MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

High-Accuracy User Identification Using EEG Biometrics

Koike-Akino, T.; Mahajan, R.; Marks, T.K.; Tuzel, C.O.; Wang, Y.; Watanabe, S.; Orlik, P.V.

TR2016-105 August 2016

Abstract

We analyze brain waves acquired through a consumer-grade EEG device to investigate its capabilities for user identification and authentication. First, we show the statistical significance of the P300 component in event-related potential (ERP) data from 14-channel EEGs across 25 subjects. We then apply a variety of machine learning techniques, comparing the user identification performance of various different combinations of a dimensionality reduction technique followed by a classification algorithm. Experimental results show that an identification accuracy of 72% can be achieved using only a single 800 ms ERP epoch. In addition, we demonstrate that the user identification accuracy can be significantly improved to more than 96.7% by joint classification of multiple epochs.

IEEE Engineering in Medicine and Biology Society

High-Accuracy User Identification Using EEG Biometrics

Toshiaki Koike-Akino, Ruhi Mahajan, Tim K. Marks, Ye Wang, Shinji Watanabe, Oncel Tuzel, and Philip Orlik

Abstract— **We analyze brain waves acquired through a consumer-grade EEG device to investigate its capabilities for user identification and authentication. First, we show the statistical significance of the P300 component in event-related potential (ERP) data from 14-channel EEGs across 25 subjects. We then apply a variety of machine learning techniques, comparing the user identification performance of various different combinations of a dimensionality reduction technique followed by a classification algorithm. Experimental results show that an identification accuracy of 72% can be achieved using only a single 800 ms ERP epoch. In addition, we demonstrate that the user identification accuracy can be significantly improved to more than 96.7% by joint classification of multiple epochs.**

I. Introduction

Biometrics based on neurological signals such as the electroencephalogram (EEG) have been of recent interest in the literature. Compared to commonly used biometrics such as fingerprint, palm vein, and iris recognition, EEG-based biometrics may be less possible to forge [1]. There is a plethora of research demonstrating the potential of EEG for user identification (one-to-many matching) as well as for authentication (one-to-one matching) of a given person in a pool of multiple people [2]–[6]. In this proof-of-concept study, we limit the scope of analysis to identification, because the extension to authentication is straightforward.

Many studies have reported applying various signal processing and machine learning techniques to EEG data for user identification and authentication, with varying accuracy. He et al. [7] used autoregressive features and a naïve Bayes (NB) classifier, achieving a half total error rate (HTER) of 6.7% for authentication on 4 subjects. Marcel et al. [2] implemented a statistical framework with a Gaussian mixture model (GMM) and maximum *a posteriori* (MAP) estimation for authenticating 9 subjects, achieving an HTER as low as 6.6%. Palaniappan [8] used a linear discriminant analysis (LDA) classifier to achieve 100% accuracy in classifying 5 subjects, whereas Nguyen et al. [9] used a support vector machine (SVM) to obtain an error rate below 7.6%.

Most of these studies used either clinical-grade or high-density EEG systems with access to 32, 64, 128, or even more channels of EEG data. While these high-cost EEG systems can provide good space-time resolution of brain activities, the mobility of the user can be too restricted for real-life scenarios, and lengthy setup time may be required. For consumer-grade EEG devices, only a few studies have reported high performance [1], [10], [11]. Ashby et al. [10] showed high classification accuracy near 100% for 5-subject authentication with low-cost EEG sensors. More recently, Abo-Zahhad et al. [1] and Chuang et al. [11] reported more than 99% accuracy by using single-channel EEG data, for 10- and 15-subject authentication, respectively.

T. Koike-Akino, T. K. Marks, Y. Wang, S. Watanabe, O. Tuzel, and P. Orlik are with the Mitsubishi Electric Research Labs (MERL), Cambridge, MA, 02139-1955 USA (e-mail: {koike, tmarks, yewang, watanabe, oncel, porlik}@merl.com).

R. Mahajan is with the Dept. of Electrical and Computer Eng., University of Memphis, Memphis, TN, 38152 USA (rmhajan1@memphis.edu). She conducted this research as an intern at MERL.

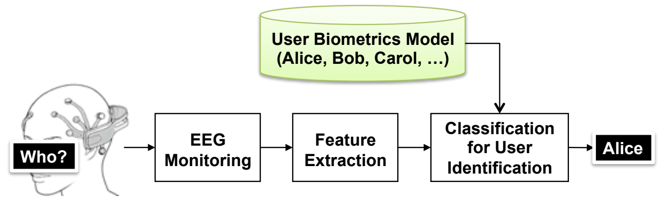


Fig. 1. General EEG-based user identification/authentication framework.

In this study, we demonstrate the feasibility of using a consumer-grade wireless EEG device, the Emotiv EPOC, for user identification in natural environments. Fig. 1 illustrates a general EEG-based identification/authentication system. Compared to existing literature, we focus on event-related potential (ERP) data, rather than EEG spectrograms, to analyze the impact of channel selection and dimensionality reduction techniques on classification accuracy. Our contributions are five-fold: 1) we show statistical significance for target vs. non-target P300 components in ERP data; 2) we compare dimensionality reduction techniques including principal component analysis (PCA) [12], partial least-squares (PLS) [13], and channel selection for feature extraction; 3) we compare a number of widely used classifiers in machine learning [14], specifically, LDA, quadratic discriminant analysis (QDA), NB, decision tree (DT), k-nearest neighbors (k-NN), SVM, logistic regression (LR), and deep neural network (DNN); 4) we demonstrate that more than 96% accuracy can be achieved for larger-scale 25-subject experiments; and 5) our analysis suggests that even larger-scale systems are feasible using multi-epoch classification.

II. Methodology

A. Event-Related Potential (ERP)

An event-related potential (ERP) is an EEG signal time-locked to a specific motor, cognitive, or sensory event. The components of an ERP waveform are usually categorized as sensory/exogenous (peaking up to $\sim$100 ms after the stimulus) and cognitive/endogenous (peaking up to 600 ms after the stimulus) [15]. The endogenous components represent how the stimulus information is processed, whereas the exogenous components are mainly governed by the stimulus characteristics. In this paper, we focus on endogenous components, in particular the P300 component, which is a positive peak that occurs roughly 300 ms after stimulus onset.

The P300 component can provide information about how a subject reacted to a stimulus. We implement a commonly used *oddball paradigm* [16] to elicit the P300, in which the subject is instructed to respond to (relatively rare) target stimuli and ignore non-target stimuli. Our approach is to identify features and inter-subject variations in this stimulus response to perform identification.

B. EEG Data Acquisition

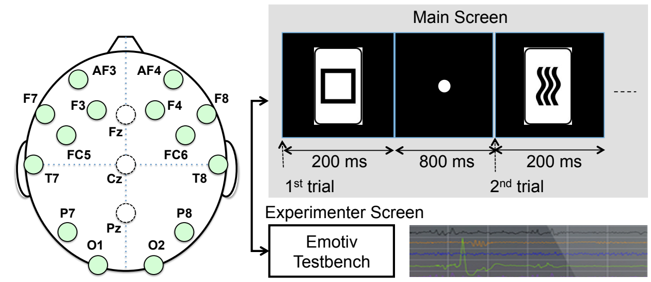
To record the brain activities from participants, we used a 14-channel referential montage-based commercial EEG device, the Emotiv EPOC (which is regarded by [17] as the best low-cost EEG device in terms of usability). Real-time EEG data were monitored at a rate of 128 samples/second.

EEG data were collected from 25 healthy adults, based on the international 10–20 standard electrode locations at AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1, and O2, depicted as green-filled circles in Fig. 2(a). The data were acquired in a naturalistic setting: a typical office environment, without any specific isolation from office workers. Each subject performed either one session (4 subjects) or two sessions (21 subjects) of the experimental task. Subjects with two sessions removed the headset and took a break (15–60 min.) between the sessions. All procedures were in accordance with the ethical standards of the responsible committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2000. Written informed consent was obtained from all participants prior to this study.

C. Experimental Protocol

In this paper, we focus on a card counting task. This is only one of many tasks that can be used for identification/authentication (for example, [11] found other tasks such as breathing to be more effective). Our goal is not to select the optimal experimental protocol, but rather to demonstrate the effectiveness of our statistical analysis and classification approach. We leave investigation of the relative effectiveness of various experimental protocols to future work.

Fig. 2(b) illustrates our experimental protocol, in which five Zener cards [18] are sequentially displayed in random order on a computer screen. Prior to data acquisition, participants are asked to select any one of the five Zener cards, which we refer to as the *target card* in this study. The task of the participant is to count how many times their target card appeared on the screen. At the end of the experiment, participants are asked how many times their target card appeared, to verify that they performed the counting task. As depicted in Fig. 2, each card (target or non-target) is displayed for 200 ms, with an inter-stimulus interval of 800 ms. One session lasts about 4.2 min., leading to about 250 trials (i.e., ~ 50 presentations of each Zener card). Participants are instructed to refrain from moving their body and eyes during the experiment by fixating eyes on the center dot on the screen during the 800 ms inter-stimulus interval.



(a) Electrode locations (b) Experimental system

Fig. 2. Experiment protocol with Zener cards [18] displayed on the main screen. The real-time EEG data are captured and monitored on the experimenter's screen using the Emotiv Testbench software.

D. Extracting Features with Reduced Dimensionality

To investigate the event-related EEG dynamics, epochs are time-locked to the stimulus presentation. Each epoch is recorded beginning 100 ms pre-stimulus and ending 700 ms post-stimulus presentation. To remove ocular artifacts, we used the blind source separation canonical correlation analysis (BSS-CCA) in the AAR toolbox [19]. After artifact removal (rejection ratio was 4.22%), we obtained 12,139 epochs in total, each with a duration of 800 ms.

The raw 14-channel ERP data for a single epoch have a size of $103 \times 14 = 1442$ dimensions. Dimensionality reduction of these ERP data is beneficial both to facilitate real-time applications and to improve classification performance. As shown in [11], even data from just a single channel may be useful for identification/authentication. The most basic dimensionality reduction is to simply select a subset of the 14 channels. We also consider the more sophisticated dimensionality reduction techniques PCA [12] and PLS [13].

PCA projects the measurement data via an orthogonal transform onto a space in which the data components are uncorrelated. The components are the dominant eigenvectors (corresponding to the largest eigenvalues) of the data covariance matrix, without considering class labels. In contrast, PLS tries to find projection vectors that maximize the covariance between the projected data and the class labels.

E. Machine Learning for Classification

We compare the effectiveness for user identification of several machine learning classification algorithms: LDA, QDA, NB, DT, k-NN, SVM, LR, and DNN [14]. The classifiers of LDA, QDA, and (Gaussian) NB are based on an assumption that the ERP data from each subject are normally distributed. Compared with QDA, LDA adds an additional assumption that the covariance of each class is identical, leading to lower-complexity classifications. For LDA, we adopt no shrinkage and no threshold offset in this paper, because they provided almost no performance improvement. Like QDA, NB allows classes to have non-identical covariance; the key assumption of NB is that the measurement variables are assumed to be conditionally independent given the class label. The DT is a nonparametric classifier using a set of simple rules with no explicit assumptions about the data distribution. The method k-NN is also non-parametric, and one of simplest machine learning algorithms. We chose the number of neighbors $k = 1$ (simple nearest neighbor), as it worked the best. For multi-class SVM classifier, we use multiple linear SVMs based on a one-vs.-one method, since other variants such as Gaussian-kernel SVM and one-vs.-all did not perform better. We also consider LR as alternative to NB and linear SVM. Our DNN has 2 hidden layers (1000 rectified linear units each), trained using an adaptive-moment stochastic gradient method.

III. Data Analysis Results

A. Significance Test

It is known that the P300 component is typically most prominent at the Fz (frontal), Cz (central), and Pz (parietal) midline scalp electrode locations, shown as white-filled circles in Fig. 2(a). However, the Emotiv EPOC headset does not provide direct access to these locations. To access these scalp locations indirectly, we averaged the channel pairs F3 and F4 (which we refer to as F), AF3 and AF4 (AF), O1 and O2 (O), and P7 and P8 (P). We used these four electrode pairs to test for statistically significant differences between the P300 responses for target vs. non-target stimuli. Fig. 3 shows the grand average ERP for all 12,139 epochs of all subjects (with no specific processing to account for inter-subject variability) for these four averaged pairs of channels (F, AF, O, and P). Statistically significant differences, found using a paired t-test with Bonferroni correction (with $n = 4$) at a significance level of $\alpha = 0.10$, are indicated by black rectangles below the ERP plots in Fig. 3. There is a significant difference between the target and non-target stimuli responses, especially for F and AF channels from 400–600 ms, which may correspond to the P300 component. The P and O channels have more complicated characteristics, yet a statistically significant difference is still observed.

The t-test for all 12,139 epochs of 25 participants showed statistical significance in the F, AF, O, and P channels. In fact, when we jointly analyze the statistics across all 14 channels at once, the required number of epochs needed to observe statistical significance can be decreased, as shown in Fig. 4. This figure shows p-value of Hotelling's t^2-test for 14-channel ERPs. Here, we randomly select 50, 100, or 200 epochs from the set of all target epochs, and the same number from the set of all non-target epochs. It can be seen that the p-value can run below a significance level of $\alpha = 0.10$ for 100 epochs at around 500 ms. Since there are 25 subjects, this indicates that 4 epochs per subject are sufficient to statistically differentiate ERPs for target versus non-target events using P300 components.

B. Dimensionality Reduction

Fig. 5 shows the percentage of variance explained by a limited number of components extracted using PCA or PLS. Observe that for PCA, about 7% of the principal components (100 components out of 1442 original feature dimensions) explain 90% of the data variance. Although PLS

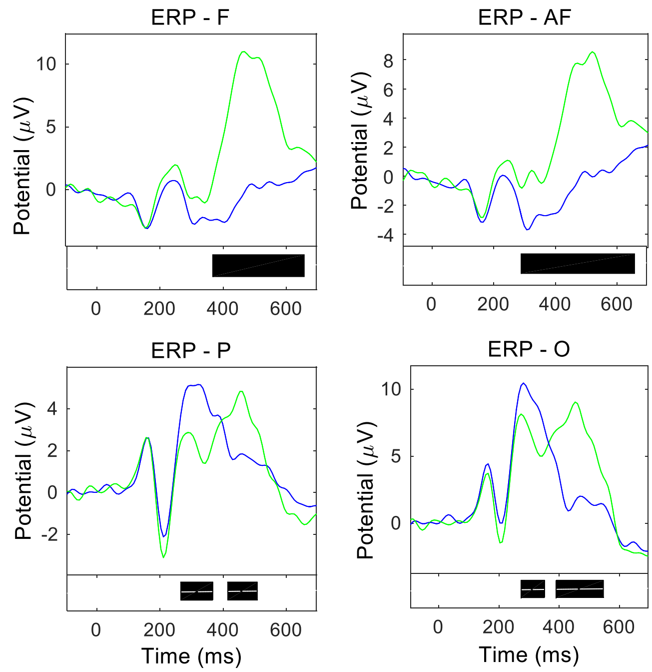


Fig. 3. Grand average ERP from frontal (F), pre-frontal (AF), occipital (O), and parietal (P) lobes (green: target card, blue: non-target card).

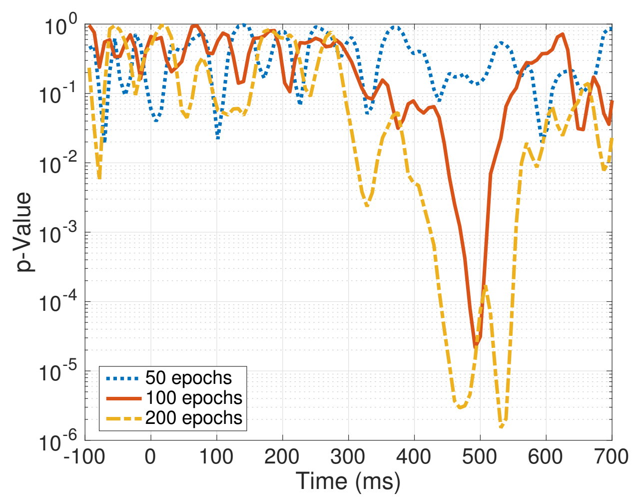


Fig. 4. p-value of Hotelling's t^2-test for joint 14-channel statistics with limited number of epochs.

accounts for a slightly lower percentage of the variance in the measurement data (X) than PCA, PLS can be more effective for regression because it is designed to explain the variance in the class label (Y). Using PLS, 65% of the variance in Y can be explained by 100 components, with almost no loss in the explained variance of X compared to PCA.

The impact of the reduced dimensionality is evaluated using several classifiers in Fig. 6, which plots the user-identification error rate (in a 10-fold cross-validation) as a function of the number of dimensions in the reduced data. Here, PLS is used to reduce the dimensionality of the data from the original 1442 dimensions for all epochs (including targets and non-targets). For two of the classification methods

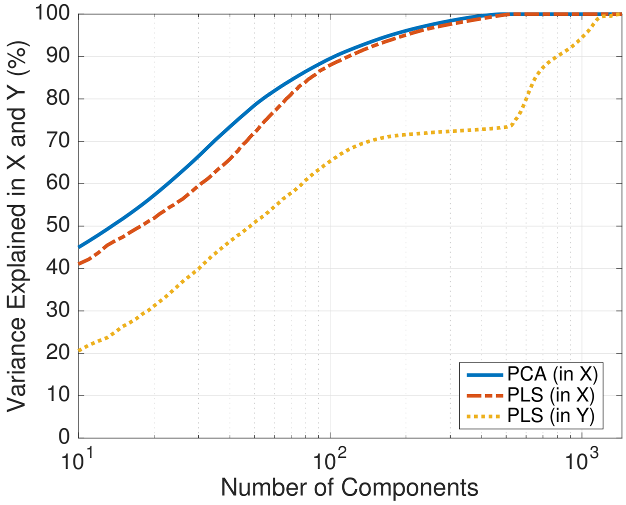


Fig. 5. Dimensionality reduction via PCA and PLS: Percentage of the variance in X (measurement data) or the variance in Y (class label for regression) that is explained by a given number of components.

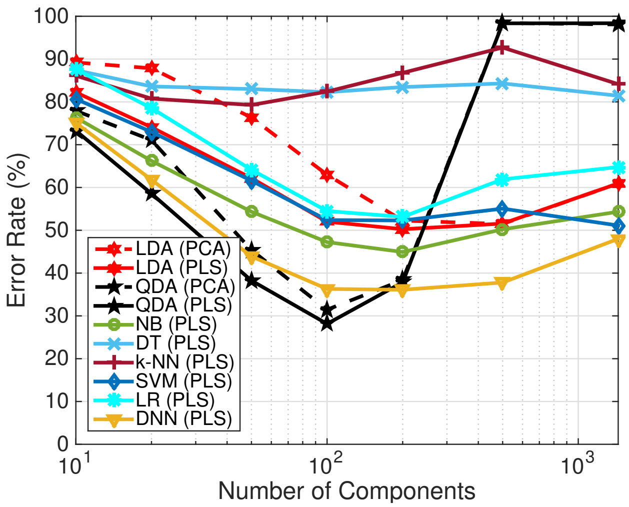


Fig. 6. User identification accuracy with various classifiers: LDA, QDA, NB, DT, k-NN, SVM, LR, and DNN.

(LDA and QDA), the performance curves with PCA are also presented for comparison. Comparing these two dimensionality reduction techniques, it is seen that PLS outperforms PCA as input to both of the classification algorithms, LDA and QDA. More importantly, the results demonstrate that dimensionality reduction is of great importance to achieve higher classification accuracy, in particular when the available training data are limited (e.g., $12{,}139/25 \approx 486$ epochs per subject in our data set). The best performance, near 72% accuracy, was obtained by QDA after using PLS to reduce the number of dimensions to 100. In our analysis, non-parametric classifiers, i.e., DT and k-NN, performed worse than other methods. In addition, SVM had almost no advantage over simpler methods of LDA, QDA, and NB. Although DNN shows relatively good performance, QDA outperforms DNN because the number of ERP measurements in our experiment was not large enough for deep learning. From this result, we focus on QDA for the following analyses.

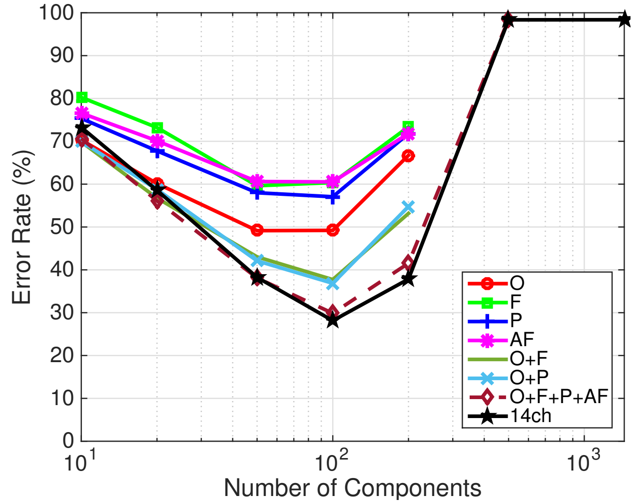


Fig. 7. Impact of channel selection for QDA with PLS.

Fig. 7 shows the performance of QDA classification with PLS dimensionality reduction when we select a subset of channels from the original 14 channels. Among the O, F, P, and AF channels, the O channels (O1 and O2) were shown to be the most effective electrodes for classification. This may be because our experiments are based on visual stimuli. The P channels were slightly better than the AF and F channels. The reason may lie in the fact that both the O and P channels have relatively complicated ERP dynamics, as is evident in Fig. 3, which might facilitate user identification. However, the combination of the O and P channels was not noticeably more effective than the combination of the O and F channels. The 8-channel combination of the O, F, P, and AF channels approaches the accuracy of the full 14-channel case.

C. Joint Classification of Multi-Epoch ERP

The above-mentioned 72% accuracy achieved using only a single epoch (800 ms) of data is remarkable for 25-subject identification, in comparison to previous work in the literature. For example, 150 seconds of total data were used with fewer subjects in [10], while 5-second spectrograms and about 10-second ERP (for 15 eye blinks) were used in [11] and in [1], respectively. The advantage of longer-duration EEG data is analyzed in Fig. 8, where we perform PLS dimensionality reduction on concatenated multiple epochs (with 4-times bootstrapping), followed by LDA or QDA for joint classification of multiple epochs. It was revealed that this multi-epoch classification can significantly improve classification accuracy. For example, LDA can achieve 96.7% accuracy with 16-epoch (12.8 seconds) classification.

Note that QDA performance degrades considerably for more than 200 components (as is evident in Figs. 6, 7, and 8). This is because LDA estimates a single covariance matrix (identical for all subjects), whereas QDA must estimate a separate covariance matrix for each subject. Thus, QDA can use only $1/25$ as much data as LDA to train each covariance matrix. This leads to rank deficiency of QDA's subject-specific data covariance matrices when the data are represented using high-dimensional features.

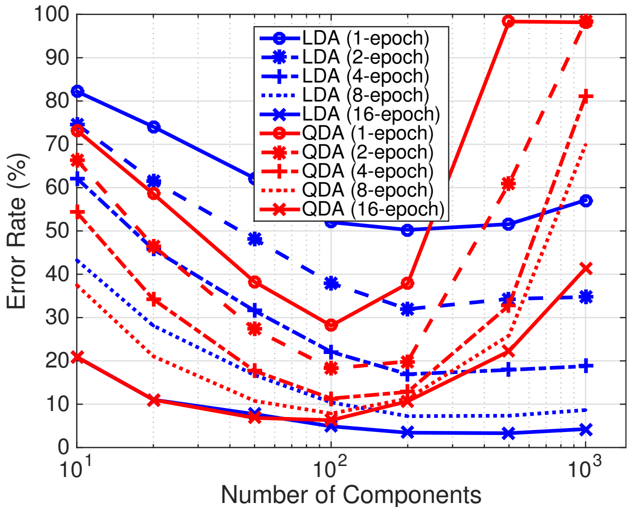


Fig. 8. Effect of multi-epoch classification for QDA and LDA with PLS.

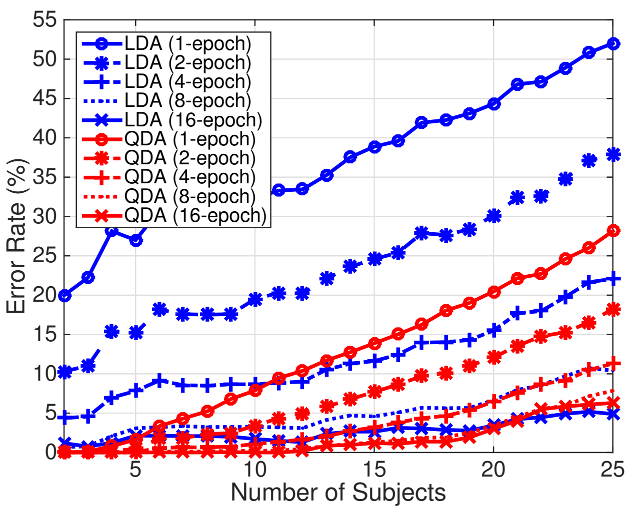


Fig. 9. Impact of the number of subjects for QDA and LDA with PLS.

We now discuss the impact of the number of subjects to be identified. In Fig. 9, the error rates obtained by LDA and QDA are plotted as a function of the number of subjects (100 PLS components). Note that the error rate is nearly linear with respect to the number of subjects to identify. Thus, building a large-scale user-identification system may be challenging. Nevertheless, multi-epoch classification is a viable countermeasure, since the error rate can be decreased almost exponentially as a function of the number of epochs (e.g., 52.0, 37.9, 22.1, 10.4, and 4.9% error rates for 25-subject LDA with 1, 2, 4, 8, and 16 epochs, respectively).

IV. Conclusion

We analyzed the use of ERPs for identification and authentication using a low-cost EEG headset. In our experiments, the t-test and t^2-test showed statistical significance in the P300 components of target vs. non-target stimulus response. We also demonstrated that dimensionality reduction plays an important role for classification. Through the comparison of several dimensionality reduction techniques and classification algorithms, we found that using only a single 800 ms epoch, PLS dimensionality reduction followed by QDA classification achieves 72% accuracy for 25-subject identification from EEG biometrics. Moreover, we demonstrated the significant advantage of multi-epoch classification, which can almost exponentially decrease error rates to achieve near 100% accuracy. The impact of channel selection for lower-cost sensing and the impact of the number of users for large-scale identification were also discussed. This study paves the way for future investigation of real-time EEG-based biometrics techniques using a wireless EEG device in non-clinical settings. More efficient experimental protocol design remains as future work.

References

[1] M. Abo-Zahhad, S. M. Ahmed, and S. N. Abbas, "A new EEG acquisition protocol for biometric identification using eye blinking signals," *Intelligent Systems and Applications*, vol. 6, pp. 48–54, 2015.

[2] S. Marcel and J. R. del Millan, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 743–752, April 2007.

[3] I. Nakanishi, S. Baba, and C. Miyamoto, "EEG based biometric authentication using new spectral features," *IEEE International Symposium on Signal Processing and Communication Systems (ISPACS)*, pp. 651–654, 2009.

[4] Q. Gui, Z. Jin, and W. Xu, "Exploring EEG-based biometrics for user identification and authentication," *IEEE Symposium on Signal Processing in Medicine and Biology (SPMB)*, pp. 1–6, 2014.

[5] J. Thorpe, P. van Oorschot, and A. Somayaji, "Pass-thoughts: Authenticating with our minds," *ACM Proceedings of the New Security Paradigms Workshop (NSPW)*, pp. 45–56, 2005.

[6] T. T. Pham, W. Ma, D. Tran, P. Nguyen, and D. Phung, "A study on the feasibility of using EEG signals for authentication purpose," *Neural Information Processing*, pp. 562–569, 2013.

[7] C. He, X. Lv, and J. Wang, "Hashing the mAR coefficients from EEG data for person authentication," *IEEE International Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1445–1448, April 2009.

[8] R. Palaniappan, "Two-stage biometric authentication method using thought activity brain waves," *International Journal of Neural Systems*, vol. 18, no. 1, pp. 59–66, 2008.

[9] P. Nguyen, D. Tran, X. Huang, and W. Ma, "Motor imagery EEG-based person verification," *Advances in Computational Intelligence*, pp. 430–438, 2013.

[10] C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein, "Low-cost electroencephalogram (EEG) based authentication," *IEEE 5th International EMBS Conf. Neural Engineering*, pp. 442–445, April 2011.

[11] J. Chuang, H. Nguyen, C. Wang, B. Johnson, "I think, therefore I am: Usability and security of authentication using brainwaves," *Financial Cryptography and Data Security*, pp. 1–16, 2013.

[12] I. T. Jolliffe, *Principal Component Analysis, 2nd ed.*, Springer, 2002.

[13] S. de Jong, "SIMPLS: An alternative approach to partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 18, pp. 251–263, 1993.

[14] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, The MIT press, 2012.

[15] S. Sur and V. K. Sinha, "Event-related potential: An overview," *Industrial Psychiatry Journal*, vol. 18, no. 1, pp. 70–73, 2009.

[16] N. K. Squires, K. C. Squires, and S. A. Hillyard, "Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man," *Electroencephalogr Clin Neurophysiol.*, vol. 38, no. 4, pp. 387–401, 1975.

[17] K. Stamps and Y. Hamam, "Towards inexpensive BCI control for wheelchair navigation in the enabled environment - a hardware survey," *International Conf. Brain Informatics*, pp. 336–345, 2010.

[18] J. L. Kennedy, "The visual cues from the backs of the ESP cards," *The Journal of Psychology*, vol. 6, no. 1, pp. 149–153, 1938.

[19] G. Gómez-Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, and W. Van Paesschen, "Automatic removal of ocular artifacts in the EEG without an EOG reference channel," *IEEE 7th Nordic Signal Processing Symposium*, pp. 130–133, 2006.