Feature Extraction and Classification of EEG Spectra of Alcoholic Subjects

Paramita Guha
Scientist
CSIR-Central Scientific Instruments Org., Chandigarh, India

Sugandh Jain
Project fellow
CSIR-Central Scientific Instruments Org., Chandigarh, India

Sunita Mishra
Principal Scientist
CSIR-Central Scientific Instruments Org., Chandigarh, India

ABSTRACT
This paper considers the modeling and simulation techniques of electroencephalography (EEG) signals. EEG signals of two different categories of subjects viz., alcoholic and normal patients are considered here. The signals are decomposed into several components using discrete wavelet transform technique to achieve different frequency bands of the brainwaves. After that different classification techniques, like, Principle Component Analysis (PCA) and Partial Least Square (PLS) to distinguish the alcoholic signals from the normal subjects. A comparative analysis is given and also further extensions are identified.

Keywords
EEG signals, PCA, PLS, classification

1. INTRODUCTION
Nervous system is the most important part of a human body. Hence, the monitoring the activities of this is necessary. There are several methods involved for this purpose, like, electroencephalography (EEG), Magnetic resonance imaging (MRI), Computed Tomography (CT), Single Photon Emission Computed Tomography (SPECT), Positron emission tomography (PET) etc. Among all these methods, EEG is quite straight forward and accurately detects the changes in brain activities. It is a tool for exploring the activities of brain functioning and neural activity. The signals are measured using electrodes placed on the scalp, which record the electrical field generated by the nerve cells. The main advantage of this technique is that it can accurately detect the brain's activity with a very high resolution (a single millisecond or less). This is a non-invasive technique which allows the investigators to get a clear access to a human brain. The EEG signal is measure as voltage changes at the scalp. The changes in voltage may occur due to several reasons like, eye movements, change in behavior, emotions, motor and memory functions [1] etc.

After applying several preprocessing techniques [2], the main challenge remains in an EEG signal is to extract the features so that a suitable classification technique can be applied. Depending upon the frequency ranges, an EEG waveform can be subdivided into several bandwidths e.g. alpha, beta, theta, delta. If the frequency of a waveform is less than 4 Hz, it is known as Delta; for frequency between 4-7 Hz, the waveform is known as Theta. The waveform having frequency between 8-15 Hz is termed as Alpha whereas Beta waveform has frequency of 16-30 Hz. The alpha and beta waveforms are quite important as they show the inhibition control, active thinking, focus in adults [3].

In this paper the EEG data for normal and alcoholic subjects are collected from an open source database [4]. The data from different position of electrodes e.g. C_3, C_4 are considered. After this, data are normalized and noises are removed. A discrete wavelet transform is applied to the processed data and they are classified into several suitable waves like, Alpha, Beta and Theta. Classification algorithms like, Principle Component Analysis (PCA) and Partial Least Square (PLS) are applied on this data and a comparative analysis is made. The paper is organized in the following way. In the next section, the methods of data collection and classification tools are discussed in detail. In Section III, numerical results along with discussion are given and finally the paper is concluded in the subsequent section.

2. MATERIALS & METHODS
2.1 Data Collection & Preprocessing
The data were collected from an open source database [4], [5]. There are two groups of subjects: alcoholic and control. Each subject was exposed to a single stimulus (S1). There were 122 subjects and each subject completed 120 trials where different stimuli were shown. The electrode positions were located at standard sites. After collecting the data, noises are removed using weighted averaging filter technique [2]. The EEG data were now analyzed with wavelet decomposition technique to obtain the alpha, beta and theta waveform which are considered as *features* for further study. After obtaining the required features, PCA, PLS techniques are applied to classify the alcoholic and control data.

2.2 Wavelet Decomposition
Wavelet transform technique is widely used in the field of telecommunications, biology, power system etc. The main advantage of this method is that it can analyze non-stationary signals while preserving their temporal locality for which it is mostly replacing Fourier transform. It decomposes a signal into several components at different resolutions. The wavelet transform involves decomposing the signal into a number of translated and dilated wavelets. A mother wavelet is selected such that it can be translated, dilated and convolve with the function of interest. In this paper, the mother wavelet is given as [5].

$$\Psi_{a.b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \qquad 1)$$

where a is the scaling parameter and b is the shifting parameter. Using this function, the wavelet function can be given as

$$W_f(a,b) = \int_{-\infty}^{\infty} f(t)\Psi_{a,b}(t)dt \qquad 2)$$

2.3 Classification Tools

2.3.1 PCA

Principal component analysis (PCA) is considered as a basis for multivariate data analysis [6]. It can be used as a method which can take into account of only useful features.In other words, it reduces the number of variables (scores) and transforms the model into a low dimensional sub-space. Thus, a smaller number of variables (loadings) can be obtained which will have almost same properties as the original one and can be interpreted like spectra. The transformation can be written in the form of

$$X = T.\,P + E$$

with T is called the scores and is of same number of columns as the original matrix; P is the loadings and has as many rows as the original matrix.

2.3.2 PLS

Partial least square regression (PLS) [7] is another reduction technique which combines features from PCA and multiple regressions. It is useful when a set of dependent variables has to be predicted from a large set of independent variables. This algorithm exploits the information from both independent and dependent variables, in such a way that suitable regression coefficientscan be obtained. In this paper, both PCA and PLS are applied to distinguish the alcoholic data from the control data

3. NUMERICAL RESULTS & DISCUSSION

In this study, we used EEG signals of normal and alcoholic patients in order to perform a comparison between the PCA and PLS. EEG recordings were divided into sub-band frequencies such as alpha, beta and theta. The electrodes can be positioned at various places over the scalp. It has been observed from the literature that the EEG data of C3 and C4 locations are related to the electrical activity in somato-sensoric and motoric brain areas and movements. Hence, we have considered only those data. The scores plots of PCA are shown in Figs. 1- 6. From the score plots of alpha, beta and theta, EEG data of both the subjects can be easily distinguished.The output data has clearly been divided into two main groups in all the waveforms i.e. alpha, beta and theta. The first group, shown in the first quadrant, represents the data for normal patients and the other group of data is clustered in the second quadrant represent the EEG for alcoholic persons.

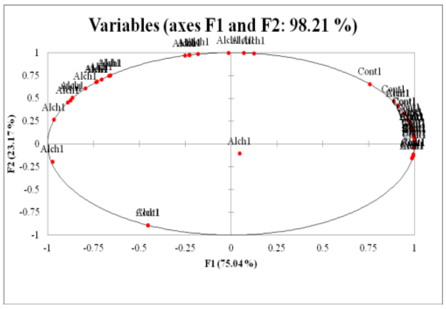

Figure 1: PCA Analysis of Alpha C_3

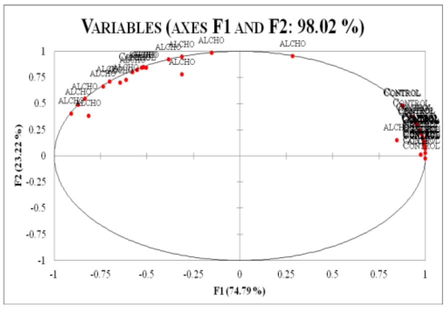

Figure 2: PCA Analysis of Alpha C_4

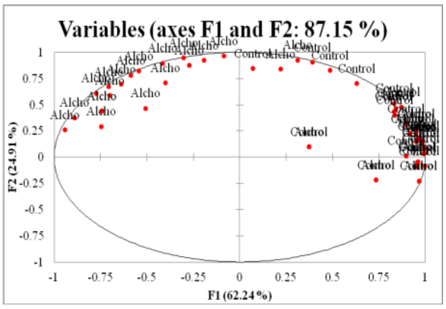

Figure 3: PCA Analysis of Beta C_3

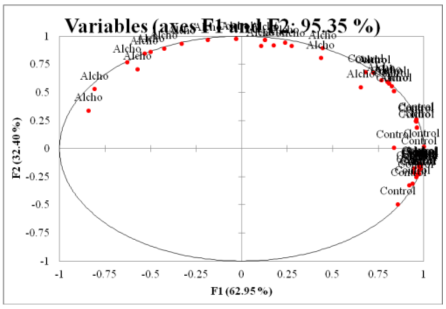

Figure 4: PCA Analysis of Beta C_4

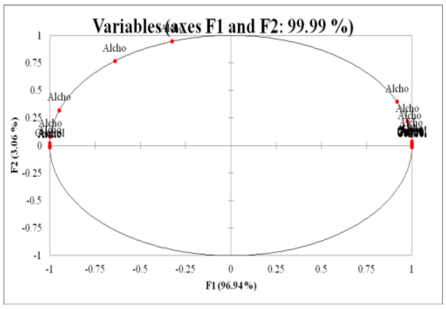

Figure 5: PCA Analysis of Theta C_3

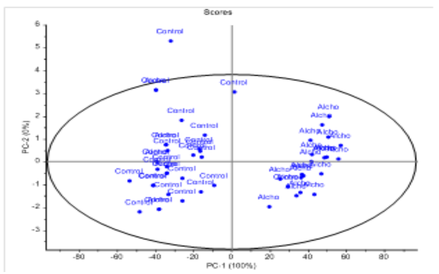

Figure 6: PCA Analysis of Theta C_4

After applying the PCA, PLS algorithm was applied to describe the relationship between the variables measured from the alcoholic and normal patients and shown in Figs. 7-12. The PLS model also divides the data into two groups as in the PCA. The first group includes all the alcoholic patients while the second group includes the normal patients for the EEG data.

Analysis of variance (ANOVA) was applied to determine the changes between alcoholic and normal patients. Results show that the alpha, beta and theta wave forms of EEG taken from C_3 and C_4 electrodes indicates significant difference between alcoholic and normal patients. The results of alpha C3 and C4 showed significant different ($p>0.05$). The same observation is for beta C_3 and C_4 but Theta C_3 and C_4 samples showed that there is no significant difference between the subjects.

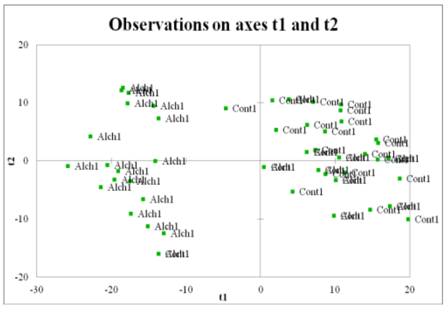

Figure 7: PLS Analysis of Alpha C_3

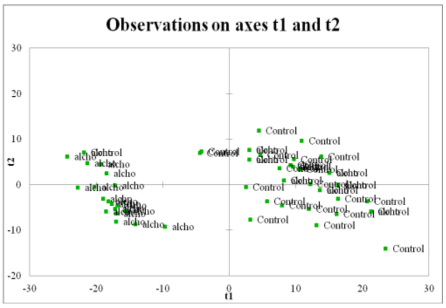

Figure 8: PLS Analysis of Alpha C_4

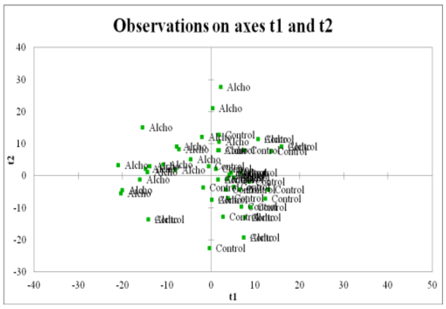

Figure 9: PLS Analysis of Beta C_3

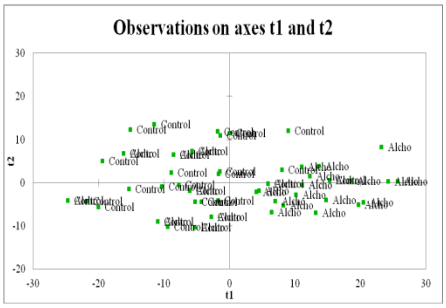

Figure 10: PLS Analysis of Beta C_4

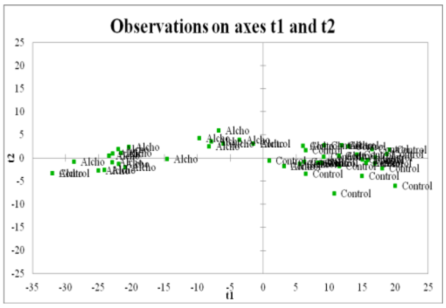

Figure 11: PLS Analysis of Theta C_3

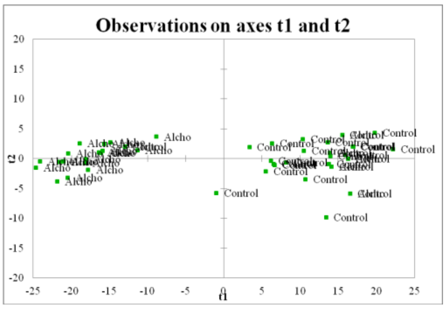

Figure 12: PLS Analysis of Theta C_4

4. CONCLUSION

In this paper, modeling and classification techniques for the EEG data are analyzed. The EEG signals for both normal and alcoholic subjects are considered. Different noises and artifacts are removed from the data. After that wavelet transform is applied to extract the frequency band from each data. It has been observed that alpha, beta and theta are the most important bands. Different classification techniques are used to differentiate the subjects based on the frequency bands. Here only PCA and PLS are used. It has been observed from the results, PCA technique can distinctly differentiate the alcoholic and normal subjects' EEG. The data have been reduced using PLS technique and it determines the regression coefficient of the reduced model in terms of the original model.

5. REFERENCES

[1] L. Behmer, L. Fournier, "Working memory modulates neural efficiency over motor components during a novel action planning task: An EEG study", Behavioural Brain Research, vol. 260, no. 1, pp. 1–7, 2014.

[2] M. Saeidi, A. Nazemipour, "Adaptive weighted averaging filter using the appropriate number of consecutive frames", International Journal of Electrical, Computer, Energetic, Electronic & Communication Engineering, vol. 4, no. 10, 2010.

[3] N. Liu, C. Chiang, H. Chu, "Recognizing the degree of human attention using EEG signals from mobile sensors", Sensors, vol. 13, 2013.

[4] *https://archive.ics.uci.edu/ml/datasets/EEG+Database*

[5] X. Zhang, H. Begleiter, B. Porjesz, W. Wang, A. Litke, "Event related potentials during object recognition tasks", Brain Research Bulletin, vol. 38, no. 6, pp. 531-538, 1995.

[6] Chesnutt, "Feature generation of EEG data using wavelet analysis", MSc Thesis, Texas Tech University, 2012.

[7] R. Kottaimalai, M. Rajasekaran, V. Selvam, B. Kannapiran, "EEG signal classification using principal component analysis with neural network in brain computer interface applications", International Conference on Emerging Trends in Computing, Communication and Nanotechnology, India, 25-26 March 2013.

[8] T. Kayikcioglu, M. Maleki, K. Eroglu, "Fast and accurate PLS-based classification of EEG sleep using single channel data", Expert Systems with Applications, vol. 42, no. 21, pp. 7825–7830, 2015.