Feature Extraction of Electroencephalogram (EEG) Signal – A Review

Wan Amirah W Azlan, Yin Fen Low

***Abstract*—This paper presents a review on signal analysis method for feature extraction of electroencephalogram (EEG) signal. It is an important aspect in signal processing as the result obtained will be used for signal classification. A good technique for feature extraction is necessary in order to achieve robust classification of signal. Considering several techniques have been implemented for extracting features in EEG signal, we only highlight the most commonly used for schizophrenia. The techniques are Hilbert-Huang transform, Principal Component Analysis, Independent Component Analysis and Local Discriminant Bases. Despite of their drawbacks, they can be applied which depends on the aim of a research, parameters and the data collected. Nevertheless, these techniques can be modified so that the new algorithm can overcome the shortcomings of original algorithm or algorithm beforehand. The modified Local Discriminant Bases algorithm is introduced in the present paper as another powerful adaptive feature extraction technique for EEG signal which is not reported elsewhere in investigating schizophrenia.**

Keywords–EEG signal, Hilbert-Huang Transform, Principal Component Analysis, Independent Component Analysis, Local Discriminant Bases

I. INTRODUCTION

Schizophrenia is a brain disorder which catches researchers' attention in understanding its nature. There are still many findings need to be explored and interpreted in terms of etiology, phenomenology, epidemiology and endophenotype of schizophrenia. Yet, numerous researches have been done to gain better understanding of schizophrenia. For example, the left hemisphere shown dysfunction or impairment upon schizophrenia [1], [2] was said to be related to the auditory processing deficits in patients, the risks of having schizophrenia in the patient's family [3], [4] and the relation of negative symptoms and positive symptoms [5], [6]. For schizophrenia patients, cognitive impairment is the core symptom. Up to this day, diagnosis of schizophrenia follows the 'Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition (DSM-IV)' published by the American Psychiatric Association. In 2013, DSM-IV is updated to DSM 5 where some changes have been made including changes in schizophrenia symptoms for a diagnostic approach [7]. Thus, an alternative method in diagnosing schizophrenia needs to be developed so that psychiatrists or doctors are able to distinguish schizophrenia with other psychiatric disorder. Regarding to this, feature extraction plays a crucial role to determine the features or information within EEG signals.

Research supported by ERGS/2013/FKEKK/TK02/UTEM/02/03 E00019.

Wan Amirah W Azlan is with the Faculty of Electronics and Computer Engineering, Universiti Teknikal Malaysia Melaka (e-mail:m021310043@student.utem.edu.my).

Yin Fen Low is with the Faculty of Electronics and Computer Engineering, Universiti Teknikal Malaysia Melaka (e-mail:yinfen@utem.edu.my)

Feature extraction is a process whereby the relevant information or characteristics from the signal is extracted so that the features can be easily interpreted. Therefore, it is a substantial process in interpreting an input signal. The information extract reflects the physiology and anatomy of the activity going on within the brain. It involved a number of variables in a large set of data, which require a large amount of memory or powerful algorithm to analyze the data. In this context, feature extraction method is needed in order to resolve these variables or information to be interpreted in a simple and accurate way.

The aim of the paper is to review the signal processing techniques which have been used to extract the features in EEG signal for schizophrenia. The concept, advantages and disadvantages for each of the techniques are being discussed. The algorithm of each technique will be presented to provide better understanding regarding the techniques. This paper consists of four sections; Section 1 gives an introduction on schizophrenia and feature extraction. Section 2 brief descriptions on the concept, pros and cons of each techniques and Section 3 shows the results and discussions. Section 4 is the conclusion.

II. FEATURE EXTRACTION TECHNIQUES

A. Hilbert-Huang Transform

The concept of Hilbert-Huang transform (HHT) is based on arbitrary continuous signal and return analytic signal from real data sequence. It will calculate the FFT of the input sequence, replace those FFT coefficients corresponding to negative frequencies with zeros and calculate the inverse FFT of the result. This technique is designed by Huang et al. (1998) [8]. It is able to determine the 'instantaneous' frequency and power of a signal and also able to analyze nonlinear and nonstationary signal [9]. Hilbert-Huang transform can retain the time information from peak time frequency analysis and possess linearity.

HHT compose of two – empirical mode decomposition (EMD) and Hilbert Transform. Both of them are the first and second stage in HHT. EMD acts in decomposing the data to components called Intrinsic Mode Functions (IMF). IMF represents the oscillation mode contained in the signal which has to be satisfied by two requirements: 1) in the overall data set, the number of extreme and zero crossings must be differ or equal at most by one. 2) at any point of the data, the mean value defined by both local maxima and minima should be zero. Procedure to extract IMF is known as sifting process. Below is the algorithm described in [8], [10].

ALGORITHM OF HHT

A) Empirical mode decomposition

1. Identify local extrema in the data set.
2. Local maxima is connected by a cubic spline line as the upper envelope.
3. The same procedure is used for local minima as the lower envelope.
4. Compute the detail, $h_t(t) = x(t) - m_t(t)$.
5. Repeat step 1-4 on h_t denotes by

$$h_{1k} = h_{1(k-1)} - m_{1k}$$

and obtained c_1;

$$c_1 = h_{1k}$$

Stopping criteria of the sifting process:

5.1)

$$SD_k = \frac{\sum_{t=0}^{T} |h_{k-1}(t) - h_k(t)|^2}{\sum_{t=0}^{T} h_{k-1}^2(t)}$$

Stop when the standard deviation is smaller than pre-given value.

5.2) S number = number of consecutive siftings when the numbers of zero crossings and extrema are equal or at most differ by one.

6. Compute residual $r_1(t) = x(t) - c_1(t)$.
7. Repeat steps 1-6 to find more IMFs until the residual are constant or monotonic function.
8. Final $x(t)$ can be decomposed into n IMFs and residual r_n represented by

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \qquad (1)$$

B) Hilbert Spectral Analysis

1. Apply Hilbert transform

$$y(t) = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{x(t')}{t-t'} dt' \qquad (2)$$

2. Apply Hilbert spectrum,

$$x(t) = \sum_{j=1}^{n} a_j(t) e^{i \int \omega_j(t) dt} \qquad (3)$$

m_1 is the mean value of local extrema while $x(t)$ is the input data. The difference between m_t and input data is IMF which denoted by $h_t(t)$. The process of sifting is repeated until the IMF (h_{1k}) is obtained by k times. To retain the physical sense of amplitude and frequency modulations, either one of the two stopping criteria is needed to stop the sifting process. The criterion is described as above. SD is referring to standard deviation while c_1 represent the first IMF obtained which can be separated from the rest of data to get r_1(residue). This r_1 is considered as new data and is subjected to the sifting process. Then, the procedure is repeated until small value of c_n or r_n is achieved. The original data $x(t)$ can be represented as the summation of n IMF with residue (equation 1).

Hilbert Transform (HT) is represented by equation 2 to find the instantaneous frequency from the IMF. Cauchy Principal Value is denoted by PV. Hilbert spectrum is performed after applying HT for each IMF using equation 3. Both a_j and ω_j are constants. Lastly, the result will be presented in time-frequency-energy space for feature extraction [9]. Limitations in HHT involve the tendency to generate uncertain results when sudden changes in frequency occur in time series and may contaminate with thousands of spline fittings in the siftings [8]. Example on HHT is [10].

B. Principal Component Analysis

Another technique that is commonly implemented in feature extraction of EEG signal is Principal Component Analysis (PCA). The concept of PCA was first introduced by Karl Pearson in early 1900 [11]. It is a simple, non-parametric technique in extracting information from confuse data [12] and express it in their commonality and differences [13]. PCA will change the data by rotation, where it will find the direction in the data with most variation [14], [15]. The first axis corresponding to the first component is rotated to the direction where the variance of the data is the most. Then, the next component will be perpendicular to the first component with the most variance and so on [16].

Besides applicable for feature extraction, it also provides the reduction of dimensionality [13] which can reduce the complex data to lower dimension and uncovered the hidden information embedded within the data [12]. PCA is also known as graphical representation in analyzing and finding patterns in dataset [17]. The algorithm of PCA is as follows [18]:

ALGORITHM OF PCA

1. Input vectors = $x_t(t = 1, \ldots, l$ and $\sum x_{t=0})$.
   For m dimension, $x_t = (x_t(1), x_t(2), \ldots, x_t(m))^T$ and m < l
2. Calculate covariance matrix $C = \frac{1}{l} \sum_t^l x_t x_t^T$.
3. Calculate eigenvalue and eigenvector $\lambda_i u_i = C u_i$ where $i = 1, \ldots, m$ and sort the eigenvectors in descending order of eigenvalue.
4. Estimate u_i components of s_t, calculate orthogonal transformation of x_t,
   $s_t = u_i^T x_t$ where $i = 1, \ldots, m$
5. Linearly transform the each vector x_t into new s_t by

$$s_t = U^T x_t$$

U is refer to $m \times m$ orthogonal matrix whose ith column u_i is ith eigenvector of the sample covariance matrix. λ_i is one of the eigenvalue of C while u_i is corresponding eigenvector. Some of the limitations of PCA are reduction of the dimension can only be achieved if the original variables are correlated, the scale is not invariant and the input data should be assumed as continuous and real. The PCA use as feature extraction for EEG analysis is as described in [13]. Several researches on schizophrenia using PCA are described in [16], [19].

C. Independent Component Analysis

Independent Component Analysis (ICA) is a method that is applied in blind source separation (BSS) problem. The concept of ICA was first introduced by Comon in 1994 [20]. It considered the fact that the signals may decompose into independent components and such where multiple sources can be assumed independent from each other [21]. Hence, the independent components of the signals are able to be extracted through this method. Suppose there are m independent components for multichannel signal x in time series. There are several different ICA algorithms available, but these three are the algorithm of ICA that are widely used - FastICA, Infomax and JADE [22], [23].

Although there are different algorithms based ICA, all of them applies the concept of x=As equation [22]**.** x is the random vector whose elements are the mixtures $x_1, \ldots, x_n$, while s is the random vector with elements $s_1, \ldots, s_n$ and A is the matrix with elements $a_{i,j}$ [24]. Details on Infomax [25], [26], FastICA [26], [27] and JADE [27] can be seen here. In EEG, different sources refer to the electrodes containing mixtures of brain activities. Thus, some researches employed ICA to extract the original signals from their mixtures. In other words, it decomposes the independent component from the mixtures of sources. These components can be signals from brain activities or artifacts generated by Transcranial Magnetic Stimulation (TMS) [24].

ALGORITHM OF ICA
1. Initialize w_i. 2. Compute Newton phase $w_i^+ = E\left(\acute{\emptyset}(w_i^T X)\right) w_i - E\left(x\emptyset(w_i^T X)\right)$ 3. Calculate Normalization $w_i = \frac{w_i^+}{\lVert w_i^+ \rVert}$ 4. For i=1, go to step 7 or else continue with step 5. 5. Perform decorrelation $w_i^+ = w_i - \sum_{j=1}^{i-1} w_i^T w_j w_j$ 6. Compute $w_i = \frac{w_i^+}{\lVert w_i^+ \rVert}$ once again, 7. Go back to step 2 if not converged or else go to step 1 with i=i+1 until all components are extracted.

The algorithm of FastICA (fICA) is described as above as in [18], [26]. Before applying fICA, the matrix X must be centered, compressed and whitened. The details can be seen here [27]. ICA is able to separate and denoising signal as well as to recover the source signal. Besides, it is also able to convert multivariate random signals into mutually independent component of a signal. However, this method is only applicable for independent signal. Advantages of ICA include allowing the identification of distinct sources of activity and able to determine the impact of the external stimuli on some specific neuronal structures. Nevertheless, a number of sources must be known to achieve good result. As in practice, perfect separation is not anticipated. One of the limitations in ICA is the energies and the order of the components for each run are unknown [22]. In [28], a modified version of ICA is designed to improve the limitation mentioned above. However, the implementation is still not common. Besides, ICA is able to estimate the original signal until indeterminacies whereby it can preserve the original signal. So, there is a possibility that the waveform extracted after performing ICA is the original signal being analysed [23]. Research on schizophrenia can be found here [29], [30].

D. Local Discriminant Bases

This particular technique firstly presented by Saito and Coifman in 1994 [31] where Local Discriminant Bases (LDB) utilizes fully decimated wavelet packet decomposition (WPD). It possess disadvantages on the sensitivity of signal translation and only suitable for stationary data [33]. When the shifted version of the same signal is applied, the discriminant features will not be consistent. This is because the decimated WPD for library construction and the discriminant measures are employed. They will estimate each coefficient individually. Despite of that, LDB is a well-accepted scheme. The LDB is depend on best basis paradigm [34] and utilizes a tree adjustment which selects a basis from the dictionary that highlighted the discrepancy among classes. Below is the LDB algorithm [34], [35].

ALGORITHM OF LDB
1. Construct time-scale energy maps Υ_l for every class l=1,…,L. 2. For $k = 0, \ldots, 2^J - 1$ set $A_{J,k} := B_{J,k}$ $\Delta_{J,k} := \sum_{m \in T_j} \mathrm{D}(\Upsilon_1(J,k,m), \ldots, \Upsilon_L(J,k,m))$ 3. For $j = J-1, \ldots, 0$ and $k = 0, \ldots, 2^j - 1$ determine the "best subset" $A_{J,k}$ by the following rule : Set $\Delta_{J,k} := \sum_{m \in \mathcal{T}_j} \mathrm{D}(\Upsilon_1(j,k,m), \ldots, \Upsilon_L(j,k,m))$ If $\Delta_{J,k} \geq \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$ then $A_{J,k} := B_{J,k}$ else $A_{J,k} := A_{j+1,2k} \cup A_{j+1,2k+1}$ and $\Delta_{j,k} := \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$ Output: $A_{0,0}$

Haris et al. (2008) has proposed shift-invariant LDB (SLDB) algorithm to preserve the information in sub-bands. Below is the algorithm of SLDB for feature extraction [31], [35].

ALGORITHM OF SLDB
1. Construct the shift-invariant WPD Υ for J levels. 2. Apply For $j = 0, \ldots, J$ and $k = 0, \ldots, 2^j - 1$ $\mathcal{F}(m) = \begin{cases} \mathcal{E}(\Upsilon_{\mathrm{j,k}}) & \text{if } B_{\mathrm{j,k}} = 1 \\ 0 & \text{else} \end{cases}$ where $\mathcal{E}(.)$ is energy of sub-bands. 3. Apply $\mathcal{F} \neq 0$ as features.

Morphological LDB (MLDB) has been proposed by Strauss et al. (2003) which employed *lattice structure* on conventional wavelet packet [34]. MLDB is also known as shape-adapted LDB as it additionally adapts the shape of the analysed wavelet packets [35]. This algorithm adopts the original LDB algorithm and makes an adjustment of the two-channel filter bank building block. By doing so, more discriminant information can be extracted among signal classes. $A_{0,0}$ as discriminatory power is used in MLDB technique to enhance the LDB algorithm which given by:

$$\hat{\vartheta} = \arg\max_{\vartheta \epsilon p^{\mathrm{K}}} \Delta_{0,0}(\vartheta)$$

Morphological shift invariant LDB (MSLDB) is another LDB based adapted technique. The algorithm is the combination of morphological adapted filter banks with shift invariant technique which proposed a parameter space that contains all paraunitary filter banks of a given order at least one vanishing moment of the high pass filter is used for the selection of local discriminant bases (LDB). In other words, it is the combination of MLDB and SLDB techniques.

A comparison has been made between LDB, SLDB, MLDB and MSLDB on extracting features of slow wave components (SCP) from TMS and acoustic-somatosensory stimulation (ASS) in [33]. The results show the features selected are better distinguished in SLDB and MSLDB compared to others. Thereby, SLDB and MSLDB convey as a powerful technique in biosignal processing for feature extraction. Thus, these techniques are worth to be implemented for feature extraction in schizophrenia patients. Such technique has been applied for tinnitus patients and a robust results is obtained in [36], [37].

III. RESULTS AND DISCUSSION

EEG data were obtained from a study on examining EEG correlates of genetic predisposition to alcoholism. It is taken from the UCI machine learning repository (https://archive.ics.uci.edu/ml/datasets/EEG+Database). We used the data of 5 subjects from each group - control and alcoholic. The EEG data were signal at electrodes PZ, FZ, CZ, OZ, O1, O2, PO1, PO2, POZ, and P8. Ten trials at each electrodes are used and averaging is performed for all electrodes. Averaging of EEG signal at all electrodes for five subjects is shown in Figure 1.

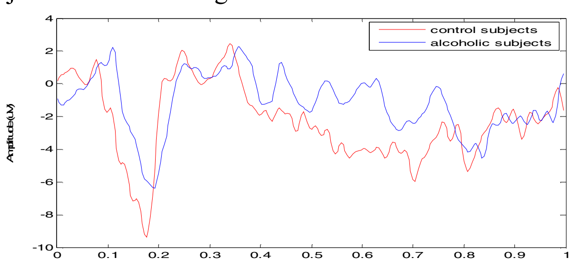
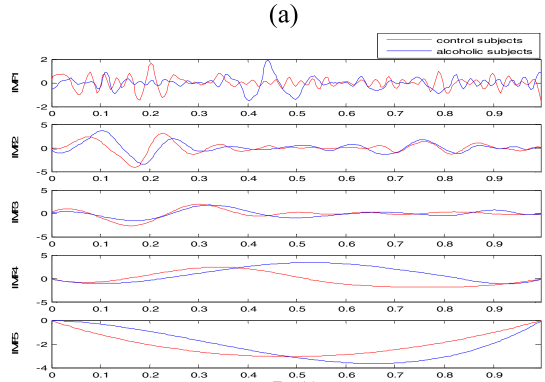


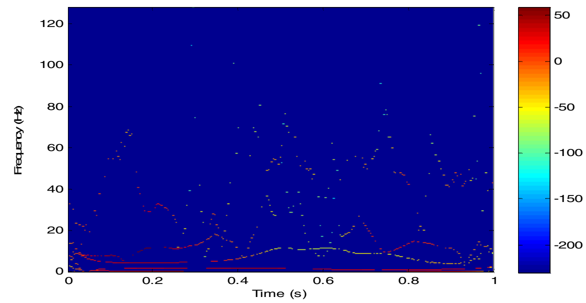
Figure 1. Grand averaged of EEG signal for control and alcoholic subjects.

A. Hilbert-Huang Transform

(a)



(b)



(c)

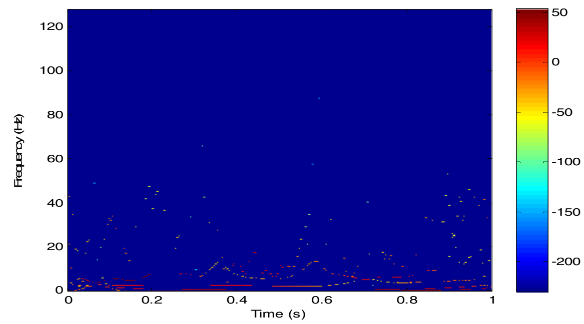


Figure 2. (a) IMFs of control and alcoholic subjects (b) Hilbert spectrum of control subjects (c) Hilbert spectrum of alcoholic subjects

B. Principal Component Analysis

(a)

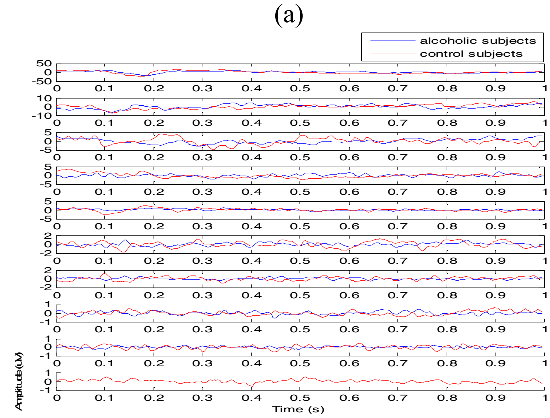


Figure 3. PCA of control subjects and alcoholic subjects.

C. Independent Component Analysis

(a)

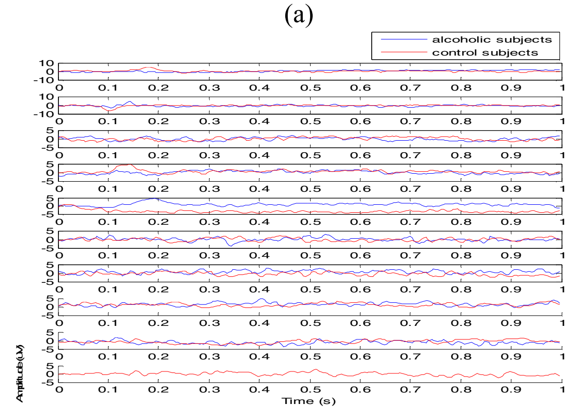


Figure 4. ICA of control subjects and alcoholic subjects.

D. Local Discriminant Bases

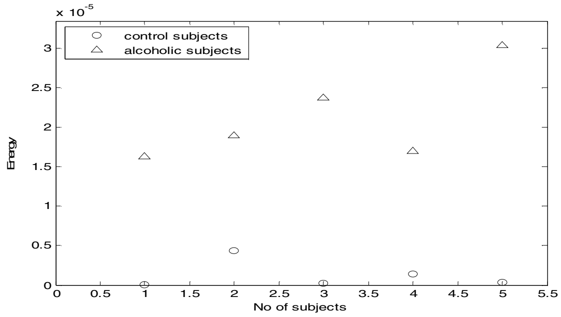


Figure 5. Feature space of control and alcoholic subjects.

From Figure 2 (a), P1 and N2 peak can be observed at IMF2 meanwhile P3 can be observed in IMF3 for both control and alcoholic subjects. We can also observe the frequency is decreasing from IMF1 to IMF5 as shown in Figure 2 (a). The decompose signal can be represented in different frequencies hence, the frequency of ERP peak is able to be locate. Higher energy distribution is noted at frequency range of 0-20 Hz for both control and alcoholic subject. However, there is high energy at 0-0.2s which frequency is more than 40 Hz as shown in Figure 2(b). The energy distribution is differ between control and alcoholic subjects due to the concentration shown by both groups. The cognitive process in control subjects tends to work compared to alcoholic subjects. This is because the alcoholic subjects are not in a good condition. In the analysis of PCA, all channels are selected for control subjects. Meanwhile for alcoholic subjects, only 9 channels are selected. This is shown in Figure 3. In the analysis of ICA, 10 channels are selected for control subjects while 9 channels are selected for alcoholic subjects as shown in Figure 4. The dimension reduction in ICA and PCA helps in simplifying the data and makes easier visualization. PCA dimension is reduced when only the eigenvectors that have large eigenvalues are choose. This is where a lot of information lies. The small eigenvalues which close to zero are usually discarded because it only has little information. Thus, in alcoholic subjects, only 9 channels are selected and 10 channels are selected in control subjects. In ICA, the independent components are extracted from mixed signal where they are assumed independent from each other. The independent means that the occurrence of each signal does not affect to each other. Thus, in alcoholic subjects, only 9 channels are independent between each other meanwhile in control subjects all 10 channels are independent between each other. The result of MSLDB is shown in Figure 5, where the energy is the feature. The first 250ms is choose due to the large difference between ERP signals of two groups as shown in Figure 1. This shows that the energy for both control and alcoholic subjects can be discriminated. So, this feature then can be used as an input for the purpose of classification.

IV CONCLUSION

In summary, all the techniques able to differentiate the control and alcoholic groups. Thus, these techniques able to extract the relevant information embedded in EEG signal. The EEG signal contains information which can be correlated with the neuropsychology and brain activity of human brain. Moreover, integrating these techniques is also possible in order to achieve a good result. Basically, feature extraction is very subjective and there are no such best feature extraction techniques. All of them can be implemented or modified according to the situation and problem faced. The LDB is not broadly used in EEG signal, but from the concept above, it is very helpful to extract features due to the discriminant properties. Hence, this technique can be further explored in extracting the EEG signal in schizophrenia.

ACKNOWLEDGMENT

The authors would like to thank Universiti Teknikal Malaysia Melaka and grant ERGS/2013/FKEKK/TK02/UTEM/02/03 E00019 for the funding. Also, a big gratitude to the author who donated the datasets.

REFERENCES

[1] U. Schall, A. Schon, D. Zerbin, S. Bender, C. Eggers and R. Oades, 'A left temporal lobe impairment of auditory information processing in schizophrenia: an event-related potential study', *Neuroscience letters*, vol 229, no 1, pp. 25-28, 1997.

[2] C. Gilmore, B. Clementz and P. Buckley, 'Stimulus sequence affects schizophrenia--normal differences in event processing during an auditory oddball task', *Cognitive brain research*, vol 24, no 2, pp. 215-227, 2005.

[3] M. Pellicciari, D. Veniero, C. Marzano, F. Moroni, C. Pirulli, G. Curcio, M. Ferrara, C. Miniussi, P. Rossini and L. De Gennaro, 'Heritability of intracortical inhibition and facilitation', *The Journal of Neuroscience*, vol 29, no 28, pp. 8897-8900, 2009.

[4] T. Rosburg, N. Boutros and J. Ford, 'Reduced auditory evoked potential component N100 in schizophrenia—a critical review', *Psychiatry research*, vol. 161, no 3, pp. 259-274, 2008.

[5] A. Sevik, A. Anil Yagcioglu, S. Yagcioglu, S. Karahan, N. Gurses and M. Yildiz, 'Neuropsychological performance and auditory event related potentials in schizophrenia patients and their siblings: a family study', *Schizophrenia research*, vol. 130, no 1, pp. 195-202, 2011.

[6] C. Shi, X. Yu, E. Cheung, D. Shum and R. Chan, 'Revisiting the therapeutic effect of rTMS on negative symptoms in schizophrenia: A meta-analysis', *Psychiatry research*, vol. 215, no 3, pp. 505-513, 2014.

[7] American Psychiatric Association (2013, May), 'Highlights of Changes from DSM-IV-TR to DSM-5' [Online]. Available: http://www.dsm5.org

[8] Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. Chyuan Yen, C. Chao Tung, H. H. Liu and N. E. Huang, 'The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis', *The Royal Society*, vol 454, pp. 903-995, 1998.

[9] N. Huang, 'Introduction to the Hilbert Huang Transform and Its Related Mathematical Problems', in *Hilbert-Huang Transform and Its Applications*, 1st ed., Singapore, World Scientific Publishing Co. Pte. Ltd., 2005, ch. 1, pp. 1-26.

[10] A. Pigorini, A. Casali, S. Casarotto, F. Ferrarelli, G. Baselli, M. Mariotti, M. Massimini and M. Rosanova, 'Time--frequency spectral analysis of TMS-evoked EEG oscillations by means of Hilbert--Huang transform', *Journal of neuroscience methods*, vol. 198, no 2, pp.236-245, 2011.
[11] K. Pearson, 'On Lines and Planes of closest fit to system of points in space', *Philosophical Magazine*, vol. 2, pp. 559-572, 1901.
[12] J. Shlens (2014, April 7), 'A tutorial on principal component analysis' [Online]. Available: http://arxiv.org/pdf/1404.1100.pdf
[13] A. Subasi and M. Ismail Gursoy, 'EEG signal classification using PCA, ICA, LDA and support vector machines', *Expert Systems with Applications*, vol. 37, no 12, pp. 8659-8666, 2010.
[14] E. ter Braack, B. de Jonge and M. van Putten, 'Reduction of TMS induced artifacts in EEG using principal component analysis.', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no 3, pp. 376-382, 2013.
[15] L. Kuncheva and J. Rodriguez, 'Interval feature extraction for classification of event-related potentials (ERP) in EEG data analysis', *Progress in Artificial Intelligence*, vol 2, no 1, pp. 65-72, 2013.
[16] A. Caprihan, G. D. Pearlson and V. D. Calhoun, 'Application of Principal Component Analysis to Distinguish Patients with Schizophrenia from Healthy Controls Based on Fractional Anisotropy Measurements', in *NeuroImage,* 2008, vol. 42, no 2, pp. 1-19.
[17] R. Kottaimalai, M. Rajasekaran Pallikonda, V. Selvam and B. Kannapiran, 'EEG Signal Classification using Principal Component Analysis with Neural Network in Brain Computer Interface Applications', in *IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology*, 2013, pp. 227-231.
[18] M. R. Vargantwar, S. M. Rajput and K. Mahajan, 'Classification of EEG using PCA, ICA and Neural Network,' *International Journal of Engineering and Advanced Technology*, vol. 1, no 1, pp. 80-83, 2011.
[19] A. Caprihan, G. Pearlson and V. Calhoun, 'Application of principal component analysis to distinguish patients with schizophrenia from healthy controls based on fractional anisotropy measurements', *Neuroimage*, vol. 42, no 2, pp. 675-682, 2008.
[20] P. Comon, 'Independent component analysis, a new concept?', *Signal Processing*, vol. 36, no 3, pp. 287-314, 1994.
[21] S. Sanei and J. Chambers, *EEG signal processing*, 1st ed. Chichester, England: John Wiley & Sons, 2007.
[22] E. Kallioniemi, 'Electroencephalography responses evoked by transcranial magnetic stimulation on the primary visual cortex', Master of Science in Technology, Aalto University, 2012.
[23] G. Naik and D. Kumar, 'An overview of independent component analysis and its applications', *Informatica: An International Journal of* Computing *and Informatics*, vol. 35, no 1, pp. 63-81, 2011.
[24] U. Unjoo Lee and W. Woo-Kyoung Yoo, 'TMS-induced EEG Artifacts Removal Methods based on Cross-Correlation Coefficients of ICA Components', *International Journal of Bio-Science and Bio Technology*, vol. 5, no 5, pp. 161-170, 2013.
[25] T. J. Sejnowski and A. J. Bell, 'An information-maximisation approach to blind separation and blind deconvolution', *Neural Computation*, pp. 1-38, 1995.
[26] D. Langlois, S. Chartier and D. Gosselin, 'An introduction to independent component analysis: InfoMax and FastICA algorithms', *Tutorials in Quantitative Methods for Psychology*, vol 6, no 1, pp. 31-38, 2010.
[27] J. Hernandez-Pavon, J. Metsomaa, T. Mutanen, M. Stenroos, H. Maki, R. Ilmoniemi and J. Sarvas, 'Uncovering neural independent components from highly artifactual TMS-evoked EEG data', *Journal of Neuroscience methods*, vol. 209, no 1, pp. 144-157, 2012.
[28] R. Korhonen, J. Hernandez-Pavon, J. Metsomaa, H. Maki, R. Ilmoniemi and J. Sarvas, 'Removal of large muscle artifacts from transcranial magnetic stimulation-evoked EEG by independent component analysis', *Medical & biological engineering & computing*, vol 49, no 4, pp. 397-407, 2011.
[29] D. Iyer, N. Boutros and G. Zouridakis, 'Single-trial analysis of auditory evoked potentials improves separation of normal and schizophrenia subjects', *Clinical Neurophysiology*, vol. 123, no 9, pp. 1810-1820, 2012.
[30] J. Sui, T. Adali, G. Pearlson, H. Yang, S. Sponheim, T. White and V. Calhoun, 'ACCA+ICA based model for multi-task brain imaging data fusion and its application to schizophrenia', *NeuroImage*, vol. 51, pp. 123-134, 2010.
[31] N. Saito and R. Coifman, 'Local discriminant bases and their applications', *Journal of Mathematical Imaging and Vision*, vol. 5, no 4, pp. 337-358, 1995.
[32] A. Harris, K. Schwerdtfeger and D. Strauss, 'Optimized shift-invariant wavelet packet feature extraction for electroencephalographic evoked responses', pp. 2685--2688, 2008.
[33] D. Strauss, W. Delb, J. Jung and P. Plinkert, 'Adapted filter banks in machine learning: applications in biomedical signal processing', vol. 6, pp. 425-428, 2003.
[34] A. Harris, K. Schwerdtfeger and D. Strauss, 'Adapted filter banks for feature extraction in transcranial magnetic stimulation evoked responses', *Medical & Biological Engineering & Computing*, vol. 49, pp. 221-231, 2011.
[35] D. Strauss, G. Steidl and W. Delb, 'Feature extraction by shape-adapted local discriminant bases', *Signal processing*, vol. 83, no 2, pp. 359-376, 2003.
[36] A. R. Harris, K. Schwerdtfeger, Yin Fen Low, D. J. Strauss, 'Effects of Transcranial Magnetic Stimulation on Auditory Attention: An Electoencephalographic Study', in 4th *International IEEE EMBS Conference on Neural Engineering*, 2009, pp. 577-580.
[37] Yin Fen Low, K. Schwerdtfeger, A. R. Harris, D. J. Strauss, 'An investigation on the effects of Single Pulse Transcranial Magnetic Stimulation in a Modified Maximum Entropy Auditory Stimulation Paradigm', in 31st *International IEEE EMBS*, 2009, pp. 678-681.