32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

Spectral Spatio-Temporal Template Extraction from EEG Signals

Sarah Ostadabbas and Roozbeh Jafari
Embedded Systems and Signal Processing Lab, Department of Electrical Engineering
University of Texas at Dallas, Richardson, TX, 75080
Email: { sxo081000, rjafari}@utdallas.edu

***Abstract*—Analysis of Event Related Potentials (ERPs) produced by brain activities can provide insight into the timing of underlying brain function. ERPs can be classified by their time/frequency characteristics and spatial location on the scalp. Traditionally, ERPs are manually located by temporally and spatially averaged EEG signals. This process is error prone and sensitive to *a priori* assumptions. Our proposed algorithm is a general neuroscience-focused data mining algorithm that performs time and frequency analysis on ERPs and automatically extracts templates corresponding to Spectral Spatio-Temporal (SST) regions exhibiting significant differences between experimental outcomes. The method uses time-aligned templates, which preserve the characteristics of the signal important to cognitive researchers. The ability of the selected signal templates to differentiate between stimulus responses has been verified using a pattern recognition procedure. SST template extraction is tested on data taken from a Go/NoGo task and shown to both find relationships consistent with published neuroscience literature as well as novel relationships.**

I. Introduction

The electrical activities of brain can be recorded by means of Electroencephalography (EEG) from the scalp. Extracting informative and discriminative features from these EEG signals is a crucial step in documenting and classifying patterns of brain activity. Some brain activities produce characteristic EEG signals, called Event Related Potentials (ERP), in reaction to an internal or external stimulus [1]. Stimulus synchronous averaging [2] emphasizes the event-related response by raising the inherently low SNR of EEG signals.

Analyzing EEG signals to find meaningful patterns is the subject of ongoing research. Traditionally, one method of interpreting EEG data is measuring peaks and their timing in ERP averages. These peaks are identified by visual inspection of averaged EEG. Then, some standard statistical techniques, such as ANOVA, are applied to validate the observations across segments of EEG data [3]. However, visual inspection of EEG signals to find the important ERP patterns is sensitive to *a priori* assumptions and ignores trial-to-trial variabilities. Recently, techniques from the signal processing community have been used for EEG pattern recognition. Two primary application areas are signal analysis and data visualization.

Signal analysis techniques, such as single-sweep analysis, automatically locate statistically significant patterns in EEG data using signal processing and pattern recognition methodologies. Single-sweep analysis deals with the estimation and parameterization of the event-related response in a single trial for the investigation of the ERP variabilities. Various algorithms for ERP single-sweep analysis have been reported in the literature [4].

Data visualization tools help researchers to explore datasets by creating visual representation of underlying relationships. Authors in [5] studied individual trials with a custom visualization tool, and then used ICA to separate ERPs, Non-event related potentials, and artifacts. In [6], authors have developed a toolbox and graphical user interface, EEGLAB, for processing and visualization of collections of single-trials from any number of channels.

These signal processing methods create an alternate representation of EEG data. However, many analysis techniques in the neuroscience community depend on a signal-based representation. Also, studies show that ERPs appear in specific spatial regions of the brain at specific times and may be most prominent in certain frequency bands [7]. In the currently available single-sweep analysis methods, either the spatial or temporal regions of interest must be known in advance. The visualization tools offer an efficient and tailored method of exploring the dataset, but locating regions containing discriminative ERPs is still a time consuming process.

We address these issues by proposing a Spectral Spatio-Temporal (SST) template extraction algorithm that automatically searches for SST regions exhibiting significant differences among experimental outcomes. To the best of our knowledge, no other research automatically finds a discriminative time-domain representation of ERPs without *a priori* knowledge of regions of interest. Our method is further distinguished by time-aligned templates used to preserve important characteristics of the signal. The ability of the selected signal templates to differentiate between stimulus responses is verified using a pattern recognition procedure.

II. Validation Dataset

The proposed algorithm is a general neuroscience-focused data mining algorithm to extract templates for finding spectral temporal-spatial differences between sets of EEG signals. We validate this algorithm with data taken from the standard Go/NoGo task [8] in which the subject is instructed to push a button if shown Go, or to do nothing if shown NoGo. Stimuli are presented for 300 ms followed by 1700 ms of blank screen. The Go/NoGo stimuli are shown in Fig. 1a.

978-1-4244-4124-2/10/$25.00 ©2010 IEEE

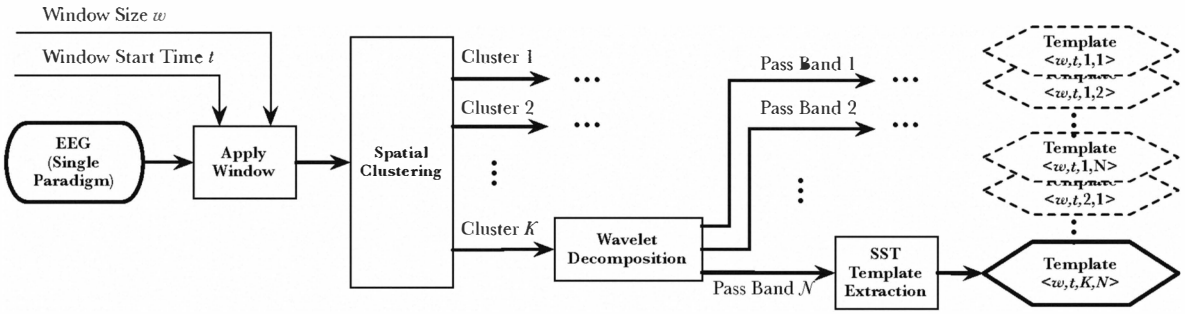


Fig. 3: SST template extraction algorithm

A. *Data Collection and Preprocessing*

Continuous EEG was recorded using a 64-electrode Neuroscan Quickcap (shown in Fig. 1b) with a Neuroscan amplifier and Scan 4.3.2 software sampled at 1 kHz. Fifteen subjects between the age of 18 and 31 completed the Go/NoGo task with 20 trials for the Go outcome and 20 trials for the NoGo outcome. The average of all Go and all NoGo trials for the Fz electrode is shown in Fig. 2.

B. *Hypothetical Results*

Neuroscientists have found two significant ERP components elicited in the NoGo condition of a Go/Nogo task related to response inhibition: NoGo-N2 and NoGo-P3. NoGo-N2 is a negative deflection constrained to the frontal scalp locations. The NoGo-P3 is a large positive deflection and is present in frontal, central, and parietal regions [8]. Both ERPs are shown in Fig. 2.

III. SST Template Extraction Algorithm

The proposed algorithm extracts representative signals, called templates, from each stimulus response for a given spectral spatio-temporal (SST) region. This collection of templates is called a template set. The templates are time-aligned averaged signals of all EEG template extraction trials. A template selection algorithm based on AdaBoost selects the most discriminative template sets from a list of semi-exhaustively extracted SST regions.

In order to cover all spectral, spatial and temporal regions of the EEG signals, we propose an algorithm to generate different SST regions and then extract template sets from each region as shown in Fig. 3. In the time domain, the EEG signals are segmented by sliding rectangular windows of various duration over the signal with small step size. In the spatial domain, the EEG is recorded from 64 electrodes with different spatial location over the scalp. The electrodes are grouped together to study the effect of different spatial regions. There are 2^{64} different combinations of electrodes. Making an exhaustive enumeration of spatial regions is impractical. Instead, a clustering algorithm is used to group electrodes into mutually exclusive clusters as explained in Section III-B.

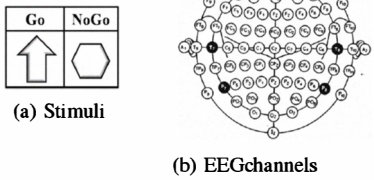
The signals in each cluster are averaged together to increase SNR. Then, a Discrete Time Wavelet (DTW) is applied to the averaged signal to decompose it into different frequency bands. Section III-C explains the frequency analysis algorithm. The output of several wavelet levels feeds to the SST template extraction part that is explained in Section III-A.

A test trial can be labeled as an experimental outcome by finding closest matching template from the template set. The most discriminative template sets extracted from different SST regions can be selected in order of importance by using AdaBoost. AdaBoost selects templates in a way that minimize information overlap. The procedure to find the most discriminative template sets is explained in Section III-D.

The EEG signals from 64 electrodes are split into the training dataset and the test dataset. The training dataset is used for SST template extraction and AdaBoost training and the test dataset is used for validation of the results. The classification accuracy of selected SST templates on test dataset is demonstrated in Section IV.

A. *Temporal Analysis*

EEG signals have low SNR, therefore SST templates are created using an averaging technique because averaging de-



(a) Stimuli

(b) EEGchannels

Fig. 1: (a) Standard Go/NoGo task stimuli, (b) Modified 10–20 standard EEG electrode position

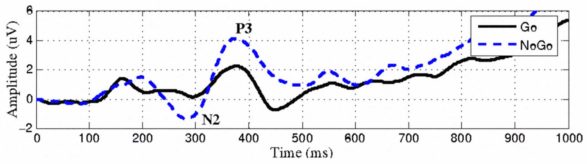


Fig. 2: EEG grand mean for all subjects on the Fz channel (bandpass filtered from 0.5 Hz to 30 Hz)

creases uncorrelated noises. Fig. 4 shows the full SST template extraction technique for a single outcome. Each template in the template set is created in an identical manner. First, SST windowed EEG signals are extracted as described in Fig. 3. Next, an initial template is created by averaging the entire training set. Then each signal is time-aligned to the template and averaged once more to obtain the final template. This is done for each experimental outcome in a SST region to create a template set.

1) Time Alignment: It is well recognized that ERPs are time-varying signals reflecting the averaged time courses of underlying neural events during cognitive processing. The cognitive processing and response mechanisms are reflected by distinct components that depended on a subject's psychological state and may exhibit on different time onsets in different trials.

Time-alignment can compensate for these offsets and is critical to template matching, as misalignment can reduce the apparent similarity between the template and the signals. Signals are aligned by maximizing a scoring function over a set of possible time offsets. Each template is expected to match in the time interval equals to 20% of template window size.

2) Template Matching: With template matching, a similarity measure is used to compare the template to a portion of the source signal. A common similarity measure is the Euclidean distance between the signal and template. The Euclidean distance is computed between the template, $\mathcal{T}$, and a portion of the signal, f, equal in duration, starting at time τ.

Matching the signal shape is more important than matching exact amplitudes, which vary due to subject differences, quality of the electrode contacts, etc. Normalizing the template and signal vector to zero-mean unit vectors constrains the matching to the surface of a hypersphere of unit radius. This constrained matching is more shape-oriented than the simple Euclidean distance. We will show that this measure produces equivalent ranking to the standard normalized cross-correlation (NCC) measure, γ described by Lewis [9].

Theorem 1. *The normalized Euclidean distance produces measures with equivalent ordering to the normalized cross correlation.*

Proof: f_n *and* $\mathcal{T}_n$ *refer to the normalized versions of* f *and* $\mathcal{T}$.

$$
\begin{aligned}
& d_0(\mathcal{T}_n, f_n, \tau) \quad \textit{Normalized Euclidean distance} \\
& \cong d_0^2(\mathcal{T}_n, f_n, \tau) \quad \textit{Equivalent with respect to ordering} \\
& = \sum_t [f_n(t+\tau) - \mathcal{T}_n(t)]^2 \\
& = \sum_t \left[f_n^2(t+\tau) - 2 f_n(t+\tau)\mathcal{T}_n(t) + \mathcal{T}_n^2(t) \right]
\end{aligned}
$$

The next step follows because f_n *and* $\mathcal{T}_n$ *are normalized.*

$$
\begin{aligned}
& = 2 - 2 \sum_t [f_n(t+\tau)\mathcal{T}_n(t)] \\
& \cong \sum_t f_n(t+\tau)\mathcal{T}_n(t) \quad \textit{Reversed order equivalent} \\
& = \frac{\sum_t [f(t+\tau) - \bar{f}_t][\mathcal{T}(t) - \bar{\mathcal{T}}_t]}{\sqrt{\sum_t [f(t+\tau) - \bar{f}_t]^2 \sum_t [\mathcal{T}(t) - \bar{\mathcal{T}}_t]^2}} \\
& = \gamma(\mathcal{T}, f, \tau) \quad \textit{Normalized cross-correlation}
\end{aligned}
\tag{1}
$$

■

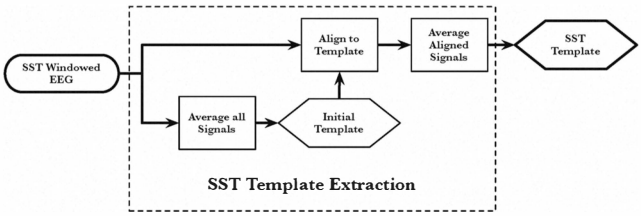


Fig. 4: Procedure for generating SST templates

B. Spatial Clustering

Electrodes in a given region often observe similar responses. The temporally windowed EEG signals are bandpass filtered from 0.5 Hz to 100 Hz and for each electrode are averaged on all trials. Averaging EEG signals recorded from these electrodes can increase SNR. Hierarchical agglomerative clustering [10] is used to group electrodes based on similarity of observed signals. The clustering initially assigns each electrode to a unique cluster. In each iteration, the closest clusters are combined until a stopping criterion is reached.

We used a complete linkage inter-cluster distance measure. The distance between clusters A and B is defined as:

$$
L(A, B) = \max_{i,j} dist(e_i, e_j) \quad i \in A, j \in B \tag{2}
$$

where $dist(e_i, e_j)$ is the distance between electrodes e_i and e_j and $\gamma(e_i, e_j, 0)$ is the normalized cross correlation function that is defined in (1) without any time-alignment.

$$
dist(e_i, e_j) = 1 - \gamma(e_i, e_j, 0) \tag{3}
$$

The termination criterion used in this experiment was $L = 0.8$. All the EEG electrodes that belong to a cluster are averaged together as input for the spectral analysis.

C. Spectral Alalysis

Typically, ERP analysis is performed in the time domain, where the amplitudes and latencies of prominent peaks in the averaged potentials are measured and correlated using information processing mechanisms. However, analysis in the frequency domain has revealed that EEG/ERP frequency components in different frequency ranges (delta, theta, alpha, beta, gamma) are functionally related to information processing and behavior [11].

The wavelet transform (WT) is an efficient time-frequency decomposition method. For discrete time signals like EEG, the Discrete Wavelet Transform (DWT) is found to yield a fast computation of the Wavelet Transform. The DWT is computed by successive lowpass and highpass filtering of the discrete-time signal. We used eight levels of Daubechies wavelet order 3 that has fewer sharp edges to take advantage of the similarity

Algorithm 1 Generating semi-exhaustively SST template sets

```
for w = W_min to W_max by W_s do {Select window sizes}
  for t = T_min to T_max − w by T_s do {Select start times}
    for f = F_min to F_max by 2^i, i ∈ {2, ..., 7} do {Select frequency bands}
      for all R_i ∈ ℜ do {Each spatial cluster}
        T_{w,t,f,i} = ∅
        for all A_j ∈ A do {For each experimental outcome}
          T_{w,t,f,i,j} = createTemplate(w, t, f, R_i A_j, data_train)
          T_{w,t,f,i} ← T_{w,t,f,i} ∪ {T_{w,t,f,i,j}}
        end for
        acc_{i,w,t} = test(T_{w,t,f,i}, data_test)
      end for
    end for
  end for
end for
```

between the typical shape of ERPs and the wavelet function. To preserve the signal shape and length, we omitted the decimation step for each filter level from the standard wavelet transformation procedure.

D. Locating the Most Discriminative Template Sets

Finding the most discriminative template sets is a two-step process. The first step is creating a large group of template sets from various SST regions. The second step is choosing the most relevant SST regions.. The original list of template sets is created semi-exhaustively by the procedure in Algorithm 1. For our study, the eight frequency bands start at $F_{max} = 512\,\text{Hz}$ then go to $F_{min} = 0\,\text{Hz}$ in step of $2^i\,\text{Hz}$, $i \in \{2, ..., 8\}$. Window sizes range from $W_{min} = 100\,\text{ms}$ to $W_{max} = 400\,\text{ms}$ in steps of $W_s = 50\,\text{ms}$. Times range from $T_{min} = 0\,\text{ms}$ to $T_{max} = 1000\,\text{ms}$ in steps of $T_s = 10\,\text{ms}$. It is important to try different window sizes: too small a window may miss relevant details, while too large a window may contain irrelevant or inconsistent signals.

Choosing the most discriminative SST regions is achieved by applying the AdaBoost algorithm on the template sets. The template matching procedure for each template set is considered as a weak classifier. A weak classifier labels a trial as one of the experimental outcomes based on the similarity measure between the trial and the templates of that outcome. In the Go/NoGo experiment, we label a given trial, $\mathfrak{f}$, **1** if it is "Go" and **-1** if it is "NoGo":

$$label(\mathfrak{f}) = sign\Big(score_{Go}(\mathfrak{f}) - score_{NoGo}(\mathfrak{f})\Big) \qquad (4)$$

For each experimental outcome, A_i, $score_{A_i}(\cdot)$ is defined as:

$$score_{A_i}(\mathfrak{f}) = \gamma(\mathcal{T}_{A_i}, \mathfrak{f}, \tau_i), \quad \tau_i \in [-R, R] \qquad (5)$$

where $\gamma(\cdot)$ is the normalized cross correlation function between the trial and the template of that outcome with time offset τ_i. We consider $R = 0.2 \times ||\mathcal{T}_{A_i}||_0$, that means the maximum time shifting is equal to 20% of template length.

AdaBoost is applied on all weak classifiers to locate the most discriminative template sets in training portion of data set. The details of AdaBoost algorithm are described by Schapire and Freund [12].

IV. Experimental Results

For template extraction, the trials from the Go/NoGo experiment were split into training and test sets. The training set contains 15 Go and 15 NoGo trials per subject. The test set contains 5 Go and 5 NoGo trials per subject. The template sets were built based on the SST template extraction algorithm described in Section III. The most discriminative template sets were selected with AdaBoost as described in Section III-D.

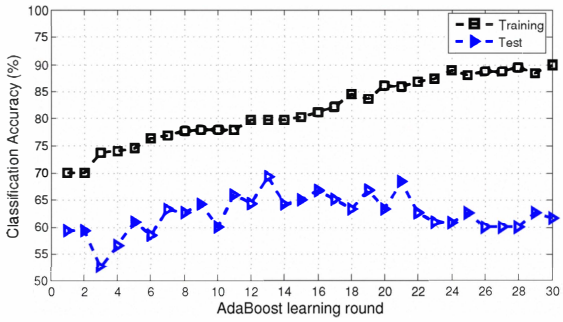


Fig. 5: AdaBoost classification accuracy at different rounds on training and test dataset

A. General SST Template Extraction Results

Table I shows the selected templates with their spectral and temporal information that were selected by AdaBoost in the first fifteen rounds. As described in Section III-D, AdaBoost will greedily select template sets based on performance on the training set. The "Acc trn each" shows the classification accuracy of individual template set on training set in each round and "Acc trn Ada" shows the overall AdaBoost performance after that round.

We choose to terminate after round $r = 15$ to avoid overfitting. In Table I, the "Acc tst each" and "Acc tst Ada" show the classification accuracy of individual template set on the test set in each round and the overall AdaBoost performance after that round, respectively. Also, Fig. 5 shows the learning rate of AdaBoost on the training data as well as

TABLE I: Selected templates by using AdaBoost

Rnd #	Window (ms)	Start (ms)	Frq. (Hz)	Acc trn each	Acc tst each	Acc trn Ada.	Acc tst Ada.
1	300	161	4 - 8	70%	59%	70%	59%
2	350	21	0 - 4	69%	63%	70%	60%
3	300	1	8 - 16	57%	53%	73%	52%
4	200	101	8 - 16	62%	54%	74%	57%
5	350	381	8 - 16	60%	57%	74%	60%
6	300	561	0 - 4	63%	59%	76%	58%
7	150	101	4 - 8	59%	60%	77%	63%
8	200	61	0 - 4	57%	56%	78%	62%
9	350	81	4 - 8	59%	53%	78%	64%
10	350	361	16 - 32	61%	58%	78%	60%
11	250	621	4 - 8	54%	53%	78%	66%
12	400	461	4 - 8	61%	57%	79%	64%
13	100	521	8 - 16	57%	58%	79%	70%
14	100	101	4 - 8	62%	57%	80%	64%
15	350	61	16 - 32	55%	54%	80%	65%

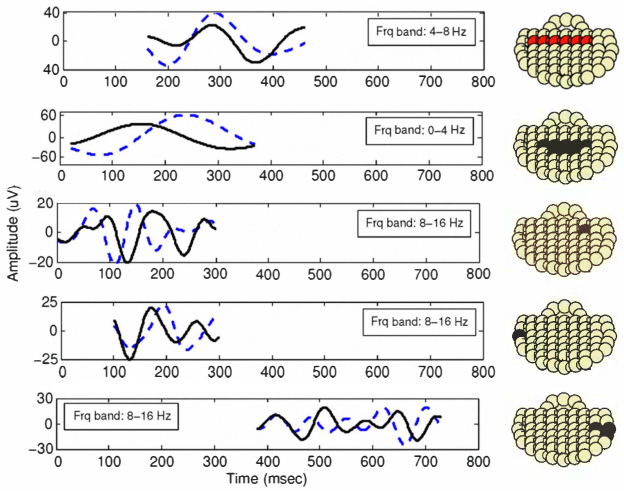


Fig. 6: The most discriminative template sets selected by AdaBoost (dash: NoGo, solid: Go)

corresponded classification accuracy on test data. The final classification accuracy after round $r = 15$ is 65%. The statistical significance of this result ($\bar{x} = 0.65$, $n = 60$) was verified using the Student's t test. This accuracy cannot be attributed to chance alone for our test set ($p = 0.008$).

It can be seen from Fig. 5, the maximum classification accuracy in the test dataset occurs after round $r = 13$. After this round, because of overfitting, the performance on the training dataset still increases while the performance on test dataset levels off until $r = 21$, and after that becomes worse.

B. Comparison to Hypothetical Results

The first five most discriminative template sets for Go and NoGo and their spatial regions are shown in Fig. 6. Recall from Section II-B, N2 is strongest over the frontal region, and P3 is strong for most other locations. The presence or absence of N2/P3 can be ascertained by visually inspecting Fig. 6 and comparing the template window location to the temporal locations specified for N2 and P3 in Fig. 2. There is a large difference between amplitudes of Fig. 2 and Fig. 6. The grand mean is produced from an average of signals which have not been aligned, and therefore have significantly less coherence than aligned signals. Less coherence results in lower amplitudes.

The first selected template set that is shown in the first row of Fig. 6 is in the Fronto-Center (FC) region and clearly shows N2 and P3 in Theta band ($4 - 8$ Hz). The second template set almost covers the Central (C) region. The NoGo-P3 and the Go-P3 in this template appear in Delta band (up to 4 Hz) and the NoGo-P3 has larger amplitude in comparison to Go-P3. The next three template sets in Fig. 6 are in a higher frequency band ($8 - 16$ Hz). This frequency band contains the whole Alpha band ($8 - 12$ Hz) and early part of Beta band ($12 - 30$ Hz). These template sets do not contain N2 or P3 but as Fig. 5 shows, their contribution in recognition results in increasing classification accuracy.

These results are consistent with the hypothesis, and further show one of the strengths of proposed template extraction algorithm for finding unexpected regions of interest. When visually inspecting the grand mean, it is tempting to choose places where the grand mean shows the most significant variation between outcomes, such as at P3 and N2. However, sometimes smaller variations are more consistent between subjects and trials. Our method is capable of locating such ERPs.

V. Conclusion and Future Work

In this paper, we introduced a signal-based template extraction algorithm to datamine relationships in EEG signals. The proposed method generates templates from each stimulus response for different time windows, spatial locations and frequency bands by using a matching algorithm based on normalized cross correlation. A template selection algorithm based on AdaBoost selects the most discriminative template sets from a list of semi-exhaustively extracted SST regions for the recognition purpose. Experiments on data from a Go/NoGo task show that it not only highlights known relationships, but also picks up differences that have significant discriminating power, even when they appear less pronounced on the grand mean. The next step in this research could be using data mining and regression technique to find a predictive equation for an independent value observed for each trial.

References

[1] G. McCarthy and C. Wood, "Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models." *Electroencephalography and clinical Neurophysiology*, vol. 62, no. 3, p. 203, 1985.

[2] G. Dawson, "A summation technique for the detection of small evoked potentials." *Electroencephalography & Clinical Neurophysiology*, vol. 6, pp. 65–84, 1954.

[3] M. Falkenstein, N. Koshlykova, V. Kiroj, J. Hoormann, and J. Hohnsbein, "Late ERP components in visual and auditory Go/Nogo tasks." *Electroencephalography and clinical neurophysiology*, vol. 96, no. 1, p. 36, 1995.

[4] J. Yordanova and V. Kolev, "Single-sweep analysis of the theta frequency band during an auditory oddball task," *Psychophysiology*, vol. 35, no. 01, pp. 116–126, 2001.

[5] T. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. Sejnowski, "Analysis and visualization of single-trial event-related potentials," *Human Brain Mapping*, vol. 14, no. 3, pp. 166–185, 2001.

[6] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[7] K. Mann, P. Backer, and J. Roschke, "Dynamical properties of the sleep EEG in different frequency bands," *International Journal of Neuroscience*, vol. 73, no. 3-4, pp. 161–169, 1993.

[8] M. Falkenstein, J. Hoormann, and J. Hohnsbein, "ERP components in Go/Nogo tasks and their relation to inhibition," *Acta Psychologica*, vol. 101, no. 2-3, pp. 267–291, 1999.

[9] J. Lewis, "Fast normalized cross-correlation," in *Vision Interface*, vol. 10, 1995, pp. 120–123.

[10] S. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.

[11] A. von Stein and J. Sarnthein, "Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization," *International Journal of Psychophysiology*, vol. 38, no. 3, pp. 301–313, 2000.

[12] Y. Schapire and Y. Freund, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, no. 5, pp. 771–780, 1999.