Proceedings of the
7th Graz Brain-Computer Interface Conference 2017 DOI: 10.3217/978-3-85125-533-1-42

# RESTING EEG-BASED SUBJECT IDENTIFICATION SYSTEM: A PRACTICAL SCENARIO FOR OFFLINE ANALYSIS

D. Kim, K. Kim

School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea

E-mail: {dhkim518, kskim}@gist.ac.kr

ABSTRACT: Rapid advancements in brain-computer interfaces (BCIs) have fostered the wide-spread application of biometric systems built on electroencephalography (EEG) measurements in recent years. While current subject identification approaches exhibit high performance, most efforts rely on limited number of EEG recording runs to validate the feasibility and practical usability of EEG-based biometric systems. For a realistic system however, quantitative assessment over long time series is essential due to the intrinsic non-stationary behaviour of EEG signals. In this work, we propose a more practical scenario of the resting EEG-based subject identification system for off-line analysis using public dataset. Moreover, we implement a subject recognition system based on widely-adopted functional connectivity measures. The system is applied to assess performance variations under consecutive EEG recording runs over time. We also provide a simple approach to overcome the performance degradation and eventually, raise several issues as potential future works relating EEG-based systems.

## INTRODUCTION

In spite of copious research in BCI technology and its applications in neuroscience and cognitive psychology, it is only recently that the possibility of EEG measurement utilization in biometric systems has emerged [1-3]. Unlike existing technologies that use fingerprints, speech, iris, and facial features as identifiers for human authentication, EEG signal stands out as a unique physiological biometric that is robust against forgery and falsification.

Various paradigms have been proposed for biometric systems based on event-related potentials (ERP) [4,5] or specific tasks related mental states [6]. Since these brainwaves are evoked after a specific time onset, the system is possible to utilize the temporal characteristics of brain signals. Nonetheless, EEG in resting state manifest beneficial advantages regarding user-friendliness in that it does not require additional stimulatory devices and any conditions/ instructions for subjects. In addition, the high potentials of biometric systems based on resting state EEG should not be undermined as these resting state signals differ considerably among subjects [7].

State-of-art studies of brain functional connectivity modulated by resting state networks have advanced our understanding of human brain functions. The traits of functional connectivity are reported to fully exploit physiological information and represent functionally coupled brain regions thus, serving as appropriate tools for EEG-based biometric systems yielding high performance.

Although a number of works based on various functional connectivity measures have been reported [8,9], the system performance were validated within a limited scenario comprising of only a single run of EEG recordings. In particular, researchers have confirmed the potential usability and feasibility of such systems by applying their proposed methodologies on divided non-overlapping EEG segments resulting from the one recording run. In the realm of practical adoption however, such biometric systems should be able to operate successfully in the presence of discontinuous signal inputs as well. Especially, the performance and reliability of these systems should be scrutinized prior to realization since EEG signals have inherent non-stationary characteristics due to various physical and mental drifts.

In this study, we mainly focus on a subject identification system among the various biometric-related topics. Given an input EEG signal, the identification system is required to find and recognize the correct subject label from a database containing various EEG information of several subjects as depicted in Fig. 1. We aim to propose a more realistic and practical scenario of an EEG-based subject identification system for off-line analysis using publicly available dataset. The proposed scenario allows us to determine whether a novel approach would perform reliably as we treat discontinuous trials over time. We then implement the wll-known EEG-based biometric system exploiting the functional connectivity measures and validate it for the proposed scenario. Furthermore, we present a simple approach to enhance the performance of the system and highlight certain issues to be considered as future work.

## METHODS

In this section, we describe the approaches undertaken. Our main focus is to investigate and comprehend the performance variation of subject identification incurred by non-stationary EEG signals as the recording runs are processed gradually. Simulations are conducted using public

DOI: 10.3217/978-3-85125-533-1-42

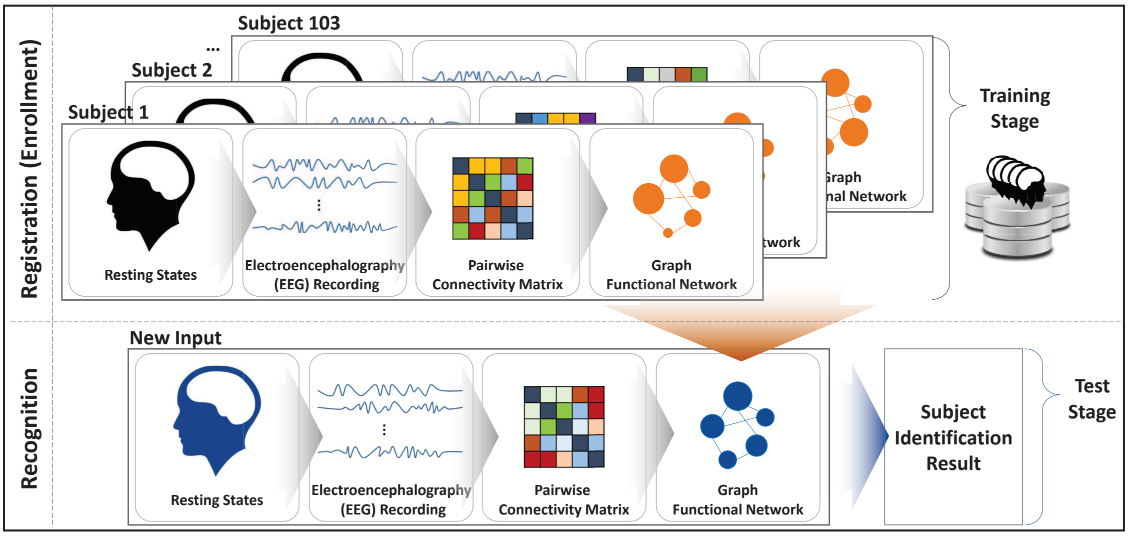


Figure 1: Schematic diagram of electroencephalography-based subject identification system.

dataset and the accuracy assessed under different spectral bands using two prominent feature extraction methods for multiple test runs over time.

### A. Dataset

The publicly available “Physionet EEG Motor Movement/Imagery Dataset” acquired by BCI2000 system from 109 subjects was used in our study [10,11]. The EEG signals were recorded from 64 scalp electrodes at the international 10-10 locations and 160 Hz sampling rate was used. The dataset consists of two baseline tasks (eye-open resting state and eye-closed resting state) and twelve motor movement (MM)/ motor imagery (MI) tasks. In our study, we used the twelve MM/I dataset were used, of which the first run was designated as the training data for an enrolment stage and the remaining eleven runs as consecutive testing data for the recognition stage of the performance evaluation process. By excluding subjects with less than fifteen conducted MM/I trials from the dataset, our experiment was finally carried out for a reduced set of 103 subjects.

### B. Pre-processing

In the signal pre-processing phase, the dataset was epoched to a range of -4 to 0 seconds with respect to the instructing MM/I target onset. We assumed that the EEG signal in this time period corresponds to the resting states awaiting the cue instruction. Since each subject produced fifteen trials, the fifteen resting state epochs prior to the MM/I trials are extracted in each run. The non-overlapping epochs of 4 seconds were treated as separate trials in the subject identification system.

The extracted epochs were band pass filtered separately to account for the effects of EEG spectral bands. The six different frequency bands that were considered are delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), gamma (30-50 Hz), and the complete range (0.5-50 Hz). The accuracy results were analysed and presented in terms of the spectral ranges of the EEG signals.

### C. Feature Extraction

Based on recently reported methods for biometric subject identification, we applied widely-used brain connectivity analysis schemes on the EEG signals. We now describe the methods to obtain inter-channel connectivity.

- **Inter-channel relation measures:** We calculated two of the most widely-used brain functional connectivity metrics namely, inter-channel Pearson’s correlation coefficient (COR) [12], and inter-channel phase lag index (PLI) [9,13]. The inter-channel COR reveals the strength of the linear relation of pairwise channels by magnitude as it varies from -1 to 1. It measures linear correlation in the time domain between two signals at zero lag. On the other hand, the PLI measure discards phase distributions that center around 0 mod $\pi$, in order to be robust against the presence of common sources. The PLI calculates the asymmetry of the distribution of phase differences between two signals varying from 0 to 1, and serves significant in volume conduction and active reference problems.

- **Eigenvalue centrality (EC):** Indicators of centrality represent the most important nodes within a graph in the context of network analysis. One such measure is EC which is used to characterize connectivity and has evidently shown high accuracy performances for brain connectivity-based biometrics [9]. We computed EC for the inter-channel Pearson’s correlation coefficient, and phase lag index which are structured as $64 \times 64$ matrices (the number of channels is 64). The EC result of each matrix was used as a feature vector. The functional connectivity with EC approach with respect to PLI showed high performance in literature [9].

### D. Feature Classification

 DOI: 10.3217/978-3-85125-533-1-42

For off-line analysis of the biometric system, dataset composed of twelve runs from 103 subjects were used. By labelling each individual as a distinct class, the system can be treated as a multi-class classification problem with 103 classes. The fifteen epochs in the first MM/I run from all subjects were allocated to training trials, while the remaining eleven runs (consisting of fifteen epochs each) were assessed in a consecutive manner.

As mentioned earlier, the inter-channel COR and PLI-based ECs were computed for each spectral band of the resting state EEG signals. Using the feature vectors extracted by EC, we then calculated the Euclidean distance for the classification process. The subject label showing the minimum Euclidean distance between the training dataset and test trial were then compared for every single test trial of the 103 subjects. We treated the fifteen epochs of all subjects in each test run as fifteen sets composed by 103 subjects' single epoch. In every test run (a total of eleven test runs), the classification procedure was repeated fifteen times for a sequence of 103 subjects' trials. The accuracy was calculated by classifying the epoch out of all the subjects. Finally, we computed a mean and a standard deviation of the accuracies in each test run.

## RESULTS

Simulations results for performance variation with consecutive multiple runs with respect to different spectral ranges and functional connectivity measures are provided in this section.

Fig. 2 illustrates the performance variations plots for consecutive test runs at six different spectral bands. In each run, fifteen trials from 103 subjects were tested to identify the correct subject label. The accuracy was calculated based on the number of correct subject identification results, and represented as the mean and standard deviation from the fifteen repeated recognition results in every run. In general, the resting EEG states in beta (13-30 Hz) and gamma (30-50 Hz) bands showed relatively higher accuracy compared to the other spectral bands. Our results are in accordance with the existing works that have reported high classification rates in these two bands [9]. Although the COR-based approach resulted in better performance as compared to its PLI counterpart in most spectral bands, the PLI-based functional connectivity feature shows highest accuracy in the gamma spectrum. However, the performance of this approach in the gamma band also decreased to $57.67 \pm 3.8\%$ over time in the worst case.

The classification matrices of eleven test runs using the PLI-based approach in the gamma band are shown in Fig. 3. Each matrix represents the estimated results for each subject given the actual labels of subject indexes. Therefore, the diagonal of each matrix shows the result of correct classification, whereas the remaining elements correspond to the mis-classification results. We observe that with increase in the number of test runs, the number of mis-classified subjects also gradually increases. For sake of comparison, we reproduced the optimistic identification results ($> 90\%$) reported by the authors in [9] by implementing the same signal processing procedure (band pass filter, PLI, EC, and Euclidean distance-based classification) and evaluated the system performance for our proposed scenario.

Apart from obtaining high accuracy in several runs, we also account for reliable and persistent performance over long time series as important assessment measures in our scenario. To improve the performance over time, we included the mean values of various combinations of the training dataset as additional features for classification. Since the EEG signals are non-stationary, even a small number of training sets may possess non-stationary traits. Hence, calculating the mean values from several EEG segments would de-noise and reduce variations of the training sets. We repeated the same simulation process after adding training features made by taking the sum of three consecutive trials, sum of five consecutive trials, and sum of all the trials. Fig. 4 displays performance variation in gamma bands for the added features. Through such simple addition of enhanced features, the PLI-based performance of the worst case ($57.67 \pm 3.8\%$) improved to $61.93 \pm 3.12\%$ (around $4.2\%$ accuracy improvement). We observe that understanding the stationary features can be a possible solution to overcome performance degradation over time.

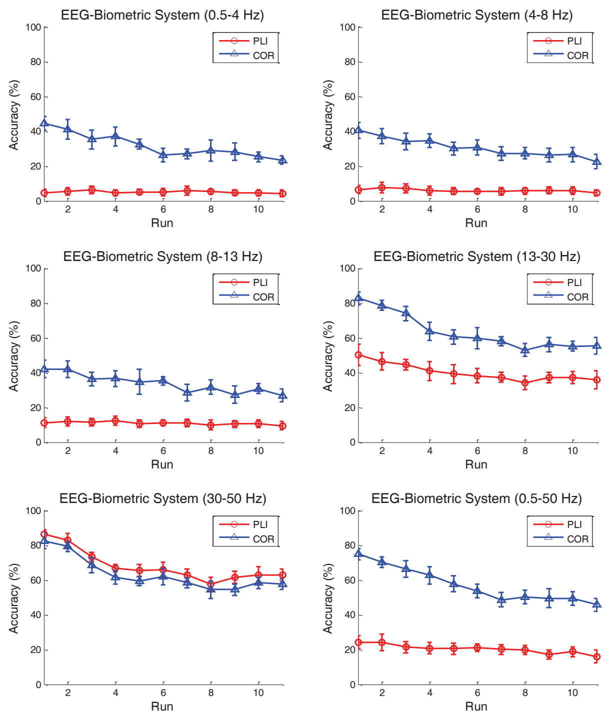


Figure 2: Performance evaluation of consecutive test runs using phase lag index (PLI) and Pearson's correlation coefficient (COR)-based approaches in six different spectral bands.

 DOI: 10.3217/978-3-85125-533-1-42

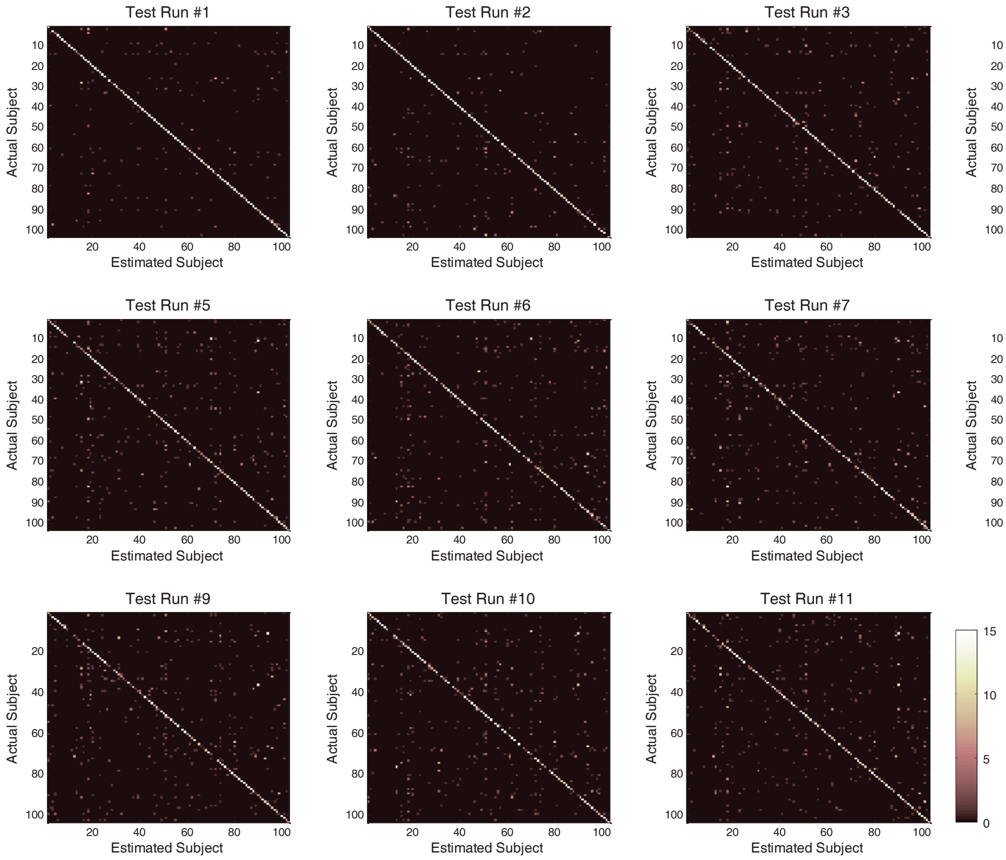


Figure 3: Classification matrices of eleven test runs in gamma band ($30 - 50$ Hz) where phase lag index-based approach is applied.

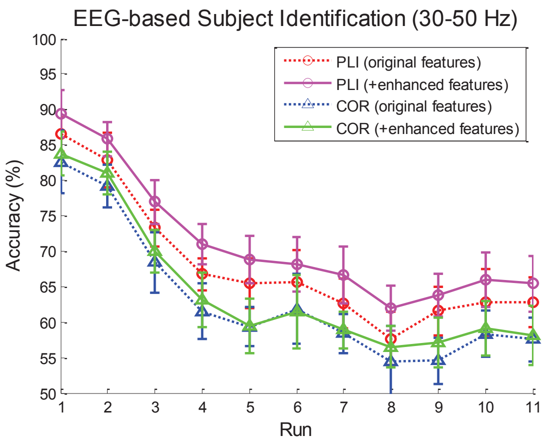


Figure 4: Performance comparison of consecutive test runs using phase lag index (PLI) and Pearson's correlation coefficient (COR)-based approaches in gamma band and comparison with enhanced feature sets.

## DISCUSSION

As EEG signals exhibit non-stationary characteristics, BCI features vary from training to testing stages during a BCI experiment. Through off-line analysis, we confirmed that the EEG-based biometric system requires a novel methodology to ensure stable operation over time.

One possible approach to enhance the permanence of such biometric systems is the development of signal processing techniques to identify stationary traits hidden in the EEG signals of individual subjects and improve accuracy via a simple averaging approach. One of the issues concerning data processing is to find a unique feature appearing individually and a common feature prominent in all the subjects. The distinct stationary signal features extracted using machine learning and pattern recognition techniques eventually result in the enhanced performance of the biometric subject identification system.

Not only would the stationary traits of EEG sources from training dataset (feature extraction step) enhance the performance, but also would the proper classifier derived by training dataset (classification step). Although linear classification based on Euclidean distance measure was adopted in the our simulation environment, other forms of non-linear classifiers have potentials to strengthen the robustness and improve the reliable accuracy over long runs. Particularly, approaches built on deep neural networks that find and learn the features directly from the data would serve more efficient when handling large volumes of data.

Another appealing approach would be designing an adaptive experimental paradigm for biometric system protocols. Though we assumed the EEG signals before instructing target onset as the resting state, subjects may

 DOI: 10.3217/978-3-85125-533-1-42

reveal different mental states, for instance, expecting the next instruction, being affected by the surrounding environment, feeling exhausted due to long experimentation time. The protocol design to maintain stationary mental states between training and test stage would therefore, enhance the biometric system.

## CONCLUSION AND FUTURE WORKS

In this paper, we demonstrated the performance evaluation of an EEG-based subject identification system for a more practical scenario via off-line analysis. We implemented the well-known functional connectivity measures for the biometric system which showed high performance in single run recording data. We further applied the state-of-art biometric system to consecutive runs over time and observed the accuracy degradation, and consequently, highly emphasized on prominent issues regarding the necessity of the realistic scenario for rational and reliable evaluation of the practical EEG-based subject identification system. Finally, We discussed possible approaches to overcome current limitations and provided research guidelines for follow-up in future works.

We expect the EEG based biometric system would be spread to be accessed gradually from limited environments requiring an advanced security (such as banks, military or companies) to personal usages. To be the practical application, the longitudinal studies exploring uniqueness and distinctiveness of the brainwaves are fundamentally required.

## ACKNOWLEDGEMENT

This research was supported by the Brain Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Science, ICT & Future Planning (NRF-2016M3C7A1905477), and partly by the "Basic Research Projects in High-tech Industrial Technology" Project provided by GIST.

## REFERENCES

[1] DelPozo-Banos M, Travieso CM, Weidemann CT, Alonso JB. EEG biometric identification: a thorough exploration of the time-frequency domain. J. Neural Eng. 2015;12(5):056019.

[2] DelPozo-Banos M, Alonso JB, Ticay-Rivas JR, Travieso CM. Electroencephalogram subject identification: A review. Expert Syst. Appl. 2014;41(15):6537-6554.

[3] Campisi P, La Rocca D. Brain waves for automatic biometric-based user recognition. IEEE Trans. Inf. Forensics Secur. 2014;9(5):782-800.

[4] Palaniappan R, Mandic DP. Biometrics from brain electrical activity: A machine learning approach. IEEE Trans. Pattern Anal. Mach. Intell. 2007;29(4):738-742.

[5] Ruiz-Blondet MV, Jin Z, Laszlo S. Cerebre: A novel method for very high accuracy event-related potential biometric identification. IEEE Trans. Inf. Forensics Secur. 2016;11(7):1618-1629.

[6] Yang S, Deravi F, Hoque S. Task sensitivity in EEG biometric recognition. Pattern Anal. Appl. 2016:1-13.

[7] Laufs H et al. Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. PNAS, 2003;100(19):11053-11058.

[8] La Rocca D et al. Human brain distinctiveness based on EEG spectral coherence connectivity. IEEE Trans. Biomed. Eng. 2014;61(9):2406-2412.

[9] Fraschini M, Hillebrand A, Demuru M, Didaci L, Marcialis GL. An EEG-based biometric system using eigenvector centrality in resting state brain networks. IEEE Signal Process Lett. 2015;22(6):666-670.

[10] Goldberger AL et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation. 2000;101(23):e215-e220.

[11] Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. IEEE Trans. Biomed. Eng. 2004;51(6):1034-1043.

[12] Riera A, Soria-Frisch A, Caparrini M, Grau C, Ruffini G. Unobtrusive biometric system based on electroencephalogram analysis. EURASIP J. Adv. Signal Process. 2008;2008:18.

[13] Stam CJ, Nolte G, Daffertshofer A. Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. Hum. Brain Mapp. 2007;28(11):1178-1193.