Artificial Neural Network with Electroencephalogram Sensors for Brainwave Interpretation: Brain-Observer-Indicator Development Challenges

Nicholas Polosky, Jithin Jagannath, Daniel O'Connor, Hanne Saarinen, Svetlana Foulke
ANDRO Advanced Applied Technology, ANDRO Computational Solutions, LLC, Rome NY,
{npolosky, jjagannath, doconnor, hsaarinen, sfoulke}@androcs.com

***Abstract*—This paper reports on challenges and opportunities associated with the development of an electroencephalogram (EEG) based personalized device for monitoring of brain activities pertaining large scale neural dynamics in the observed and providing relevant feedback to the observer. The envisioned device interprets signals and categorizes them on classes of typical responses. This could enable a speechless interaction between an observer and a participant wearing the device. This framework is different from the brain-computer-interface (BCI) framework as it focuses on indicators relevant to the human observer, brain-observer-indicator (BOI). Sensors detect resting states of the brain with associated patterns, synchrony between regions, and spectral changes in response to a cognitive event. A cognitive event results in notable changes in the associated patterns of electrical potentials. Recognition of these patterns has a broad application base, if the pattern-activity mechanism is characterized and recognized. The scope of the project includes development of a smart interaction support system BOI, relying on utilization of an EEG toolkit and an artificial neural network for personalization. The objective is to develop software that will support applications requiring feedback (i.e., training), along with a method for obtaining statistical data on the associated brain activity for engineering studies geared to improve signal acquisition and device performance. The findings from preliminary stages of the project are encouraging but indicate multiple challenges that must be addressed including a trade between a reduction of noise and complexity of classification software, definition of classes and recognition of classes and patterns, and development of an effective training data set acquisition strategy.**

***Keywords*—Artificial Neural Network, Electroencephalogram.**

I. Introduction

Sometimes we wish we could look into people's minds to obtain some feedback. A few examples of such situations include a conversation with a preoccupied person, a questioning session of a person with an impaired verbal communication ability, or a training session where learners have various degree of exposure to the subject matter. Both sides (talking-listening, teaching-learning) would benefit if some degree of feedback was available for the observer to understand the state of mind of the observed participant. The neural excitement within the brain creates large number of electrical dipoles emitting electric signals that can be measured and analyzed to identify patterns, the feedback regarding the participant's engagement in the subject can be assessed (Fig. 1).

The major categories of brain wave oscillations, including Alpha, Beta, Gamma, Theta and Delta waves, are briefly described in Table I. The patterns of oscillations can be used to assess participant's concentration. Electroencephalogram (EEG) measures brain's ongoing electrocortical activity recorded from the scalp [1]. Event-related-potential (ERP) reflects the changes in electrical activity in response to stimuli or event [2], and is estimated from EEGs. EEG detects resting states of the brain with associated patterns and rhythms, synchrony between regions, and spectral changes in response to a cognitive event. Deflections in the ERP reflect specific aspects of cognitive processes. Measures used in ERP research (scalp topographic distribution, polarity, amplitude, latency, etc.) may provide important insight about perceptual, cognitive, and motor functions in normal and in psychopathological conditions [3]. The placement of sensors and the number of sensors necessary to capture consistent and prominent features, depends of the manifestations of the event on the scalp. There is a tremendous body of research relating EEGs and ERPs to events [3]–[6]. And there is equally impressive body of literature examining brainwave oscillations patterns, the balance between β, α, θ and δ waves, and relating these to activities and outcomes [7]–[9].

TABLE I: Types of Brain Waves

Type	Frequency	Significance
Delta, δ	0.5-3.5 Hz	in deep sleep
Theta, θ	3.5-7.5 Hz	drowsy, daydreaming, light sleep
Alpha, α	8-14 Hz	relaxed state, not actively thinking
Beta, β	14-30 Hz	interacting, concentrating, solving problems
Gamma, γ	30-70 Hz	perception, cognition, memory access.

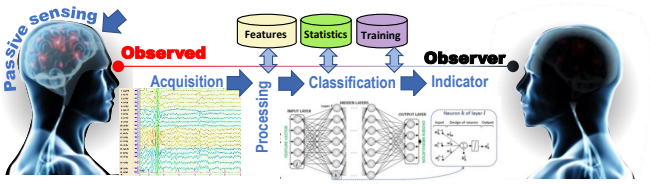

Fig. 1: Brain-Observer-Indicator framework.

In the absence of an explicit task, the brain shows a temporally coherent activity - "resting state", the default-mode network associated with daydreaming, free association, stream of consciousness [10]. Meditators have been known to show high activity of α brainwaves accompanied by β, θ and even δ waves that were about half the amplitude of the α waves [11]. A specific balance between β, α, θ and δ brainwaves with the strongest amplitude in α is speculated to enable the individuals to reach peak performance of their cognitive, creative and athletic abilities through "brainwave training" [11]. The strength of θ-γ coupling is reported to increase during learning [12]. α rhythm is the most prominent component of the vast majority of human EEG records and is considered to be dominant rhythm. There is ample evidence

978-1-5386-2275-9/17/$31.00

Fig. 2: Brainwave activity in 8-channels displayed in the GUI and the observed wearing headset.

that the frequency of α rhythm represents neurophysiological mechanisms directly related to individual differences in information processing. It is less influenced by extracerebral factors (skull thickness or conductance) and its variation is attributed to the variation in brain function [7].

Over the last decade several schools in the USA have begun to utilize neurofeedback for the special education of children with attention and learning disorders, based on a trend that such kids show low levels of activity in frontal brain areas, with an excess of θ waves and deficit of β waves [13], [14]. Even though EEG shows a high inter-individual variability [7], patterns are revealing factors and BOI enablers.

The intent of BOI is to enable an observer to obtain a real-time feedback from the observed. For example, a feedback regarding the participant's thinking is derived from measurements, and assessed to evaluate such indicators as concentration (boredom, flow, frazzle). These indicators are important to the observer for improving the quality of interaction. The methodology is to utilize databases and toolkit measurements to develop the framework and software architecture with an artificial neural network (ANN) in its core to enable personalization.

II. Challenges

There are several challenges that need to be addressed. The discernible brainwave patterns must be identified and related to the typical anticipated response categories (a multidimensional problem). Effective signal processing approaches for pattern recognition in the environments with varying noise must be developed. The effective ANN training and setup criteria must be developed. Due to tremendous recent developments in the fields of machine learning and BCI technology [15], we identified BCI components suitable for a breadth of BOI applications. BOI consists of an array of compact EEG sensors, supported by a computing board with software for control, processing and analyses.

A. Noninvasive Sensors

We use OpenBCI and BCILAB hands-on tools for initial development and feasibility studies [16]. OpenBCI carries the hardware and software necessary for the prototype implementation, and offers the application programming interface (API) and Programming Environment with the Software Development Kit [17]. Simulation tools within the MATLAB and Octave environments with the standard EEG data available online [18]. In our setup, shown in Fig. 2, we passively collect new data using OpenBCI toolkit with various number-of-sensors and sensor placement. Fig. 2 depicts a snapshot of the GUI and a photograph of the observed wearing the headset with eight EEG sensors. We use Ultracortex Mark IV headset with the Cyton Biosensing Board to record research-grade brain activity, muscle activity, and heart activity, with 16 channels of EEG sampling from up to 35 different sensor locations from the internationally accepted 10-20-System for electrode placement [16]. Dry electrodes are placed on forehead and scalp, and ear-clip electrodes are clipped to ear-lobes to establish reference.

A relatively small number of sensors gives only a coarse resolution depiction of the scalp potentials; however, the processing load reduces. The potentials on the scalp are composites combined from many signal sources, so even small number of sensors would pick up prominent signatures. Ideally, we strive to find the setup with the smallest number of electrodes sufficient to produce the classification data in a given application. Various number of electrodes and their placement affect the classification patterns and features. It is speculated that an electrode capable of assessing its relative location on the scalp and providing this information to the classification software, would improve the classification outcome and simplify the classification software. In the current configuration classification results significantly vary based on how the head-gear is aligned. Achieving a consistency in electrode placement is an unresolved challenge.

B. Brainwave Patterns

There are known brain wave patterns and a body of literature with interpretations. However, there are could be deviations from the typical interpretations of such patterns as well as changes of the interpretation itself. For example, according to [9], because the amplitude of α oscillation is suppressed by visual stimuli and enhanced during mental calculation and working memory, they were thought in the past to reflect idling or inhibition of task-irrelevant cortical areas. However, recently these dynamics have gained an attribution to the mechanisms of attention and consciousness. Simultaneous α, β and γ frequency band oscillations may indicate unified cognitive operations coordinating the selection and maintenance of neuronal object representations during working memory, perception and consciousness [9]. γ oscillations are thought to transiently link distributed cell assemblies that are processing related information, a function important to perception, attentional selection and memory [19]. Fig. 3 depicts an example of the indicators for a learner's state of mind for education application of BOI. The Processes are subjective. The Waves change either in amplitude relative to a background affecting the balance among the bands or in the degree of coupling between specific bands. Among the challenges in the classification framework are

- identification of relevant classes of patterns
- inventory of typical events-processes-library
- strategy for calibration or algorithm learning as applicable

C. Classification Preprocessing

The four major components of BOI system are signal acquisition, pre-processing, classification, and the application

Process \ Waves	δ	θ	α	β	γ	
Deep sleep	↑					
Light sleep, daydreaming		↑				
Relaxed, not thinking			o			
Visuals			↓			
Mental calculation			↑			
Concentrating, solving problems				↑		
Attention, information processing					↑	
Working memory			α-β-γ			balance
Learning		θγ			θγ	coupling
Boredom		↑	?	↓		
Flow	↑	↑	↑	↑		balance
Frazzle		↓	?	↑	?	f(stress)
Thoughtless overactivity		↑				
Meditation	δ + θ + 2α + β					balance

Fig. 3: Examples of states for classification

interface. Preprocessing plays a vital role, especially in the filtering and removal of artifacts from EEG signals which consist of signals from a non-cerebral origin and can be categorized on physiological and non-physiological artifacts [20]. Independent Component Analysis (ICA) provides a potential to the removal of such artifacts by separating a set of linear mixed signals into a set of independent components (i.e., movement of the head) [21]. Non-physiological artifacts comprise noise from the environment (i.e. electrical equipment). The choice of preprocessing technique depends on application. Feature extraction is a critical step [22] that can be performed in spatially, spectrally, and temporally. Spatial features relate to physical locations on the scalp (not signal origins). Spectral information provides the frequency details related to the state of mind. Due to the anticipated variability in EEG signals/features we focus on machine learning approaches for classification.

D. Machine learning

Many machine learning algorithms are based on training the system to produce a solution. The type, size, and non-uniformity of the training data set can affect the accuracy and practical effectiveness of the system. A properly designed procedure for obtaining a training set can enable robust performance.

Several approaches have been studied to address BOI classification [23]. Each studied system was constrained to a specific task due to the immense amount of data. The dimensionality of data is a challenge accentuated around the necessity of classifying features in the time-series data. EEG signal data are often time dependent which adds to the dimensionality of the feature vectors of static classifiers. To mitigate this problem, dynamic classifiers are used. Common examples of static classifiers are Artificial Neural Networks (ANN) and Support Vector Machines [24]. Some dynamic techniques use a Hidden Markov Model (HMM). HMMs assume the system is modeled as a Markov process with unobservable states and develop transition probability distributions for time-series data. Several adaptations have been made to create dynamic instances of ANNs: Time Delay Neural Network (TDNN), Gamma Dynamic Neural Network (GDNN), and Recurrent Neural Network (RNN). TDNNs attempt to achieve dynamics by delaying certain sets of inputs and using them in conjunction with inputs that occur later in time. The process of deciding which inputs are delayed is arbitrary and application specific. In general, TDNNs have a lower depth of memory but a higher resolution of that recent memory. GDNNs have parameters that can be learned, which allow the network to trade memory depth for memory resolution, and vice-versa. RNNs allow the network to keep an internal state that is passed from time step to time step. Previously we have successfully developed signal classification techniques for wilress communication systems [25], [26] and even employed ANN for the same [27]. Similar challenges were addressed in classifying features in multi-frequency, multi-modulation, and diverse signal-to-noise environments. Although HMMs are widely used due to their simplicity in contrast to other dynamic algorithms, they have several drawbacks. RNNs can achieve a much larger state space and can model higher order relationships. Furthermore, RNNs have been shown to outperform TDNNs and GDNNs in various classification tasks in [28]. There has been recent work on RNNs to further improve their effectiveness as well as allow them to retain memory for longer periods of time. For the reasons mentioned above, the classification algorithm considered for the notional BCI is a Long-Short Term Memory Network (LSTM), which is a specific type of RNN.

III. Application

A. Feedback to teacher in an educational setting

In an education setting, a BOI will transform the way student and instructor interact and will improve the efficiency of learning (the seamless and honest feedback to teacher will enable teacher to adjust their teaching approach).

A teacher will have a tool to assess if the students are taking in the material and to manage performance-stress relationship to improve the quality of teaching. Based on literature survey, we assume several representative groups of the mind state relevant to education setting (Fig. 3):

- *relaxed not thinking* - α waves at a resting state
- *daydreaming* - a resting state network with θ and α waves
- *thoughtless over-activity* - increase in θ
- *boredom* - increase in θ waves, decrease in β
- *thinking* - increase in α or β waves
- *remembering* - balance between α, β and γ waves
- *learning* - increased strength of θ-γ coupling
- *flow* - peak performance with balanced β, α, θ and δ
- *frazzle* - increase in β waves, decrease in θ
- *meditation* - α accompanied by weaker β, θ and δ waves

The balance and coupling between different frequency bands are prominent features. However, approximate balance is a fuzzy feature: difference between flow and daydreaming is in the amount of contributing β waves, while a slight increase in θ and α indicate a loss of focus. In addition, a degree of stress plays a significant role in the activations of brain networks. The neurobiology of frazzle indicates that frazzle arises from the nervous system's plan for crisis [29]. During this state, control from the brain's executive center in the prefrontal area (behind the forehead) shifts to the more primitive emotional circuitry in mid-brain (between the ears). For learners this means the more anxious they are, the less they register their lessons.

Fig. 4 provides an example of identification of relevant classes under the constraints of the application in educational

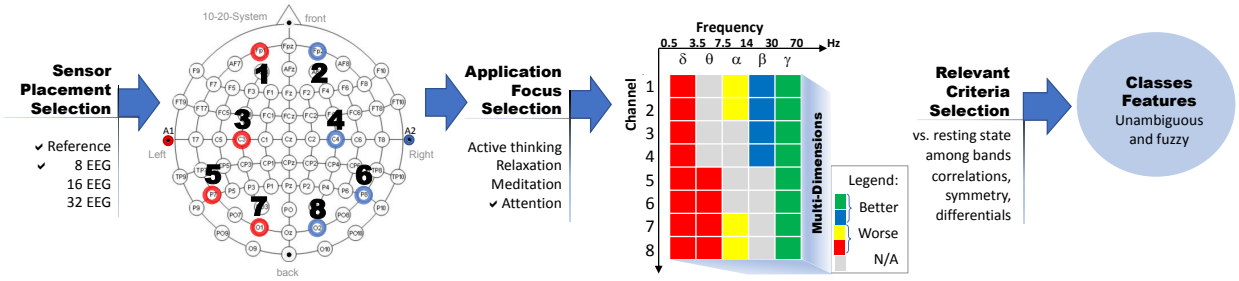

Fig. 4: Identifying criteria for classification

setting. The number of sensors and their locations in the 10-20-System grid are determined first. These are selected to detect changes in the projected to the scalp potentials due to activity originated approximately in the visual cortex, visual information processing regions, in the prefrontal area and mid-brain. Next we select application specifics, such as attention to visuals or active thinking or meditation/concentration. The observed begins to pay attention and learn, while the observer monitors the brainwave activity. Dominant δ and θ waves will give away that the observed is not paying attention.

Such scenario could be classified using a traditional classification tree if variability was limited. There are several major contributors to the variability: the stress level, the degree of exposure to the subject, the intention (pass the test and forget, or retain the information) and attitude (which could be influenced by peers or environment). There are differences of information retention based on the degree of exposure (new, familiar from the short term memory, known from the long term memory), which activates different networks and results in different patterns of oscillations.

A personalized setup of the BOI is possible with ANN. It does not depend on the number of sensors, but depends on the quality of training data set. A training session must be executed to obtain the brainwave activity signals from the resting state, learning and retention states with a different degree of participant engagement from boredom to flow to frazzle or under different stress conditions. We set up an ANN training session for low stress environment and attempted to initiate these states in the participant to generate a sufficient amount of data for ANN training. The training strategy is transferable to other participants.

IV. Design Concept

A. System Overview

The system architecture components relevant to the software development are depicted in Fig. 5. The main software blocks are the signal acquisition and assimilation, the signal processing and feature extraction, the classification, the reasoner, and the interface to a communication system. The reasoner performs analyses of classified data and produces the response indicator decision. The communication system delivers the concise relevant information to external devices (mobile phone, personal computer, etc.).

The classification system can use a variety of classification approaches to perform a capability assessment and trade study for performance versus system resources. We reduce the number of classes in the library to three: not thinking, mental calculation, and learning. This limits the waves of interest and the features to the amplitude change in θ and α waves, and to the θ-γ coupling. We intend to gradually expand the classification library to include other identified states, both the unambiguous states and the fuzzy states.

B. Experiment Implementation

For initial experiments we focused on a LSTM network approach to the classification task. The Nesterov-Accelerated Adaptive Moment Estimation (NADAM) algorithm was used to optimize the LSTM network during the training phase of our experiments. LSTM networks are a more specific type of a broader category of artificial neural networks called RNN, and were presented as an improvement to the vanilla RNN when they were first introduced to machine learning field. RNNs perform learning tasks related to sequence representation, which include tasks such as classification and generation. One way to conceptualize RNNs, is to imagine a sequence of feed-forward networks, one for each time step in the input sequence, with connections from inputs to outputs as well as connections across time steps. LSTMs have similar connections between time steps, additionally these networks have internal cell states, which are shared across time steps, that allow the network to learn time dependencies between the input features at different times in the input sequence. LSTMs incorporate the idea of four different gates, namely the input, forget, output, and block input gate. These gates essentially control information flow within the LSTM cell. The use of these gates allow the LSTM to learn long term as well as short term dependencies in the input sequence. The way these gates interact with the information contained in an LSTM cell can be characterized using the formulas that the concepts are based on, thus a mathematical outline of the LSTM structure follows. Input weights will be denoted as W, recurrent weights will be denoted as R, bias weights will be denoted as b, and the subscript will determine to which gate the associated weight belongs to. The subscripts used will be i, f, o, g, denoting the input, forget, output, and block input gates respectively. Superscripts will be used to denote the time step to which the variable belongs. The logistic sigmoid function will be denoted using $\sigma(x)$ and is given explicitly below:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{1}$$

Pointwise vector multiplication is denoted with $\odot$. The input and output vectors at a time t are given as x^t, y^t respectively.

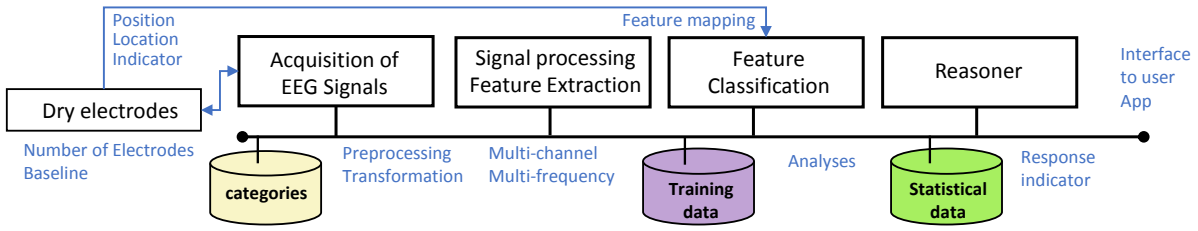

Fig. 5: System Architecture

Lastly, the cell state at a time t is given as c^t. The definition of these variables allows for the interpretation of the forward propagation through an LSTM network. The weighted sum and activation of the block input gate is given below, respectively:

$$\bar{g}^t = W_g x^t + R_g y^{t-1} + b_g \qquad g^t = tanh(\bar{g}^t) \tag{2}$$

The *tanh* function is used here as the activation function for the block input gate; however, any pointwise non-linear function can be substituted. The weighted sums and activations for the input, forget, and output gates are all computed in a similar manner and are given below using the same convention as above:

$$\bar{i}^t = W_i x^t + R_i y^{t-1} + b_i \qquad i^t = \sigma(\bar{i}^t) \tag{3}$$

$$\bar{f}^t = W_f x^t + R_f y^{t-1} + b_f \qquad f^t = \sigma(\bar{f}^t) \tag{4}$$

$$\bar{o}^t = W_o x^t + R_o y^{t-1} + b_o \qquad o^t = \sigma(\bar{o}^t) \tag{5}$$

Following the calculation of the gate activations the new cell state is computed and given as:

$$c^t = z^t \odot i^t + c^{t-1} \odot f^t \tag{6}$$

Finally, the output of the LSTM cell at a time t is,

$$y^t = tanh(c^t) \odot o^t \tag{7}$$

Where again, the *tanh* function may be substituted with any pointwise non-linear function. In our implementation we add a *softmax* layer which takes the output vector y^t as an input and places a distribution on it so it can easily be interpreted for a classification task.

The backpropagation of error through a recurrent network differs from a static network in that the error gradient at the future time steps impacts the error gradient at past time steps in the sequence. The standard backpropagation algorithm can be modified to the backpropagation through time (BPTT) algorithm and can subsequently be applied to the LSTM structure. The BPTT algorithm is outlined in the following as it applies to the LSTM network defined above. The error gradient passed down from the layer above will be represented using Δ^t. At the upper most LSTM layer this corresponds to $\frac{\delta L}{\delta y^t}$ where L is the loss function being used to minimize the error of the network. Thus, the error gradient with respect to the LSTM output vector within the LSTM block is defined as:

$$\delta y^t = \Delta^t + R_g^T \delta g^{t+1} + R_i^T \delta i^{t+1} + R_f^T \delta f^{t+1} + R_o^T \delta o^{t+1} \tag{8}$$

The error gradient with respect to the output gate can subsequently be computed with:

$$\delta o^t = \delta y^t \odot h(c^t) \odot \sigma'(\bar{o}^t) \tag{9}$$

Where h is the pointwise non-linear function used in the network's forward propagation, in our case, *tanh*. The error gradient with respect to the cell state in the LSTM is given as:

$$\delta c^t = \delta y^t \odot o^t \odot h'(c^t) \odot + \delta c^{t+1} \odot f^{t+1} \tag{10}$$

The error gradients with respect to the forget, input, and block input gates are given below, respectively:

$$\delta f^t = \delta c^t \odot c^{t-1} \odot \sigma'(\bar{f}^t) \tag{11}$$

$$\delta i^t = \delta c^t \odot z^t \odot \sigma'(\bar{i}^t) \tag{12}$$

$$\delta g^t = \delta c^t \odot i^t \odot h'(\bar{g}^t) \tag{13}$$

To compute the error gradient with respect to the gate weights, we find the outer products of the error gradient with respect to the gate and the input, and sum the over time steps. The equations below describe this process, and the $_*$ denotes any one of i, f, o, g:

$$\delta W_* = \sum_{t=0}^{T} \delta_*^t \otimes x^t \tag{14}$$

$$\delta R_* = \sum_{t=0}^{T-1} \delta_*^{t+1} \otimes y^t \tag{15}$$

$$\delta b_* = \sum_{t=0}^{T} \delta_*^t \tag{16}$$

The error gradients with respect to the weights computed above can be used to update the weights within the network and ultimately train the LSTM model. The updating of the model parameters often occurs through a gradient descent process. There are many adaptations and improvements to the standard gradient descent algorithm that can be used; however, in our implementation we use the NADAM algorithm. The NADAM algorithm, which is a combination of the Nesterov's Accelerated Gradient (NAG) algorithm and the Adaptive Moment Estimation (ADAM) algorithm, has been shown to achieve a lower training and validation loss compared to other learning algorithms in [30]. In the algorithm discussed below, $f_t(\phi)$ denotes the loss function at time step t parameterized by ϕ. The gradient of the loss function with respect to the parameters at time step t is given as g_t, and the learning rate is represented by η.

The NADAM algorithm, given in Algorithm 1, incorporates a momentum term, which accumulates a sum over previous gradients and then multiplies them by some constant decay factor (μ). A velocity term, n_t, is used which makes use of a sum of previous squared gradients multiplied by an exponentially decaying term, ν. This allows the resulting combined algorithm to make use of previous parameter updates, previous gradients, and squared gradients similar to the ADAM

Algorithm 1 NADAM

$g_t \leftarrow \nabla_{\phi_{t-1}} f(\phi_{t-1})$
$\hat{g} \leftarrow \frac{\hat{g}t}{1-\prod_{i=1}^{t}\mu_i}$
$m_t \leftarrow \mu\, m_{t-1} + (1-\mu) g_t$
$\hat{m}_t \leftarrow \frac{m_t}{1-\prod_{i=1}^{t}\mu_i}$
$n_t \leftarrow \nu\, n_{t-1} + (1-\nu) g_t^2$
$\hat{n}_t \leftarrow \frac{n_t}{1-\nu^t}$
$\overline{m}_t \leftarrow (1-\mu_t)\hat{g}_t + \mu_{t+1}\hat{m}_t$
$\phi_t \leftarrow \phi_{t-1} - \eta \frac{\overline{m}_t}{\sqrt{\hat{n}_t}+\varepsilon}$

algorithm. An initial momentum step is incorporated as well, gaining the advantage of a higher quality subsequent gradient step as seen in the NAG algorithm. The NADAM algorithm was used to update the weights of the network used in our implementation.

C. Experimentation platform

Data acquisition is an incremental part in experimentation process. Initial experiments were mainly focused on classifying data that was acquired from an online EEG signal database. Since, we have shifted our focus to acquiring EEG signals ourselves, using commercial hardware to record our own brain activity. The Ultracortex Mark IV EEG Headset was used in 8-sensor configuration to obtain samples. The Cyton Biosensing Board samples these EEG signals at 250Hz and relays the sampled data wirelessly back to the computer using the OpenBCI USB dongle which make use of RFDuino radio modules. The Cyton Biosensing Board is also capable of communicating wirelessly with any mobile device that is compatible with Bluetooth Low Energy. Once the data is transmitted to the computer it is then visualized by the user within the OpenBCI GUI. From the OpenBCI GUI, the signals are plotted to examine signals in many different ways including a time series, FFT plot, band power, and activity over a head plot. This allows us to examine and visualize data. We have observed the exemplary changes in brain wave amplitudes corresponding to the "not thinking" and the "mental calculation" processes. We noted an increase in wave coupling corresponding to the subjective "learning" category. Such data, if repeatable, are expected to be effective for training ANN. Monitoring 8-channels with 3-bands θ, α and γ and estimating 2-features (amplitude-change and coupling-change) appears to provide sufficient information for classification.

V. Conclusion

We defined classes and a framework for BOI pattern recognition in education setting, suitable for the tree- and ANN-classifiers. We observed exemplary discriminating changes in the wave bands (corresponding to three processes: not thinking, mental calculation, and learning) real-time on the GUI. We setup ANN and are developing training data set acquisition strategy. The BOI framework is ready for testing and expansion. The findings are encouraging but multiple identified challenges still need to be addressed.

References

[1] "An open source EEG related papers and tutorials," http://openeeg.sourceforge.net/doc/links.html.

[2] in *Quantitative EEG, Event-Related Potentials and Neurotherapy*. San Diego: Academic Press, 2009, pp. xxxi – lviii.

[3] E. Sokhadze, M. F. Casanova, E. L. Casanova, E. Lamina, D. Kelly, and I. Khachidze, "Event-related potentials (ERP) in cognitive neuroscience research and applications," *Neuroregulation*.

[4] W. Tatum, *Handbook of EEG Interpretation*. Demos Medical Publishing, 2007.

[5] T. Budzynski, *Introduction to quantitative EEG and neurofeedback. Advanced theory and applications.* Academic Press, 2009.

[6] L. Hirsch and R. Brenner, *Atlas of EEG in Critical Care*. Wiley, 2010.

[7] A. Anoukhin and F. Vogel, "EEG Alpha Rhythm Frequency and Intelligence in Normal Individuals." *Intelligence*, vol. 23, pp. 1–14, 1996.

[8] T. Boynton, "Applied research using alpha/theta training for enhancing creativity and well-being." *of Neurotherapy*, vol. 5.

[9] S. Palva and J. M. Palva, "New vistas for alpha-frequency band oscillations," *Trends in Neurosciences*, vol. 30, no. 4, 2007.

[10] G. Deco, V. Jirsa, A. McIntosh, O. Sporns, and R. Kotter, "Key role of coupling, delay and noise in resting brain fluctuations," *Proc National Academy of Sciences*.

[11] A. Wise, *The High-Performance Mind: Mastering Brainwaves for Insight, Healing, and Creativity*. TarcherPerigee, 1996.

[12] A. Torta, R. Komorowskie, J. Mannsf, N. Kopellc, and H. Eichenbaume, "Theta-gamma coupling increases during the learning of item-context associations," *PNAS*, vol. 106, no. 49, pp. 20 942–20 947, 2009.

[13] M. Foks, "Neurofeedback training as an educational intervention in a school setting," *Educational and Child Psychology*, vol. 22, no. 3, pp. 67–77, 2005.

[14] "The future of ADHD treatment lies in drug-free BCI therapy," http://neurogadget.net/2012/12/18/the-future-of-adhd-treatment-lies-in-drug-free-bci-therapy/6539.

[15] Y. Wang and T.-P. Jung, *Improving Brain–Computer Interfaces Using Independent Component Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 67–83.

[16] "BCILAB - SCCN," https://github.com/sccn/BCILAB.

[17] "OpenBCI.Org," http://docs.openbci.com/Tutorials/01-Cyton_Getting%20Started_Guide.

[18] "BNCI Horizons 2020," http://bnci-horizon-2020.eu/database/data-sets.

[19] L. Colgin, T. Denninger, M. Fyhn, T. Hafting, T. Bonnevie, O. Jensen, M.-B. Moser, and E. Moser, "Frequency of gamma oscillations routes flow of information in the hippocampus," *Nature*, vol. 462, no. 19, 2009.

[20] G. Sridhar and P. Rao, "A Neural Network Approach for EEG classification in BCI," *International Journal of Computer Science and Telecommunications*, vol. 3, no. 10, pp. 44–48, 2012.

[21] T. Raduntz, J. Scouten, O. Hochmuth, and B. Meffert, "EEG artifact elimination by extraction of ICA-component features using image processing algorithms ," *Journal of Neuroscience Methods*, vol. 243.

[22] P. Sarma, P. Tripathi, M. P. Sarma, and K. K. Sarma, "Pre-processing and Feature Extraction Techniques for EEGBCI Applications," *ADBU Journ. of Engineering Tech.*, vol. 5, no. 1, 2016.

[23] A. Barreto, A. Taberner, and L. Vicente, "Classification of spatio-temporal EEG readiness potentials towards the development of a brain-computer interface," *IEEE Proceedings of the Southeastcon '96. Bringing Together Education, Science and Technology.*, 1996.

[24] R. Palaniappan, R. Paramesran, S. Nishida, and N. Saiwaki, "A new brain-computer interface design using fuzzy ARTMAP," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 3, pp. 140–148, 2002.

[25] S. Foulke, J. Jagannath, A. Drozd, T. Wimalajeewa, P. Varshney, and W. Su, "Multisensor Modulation Classification (MMC): Implementation Considerations – USRP Case Study," in *Proc. of IEEE Military Communications Conference (MILCOM)*, Oct 2014.

[26] J. Jagannath, D. O'Connor, N. Polosky, B. Sheaffer, L. N. Theagarajan, S. Foulke, P. K. Varshney, and S. P. Reichhart, "Design and Evaluation of Hierarchical Hybrid Automatic Modulation Classifier using Software Defined Radios," in *Proc. of IEEE Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, Jan 2017.

[27] J. Jagannath, N. Polosky, D. O'Connor, L. Theagarajan, B. Sheaffer, S. Foulke, and P. K. Varshney, "Artificial neural network based automatic modulation classification over a software defined radio testbed," in *Proc. of IEEE Intl. Conf. on Communications (ICC)*, 2018 [under review].

[28] B. G. Horne and C. L. Giles, "An experimental comparison of recurrent neural networks," *Neural Information Processing Systems*.

[29] A. Arnsten, "The biology of being frazzled," *Science*, vol. 280, pp. 1711–1713, 1998.

[30] T. Dozat, "Incorporating nesterov momentum into adam," in *Proc. of International Conference on Learning Representations*, San Juan, Puerto Rico, Feb 2016.