

EEG Based Biometric Identification with Reduced Number of Channels

Brijraj Singh, Sudhakar Mishra, Uma Shanker Tiwary

Department of I.T., IIIT Allahabad, Allahabad, India

brijraj.08@gmail.com, smsmishra46@gmail.com, ustiwary@gmail.com

Abstract— EEG based biometric system can be used for authentication, with advantages like confidentiality retention and forgery prevention. Signals which are taken from maximum brain regions show some sort of unique information that can be used for extracting the subject dependent pattern. This paper presents an approach to find the relationships among signals generated in different brain regions which give birth to unique patterns. A bivariate measure, Magnitude Squared Coherence (MSC) is selected as a feature which is insensitive to the random changes in the amplitude of the signals (because of circadian rhythm). We are trying to optimize the number of EEG channels to be considered for the identification without compromising the accuracy. An experiment is performed on all the possible combinations of channels and on accomplishment of 100% accuracy channels are reduced one by one. This incremental approach is followed till we get accuracy less than 95%. K-nearest neighbour (K=1), a distance based classifier, which worked well with very high dimensional data with limited number of samples per class is used here. 100% accuracy on 108 subjects with eye open resting state was previously claimed using 64 channels, whereas the same accuracy is obtained here on 109 subjects by selecting only 10 channels. Obtained result lead us to conclude that 10 channels can be used conveniently in confidential environment for biometric identification.

Keywords— EEG, Magnitude squared coherence, authentication, distance based classifier, 10 channel EEG, k-nearest neighbour.

I. INTRODUCTION

EEG based identification system has significant importance in the field of biometrics authentication. In this system a user can be traced while working in very confidential environment. These systems have the advantage of confidentiality and difficult to copy and steal. It has been reported that EEG signals show distinctness among individuals because of its dependence on anatomical and functional brain properties [5] [6]. This property has a great impact in the development of more reliable biometric system. However, EEG based human recognition technique has been less worked area and needs some improvements as far as individual's state and number of channels is concerned. An effort has been made to address these issues in the present work.

The first issue while dealing with EEG based system is to get the common brain state at which the signals can be recorded (registration), so that the similar state can again be attained at the time of identification. For that reason eye open resting

state is selected by putting the subject in complete rest. Resting state has advantages over other states as it disallows active movement and user is not required to memorise how he performed certain actions at the time of registration. Ocular blinking is avoided in order to protect signals from getting contaminated. Here data is acquired through 64 channels at 160 Hz. EEG machine is placed on the scalp which follows the standard 10-10 montage system and then experimentally selects only few of those channels which get maximum of brain information and ignores the rest. While taking data it is to be confirmed that there is no electrical interference because of any heavy nearby electrical machines, as well as the subject is completely relaxed in order to get rid of all previous mental activities. To achieve this one can cover the room completely with dim light with complete silence. The data was pre-processed to eliminate noise. Originally data is received with the sampling rate at 160 Hz. Since human brain normally generates signals up to 40-50 Hz so data is required to be passed through a low pass filter in order to remove noises. Afterwards relationships among all the signals (coherence) are calculated and are made as features. In an attempt to address the variability of amplitude from subject to subject a bivariate measure, called Magnitude Squared Coherence (MSC), is used. This measure depends on phase constancy while one area of brain interacts with the other [4]. Features are found in the form of vectors corresponding to each class and a distance based classifier (K-nearest neighbour) is used to find the nearest class by calculating the Euclidean distance from testing data. Accuracy with 10 channels confirms the notion that there are various redundant channels in 64 channel EEG device whose impact can be ignored in order to find the distinctive features of each individual. This lesser number of channels will increase the usability of the device for biometric identification and decrease the number of calculations to make it fast.

II. PREVIOUS WORK

Many of the researchers have done some work in the field of identification using brain signals. D. La Rocca et al [4] has done work by carrying spectral coherence as a feature, and has shown very good recognition accuracy with 64 channels. This work proves that the spectral coherency feature is the property of entire brain regions and their relationship with each other can be used to characterize individuals. They have fused many

channels together and shown the effect on the result. They have also compared the results produced by considering coherence as a feature with the results produced by considering power spectrum density(PSD) as a feature. Se'bastien Marcel and Jose' del R. Milla'n [1] has used the posteriori model adaptation technique for training while considering individual data in the form of Gaussian mixture model. M. Poulos and M. Rangoussi [13] have used neural network (Learning vector quantizer) by considering only the alpha rhythm of EEG signal, to reach 80 to 100 % correct classification and claimed that EEG indeed carries genetic information. De Vico Fallani F. et al [2] have used PSD as the feature and used naïve classifier and k-fold cross validation and got a recognition rate of 78% during eye open resting state.

III. DATASET AND PREPROCESSING

In this work, the EEG signals are taken from online available Physionet database [8] [9], where eye opened data on resting state for 1 minute for 109 healthy subjects are used, Each signal is obtained at 160 Hz. Since most of the human brain activities are limited up to somewhere between 40 Hz to 50 Hz [3] and higher values signifies noise, therefore a low pass filter is used to restrict the input frequency up to 50 Hz. Butterworth low pass filter of 5th order is used here for this purpose. For data acquisition 64 channels EEG device has been used following the common 10-10 montage system. Acquired data from all 64 channels was reanalysed on experimental basis and then only 14, 10, 6 and 5 most prominent channels were selected and accuracy was measured. It is already known that most of the channels just carry some sort of common information leading to redundancy [14]. One minute input signal was segmented in 6 parts of 10 seconds each, by which 5 segments were used for training purpose and rest one segment was used for testing purpose.

IV. THEORY

EEG has some genetic information because of anatomical and functional traits of mind. This theory has already been proved [5][6]. Entire brain regions cumulatively lead to uniqueness. When two different brain regions are involved in similar activities, they are supposed to exchange information [15][16]. Therefore, we need to use statistical interdependent technique, like coherence, in order to find relation among all brain regions. This relation depicts the distinctiveness of that particular brain. Hence, it is possible to use this property for biometric identification.

A. Magnitude Squared Coherence (MSC)

Unlike power spectrum density which is a univariate measure, and depicts the property of each channel individually, the spectral coherence is a bivariate measure. MSC is insensitive to amplitude changes in the EEG oscillations [12]. As changes in amplitude of EEG of the same person during the same condition is due to physiological circadian rhythm [10][11]. Coherence (C_{xy}) between two signals x, y at frequency 'f' is

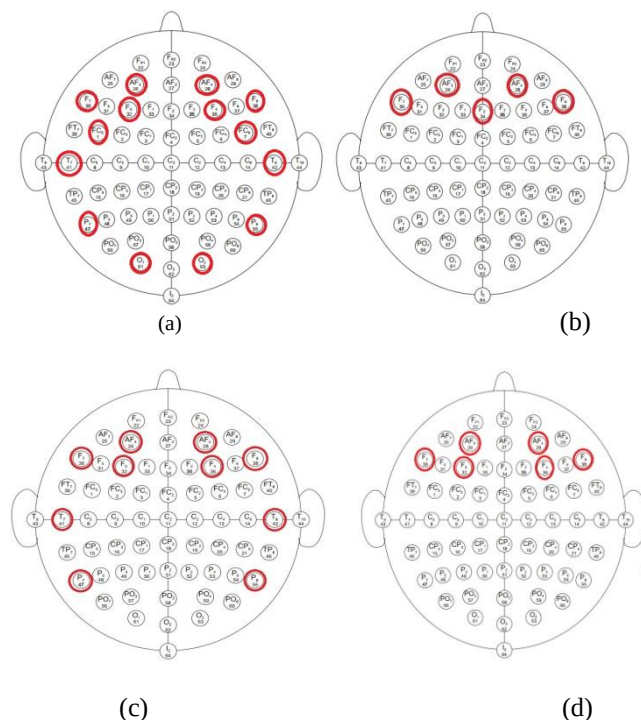


Figure1. Places of electrodes on scalp

calculated as follows:

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)}$$

Where P_{xy} is cross spectra, while P_{xx} and P_{yy} are the respective auto spectra. The value of MSC lies between 0 and 1. If two signals are very much similar on the basis of certain characteristics their coherence value will be 1, while if they are completely different their value will be 0 and for other cases it will be between 0 and 1. For calculating the coherence, window size is kept at 160 (the sampling frequency) and Number of points for FFT are set as 100 in order to have unit frequency resolution.

B. Channel Selection

Out of all 64 channels, AF3-26, F7-30, F3-32, FC5-1, T7-41, P7-47, O1-61, O2-63, P8-55, T8-42, FC6-7, F4-36, F8-38, AF4-28 are the 14 selected channels which cover Frontal lobe, Temporal lobe, Occipital lobe and Parietal lobe. AF3-26, F7-30, F3-32, T7-41, P7-47, P8-55, T8-42, F4-36, F8-38, AF4-28 are selected 10 channels and AF3-26, F7-30, F3-32, F4-36, F8-38, AF4-28 are the selected 6 channels and AF3-26, F7-30, F3-32, F8-38, AF4-28 are selected 5 channels. These 10 or 6 channels are selected on the basis of experiment which provides minimum redundancy in EEG signals with better accuracy.

C. Classifier

An appropriate classifier is that which can correctly predict the class of testing data given what it knows. We need to train the classifier for maximum variations of the input. These

maximum variations are termed as maximum learning. More the variations a classifier learns better the output it produces. However, in this case only a limited number of samples (5 for training) per individual is available.

There are many classifiers and according to the situation we select the one. Distance based classifier (DBC) works in many of the situations when sampled training data is not too much while there are many classes. DBC is very simple to use. Here we just need to calculate the distance of test data with all of the existing points and simply find out the class which is nearest.

V. METHODOLOGY

Acquired signal was of 1 minute per individual which was broken down into 6 segments of 10 seconds each. 5 segments were used for training purpose and remaining 1 segment for testing purpose. In one minute data each individual generates $160 \times 60 = 9600$ values per channel (160 Hz), where 1600 values belong to each segment. The low pass filtered signal for each pair of channel has sampling frequency as 100 and accordingly number of points for FFT has been set. All steps are shown in figure 2 with the help of a flow diagram.

A. Feature Selection

Coherence between two signals is considered as feature value. Therefore MSC (magnitude squared coherence) between every two channels is calculated. In the case of EEG signals if two oscillatory signals have stable phase relation with each other, the coherence value will be maximum i.e. 1, while in highly random phase [7], the value will be 0. For 'n' channels, the number of combinations is given by $n(n-1)/2$. When the value of 'n' is 14, 10 and 5, there are 91, 45 and 10 combinations respectively. Since brain signals up to 40 Hz are only considered, therefore we have taken only 40 values of the coherence. For each individual the size of feature vector will be $45 \times 40 = 1800$ (at $n=10$) features per sample and we have 5 samples of each individual which makes a total of 545 number of sampled feature points ($109 \times 5 = 545$). Therefore feature matrix is of size 1800×545 , and the testing matrix is of size 1800×109 (1 sample per individual). After calculating the coherence between every two channels we calculated z-score of this in order to normalize the distribution [12].

B. Classification

As, number of samples were limited and we need to try incremental classifier architecture, a distance based classifier (k-nearest neighbour) was chosen. In distance based classifier there are 5 points for each of the 109 classes, resulting in a total of 545 training points.

When a testing sample was collected we again calculated $40 \times 45 = 1800$ feature values exactly in the same way as for training samples. Then the Euclidean distances between 1800 testing feature points (X_i s) of the testing individual and the feature points of each (109×5 samples/ individual = 545) sample were calculated using the formula stated below.

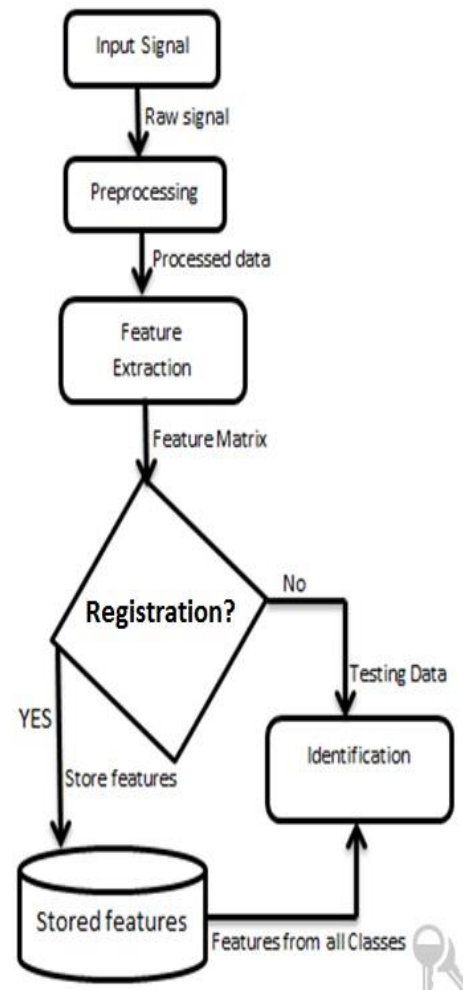


Figure 2. Flow diagram of all process

$$d_{j,k} = \sqrt{\sum_{i=1}^{1800} (X_i - Y_{i,jk})^2}$$

Where X is one of the testing points and Y_j is the training feature point for k^{th} sample of j^{th} individual. j lies between 1 to 109. $d_{j,k}$ is the Euclidean distance between testing point and k^{th} sample of j^{th} training point.

VI. RESULTS AND DISCUSSION

Now all the distances were compared and nearest one was selected which depicted the similarity between two points. We had a total of 545 points, therefore to get the result two approaches were followed: 1) To find the nearest point from each 5 samples of each class(individual) and then select the final minimum value denoted by equation 1(KNN with $k=1$). 2) Find distances of testing data feature points from the calculated mean value of each feature point from all the 5 samples of each class. Then get the minimum (nearest) distance class (assuming points to follow Gaussian distribution) which we call approach 2.

$$d_j = \min(d_{j,k}) \quad (1)$$

$$d_j = \sqrt{\sum_{i=1}^{1800} (X_i - Y_{i,jmean})^2} \quad (2)$$

Where, j lies from 1 to 109 and k lies from 1 to 5. $jmean$ is the mean of all 5 samples for j^{th} individual. Equation (1) is based on approach 1 and equation (2) follows approach (2). The testing sample belongs to the closest class j (which ranges from 1 to 109) such that:

$$j = \arg \min(d_j) \quad (3)$$

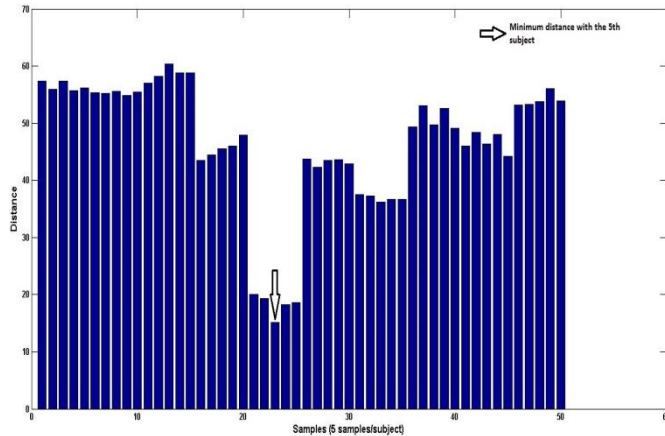


Figure 3. Testing of 5th subject

Fig.3 shows that when 5th subject is tested, its distance is convincingly closest with 23rd point while we had 5 points per class. Here approach1 (to get min distance) seems pretty straight forward. Approach 1 is giving better result than approach2 because we have limited number of training samples (5 for each) while there are too many number of classes. Hence, it is quiet probable that one among the 5 points may behave differently than others which will draw the center away and will misclassify it. Both the approaches for 109 subjects are tested and results are tabulated below.

Num. of Channels	Accuracy (109 Subjects)			
	64	10	6	5
Proposed Approach 1	100%	100%	97.24%	95.4%
Proposed Approach 2	100 %	99.1%	91.7%	89%

Table 1: Observed Accuracy

Previous research [4] has shown 100% accuracy by selecting 64 channels with the assumption that samples follow Gaussian distribution. This study shows that if we select only 10 channels, then also we get 100% accurate identification by applying nearest neighbour classifier. The result also shows that by selecting 5 channels we can get more than 95% accuracy.

VII. CONCLUSION

EEG based biometric system has tremendous advantages over other existing biometric techniques as it is very difficult to forge. It has to get user's willingness to authenticate therefore forceful authentication is not possible. Despite having practical importance of EEG based system, it is not very easy to get the user's noise free data as electrodes are so susceptible to noise that taking EEG data in common environment needs lots of effort. Besides these environmental effects human's own artifacts (heartbeat, breathing, etc.) play negative role in the process. We need to take common resting state very carefully as in the presence of retention of previous brain activity (mood) it will have higher probability to get misclassified.

Signals from all main portion of brain regions are required in order to get more subject dependent information [13][4]. 95.4% accuracy with 5 channels of EEG proves the idea that covering whole brain is more significant rather than covering each portion intensively. Hence 10 channels which cover all frontal, parietal, and temporal lobes, give same accuracy (100%) as 64 channels do.

References

- [1] Marcel, S.; Millan, J.D.R., "Person Authentication Using Brainwaves (EEG) and Maximum A Posteriori Model Adaptation," *Pattern Analysis and Machine Intelligence*, IEEE Transactions on , vol.29, no.4, pp.743,752, April 2007
- [2] Campisi, P.; Scarano, G.; Babiloni, F.; DeVico Fallani, F.; Colonnese, S.; Maiorana, E.; Forastiere, L., "Brain waves based user recognition using the "eyes closed resting conditions" protocol," *Information Forensics and Security (WIFS)*, 2011 IEEE International Workshop on , vol., no., pp.1,6, Nov. 29 2011-Dec. 2 2011
- [3] D. Mantini, M. G. Perrucci, C. D. Gratta, G. L. Romani, and M. Corbetta, "Electrophysiological signatures of resting state networks in the human brain," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 104, no. 32, pp. 13170–13175, Aug. 2007
- [4] La Rocca, D.; Campisi, P.; Vegso, B.; Cserti, P.; Kozmann, G.; Babiloni, F.; De Vico Fallani, F., "Human Brain Distinctiveness Based on EEG Spectral Coherence Connectivity," *Biomedical Engineering, IEEE Transactions on* , vol.61, no.9, pp.2406,2412, Sept. 2014.
- [5] J. Berkhout and D. O. Walter, "Temporal stability and individual differences in the human EEG: An analysis of variance of spectral values," *IEEE Trans. Biomed. Eng.*, vol. BME-15, no. 3, pp. 165–168, Jul. 1968.
- [6] H. Van Dis, M. Corner, R. Dapper, G. Hanewald, and H. Kok, "Individual differences in the human electroencephalogram during quietwakefulness," *Electroencephalogr. Clin. Neurophysiol.*, vol. 47, pp. 87–94, 1979.
- [7] P. L. Nunez, *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford, U.K.: Oxford Univ. Press, 2006.
- [8] Database physionet bci. [Online]. available:<http://www.physionet.org/pn4/eegmmidb/>.
- [9] Bci2000 system. [Online]. Available: <http://www.bci2000.org>
- [10] F. G. Andres and C. Gerloff, "Coherence of sequential movements and motor learning," *J. Clin. Neurophysiol.*, vol. 16, no. 6, pp. 520–527, 1999.
- [11] D. Aeschbach, J. R. Matthews, T. T. Postolache, M. A. Jackson, H. A. Giesen, and T. A. Wehr, "Two circadian rhythms in the human electroencephalogram during wakefulness," *Amer. J. Physiol. Regul. Integr. Comp. Physiol.*, vol. 277, no. 6, pp. R1771–R1779, 1999.
- [12] A. Amjad, D. Halliday, J. Rosenberg, and B. Conway, "An extended difference of coherence test for comparing and combining several independent coherence estimates: Theory and application to the study

- of motor units and physiological tremor,” *J. Neurosci. Methods*, vol. 73, no. 1, pp. 69–79, 1997.
- [13] M. Poulos, M. Rangoussi, and N. Alexandris, “Network based person identification using EEG Features”, *IEEE Conf.*, 1999
- [14] Lin He, Zhenghui Gu, Yuanqing Li, Zhuliang Yu “Classifying Motor Imagery EEG Signals by iterative channel elimination according to compound weight” *Springer-Verlag Berlin Heidelberg* 2010
- [15] T. Akam and D.M. Kullmann, “Oscillations and filtering networks support flexible routing of information,” *Neuron*, vol. 67, no. 2, pp. 308–320, 2010.
- [16] L. Astolfi, F. Cincotti, D. Mattia, S. Salinari, C. Babiloni, A. Basilisco, P. M. Rossini, L. Ding, Y. Ni, B. He, M. G. Marciani, and F. Babiloni, “Estimation of the effective and functional human cortical connectivity with structural equation modeling and directed transfer function applied to high-resolution EEG,” *J. Magn. Reson. Imag.*, vol. 22, no. 10, pp. 1457–1470, 2004.