2010 International Conference on Pattern Recognition

EEG-based Personal Identification: from Proof-of-Concept to A Practical System

Fei Su[1], Liwen Xia[1], Anni Cai[1], Yibing Wu[2], Junshui Ma[3]

[1] *Beijing University of Posts and Telecommunications, Beijing, China*

[2]*Easymonitor Technology Ltd.,Co. Beijing,China*

[3] *Merck Research Laboratories, Merck & Co, Inc., Rahway, NJ, USA*

sufei@bupt.edu.cn, xialiwen@bupt.cn, annicai@bupt.edu.cn, wg6122@163.com, junshui_ma@merck.com

***Abstract*—Although the concept of using brain waves, e.g. Electroencephalogram (EEG), for personal identification has been validated in several studies, some unanswered practical and theoretical questions prevent this technology from further development for commercialization. Based on a well-designed personal identification experiment using EEG recordings, this study addressed three of these questions, which are (1) feasibility of using portable EEG equipment, (2) necessity for controlling factors influencing EEG, (3) the optimal set of features. With our understanding of the answers to these questions, the EEG-based personal identification system we built achieved an average accuracy of 97.5% on a dataset with 40 subjects. Results of this study provided supporting evidence that EEG-based personal identification from proof-of-concept to system implementation is promising.**

Keywords-EEG; biometircs; personal identificaiton; portable EEG recording equipment

I. INTRODUCTION

In the past ten years, several studies have been proposed using brain waves, e.g. EEG as a biometric modality [1-11]. Compared with face, fingerprint, voice, iris, signature or other widely used biometrics, EEG has two distinct advantages. First, the active EEG must come from a living individual with a normal mental state, and an aggressor cannot force the person to provide the ideal EEG signals as those recorded in normal states. Second, EEG is hard to mimic. Therefore, EEG can be an excellent complement modality to the existing biometric systems which are prone to forgery [12]. However, several unanswered practical and theoretical questions prevent this technology from further development for commercialization.

Almost all the published studies on EEG-based biometrics used EEG data recorded by medical equipment. The conventional EEG recording system for medical use is expensive and time consuming for recording, which makes it inconvenient to be used in personal identification. As we know, properly preparing human scalp for this kind of EEG recording equipment remains a time-consuming process. Furthermore, the wet-electrode used by the EEG recording system needs to apply conductive gel between electrodes and the human scalp. These restrictions hinder the applications of EEG-based personal identification system in practice. Meanwhile, with the progress in the development of EEG recording technology, some portable EEG recording systems are emerged as the peripheral equipments to allow people use their EEG to play computer games or control toys [13]. However, although this kind of product has been successfully demonstrated to allow people to use their brain waves to control a computer game, no evidence is available for us to know whether its signal quality is good enough for personal identification.

Some people doubt EEG can be a practical biometric modality because many factors, including diet, time of day etc., can alternate a person's EEG signal, and it is impossible to control these factors in a practical personal identification system. However, whether these factors need to be controlled to achieve high performance in an EEG-based biometric system is unknown. Also, some other people concerned the long EEG collection time required in some existing publications could damage its practical usage. However, the minimum EEG recording time to achieve acceptable system performance is unknown.

Different feature sets are used in different publications. The performance of these feature sets is not compared in the same context. Therefore, people still have no clear idea of which feature set to start with when implementing an EEG-based biometric system.

Based on a well-designed personal identification EEG recording experiment, this paper addresses the questions previously mentioned. Hopefully, by answering these questions, we can reenergize the effort to move this technology from proof-of-concept to system implementation.

The rest of the paper is organized as follows. In section 2, the data collection method is described. Section 3 gives the related efforts. Section 4 provides our method and the detail experiments. Finally, the conclusion is given in section 5.

II. DATA COLLECTION

All EEG signals used in this paper were recorded using HXD-I portable equipment (see Fig. 1), which collects signals from FP1 electrode without requiring any skin preparation or conductive pastes. The sampling rate is 200Hz, and reference sensors are placed at both earlobes. Therefore, it is quite unobtrusive, fast and easy to place, which makes it possible to use EEG in personal identification.

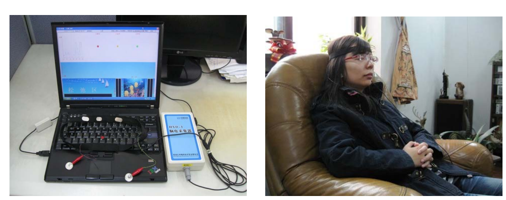

Figure 1. Portable EEG recording equipment and the recording scene.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.908

The EEG data collection procedure strictly followed a well-documented Standard Operation Procedure (SOP). 40 healthy volunteers' EEG signals were collected (29 male and 11 female). All volunteers were screened to ensure that they are healthy and not under any medication. The recording environment is quiet, normal temperature and daylight. Volunteer sits on a comfortable sofa. After a standard instruction was read, a segment of five-minute restful EEG signal was recorded when the volunteer keeps his/her eyes close (see Fig. 1).

The data collection follows a two-period crossover procedure. Each volunteer's EEG was recorded in two separate days, which was considered as two periods. In one period, a cup of pure water was drunk. In the other period, the same amount of coffee was drunk at the same time of the day. Here we use coffee to represent a diet that potentially has a big impact on EEG. In each period, 6 EEG sessions were recorded at different time points, including a pre-dose session and five post-dose sessions. Immediately after the pre-dose EEG recording session was finished, the volunteer was instructed to drink a cup of coffee or water, i.e. conducting the drinking event. Then, five post-dose sessions were recorded at 0.5 hour, 1 hour, 1.5 hours, 2 hours, and 2.5 hours after the drinking event. We purposely record EEG at different times of day to investigate the impact of a person's circadian effect on EEG signal. Each five-minute recording is an EEG event here. There are totally 480 EEG events (i.e. 1 event per EEG session, 6 sessions per period, 2 periods per subject, and 40 subjects) in our dataset.

III. RELATED EFFORTS

A. Literature Review

Related research could be classified in two categories including EEG-based and Visual Evoked Potential (VEP)-based [10]. EEG-based methods use restful EEG signals recorded in case the subject opens/closes his/her eyes in relaxation [1-6], the other is to use the response when some mental tasks or stimuli are given to the subject [7-11]. In practical personal identification, we think EEG-based method is more convenient than VEP-based system where additional display terminal is necessary, and some mental tasks or stimuli should be arranged. Therefore, we only focus on EEG-based biometrics. In [1] and [2], Poulos used autoregressive (AR) model parameters or FFT based spectral analysis with a Learning Vector Quantizer (LVQ) network to classify an unknown EEG to one of four individuals, obtaining the accuracy around 80% to 100% depending on the individual. In [3], the EEG's second order statistics were computed using AR models of various orders. Discriminant functions applied to the model coefficients were used to examine the degree to which the subjects in the data pool could be identified. They obtained 80% accuracy rate in condition that half of the data were applied to compute the discriminant functions including 40 subjects. In [4], the beta waves (a frequency range of 14 to 50 Hz) were chosen as they are generated during functions like analytical problem solving, decision making and processing information. The extracted beta waves were processed using the Welch algorithm to extract power spectral density features, and a simple feed forward network with three layers was used to classify the six individuals using three mental tasks (relax, multiplication, reading). A maximum average classification of 97.5 % is achieved. Markus et al. [5] validated the test-retest reliability of resting EEG spectra as a statistical signature of persons through the generalized linear model (GLM). Chisei et al. [6] employed the variance of spectral power combined with the non-dominant region of the power spectrum as the feature vector used in personal verification. A user claims a unique identifier and then the EEG is measured. If the difference between the input and template feature vector is smaller than a threshold, the user is recognized as genuine. They obtain the maximum accuracy of 79% on a dataset of 23 subjects.

B. Feature Evaluation

In the published papers, AR model parameters and the spectrum of specific frequency band are common used features in EEG-based biometrics [1-4]. We implemented the methods given in [1-4] individually on our dataset to compare the performances of these feature sets in the same context. For validating the feature sets impartially, 1-NN classifier is used here.

There are total 40 subjects in our experiments. 12 EEG recordings per subject are used. For each subject, we randomly select 50 percent recordings for training, and use the remaining 50 percent recordings for testing in our experiments. The training and testing sets are generated randomly 100 times.

The experimental results using the methods in [1-4] on our dataset are shown as box-plots in Fig. 2. In [1], the recorded signal was spectrally analyzed by Fourier Transform and AR model between 7.5Hz to 12.5Hz was used. In [2], the spectral values of the EEG signal were broken into three overlapping sub-bands, namely [7-10Hz], [8-11Hz] and [9-12Hz], in order to investigate whether one of them was informative enough to represent the whole EEG signal for personal identification. In [3], 15 order AR model computing through Lattice Equivalent Model and Levinson Recursion was used. In [4], because three mental tasks were included, authors considered only the beta frequency band (14~50Hz) using Welch methods.

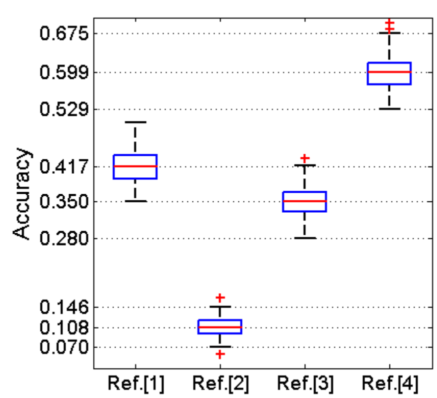

Figure 2. Performance comparison of different methods.

IV. Our Method and Experimental Results

A. *Proposed Feature Set*

Because we plan to use classifiers, e.g. K-nearest neighbor (K-NN) or Support Vector Machine (SVM), which can tolerate high feature dimension, we decide to use a feature set combining the AR model prediction coefficients and Power Spectrum Density (PSD). Also, we use the whole spectrum to construct the AR model instead of using a frequency band, as was done in [1]. The performance of an AR model with an order ranging from 10 to 50 was tested, and the optimal order is found to be 19, which is very close to the result suggested in [3].

In [2] and [4], the PSD at a particular frequency range were used, which was justified by some biological reasons. In contrast, the PSD of the full frequency range was considered here. However, because the PSD below 4Hz is frequently contaminated by ocular artifacts, and the PSD above 33Hz is affected by a system specific notch filter, we finally decide to incorporate the PSD at frequency range from 5 to 32Hz into our final feature sets, and each frequency bin was presented by four points. At last, the dimension of our feature vector is 127 (19 AR coefficients and 27 frequent bins with four points per frequency).

B. *Impacts of the EEG Recording Factors*

1) *Diet and Circadian Effects Evaluation*

For clarity, diet effect is evaluated first using a simple T-test and a Šidák multiple testing correction. We calculated the difference of the EEG power spectra between water and coffee for each session separately. The null hypothesis H0 was assumed to be that the spectral differences at each frequency bin obey zero-mean normal distribution. If the p-value is lower than the corrected significance level, H0 will be rejected. The result indicated that the main coffee effect is among 0.5 to 2 hours after the drinking events, but it has become weaker later.

Further studies have been conducted to give a quantitative measurement of diet and circadian effects on the identification performance through cross-training with effected and non-effected sessions. The results showed that almost 10% decrease in the performance was found by either the diet effect or the person's circadian. The results obtained above suggest that to get an optimal performance and make the EEG-based personal identification move from proof-of-concept to a real system implementation, more diverse samples (different diet, times of day, etc.) should be included in the training set.

2) *Recording Durations*

The system performance under different EEG signal recording durations is estimated to find out the optimal balance between shorter recording time and higher accuracy. Eight different EEG signal recording durations from 30 seconds to 3 minutes with interval of 30 seconds are investigated. The performance versus different recording durations is shown in Fig. 3. For each 5-minute recording, we randomly cut out 50 different segments of signal to simulate the data with a specific recording duration. As the same, 100 splits of the training and testing sets are randomly generated each time. Therefore, the averaged accuracy rate of 100 splits is computed for each time using a K-NN classifier with feature reduction by FDA, and then the overall accuracy on a specific recording duration is obtained after averaging the results of 50 times.

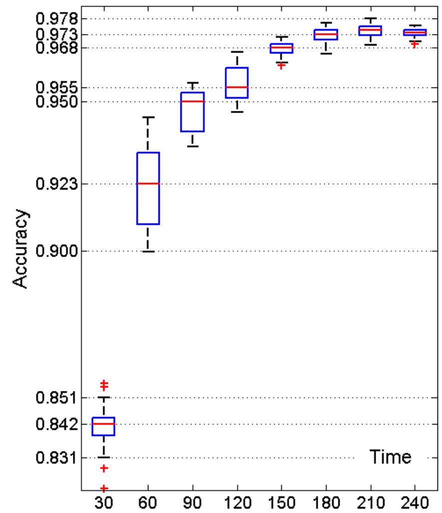

Figure 3. Accuracy versus different recording durations.

As we expected, Fig. 3 suggests that longer recording duration achieves higher accuracy, but after 3 minutes it seems relatively stable. Furthermore, too long recording duration renders the system impractical in real applications. All the experimental results given in this paper are based on 3-minute recording duration. Of course, there is a trade-off between the performance and the recording duration time according to the different applications.

C. *System Performance Evaluation*

The system performance under three classifiers is evaluated using our feature set. The selected classifiers are Kohonen's LVQ Network, multi-class SVM, and Fisher's linear discriminant analysis combining (FDA) with K-NN classifier (k=1 in our experiments). The performance is estimated using a hold-out method, which evenly splits dataset to training and testing sets 100 times, and reports the average accuracy over 100 splits. The result is shown as box-plots in Fig. 4.

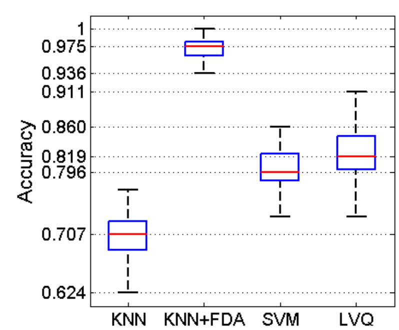

Figure 4. Box plots of classification accuracy.

Fig. 4 shows that K-NN, along with feature reduction by FDA, is the best classifier in our experiments using the proposed feature set. The average classification accuracy 97.5% is obtained on a dataset of 40 subjects. Such a high accuracy was achieved when all the data without any constrain on diet and times of a day are used.

When K-NN alone is used as the case in Figure 2, an accuracy of 70% is achieved using our feature set, which is obviously higher that the results shown in Figure 2 (The highest one is only 59.9%). This suggests our feature set is better than the proposed features in [1-4].

V. CONCLUSION

In this paper, it was demonstrated that the single channel EEG signals recorded by the portable equipment can be used for personal identification. However, the diet and circadian factors would diminish the accuracy of EEG-based personal identification to a certain extent, and the more diverse samples (different diet, times of day, etc.) should be included in the training set to get a better performance.

Combining AR model parameter and power spectrum density was demonstrated to be a set of high performance features. The overall system performance achieved an average accuracy of 97.5% on a dataset of 40 subjects. Results of this study provided supporting evidence that EEG-based personal identification was ready to move from proof-of-concept to practical system implementation.

In the future, we will evaluate the aging effect and other effects on this new biometric modality, and continue our research on the feasibility to combine more discriminating features to reduce the needed recording time.

ACKNOWLEDGMENT

This work is supported by The Key Project of The Ministry of Education of P. R. China (108012) and The National Natural Science Foundation of China (90920001, 60772114).

REFERENCES

[1] M. Poulos., et al. Person Identification based on Parametric Processing of the EEG. *The 6th IEEE International Conference on* Electronics, Circuits and Systems, 99(1):283-286, 1999.

[2] M. Poulos, et al.. Neural Network based Person Identification using EEG Features. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 99(2):1117-1120, 1999.

[3] R. B. Paranjape, et al.. The Electroencephalogram as a Biometric. *Proc. Canadian Conference on Electrical and Computer Engineering*, 01(2):1363-1366, 2001.

[4] C.R Hema, et al.. Brain Signatures: A Modality for Biometric Authentication, *International Conference on Electronic Design 2008,* 08:1-4, 2008.

[5] N.Markus, et al.. Test–retest Reliability of Resting EEG Spectra Validates a Statistical Signature of Persons, *Clinical Neurophysiology*, 07(118):2519-2524, 2007.

[6] M. Chisei., et al.. Biometric Person Authentication using New Spectral Features of Electroencephalogram (EEG), *International Symposium on Intelligent Signal Processing and Communication Systems*, 2008.

[7] Sun Shiliang, Multitask Learning for EEG-based Biometrics, *19th International Conference on Pattern Recognition,* 08:1-4, ICPR 2008.

[8] *Chen He,* et al.. Hashing the mAR Coefficients from EEG data for Person Authentication, *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009.* 09:1445-1448, 2009.

[9] A. Yazdani, et al.. Fisher Linear Discriminant Based Person Identification Using Visual Evoked Potentials, *IEEE International Conference on ICSP,* 08:1677-1680, 2008.

[10] R.Palaniappan, et al.. Biometrics from Brain Electrical Activity: A Machine Learning Approach, *IEEE Trans. On PAMI*, 07(4):738-742, 2007.

[11] S.Marcel, Person Authentication Using Brainwaves (EEG) and maximum A Posteriori Modal Adaptation, *IEEE Trans.On PAMI,* 07(4):743-748, 2007.

[12] A.K. Jain, "Biometric System Security," http://biometrics.cse.msu.edu, 2006.

[13] http://www.neurosky.com/