long hair. They were asked to read both the Stroop slides as shown in Fig. 1 and Fig. 2. EEG signals were recorded while reading the slides.

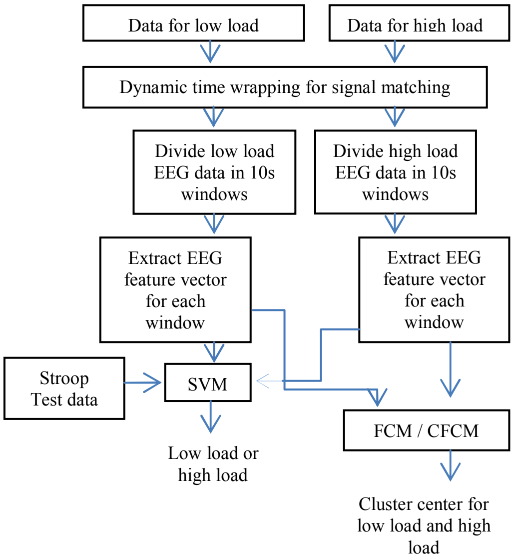

Fig. 8. Flowchart for overall process.

B. Experiment 2

In this experiment the subjects who have already been participated in experiment 1 were asked to take part in Logical reasoning test as described in section IIB. Each participant completed both the sessions of low and high cognitive tasks with a short break of 30 seconds between each session.

C. Experiment 3

All the participants were next asked to type 5 different phrases [28] using the on-screen keyboard layouts shown in Fig. 5 and Fig. 6. After typing a phrase using a particular layout, the participants were asked to take a short brake of 2 mins. EEG signals were recorded while typing the phrases.

V. RESULTS AND PERFORMANCE ANALYSIS

In this section we have described the metric used for the performance analysis of the algorithms, the selection of Fuzziness Control parameter m and it's convergence over iterations using CFCM and the performance comparison of Supervised (SVM) and Unsupervised (FCM and CFCM) algorithms while classifying the EEG signals into two levels of cognition load.

A. Performance metric

The performance metric used for analyzing the output of both supervised and unsupervised learning algorithm is the $F_{measure}$ (1) which is defined as the harmonic mean of precision (P) and recall (R).

$$F_{measure} = \frac{2 \cdot P \cdot R}{P + R} \quad (1)$$

Precision (2) is defined as the ratio of number of data points truly belonging to a cluster (TP) and the total number of data points identified to belong to the same cluster (TP+FP). Where FP represents the false positive, i.e. the total number of data points which are falsely identified to belong to the said cluster.

$$P = \frac{TP}{TP + FP} \quad (2)$$

Recall (3) is the ratio of true positive (TP) and the total number of data points that actually belong to that cluster (TP+FN). Where FN, the false negative is the total number of data points which are wrongly identified as not belonging to the same cluster.

$$R = \frac{TP}{TP + FN} \quad (3)$$

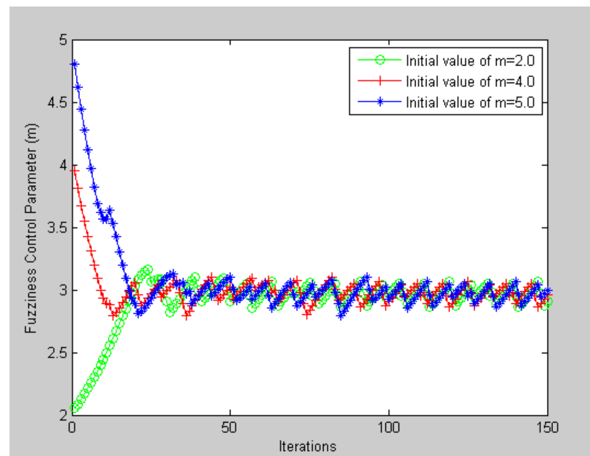

Fig. 9. Automatic tuning of Fuzziness Control Parameter (m) using CFCM with different initial value of m for Onscreen Keyboard dataset.

B. Choice of m for EEG data set

An analysis on the automatic update of fuzziness control parameter (m) is done over iterations with different initial values of m ranging from 2, 4 to 5 using CFCM algorithm. It has been observed from Fig. 9 that the value of m for EEG dataset converges within a small range from 2.8 to 3.2. Therefore a suitable choice for fuzziness control parameter (m) for EEG data set can be 3. So the traditional FCM algorithm is evaluated with a m value set at 3.

C. Analysis of the result of Supervised and Unsupervised learning algorithm

The performances of traditional FCM, CFCM and standard SVM classifier on Onscreen Keyboard and Logical reasoning dataset are shown in Fig. 10 and Fig. 11 respectively. From the figures it is evident that the performance obtained in unsupervised approach (FCM and CFCM) is comparable with that obtained by supervised learning algorithm (SVM).

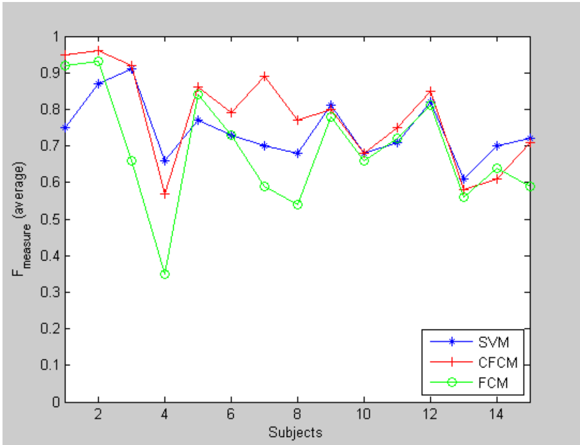

Fig. 10. Performance comparison between CFCM, FCM and SVM for On-Screen keyboard dataset.

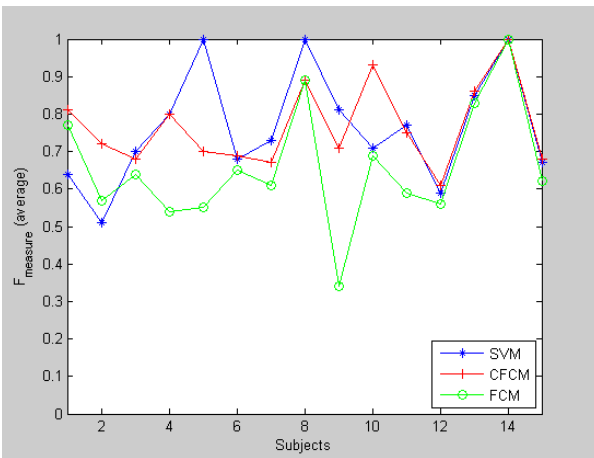

Fig. 11. Performance comparison between CFCM, FCM and SVM for Logical Reasoning dataset.

Table I give the average and standard deviation of $F_{measure}$ for all the subjects. It can be seen that the overall performance in terms of average $F_{measure}$ for CFCM and SVM are comparable. Moreover, the standard deviation for CFCM and FCM for both the datasets remains almost same whereas it varies drastically for SVM. It indicates that the issue of normalization across datasets does not affect the unsupervised (CFCM) learning.

If normalization issues arising due to non-homogeneity of human scalp and spatial drift in successive trials are to be handled properly using a supervised learning model, then it imposes an extra overhead of data normalization and preprocessing of training data. This also requires the construction of a training model every time there is a need for measuring the cognitive loads of individuals. This may not be feasible in real life applications as cited before, where continuous monitoring of cognitive loads of different users is essential to judge their cognitive ability. Therefore unsupervised approach which does not require any training phase and hence does not suffer from normalization issues can be a better alternative where there is a practical need for more real time monitoring of cognitive loads.

TABLE I. PERFORMANCE COMPARISION BETWEEN SVM, FCM AND CFCM

Stimulus	$F_{measure}$					
	SVM		***FCM***		***CFCM***	
	Avg.	***Std.***	***Avg.***	***Std.***	***Avg.***	***Std.***
On Screen Keyboard	0.743	0.082	0.688	0.155	0.779	0.129
Logical Reasoning	0.764	0.149	0.657	0.162	0.767	0.111

VI. CONCLUSION

This paper discusses the problems associated with the generation of a Universal model to analyze the cognitive loads of different individuals due to normalization issues of the EEG signal. Therefore it proposes an unsupervised approach for measuring the cognitive load which is a better alternative than the supervised one as it requires no training and hence is independent of the normalization problems. It is further experimentally proved that the unsupervised algorithm is a more feasible solution in practice where there is need for continuous measurement of cognitive load. However it is clear from the experimental results that CFCM will be a better choice, if unsupervised approach for two-level classification of cognitive loads is adopted. The complexity of the CFCM is higher compared to FCM, hence as part of future work, there is a need to reduce the complexity for real-time embedded implementation in portable solutions. Though the present approach has been justified with only two class clustering problem, it can further be employed for a multi-level cognitive load analysis problem in a seamless manner.

ACKNOWLEDGMENT

We would like to thank all the subjects participated in the experiments conducted and Mr. Debanjan Majumdar for his valuable contribution in recording the EEG signals.

REFERENCES

[1] http://en.wikipedia.org/wiki/Cognitive_load, last accessed on 29-July-2013.

[2] J. Sweller,.: The Cambridge handbook of multimedia learning. Cambridge University Press, 2005, ch. “Implications of cognitive load theory for multimedia learning,” pp. 19–30.

[3] John Sweller, "Cognitive load during problem solving: Effects on learning." *Cognitive science* 12, no. 2 (1988): 257-285.

[4] Lini, Sami, Christophe Bey, Sylvain Hourlier, Bruno Vallespir, Axel Johnston, and Pierre-Alexandre Favier. "Evaluating ASAP (Anticipation Support for Aeronautical Planning): a user-centered case study." In Proceedings of the 17th International Symposium on Aviation Psychology. 2013.

[5] T. Taylor, A. K. Pradhan, G. Divekar, M. Romoser, J. Muttart, R. Gomez, A. Pollatsek, and D. L. Fisher. "The view from the road: The contribution of on-road glance-monitoring technologies to understanding driver behavior." Accident Analysis & Prevention (2013).

[6] Moos, Daniel. "Examining hypermedia learning: The role of cognitive load and self-regulated learning." Journal of Educational Multimedia and Hypermedia 22, no. 1 (2013): 39-61.

[7] Creemers, Bert, Leonidas Kyriakides, and Panayiotis Antoniou. "Improvement of Teaching by Mastering Specific Competences: The Competency-Based Approach." In Teacher Professional Development for Improving Quality of Teaching, pp. 13-28. Springer Netherlands, 2013.

[8] M. H Leanne, and et all in "Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy", Proceedings of the SIGCHI Conference 2009 on Human Factors in Computing Systems, pp2185-2194,ISBN:978-1-60558-246-7, doi:10.1145/1518701.1519035.

[9] S. Feinberg, M. Murphy, "Applying cognitive load theory to the design of web-based instruction", Proceedings of IEEE professional communication society international professional communication conference and Proceedings of the 18th annual ACM international conference on Computer documentation: technology & teamwork, pp 353 - 360, ISBN:0-7803-6431-7.

[10] http://www.emotiv.com/epoc/features.php - last accessed on 29-July 2013.

[11] http://www.neurosky.com/Products/MindWave.aspx - last accessed on 29-July 2013.

[12] Y. Wang, S.Makeig, "Predicting Intended Movement Direction Using EEG from Human Posterior Parietal Cortex" in HCI international, San Diego CA, July 2009.

[13] D. Huang, P. Lin, Fei Ding-Yu., X. Chen, O. bai, "Decoding human motor activity from EEG single trials for a discrete two-dimensional cursor control" in Journal of Neural Engineering, vol. 6 046005, doi:10.1088/1741-2560/6/4/046005.

[14] C. Nuper, R. Gernot, R. Scherer, G. Pfurtscheller, "Motor imagery and EEG-based control of spelling devices and neuroprostheses", Progress in Brain research, Elsevier,vol. 159, 2006, pages393-409, DOI: 10.1016/S0079-6123(06)59025-9.

[15] F. Lotte, J. Van Dille Dirk, L. Koole, "Tuning down the emotional brain: An fMRI study of the effects of cognitive load on the processing of affective images", NeuroImage 2009.

[16] E. T. Solovey, P. Schermerhorn, M. Scheutz, A. Sassaroli, S. Fantini,R.J.K. Jacob, "Brainput: Enhancing Interactive Systems with Streaming fNIRS Brain Input", in CHI12, May 5-10, 2012, Austin, Texas, USA.

[17] E. E. Smith, J. Jonides, R.A. Koeppe, "Dissociating verbal and spatial working memory using PET", in Cerebral Cortex, 1996.

[18] J. Klingner, B. Tversky, P. Hanrahan: "Effects of visual and verbal presentation on cognitive load in vigilance, memory and arithmetic tasks". J. of Psyhophysiology 48, 3 (2011), 323–332.

[19] Y. Shi, N. Ruiz, R. Taib, E. Choi, F. Chen: "Galvanic skin response (gsr) as an index of cognitive load". In CHI '07 extended abstracts (2007), ACM, pp. 2651–2656.

[20] C. E. Parker, J. Gorin, and S. Bechard. "Adapting Reading Test Items: Decreasing Cognitive Load to Increase Access for Students with Disabilities." Lessons Learned in Federally Funded Projects that Can Improve the Instruction and Assessment of Low Performing Students with Disabilities (2013): 17.

[21] F. Lotte, C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified Theory and new Algorithms", in IEEE transaction on bio-medical Engineering vol. x, pages 355-362, 2011.

[22] A. Sinharay, D. Chatterjee, A. Sinha, "Evaluation of different onscreen keyboard layouts using EEG signals", to be published in SMC 2013, October 13-16, 2013.

[23] P. Chandler, J. Sweller: "Cognitive load theory and the format of instruction". Cognition and Instruction 8 (1991), 293–332.

[24] P. Hjemdahl, U. Freyschuss, A. Juhlin-Dannfelt, B. Linde, "Differentiated sympathetic activation during mental stress evoked by the Stroop test" in Acta Physiol Scand Suppl [1984, 527:25-29], (PMID:6589948).

[25] D. Chatterjee, A. Sinharay, A. Konar, "EEG-Based Fuzzy Cognitive Load Classification during Logical Analysis of Program Segments", to be published in IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), July 7-10, 2013.

[26] A. Pal, D. Chatterjee, D. Kar, "Evaluation and Improvements of on-screen keyboard for TV and set -top box", 15th IEEE International symposium on Consumer Electronics, June 2011, Singapore.

[27] D. Chatterjee, A. Sinha, A. Pal, A. Basu, "An Iterative Methodology to Improve TV Onscreen Keyboard Layout Design through Evaluation of User Studies", Scientific and academic publishing journal, Advances in computing, DOI:10.5923/j.ac.20120205.03, pages 81-91.

[28] I. S. MacKenzie, and R. W. Soukoreff, "Phrase sets for evaluating text entry techniques," in CHI'03 extended abstracts on Human factors in computing systems, Lauderdale, USA, 754-755, 2003.

[29] H. C. Huang, and B. H. Jansen. "EEG waveform analysis by means of dynamic time-warping." International journal of bio-medical computing 17, no. 2 (1985): 135-144.

[30] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, Klaus. Mller "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis", in IEEE Signal Processing Magazine, vol. 25, 2008.

[31] S. Gudmundsson, P. Runarsson, T. Sigurdsson, S. Eiriksdottir, K. Johnsen, "Reliability of quantitative EEG features", in Clinical Neurophysiology 118 (2007) 21622171, August 2007.

[32] V. Carmen, K. Nicole, B. Blankertz, A. Schlgl, "Time Domain Parameters as a feature for EEG-based Brain-Computer Interfaces" in Neural Networks 22 (2009) 1313-1319.

[33] W. Klimesch, "EEG alpha and Theta oscillations reflect cognitive and memory performance: a review and analysis" in Brain Research Reviews 29, 1999.

[34] H. Sijuan, "Feature extraction and classification of EEG for imagery movement based on mu/beta rhythms", in Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on 16-18 Oct, 2010.

[35] D. Das, A. Sinha, K. Chakravarty, A. Konar, "Stabilization of Cluster Centers over Fuzziness Control Parameter in Component-wise Fuzzy c-Means Clustering", to be published in IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), July 7-10, 2013.