Electroencephalography (EEG) Classification of Cognitive Tasks Based on Task Engagement Index

Joseph K. Nuamah and Younho Seong  
Industrial and Systems Engineering Department  
North Carolina A&T State University  
Greensboro, NC 27401, USA  
Email: jknuamah@aggies.ncat.edu

Sun Yi  
Mechanical Engineering Department  
North Carolina A&T State University  
Greensboro, NC 27401, USA

Abstract—The application of autonomous systems is on an increase, and there is the need to optimize the fit between humans and these systems. While operators must be aware of the autonomous systems dynamic behaviors, the autonomous systems must in turn base their operations, among other things, on an ongoing knowledge of operators’ cognitive state, and its application domain. Psychophysiology allows for the use of physiological measurements to understand an operator’s behavior by noninvasively recording peripheral and central physiological changes while the operator behaves under controlled conditions. Electroencephalography (EEG) is a psychophysiological technique for studying brain activation. In the present study, EEG task engagement index, defined as the ratio of beta to (alpha + theta), are used as inputs to an artificial neural network (ANN) to allow identification and classification of mental engagement. Six separate feedforward ANN with single hidden layer were trained by backpropagation with a group of similar operators used to train the classifier. Afterwards, the trained classifier is continuously fed physiological data during performance of a task. The classification accuracy across the six participants was 88.67%. The results show that differences in cognitive task demand do elicit different degrees of mental engagement, which can be measured through the use of the task engagement index.

Keywords—Artificial neural network (ANN); cognitive tasks; electroencephalography (EEG); short term Fourier transform (STFT); task engagement index

1. INTRODUCTION

The application of automated and autonomous systems is on an increase, and there is the need to optimize the fit between humans and these systems. For example, in an autonomous or highly automated work environment where the operators’ role is primarily monitoring the system, there is very little overt operator performance to observe in order to determine whether or not the operator needs adaptive aiding. Adaptive aiding improves performance of system by providing automation when the operator needs it [1]. Furthermore, in such environments, the functional state of operators may be monitored in order to determine when to provide the aid. Previous research suggests that physiological measures such as EEG ([2], [3]) may be sensitive to unpredictable changes in cognitive load making them useful for adaptive automation. EEG signals represent summed postsynaptic potentials of neurons firing at a rate of milliseconds. Graphically, an EEG is a graph of the time varying voltage difference between an active electrode attached to the scalp and a reference electrode [4].

EEG may be analyzed in the time (event related potentials), frequency (spectral) or time-frequency domain. Event related potentials (ERPs) represent the brain’s neural response to specific sensory, motor or cognitive events. They are computed by averaging EEG epochs time-locked to a specific event [5]. However, ERPs fail to capture brain activity related to stimulus processing that is not time-locked to stimulus onset [6]. The frequency bands in an EEG signal are delta (less than 4Hz), theta (4–8Hz), alpha (8–12Hz), beta (12–30Hz), and gamma (30–80Hz) [6]. During highly controlled cognitive tasks, the spectral composition of EEG changes in response to changes in task difficulty or level of alertness [7]. Spectral analysis involves quantifying amplitude or power of a certain frequency band. This approach is able to capture brain activity related to stimulus processing that is not time-locked to stimulus onset. Time-frequency analysis involves quantifying the spectral energy over time. It maps the 1-dimensional EEG time-domain signal into a 2-dimensional function of time and frequency.

In order to determine the operator functional state, physiological measures recorded during task performance are combined by a classifier. Classifiers such as artificial neural networks (ANN) have been used to obtain estimates of operator functional state. Data that represent the cognitive states of interest from each operator separately or data obtained from a group of similar operators are used to train the classifier. Afterwards, the trained classifier is continuously fed with physiological data during performance of a task. The classifier uses these data to estimate the operator functional state. When the classifier detects that the operator needs assistance, it notifies the system so that the operator will be provided the right automation [1].

Pope, Bogart, and Bartolome [8] developed the first brain-based adaptive system that used an EEG-based task engagement index. The system used a bio-cybernetic loop that was formed by changing levels of automation in response to changes in mental workload demands. Of the candidate indices that they evaluated, Pope, Bogart, and Bartolome [8] found that the engagement index based on the ratio beta to (alpha + theta) was the most sensitive. Further studies undertaken by Freeman, Mikulka, Prinzel, and Scerbo [9] confirmed the effectiveness of the beta to (alpha + theta) index. They reported improved
performance in a vigilance task when the engagement index was used to drive changes in the stimulus presentation. In another study, Berka et al. [10] found that engagement and workload increased significantly during encoding of verbal and image-learning and memory test compared with recognition or recall period.

Other researchers have explored the task engagement index in determining the cognitive states of human participants. For example, in one study Szafrir and Mutlu [11] employed the task engagement index formula to study how participants interacted with an adaptive robotic agent that monitored attention in real time and adapted its behavior to improve the discourse. The agent monitored attention levels of participants and detected attention drops. In another study Szafrir and Mutlu [12] designed and implemented a system that employed the engagement index to monitor participants’ attention to educational material in real time. This adaptive system suggested the optimal review topic. Huang et al. [13] presented an EEG-augmented reading system that monitored the engagement levels of children in real time, and provided contextual BCI training sessions to improve the children’s reading engagement. McMahan, Parberry, and Parsons [14] compared three different EEG engagement indices (frontal theta, ratio of beta to (alpha + theta), ratio of frontal theta to parietal alpha) by assessing user engagement during various video game modalities. They found beta to (alpha + theta) to be the best algorithm for calculating the engagement levels of players playing video games.

Research has shown that classifiers using physiological features are able to determine the level of cognitive activity in tasks with a high level of accuracy. King, Nguyen, and Lal [15] employed ANN that used the magnified gradient function (MGF) technique to detect and classify early driver fatigue. The MGF reduced the time required for training by modifying the standard back propagation (SBP) algorithm. The ANN classified professional driver fatigue with 81.49% accuracy and non-professional driver fatigue with 83.06% accuracy. Khare, Santhosh, and Anand [16] compared the performance of five ANN methods (Gradient Descent Back Propagation, Levenberg-Marquardt, Resilient Back Propagation, Conjugate Learning Gradient Back Propagation, and Gradient Descent Back Propagation with momentum) for classification of planning of right hand movement with respect to an awake relaxed state. Wavelet packet transform (WPT) was used for feature extraction of the relevant EEG signals. The Resilient Back Propagation method showed the highest performance with the accuracy of 95%. Chai, Tran, Craig, Ling, and Nguyen [17] examined the classification between fatigue and alert states using an autoregressive (AR) model-based power spectral density (PSD) as the features extraction method and fuzzy particle swarm optimization with cross mutated of artificial neural network (FPSOCM-ANN) as the classification method. Using 32-EEG channels, they obtained a classification accuracy of 80.51%. Classification using fewer (eleven frontal) EEG channels resulted in 75.65% classification accuracy. Belakhdar et al. [18] detected the occurrence of driver drowsiness onset based on the Artificial Neuronal Network (ANN) and using only one EEG channel. They conducted an experiment on ten human participants using nine features computed from one EEG channel using the Fast Fourier Transform (FFT). After introducing these features in an ANN classifier, Belakhdar, Kaaniche, Djenel, and Ouni [18] obtained a classification accuracy rate of 86.1% and 84.3% of drowsiness and alertness detection.

Although previous studies are suggestive of the potential of classifiers using physiological features to determine the level of cognitive activity in tasks, some questions still remain. Of particular interest is the question of whether or not task engagement indices (TEIs) can be used as features for classification of cognitive tasks. In particular, it has yet to be demonstrated that TEIs can serve as features to optimally discriminate recorded brain signals. We address this issue by employing TEIs as input to ANN to allow identification and classification of mental engagement. We propose that differences in task demand will elicit different degrees of mental engagement, which could be measured through the use of the task engagement index.

This paper is outlined as follows. First, we describe the EEG data acquisition and pre-processing, time-frequency representation, and classification processes under the methodology section. Next, we present the results and discussion. Finally, we draw conclusions.

II. METHODOLOGY

A. EEG Data Acquisition and Pre-processing

The EEG data used in this study was originally collected by Keirn and Aunon [19], and was downloaded from the website of the Brain-Computer Interfaces Laboratory of the Colorado State University [20]. The participants were seated in an Industrial Acoustics Company sound controlled booth with dim lighting and noise-less fan (for ventilation). An elastic electrode cap was used to record EEG signals from positions C3, C4, P3, P4, O1, and O2 (shown in Figure 1), defined by the 1020 system of electrode placement [21]. The electrodes were connected through a bank of Grass 7P511 amplifiers and bandpass filtered from 0.1–100Hz. Data was and recorded at a sampling rate of 250 Hz with a Lab Master 12 bit A/D converter mounted in an IBM-AT computer. Eye blinks were detected by means of a separate channel of data recorded from two electrodes placed above and below the participants left eye. Continuous EEGs were recorded for 10s during each task. For each of these tasks, participants were instructed to not vocalize nor gesture in any way. In the present study, data from six participants performing five mental tasks were analyzed. The five tasks are baseline, multiplication, letter composition, geometric figure rotation, and visual counting. For each task, participants were instructed not to vocalize or gesture in any way. The baseline task was an eye opened task in which participants were asked to relax as much as possible. This task served as a reference task against which the other tasks were to be compared. In the multiplication task, participants were given nontrivial multiplication problems, such as 49 times 78, and were asked to solve them without vocalizing or making any other physical movements.

In the letter composition task, participants were instructed to mentally compose a letter to a friend or relative without
vocalizing. In the geometric figure rotation task, participants were asked to visualize a particular three-dimensional block figure being rotated about an axis. In the visual counting task, the participants were asked to imagine a blackboard and to visualize numbers being written on the board sequentially.

Data was recorded for 10 seconds during each task and each task was repeated five times per session. With a 250 Hz sampling rate, each 10 second trial produced 2,500 samples per channel. Data from six participants were used. All data analyses were performed offline with MATLAB [22], EEGLAB [23], and custom code.

**B. Time-Frequency Representation**

It is extremely important to find a suitable representation of EEG signals in order to reliably discriminate and understand extracted relationships [25]. The original recorded EEG data used in the present study were time domain signals. Several techniques exist for spectral estimation and representation of EEG signals. Among these is the Fast Fourier Transform (FFT) which allows for the efficient estimation of the component frequencies in data from a discrete set of values sampled at a fixed rate. However, for signals whose frequencies change in time (like EEG), the FFT has disadvantages. The FFT cannot provide simultaneous time and frequency localization, which means that the power spectrum does not provide information about when certain frequencies occur in the signal. It is therefore not very useful for analyzing time-variant, nonstationary signals like EEG. This problem is overcome by using the Short Term Fourier Transform (STFT). The STFT segments the signal into narrow time intervals, and takes the Fourier Transform (FT) of each segment. Each FT provides the spectral information of a separate time-slice of the signal, providing simultaneous time and frequency information. The STFT was used in the present study. It was used to estimate the power spectrum of the EEG bands. This was implemented with the spectrogram function in MATLAB. The spectrogram computes an FFT-based spectral estimate over each sliding window and allows for the visualization of how the frequency content of the signal changes over time. The spectrogram function divides a signal into segments. Long segments or windows (also known as narrowband spectrogram) provide better frequency resolution whereas short segments (also known as wideband spectrogram) provide better time resolution. In the present study, the data was segmented into quarter-second windows with a 50% overlap of the previous segment and 50% of the next. The Hann window which has good frequency resolution and reduced spectral leakage was used. The log power, calculated as, $10 \log_{10}(\text{power})$ of the data in each window was computed for the theta band (4-8Hz), alpha band (8-12Hz), and beta band (12-30Hz).

Fig. 3 shows the 3-D spectrogram visualization of the multiplication task for participant 1. The yellow colors indicate frequency content with higher power while the blue colors indicate frequency content with very low power. For each participant under each of the five cognitive task, for each trial, and for each of the six channels (C3, C4, P3, P4, O1, and O2), the task engagement index given by Equation (1) from Pope, Bogart, and Bartolome [8] was computed. This resulted in a total of 6 features (task engagement indices) per trial per task. In order to avoid features in greater numeric ranges dominating those in smaller numeric ranges [26] each feature was linearly scaled to the range (0, 1).

$$TEI = \frac{\text{beta power}}{\text{(alpha power + theta power)}}$$  

(1)
**C. Classification**

First, data for each participant were classified separately. Six separate feedforward ANN each with single hidden layer trained by backpropagation were designed to classify the five tasks (baseline, multiplication, letter, geometric figure rotation, and visual counting) for each participant. Second, a single feedforward ANN was designed to classify the five tasks for all participants. The trainbr network training function in MATLAB was used for all the networks. Trainbr generally works best when the network inputs and targets are scaled so that they fall approximately in the range (1, 1) [22]. It updates the weight and bias values according to Levenberg-Marquardt optimization. An ANN may suffer from overfitting, a situation where the network memorizes the training samples, but performs poorly on new datasets. Regularization may be used to improve ANN generalization and avoid overfitting. The trainbr network function used in the present study uses Bayesian regularization which takes place within the Levenberg-Marquardt algorithm. It minimizes a combination of squared errors and weights, and then determines the correct combination in order to produce a network that generalizes well [22]. The performance function used for all networks was the mean square errors (MSE). The MSE is the distance between the ANNs estimate of the test data and the actual test data. The data for each participant were analyzed separately. The default dividerand function, which randomly separates targets into three sets (training, validation, and testing), was used for all networks. For the training of each network, the algorithm was allowed to run until the effective number of parameters has converged. Each training stopped with the message “Maximum MU reached.”

The best performing single layer feedforward ANN (with least MSE) for each participant was determined by varying the number of hidden nodes from 10 to 50 in steps of 5. A total of ten ANN was analyzed for each participant. Out of the ten ANN analyzed, the one with the least MSE was chosen and its classification accuracy was obtained.

To determine the best performing ANN (with least MSE) for all participants, the number of hidden nodes of single layer feedforward ANN and two layer feedforward was varied from 10 to 50 in steps of 5 for the first layer, and from 5 to 20 in steps of 5. For each n-layer (n = 1 and 2) feedforward ANN, a total of forty ANN was analyzed. Classification accuracy for the best performing ANN was obtained.

### III. RESULTS AND DISCUSSION

Table I shows the classification results of the single layer ANN for each participant. The classification accuracy of 64% for participant 6’s ANN with 30 neurons in its hidden layer was the highest. It also had the highest number of epochs, and highest MSE of 0.1223. The classification accuracy of the ANN for participants 1, 3, and 4 is 100%. The ANN for participant 5 produced the lowest classification accuracy of 62% with 50 neurons in its hidden node. The average classification accuracy across all participants was 72.3%.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Number of Hidden Nodes</th>
<th>Epoch</th>
<th>MSE</th>
<th>Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>89</td>
<td>3.80 x 10^-16</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>4750</td>
<td>0.08</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>115</td>
<td>6.90 x 10^-12</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>125</td>
<td>1.27 x 10^-12</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>311</td>
<td>2.18 x 10^-16</td>
<td>96</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>5238</td>
<td>0.12</td>
<td>64</td>
</tr>
<tr>
<td><strong>Average Classification Accuracy</strong></td>
<td><strong>88.67</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The performance plot and network diagram of ANN for participant 1 are shown in figures 4a and 4b respectively.
The confusion matrix for the ANN designed for participant 1 and participant 2 are shown in figures 5a and 5b respectively.

![Confusion Matrix for ANN for participant 1](Fig 5a. Confusion Matrix for ANN for participant 1)

![Confusion Matrix for ANN for participant 2](Fig. 5b. Confusion Matrix for ANN for participant 2)

Task engagement is defined as effortful striving towards task goals [27]. The TEI measures how cognitively engaged a person is in a task. In the present study, each participant was presented with five different cognitive tasks. Each task was expected to elicit a different degree of mental engagement. We used TEIs as inputs to ANNs to allow identification and classification of mental engagement. The average classification accuracy across the six participants was 88.67%. The classification accuracies for participants 2 and 3 were relatively low. This is partly because an ANN is data driven and is prone to small datasets. Both participants had the least datasets of size 25. In addition, it took a relatively longer time for their networks to converge, 4750 epochs for participant 2, and 5238 for participant 3. The datasets for the remaining ANN were relatively large; 50 for participants 2, 3 and 5, and 75 for participant 4.

Our results show the sensitivity of TEI to different types of cognitive processing demand. Furthermore, our results show the extent to which TEI can be used to distinguish different cognitive tasks. The ability to unobtrusively and continuously monitor operator mental states in operational environments where operator movements and physical activity are somewhat constrained, could be beneficial in finding more efficient and effective methods for humans to interact with technology [10]. EEG indices have been found to be reliably associated with varying levels of decision making tasks, and thus show promise as candidates for directing adaptation in augmented cognition [28]. One domain where the classification of TEIs can be applied is adaptive decision aids. Classification results may be used to modify an operator’s decision making task via adaptive aiding with the goal of enhancing overall performance [29]. The decision aid may adapt to an operator in a particular cognitive state at a specific time. Decision making tasks can adaptively be allocated between the decision aid and the operator so as to minimize task delay time, subject to maintaining human workload at a satisfactory level [30].

IV. CONCLUSION

Differences in cognitive task demand elicit different degrees of mental engagement. These differences can be measured through the use of the task engagement index which is correlated with task demands. The task engagement index may serve as features for identification and classification of mental engagement of cognitive tasks. Our study contributes to the elaboration of the utility of EEG indices in general, and task engagement index, in particular to optimizing the fit between humans and autonomous systems.

ACKNOWLEDGMENT

This material is partially based on research sponsored by Air Force Research Laboratory and OSD under agreement number FA8750-15-2-0116, and partially by US Army RDECOM (W911NF-13-1-0118). The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Research Laboratory and OSD or the U.S. Government.

REFERENCES


