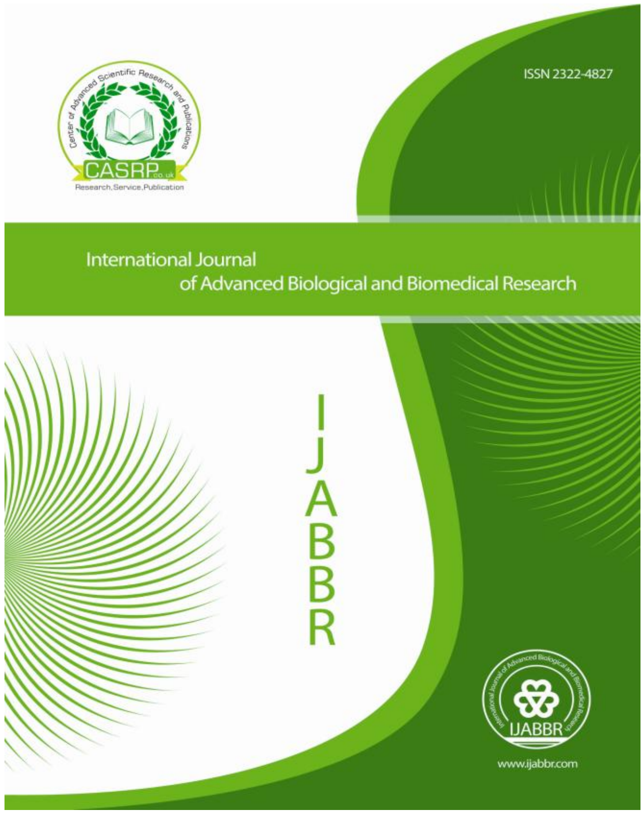
ISSN 2322-4827
CASRP
Research.Service.Publication
International Journal
of Advanced Biological and Biomedical Research
IJABBR
IJABBR
www.ijabbr.com

Available online at www.casrp.co.uk/journals

CASRP PUBLISHER

International journal of Advanced Biological and Biomedical Research 4(2) (2016) 117–123

doi: 10.18869/IJABBR.2016.117

Review Article Open Access

A review on EEG based brain computer interface systems feature extraction methods

Nazlar Ghassemzadeh[1], Siamak Haghipour[2,]*

[1]MS Student, Department of Biomedical Engineering, Tabriz Branch , Islamic Azad University Tabriz , Iran.
[2]Assistant Professor, Department of Biomedical Engineering, Tabriz Branch, Islamic Azad University Tabriz, Iran.

Abstract

The brain – computer interface (BCI) provides a communicational channel between human and machine. Most of these systems are based on brain activities. Brain Computer-Interfacing is a methodology that provides a way for communication with the outside environment using the brain thoughts. The success of this methodology depends on the selection of methods to process the brain signals in each phase Feature extraction is one of the most important stages in distinguishing of brain activities from EEG. New features are produced by primary features. Today, in the field of EEG signal processing methods for the best feature extraction are so important. In this article we mentioned EEG based brain computer interface (BCI) systems feature extraction such as Principle Component Analysis (PCA), Linear Discriminant Analysis(LDA), Independent Component Analysis (ICA), Mutual information theory (MI), Empirical Mode Decomposition(EMD), High–order frequency component, Wavelet Transform, Common Spatial Pattern (CSP), Complex Band Power (CBP).

Keywords: Feature extraction, EEG, Brain Computer Interface (BCI), Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Mutual Information theory (MI), Empirical Mode Decomposition (EMD), Common Spatial Pattern (CSP), Complex Band Power (CBP).

* **Corresponding author:** Assistant Professor, Department of Biomedical Engineering, Tabriz Branch, Islamic Azad University Tabriz, Iran.

Received 19 March 2016 Accepted 18 April 2016 Available online 25 April 2016
iThenticate screening 22 March 2016 English editing 15 April 2016 Quality control 21 April 2016

1. Introduction

Recently, many attempts have been done to use the electroencephalogram (EEG) as a new communication channel between human brain and computer. This new communication channel is called EEG-based brain-computer interface (BCI). To date, different types of BCI were suggested by different research groups (Erfanian and Erfani, 2004). BCIs are communication systems, which enable users to send commands to computers by using brain activity only; this activity being generally measured by EEG (Arvaneh et al., 2011; Pfurtscheller and Neuper, 2001). BCI are generally designed according to a pattern recognition approach, i.e., by extracting features from EEG signals, and by using a classifier to identify the user's mental state from such features (Arvaneh et al., 2011; Pfurtscheller and Neuper, 2001; Lotte et al., 2007). Brain computer interfaces are devices intended to help disabled people to communicate with a computer using the brains' electrical activity (Ting et al., 2008). The most important part of BCI system is EEG signal processing, which include preprocessing, feature extraction and classification. Feature extraction plays a vital role in these systems (Kolodziej et al., 2012).

2. EEG feature extraction

In pattern recognition, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy (Nandish et al., 2012).

3. EEG feature extraction methods

3.1. Principle Component Analysis (PCA)

PCA was invented in 1901 by Karl Pearson and later developed independently by Harold Hotelling in 1930 (Jackson, 1991). The PCA transforms the correlated vectors into linearly uncorrelated vectors. These uncorrelated vectors are called as "Principal Components" (Jackson, 1991; Feng et al., 2009). This is a classical method of Second Order Statistics. It depends on decomposition of covariance matrix. PCA helps in reduction of feature dimensions. Ranking will be done by using PCA based on the variability of the signal properties. This ranking helps in classification of the data. The application of PCA in a BCI system yields best classification results (Tomas, 2000). The PCA is well but it is not as well as ICA (Bruce et al., 2003).

PCA is s a pre-processing technique as well as a feature extraction method. It is a powerful tool for analyzing and for dimension reduction of data without loss of information (Srinivasulu et al., 2012). Using PCA the information present at all the time series multi channel is extracted as principal components. By eliminating the artifacts and by forming the principal components PCA reduces the dimensions of signals (Rajya et al., 2014; Jung et al., 1998; Araki et al., 2005).

3.2. Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA) is a well known feature reduction technique (Lotte et al., 2007). LDA is used to find a linear combination of features that can better separate two or more classes. The LDA finds such direction *a* that provide maximum linear separation of classes. An example of a data projection on directions *a* and *b* is given in fig. 1. There are many possibilities for finding directions but only some are optimal for data discrimination (Kolodziej et al., 2012).

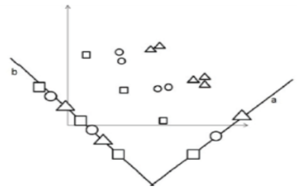

Fig. 1. LDA finds a direction (a) that maximize the separation of data (Kolodziej et al., 2012).

3.3. Independent Component Analysis (ICA)

ICA was first applied to EEG by Makeig et al. in 1996 (Delorme and Makeig, 2004). ICA separates the artifacts from the EEG signals into independent components based on the characteristics of the data without relying on the reference channels. The data in the recorded trails, each channel data and the frontal data are also preserved during the ICA artifact removal (Srinivasulu et al., 2012). The ICA algorithm decomposes the multi-channel EEG data into temporal independent and spatial-fixed components. It is computationally efficient. ICA shows high performance when the size of the data to decompose is large (Jung et al., 1998). ICA requires more computations to decompose signals (Srinivasulu et al., 2012; Araki et al., 2005). EEGLAB supports various types of ICA algorithms (nearly 20 algorithms) and most used algorithms are Joint Approximate Decomposition of Eigen matrices (JADE), fixed point ICA, Informix Bhattacharya et al., 2004). ICA can also be used as a feature extraction method. ICA forms the components that are independent to each other. From the components essential features were extracted using ICA. An important application of ICA is Blind Source Separation. This helps in identifying the independent signals and also noise separation from brain signals. Blind Source Separation (BSS) of acoustic signals are referred to as Cocktail party problem means separation of a no. of independent components from a set of un-controlling records (Rajya et al., 2014; Torseet al., 2012).

3.4. Mutual Information theory (MI)

Mutual information is a non-parametric measure of relevance between two variables. Shannon's information theory provides a suitable formalism for quantifying this concepts. Assume a random variable *X* representing continuous-valued random feature vector, and a discrete-valued random variable *C* representing the class labels. In accordance with Shannon's information theory, the uncertainty of the class label *C* can be measured by entropy *H*(*C*) as (Erfanian et al., 2011).

$$H(C) = -\sum_{c=1}^{Nc} p(c) \log p(c) \qquad (1)$$

Where *p*(*c*) represents the probability of the discrete random variable *C*. The uncertaintyabout *C* given a feature vector *X* is measured by the conditional entropy as: (Erfanian et al., 2011).

$$H(C|X) = -\int_x p(x)\left(\sum_{c\in C} p(c|x)\log(p(c|x))\right)dx \qquad (2)$$

Where $p(c|x)$ is the conditional probability for the variable *C* given *X*. In general, the conditional entropy is less than or equal to the initial entropy. It is equal if and only if one has independence between two variables *C* and *X*. The amount by which the class uncertainty is decreased is, by definition, the mutual information, *I*(*X*;*C*) = *H*(*C*) − $H_{(C|X)}$, and after applying the identities *p*(*c*,x) = $p(c|x)$*p*(x) and $p(c) = \int_x p(c,x)dx$ can be expressed as: (Erfanian et al., 2011).

$$I(X:C) = \sum_{c\in C}\int_x p(c,x)\log\frac{p(c,x)}{p(c)p(x)}dx \qquad (3)$$

If the mutual information between two random variables is large, it means two variables are closely related. Indeed, MI is zero if and only if the two random variables are strictly independent (Erfanian et al., 2011).

3.5. Emprical Mode Decomposition (EMD)

Empirical Mode Decomposition is one of the method of feature extraction. It is advantageous compared to the other methods. EMD method is able to decompose a complex signal into a series of intrinsic mode functions (IMF) and a residue in accordance with different frequency bands. EMD is self-adaptive because the IMF works as the basis functions determined by the signal itself rather than what is pre-determined. It is highly efficient in non-stationary data analysis (Davis, 2012; Jian-Da and Wu, 2011).

3.6. Wavelet Transform (WT)

Wavelet transform forms a general mathematical tool forsignal processing with many applications in EEG data analysis (Glavinovitch et al., 2005; Johankhani et al., 2006; Dimoulas et al., 2007; Selesnick et al., 2005 ; Nazareth et al., 2006) as well. Its basic use includes time-scale signal analysis, signal decomposition and signal compression. The set of wavelet functions is usually derived from the initial (mother) wavelet $h(t)$ which is dilated by value a = 2m, translated by constant b = k 2m and normalized so that

$$h_{m,k}(t) = \frac{1}{\sqrt{a}}\, h\left(\frac{t-b}{a}\right) = \frac{1}{\sqrt{2^m}}\, h\left(2^{-m}\, t - k\right) \qquad (4)$$

For integer values of m, k and the initial wavelet defined either by the solution of a dilation equation or by an analytical expression (Daubechies, 1990; Newland, 1994). Both continuous or discrete signals can be then approximated in the way similar to Fourier series and discrete Fourier transform. In case of a sequence *{x*(*n*)*}N*−1 *n*=0 having N = 2s values it is possible to evaluate its expansion:

$$x(n) = a_0 + \sum_{m=0}^{s-1} \sum_{k=0}^{2^{s-m-1}-1} a_{2^{s-m-1}+k}\; h(2^{-m}n-k) \qquad (5)$$

Wavelet transform coefficients can be organized in a matrix T withits nonzero elements forming a triangle structure.

$$\begin{bmatrix} a_{2^{s-1}} & & a_{2^{s-1}+1} & & & \cdots & & & & a_{2^s-1} \\ & a_{2^{s-2}} & & & & \cdots & & & a_{2^{s-1}-1} & \\ & & a_4 & & a_5 & \cdots & a_6 & & a_7 & \\ & & & a_2 & & & & a_3 & & \\ & & & & & a_1 & & & & \\ & & & & & a_0 & & & & \end{bmatrix}$$

With each its row corresponding to a separate dilation coefficient$m$. The set of N = 2^s decomposition coefficients $\{a(j)\}_{j=0}^{N-1}$ of the wavelet transform is defined in the wayformally close to the Fourier transform but owing to thegeneral definition of wavelet functions they can carry differentinformation. Using the orthogonal set of wavelet functions theyare moreover closely related to the signal energy (Newland, 1994; Glavinovitch et al., 2005).

3.7. Common Spatial Pattern (CSP)

The CSP algorithm is often used to optimally discriminate between two classes of EEG data based on simultaneous diagonalization of two covariance matrices (Ramoser et al., 2000). A brief description of CSP is given in this section. Given that, the preprocessed EEG data in a single trial are represented as matrix X of size N ×T, where N is the number of channels used and T is the number of samples recorded in each trial from each channel. The CSP projection matrix W is used to obtain the spatially filtered EEG signal as in (6) (Erfanian and Erfani, 2004).

$$Z=WX \qquad (6)$$

The rows of W are the stationary spatial filters and columns of W^{-1} represent the common spatial patterns. The normalized spatial covariance matrix of the EEG data is computed as (7).

$$C= \frac{XX^T}{trace\,(XX^T)} \qquad (7)$$

Where X^T denotes the transpose of matrix X , and tr (·) represents its sum of diagonal elements of two classes 1 and 2. CSP analysis aims to simultaneously diagonalize these matrices by designing W such that it satisfies (8).

$$\lambda_1 = w\, C_1 w^T \qquad (8)$$
$$\lambda_2 = w\, C_1 w^T$$

Where λ_1 and λ_2 are diagonal matrices and satisfies (9).

$$\lambda_1 + \lambda_2 = 1 \qquad (9)$$

The CSP projection matrix is determined by eigenvalue decomposition approach. Only a small number of signals j can efficiently discriminate the classes when used to train a classifier. The signals Z_P (p=1 to 2j) that maximize discrimination are the ones associated with the largest $\lambda_1$and λ_2 , which are the first and last j rows of Z. The feature vectors are obtained as in the following equation (Ramoser et al., 2000).

$$f_p = \log\left(\frac{var\ (Z_p)}{\sum_{i=1}^{2j} var\ (Z_i)}\right) \qquad (10)$$

The log transformation approximates the normal distribution of data (Erfanian and Erfani, 2004).

3.8. Complex Band Power (CBP)

Now, a brief review of the derivation of CBP features isprovided. Amplitude and phase features are extracted from Laplacian-filtered EEG signals using a sliding Hamming window to which an FFT was applied. At the sampling rate of 250 Hz used for recording. This window is then 64 samples wide. This window approximately produces equally-spaced frequency bands of 250/64= 3.9063 Hz (4 HZ). As a result, each sample, primarily, effectively represents the frequency content in each frequency range of f-(+)(1/2)w where w is the spacing of, or the width between, the frequency samples. Then subset of these coefficients, S(f) representing the frequency bands of interest mentioned below, have been further considered in (11) and (12).

$$\alpha_f = |X_f| \qquad f \in S_f \qquad (11)$$
$$\varphi^{\wedge} = \frac{\Delta\varphi_f}{\Delta t} = \varphi_f - \varphi_{f-1} \qquad (12)$$

Eight equally-spaced frequency bands were considered. These bands were 4–8 Hz, 8–12 Hz, 12–16 Hz, 31–35 Hz. Then the magnitude and phase of the frequency bands were extracted. Because an absolute phase is not meaningful without any reference in equation (4) delta phase were considered. Finally, the phase and amplitude features must be smoothed using a 1-s moving average finite impulse response filter (Townsend et al., 2006).

$$\varphi^{\wedge} = \frac{\Delta\varphi_f}{\Delta t} = \varphi_f - \varphi_{f-1} \qquad (13)$$

It is known that signals from electrodes C3, C4, and Cz reveal the most important signals (Obermaier et al., 2001). These electrodes and the four immediately surrounding electrodes in each case (Fig. 1) could be used to provide reasonable results. According tothese, only 15electrode were used for feature extraction (Townsend et al., 2006; Manoochehri et al., 2010).

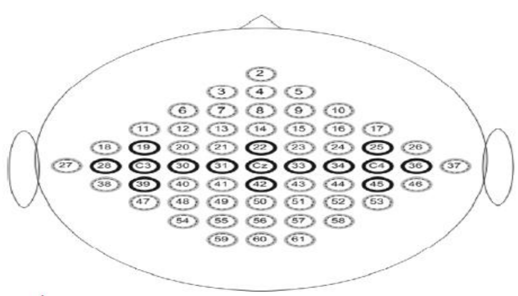

Fig. 2. Electrode montage worn by the subject consists of sixty electrodes(CBP method) (Manoochehri et al., 2010).

4. Conclusion

For several years, many efforts have been done to use the electro-encephalogram (EEG) as a new communication channel between human brain and computer. This new communication channel is called EEG-based brain–computer interface (BCI). Brain -Computer interfaces (BCIs) are communication systems, which enable users to send commands to computers by using brain activity only; this activity being generally measured by Electroencephalography (EEG). BCI are generally designed according to a pattern recognition approach, i.e., by extracting features from EEG signals, and by using a classifier to identify the user's mental state from such features. The BCI provides an additional output channel from brain, and uses the neuronal activity of brain to control effectors such as robotic arm or wheel chair; or to restore motor abilities of paralyzed or stroke patients. A brain–computer interface (BCI) acquires brain signals, extracts informative features, and translates these features to commands to control an external device. Efforts have been dedicated to the improvement of the accuracy and capacity of this EEG-based communication channel. Several factors may affect the performance of the BCI. These factors include the brain signal used as the input of the BCI, the signal processing methods used for feature extraction and classification. This paper investigates the several methods of EEG signals feature extraction which provide us to have favorable and desirable BCI systems with higher accuracy and resolution in a short time.

References

Antara, B., Bawane, N.G., Nirkhi, S.M., 2004. Brain computer interface using EEG signals, and available at: http://www.schalklab.org/sites/default/files/misc/Abrain-computer interface using electrocorticographic signals in humans.pdf.

Araki, S., Sawada, H., Mukai, R., Makino, S., 2005. A novel blind source separation method with observation vector clustering, Proc. of IWAENC.

Arvaneh, M., Guan, C., Keng Ang, K., Quek, C., 2011. Optimizing the channel selection and classification accuracy in EEG-Based BCI, IEEE Trans. Biomed. Eng., 58(6), June.

Bruce, A., Draper, Kyungim Baek, Marian Stewart Bartlett, Ross Beveridge, J., 2003. Recognizing faces with PCA and ICA, computer vision and image understanding 91, available at www.ComputerScienceWeb.com.

Daubechies, I., 1990. The wavelet transform, time-frequency localizationand signal analysis. IEEE Trans. Inform. Theory., (36), 961–1005, September.

Davis, N.V., 2012. Feature extraction using empirical modedecomposition of speech signal. Int. J. Eng. Trend. Technol., 3(2).

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Method., Vol. 134.

Dimoulas, C., Kalliris, G., Papanikolaou, G., Kalampakas, A., 2007. Long-Term signal detection, segmentation and summarization using wavelets and fractal dimension: A bioacoustics application in gastrointestinal-motility monitoring. Comput. Biol. Med., 37(4), 438–462.

Erfanian, A., Erfani, A., 2004. EEG-based brain-computer interface for hand grasp control: feature extraction by using ICA, Proc, 9th Annual Conf. Int'l. Functional ElectricalStimulation Society.

Erfanian, A., Erfani, A., 2004. EEG-based brain-computer interface for hand grasp control: feature extraction by using ICA. Proc, 9th Annual Conf. Int'l. Functional Electrical Stimulation Society.

Erfanian, A., Oveisi, F., Shadvar, A., 2011. Feature extraction by mutual information basedon minimal-redundancy-maximal-relevance criterion and its application to classifying EEG signal for brain-computer interfaces. ISBN, 978-953-307-175-6, February.

Feng Gu, Julie Greensmith, Robert Oates, Uwe Aickelin, 2009. PCA 4 DCA: The application of principal component analysis to the dendritic cell algorithm, In 9th Annual Workshop on Computational Intell. (UKCI), 7-9 Sept.

Furtscheller, P., Neuper, C., 2001. Motor imagery and direct brain-computer communication. Proc IEEE, 89(7), 1123–1134.

Glavinovitch, A., Swamy, M.N.S., Plotkin, E.I., 2005. Wavelet-Based segmentation techniques in the detection of microarousals in the sleep EEG. In 48th Midwest Symposium on Circuits and Systems, 1302–1305, IEEE.

Glavinovitch, A., Swamy, M.N.S., Plotkin, E.I., 2005. Wavelet-Based segmentation techniques in the detection of microarousals in the sleep EEG. In 48th Midwest Symposium on Circuits and Systems, 1302–1305, IEEE.

Jackson, J.E., 1991. A User's guide to principal components (Wiley).

Jian-Da, I., Wu, Yi-Jang Tsai, 2011. Speaker identification system using empirical mode decomposition and an artificial neural network. Expert Systems with Applications, 38, 6112–6117.

Johankhani, P., Kodogiannis, V., Revett, K., 2006. EEG signal classification using wavelet feature extraction and neural networks. In IEEE John Vincent Atanasoff International Symposium on Modern Computing (JVA06), 120–124, IEEE.

Jung, T.P., Makeig, S., Humphries, C., 1998. Extended ICA removes artifacts from electroencephalographic recording. Advances in Neural Inf. Processing Systems, Cambridge, Vol. 10.

Kolodziej, M., Majkowski, A., Rak, R.J., 2012. Linear discriminate analysis as EEG features reduction technique for brain-computer interfaces. ISSV 0033-2097, R.88NR 3a.

Lotte, F., Congedo, M., L´ecuyer, A., Lamarche, F., Arnaldi, B., 2007. A review of classification algorithms for EEG-based brain-computer interfaces. J. Neural. Eng., 4, R1–R13.

Manoochehri, M., Hassan Moradi, M., Pfurtscheller, G., 2010. Are all features extraction methods subject_independent. (ICBME2010), 3-4 November.

Nandish, M., Michahial, S., Hemanth, P., Ahmed, F., 2012. Feature extraction and classification of EEG signal using neural network based techniques. ISSN: 2277-3754, (IJEIT), 2(4), October.

Nazareth, P., Castellanos, Valeri, A., Makarov, 2006. Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis. J. Neurosci. Method., 158(2), 300–312.

Newland, D.E., 1994. An introduction to random vibrations, spectral and wavelet analysis. Longman Scientific & Technical, Essex, U.K., thirdedition.

Obermaier, B., Neuper, C., Guger, C., Pfurtscheller, G., 2001. Information transfer rate in a five-classesbrian-computer interface. IEEE Trans. Neural Syst. Rehab. Eng., 9(3), 283–288, Sep.

Rajya Lakshmi, M., Prasad, T.V., Chandra Prakash, V., 2014. Survey on EEG signal processing methods. ISSN: 2277 128X, 4(1), January.

Ramoser, H., Muller Gerking, J., Pfurtscheller, G., 2000. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEET rans. Rehab. Eng., 8(4), 441–446, Dec.

Selesnick, I.W., Baraniuk, R.G., Kingsbury, N.G., 2005. The Dual-Tree Complex Wavelet Transform. IEEE Signal Processing Magazine, 22(6), 123–151.

Srinivasulu, A., Sreenath Reddy, M., 2012. Artifacts removing from EEG signals by ICA Algorithms, IOSR J. Electric. Electron. Engg., 2(4), Sep-Oct.

Ting, W., Zheng, Y.G., Hua, Y.B., Hong, S., 2008. EEG feature extraction based on wavelet packet decomposition for brain computer interface. Measurment., 8 August.

Torse, D.A., Maggavi, R.R., Pujari, S.A., 2012. Nonlinear blind source separation for EEG signal pre-processing in brain-computer interface system for epilepsy. Int. J. Comp. App., 50(14).

Townsend, G., Graimann, B., Pfurtscheller, G., 2006. A comparison of common spatial patterns with complex band power features in a four-class BCI experiment. IEEE Trans. Biomed. Eng., 53, 642–651.

Zeman, T., 2000. BSS-Preprocessing Steps for Separation Improvement, May, Available at http://noel.feld.cvut.cz/vyu/prj-czs-asi/zeman/pca.pdf

How to cite this article: Ghassemzadeh, N., Haghipour, S., 2016. A review on EEG based brain computer interface systems feature extraction methods. International journal of Advanced Biological and Biomedical Research, 4(2), 117-123.