Towards EEG Biometrics: Similarity-Based Approaches for User Identification

Qiong Gui, Zhanpeng Jin
Department of Electrical and Computer Engineering
Binghamton University, SUNY
Binghamton, NY 13902-6000
{qgui1,zjin}@binghamton.edu

Maria V. Ruiz Blondet, Sarah Laszlo
Department of Psychology
Binghamton University, SUNY
Binghamton, NY 13902-6000
{mruizbl1,slaszlo}@binghamton.edu

Wenyao Xu
Department of Computer Science and Engineering
University at Buffalo, SUNY
wenyaoxu@buffalo.edu

Abstract

EEG brainwaves have recently emerged as a promising biometric that can be used for individual identification, since those signals are confidential, sensitive, and hard to steal and replicate. In this study, we propose a new stimuli-driven, non-volitional brain responses based framework towards individual identification. The non-volitional mechanism provides an even more secure way in which the subjects are not aware of and thus can not manipulate their brain activities. We present our preliminary investigations based on two similarity evaluation approaches: Euclidean Distance (ED) and Dynamic Time Warping (DTW) methods. We investigate the performance of our proposed methodology using four different visual stimuli and the potential impacts from four different EEG electrode channels. Experimental results show that, the Oz channel provides the best identification accuracy for both ED and DTW methods, and the stimuli of illegal strings and words seem to trigger more distinguishable brain responses. For ED method, the accuracy of identifying 30 subjects could reach over 80%, which is better than the best accuracy of about 68% that can be achieved by DTW method. Our study lays a solid foundation for future investigation of innovative, brainwave-based biometric approaches.

1. Introduction

Over the past few decades, biometric approaches have gained dramatically increasing interest for individual identification and authentication, since they are closely associated with an individual's physiological or behavioral features. Some of those features, such as fingerprint, face, iris and voice [28], have been extensively investigated and proved to be scientifically unique across the entire human population, which result in very promising biometrics in cyber security domain. For instance, nowadays it has been very popular for the fingerprint technology in smartphones, homeland security, and forensics. However, those existing biometric characteristics still suffer from various limitations and weaknesses, far from perfect. For example, although DNA is the ultimate unique code of individuals, the sample of DNA is really difficult to collect and can be stolen; the faces of each individual are largely different, but the appearance of faces can change a lot as people getting older; the fingerprint is a popular biometric measurement with a high matching accuracy, however it can be faked [18] or obtained by force. It has been reported that a violent gang in Malaysia chopped off a car owner's finger to get round the vehicle's hi-tech security system [12]. Therefore, it is highly desired to seek new biometric approaches that can possibly overcome those limitations.

Recently, electroencephalogram (EEG) based biometrics, representing the unique human brain activities, have emerged as a new and promising way for labeling each individual person [24]. EEG records the brain's electrical activity and can be obtained by measuring the voltage fluctuations on the scalp surface with simple placement of the electrodes on the skin [16]. The brain activities are inherently determined by the person's unique pattern of neural pathways and closely associated with each individual's unique memory and knowledge base, thus it is impossible to imitate others' brain activities [2, 24]. More usefully, those brain signals can be influenced by mood, stress and mental state of the individual [17] which makes them very difficult to be obtained under force and threat. Furthermore, it has been recognized that the brain signals are also related to the subject's genetic information, making them unique for each individual [25,33,37] and stable over time [21]. Given all aforementioned advantageous characteristics in security and reliability, brain signals have thus been proposed as an identification and authentication biometric [14, 34]. Recently, stimuli-driven non-volitional ("passive") brain responses have been particularly explored as a more secure biometric [6,29,30], compared to the conventional "active" brain activities.

In this paper, we present an EEG-based user identification and authentication framework based on similarity assessment approaches: Euclidean Distance (ED) based approach and Dynamic Time Warping (DTW) based approach. We also seek to investigate the potential impacts and performance of adopting different types of visual stimuli and focusing on a particular subset of EEG channels. Specifically, we analyzed 4 EEG electrode channels under 4 types of stimuli from 30 human subjects. The rest of the paper is organized as follows: Section II gives a brief introduction to the related work. Section III introduces the ED method and DTW method and the mechanism for making the final identification decision. Section IV describes the experimental setting for EEG data collection, and then discusses the experimental results. Section V concludes our research work and results.

2. Related Work

Existing research has demonstrated that the EEG brainwave signals can be used as a viable biometric for individual identification and authentication. The methodological flow in most of prior work can be categorized into the following steps: EEG acquisition, preprocessing, feature extraction, and classification. The preprocessing stage seeks to mitigate the impacts of noises and artifacts, usually through various filtering strategies, common reference removal, or independent component analysis (ICA) [8, 17]

The feature extraction and classification stages are closely relevant to each other and have been extensively investigated using a variety of approaches. Among those approaches, the support vector machine (SVM) has been widely employed due to its superior nonlinear classification capability. For instance, Ashby *et al*. [1] extracted the autoregressive (AR) coefficients, power spectral density (PSD), spectral power (SP), interhemispheric power difference (IHPD) and interhemispheric channel linear complexity (IHLC) from the filtered data and used the linear SVM classifier for authentication on 5 individuals and got the false rejection rate (FRR) of 2.4% to 5.1%, and the false acceptance rate (FAR) of 0.7% to 1.1%. Other similar studies have also been done using linear SVM [3, 22] or other SVM variants, like Gaussian Kernel SVM [35], Polynomial Kernal SVM [36], radial basis function (RBF) SVM [5].

Neural network (NN) is another popular classifier used in human identification and authentication. At the early stage of EEG-based biometric, learning vector quantizer (LVQ) was adopted by researchers. Poulos *et al*. [26, 27] proposed a linear rational model of ARMA type to fit the alpha band EEG signals and also used the LVQ NN. For the 75 people being tested, to distinguish a specific person from others, correct classification scores of LVQ classifier in the range of 72% to 84% were obtained. Other researchers have been using the classic feed-forward, back-propagation neural network (NN) in their EEG-based identification studies, achieving a wide range of accuracy levels [9–11, 15, 19, 36]. For example, Palaniappan [23] used visual evoked potential (VEP) signals to identify 20 individuals by the NN classifier and Shedeed [32] used the NN to identify 3 subjects based on fast Fourier transform (FFT) and wavelet packet decomposition (WPD) from 4 channels and got an correct classification rate from 66% to 93%.

3. Method

3.1. Euclidean Distance Based Method

Euclidean distance is a way to show the ordinary distance between two points in Euclidean space. For two time series, the Euclidean method gets the final distance score by accumulating all the distance between two points using Eq. 1 with aligning the *n*-th point of one time series with the *n*-th point of the other time series shown in Figure 1. If the distance value is small, it in some sense indicates that the two points share a lot of similarities. Otherwise, the two time series seem to be dramatically different from each other. However, a weakness of ED method is that, if there are time delays or misalignments between the two time series even though they still have similar shapes, ED method will produce a poor similarity value and can hardly reflect the inherent similar characteristics between the two evaluation subjects.

$$ED = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_N - q_N)^2} \tag{1}$$

where p and q are two trails, $\{p_1, p_2, \ldots, x_N\}$ and $\{q_1, q_2, \ldots, q_N\}$ are elements in two trails with N samples respectively.

3.2. Dynamic Time Warping (DTW) Based Method

DTW is a technique to find the optimal alignment between two time series if one time series may be "warped" non-linearly by stretching or shrinking it along its time axis, shown as Figure 2. For example, at time *n*, although the blue time series starts to increase, the green time series still keeps its previous value. But the green time series starts to increase at time *n+2*. For DTW method, it can synchronize time point *n* of blue time series with time point *n+2*

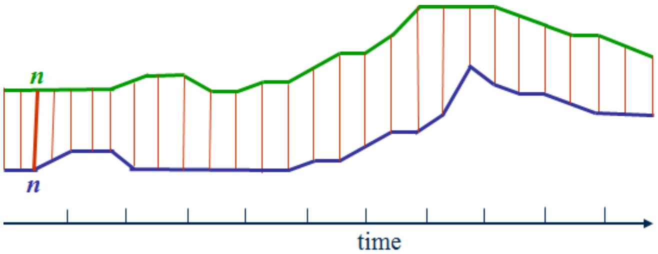

Figure 1. An example of Euclidean distance based analysis

of the green time series. It is the advantage of DTW that it can provide a more intuitive similarity measurement and allow similar shapes to match even they are out of phase along the time axis. DTW technique has been conventionally used in speech recognition, data mining, gesture recognition, robotics, manufacturing and medicine [13]. In recognizing the potential time lags and latencies involved in the biological brain responses to corresponding visual stimuli among different human subjects, we proposed to adopt DTW to identify the similarity among EEG patterns being aware of the time misalignment.

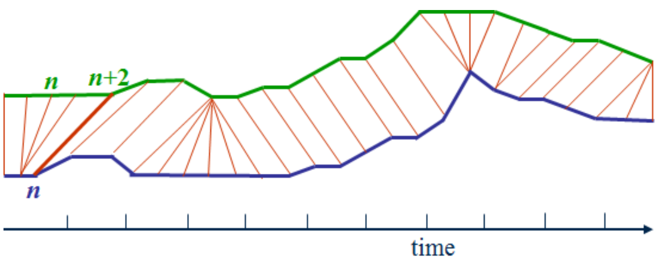

Figure 2. An example of DTW based analysis

As shown in Figure 3, for two time sequences with different lengths of *M* and *N* respectively, the standard DTW first calculates a *M*-by-*N* matrix in which every element is the minimum distance by searching three elements around it represented by the blue arrows (as Eq. 2) [20]. The optimal path shown by the red dot is determined by the path which has the minimal distance. However, this method is very time and space consuming with the complexity of $O(N^2)$.

$$\begin{aligned} DTW(m,n) := cost + min\{&DTW(m-1,n), \\ &DTW(m,n-1), \\ &DTW(m-1,n-2)\} \end{aligned} \tag{2}$$

An improved FastDTW approach was recently proposed with time and space complexity of $O(N)$ when the radius is less than N [31]. The FastDTW has three key operations: coarsening, projection and refinement. The main idea of FastDTW is to first create all of the resolutions that will be evaluated using coarsening. For the lowest resolution

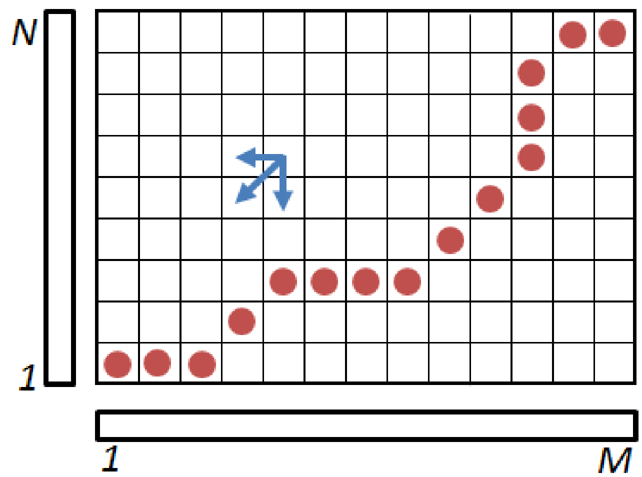

Figure 3. An example DTW grid

time series, the standard DTW algorithm is performed to find the optimal warp path. After the warp path is found for the lowest resolution, it is projected to the next higher resolution and the warp path is considered as an initial guess for a higher resolution's minimum-distance warp path. To refine the projected path, a constrained DTW algorithm is executed with the constraint that only cells in the projected warp path are evaluated. This will find the optimal warp path through the area of the warp path that was projected from the lower resolution. However, the entire optimal warp path may not be contained within projected path. Therefore, additional number of cells on each side of the projected path controlled by a parameter called "radius" will also be evaluated when refining the warp path in order to increase the chances of finding the optimal solution.

3.3. Identification Decision

For both ED and DTW methods, they use the minimal distance values to indicate the similarity between the comparison subjects and then make the decision for acceptance/rejection of the individual identification. To this end, we need to first collect some brainwave patterns from each subject and set them as the reference patterns with subject labels. For a specific person, the EEG patterns responding to the same visual stimulus more likely remain similar in different experimental trials over the time, while for different individuals, their brainwave patterns are supposed to be different from others even facing the same stimulus. So when a new pattern is collected and needs to be analyzed to identify the potential owner of this specific brainwave segment. We will first compare it with the patterns already known and calculate the distance between the new pattern and all the stored reference patterns. The final decision will be the subject that the smallest distance has reached.

4. Experiments and Results

4.1. EEG Data Collection

The raw EEG signals were collected from 30 adult participants (14 females, age range 18-25, mean age 19.53) using "EASY CAP" device (Ammersee, Germany) [4] from four electrode sites (Pz, O1, O2, Oz) around the area of the left superior temporal lobe, which are believed to better reflect each individual's cognitive behaviors associated unique semantic memory processes [7]. The data was sampled at 500 Hz. 1.1 seconds of raw EEG signals were recorded, which made 550 samples for each channel. In this experiment, the participants were asked to silently read an unconnected list of texts which included 75 *words* (e.g., BAG, FISH), 75 *pseudowords* (e.g., MOG, TRAT), 75 *acronyms* (e.g., MTV, TNT), 75 *illegal strings* (e.g., BPW, PPS), and 150 instances of their own names [29]. By showing the subjects with the stimuli, the non-volitional ("intuitive") brain activities were recorded. Each human subject was tested twice: one test was used for training and the other was used for testing purpose. In this paper, the brain responses from all the four channels and stimuli were processed and analyzed.

Since the raw EEG signals are noisy, it is common to average many trials together which can get rid of the random brain activities but keep the event-related potentials (ERPs). Thus the EEG signals were first ensemble averaged for 50 individual measurements. Figure 4 shows the ERPs of Subject 10 with the stimuli of *illegal strings*. The first two patterns were two trials from one test and the last two patterns were two trails from the other test. From the plot, the morphology of the patterns were very similar. Furthermore, the patterns were persistent during different tests which was an important indicator that the similar EEG patterns (i.e., non-volitional brain responses) could be captured in any later tests. Figure 5 shows that patterns from different human subjects. The four patterns were from Subject 8, 10, 24 and 25 respectively. It can be observed that the patterns were quite different among those four human subjects. Such observations testify the psychological rationale about the uniqueness of people's non-volitional brain responses, even to the exactly same stimuli, and also imply the feasibility of recognizing an individual through identifying the similarity between the unknown EEG brainwave segment and the reference pattern.

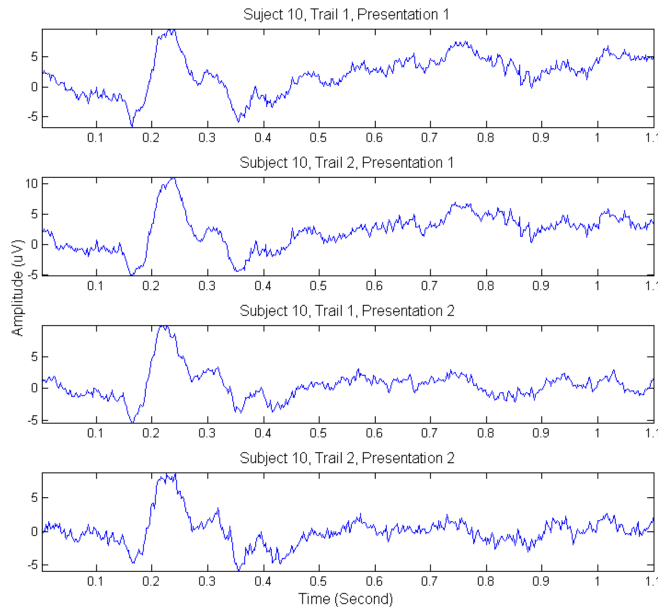

Figure 4. EEG patterns of one subject (stimuli: illegal strings; channel: Oz)

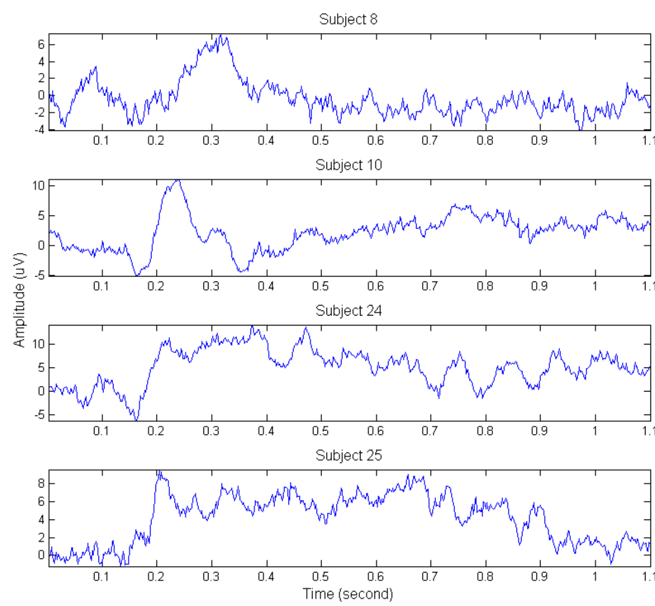

Figure 5. EEG patterns from four different subjects (stimuli: illegal strings; channel: Oz)

4.2. Experimental Results

In the experiment, four stimuli and four channels were analyzed to identify and recognize all the 30 human subjects. For the training set, 10 trials of each subject were randomly chosen as the reference pattern. 20 trials of each subject were randomly selected to evaluate the identification performance. Moreover, since the EPRs had some drift, the data was first normalized by mean and standard deviation before the methods were applied. The accuracy defined as total number of correct classifications over total number of testing trials that were used for performance evaluation.

Table 1 presents the accuracy of Euclidean Distance based approach for different EEG electrode channels and stimuli. To identify all the 30 human subjects, the minimum accuracy was 53.33% by acronyms using the Pz channel.

The maximum accuracy was 81.17% using illegal strings and the Oz channel. From the table, the Oz channel seemed to lead to higher accuracy compared to the channels Pz, O1, O2. Channels O1 and O2 seemed to have similar performance given their symmetric locations. Except for the illegal strings stimuli, the Pz channel had the worst accuracy. The illegal strings stimuli performed better than other stimuli with the accuracy of more than 70%, and the words stimuli were second to the illegal strings in the performance list with the accuracy of around 70%. Acronyms had the accuracy of less than 60%, the worst among all four different types of stimuli.

The results of DTW-based method are shown in Table 2. Similar to ED- based method, the minimum accuracy of 33.83% was also obtained by acronyms and the channel Pz, while the maximum accuracy of 67.17% was achieved from the stimuli of illegal strings and the channel Oz. Oz channel also showed its superior performance and higher accuracy compared to other channels. Channels O1 and O2 had similar performance. Pz channel performs better than O1 and O2, under the stimuli of illegal strings and words. Moreover, except Oz channel, the stimuli of words seemed to have higher accuracy than other stimuli.

4.3. Discussion

In this study, we proposed a new user identification methodological framework using the similarity-based approaches, leveraging the non-volitional brain activities which are believed to be associated with and reflect people's unique memory and knowledge. In the data collection stage, acronyms, illegal strings, words and pseudowords were presented to the human subjects. The intuitive responses of each individual when reading those stimuli were captured. Since the occipital region is the visual center of human brain to process memory and knowledge, in the experiment, channels from this region have been investigated for performance. The experimental results demonstrated that, the channel Oz showed stronger distinguishing capability compared to other channels around this region, which can be utilized in future study to understand the interpret the specific subregion of the brain in such cognitive tasks. As for the different visual stimuli, illegal strings which were not familiar by people and the words which were well understood by people seemed to make the subjects to had more distinguishable brain responses, than acronyms and pseudowords.

5. Conclusion

In this paper we focused on a preliminary study using non-volitional EEG brainwaves as a biometric, based on two similarity evaluation approaches, Euclidean Distance method and Dynamic Time Warping method, to identify 30 human subjects. For Euclidean Distance method, the accuracy can reach 81.17% and the channel Oz showed better performance than other channels. The stimuli of illegal strings and words seemed to be able to trigger more distinguishable brain activity patterns among different individuals and thus lead to better identification accuracy. In general, ED method performed better than DTW. This study represents an early stage research effort which still suffers many limitations and drawbacks. In the future, we will explore other types of visual stimuli and investigate more robust classification approaches that can provide better performance while remain as a computationally efficient manner.

6. Acknowledgements

This research was supported by NSF grants SaTC-1422417, SaTC-1423061, and BCS-1252975, and the Binghamton University Interdisciplinary Collaboration Grant.

References

[1] C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein. Low-cost electroencephalogram (EEG) based authentication. In *5th Int'l IEEE/EMBS Conf. Neural Engineering (NER)*, pages 442–445, April 2011.

[2] D. S. Bassett and M. S. Gazzaniga. Understanding complexity in the human brain. *Trends in Cognitive Sciences*, 15(5):200 – 209, 2011.

[3] K. Brigham and B. Kumar. Subject identification from electroencephalogram (EEG) signals during imagined speech. In *Fourth IEEE Int'l Conf. Biometrics: Theory Applications and Systems (BTAS)*, pages 1–8, Sept 2010.

[4] EASYCAP GmbH. 74-Channel Electrode Arrangement of EC80. `http://www.easycap.de/easycap/e/products/products.htm`. [accessed Dec. 15, 2014].

[5] A. Ferreira, C. Almeida, P. Georgieva, A. Tom, and F. Silva. Advances in EEG-based biometry. In A. Campilho and M. Kamel, editors, *Image Analysis and Recognition*, volume 6112 of *Lecture Notes in Computer Science*, pages 287–295. Springer Berlin Heidelberg, 2010.

[6] Q. Gui, Z. Jin, and W. Xu. Exploring EEG-based biometrics for user identification and authentication. In *Proc. IEEE Signal Processing in Medicine and Biology (SPMB)*, pages 1–6, Dec. 2014.

[7] E. Halgren, R. P. Dhond, N. Christensen, C. V. Petten, K. Marinkovic, J. D. Lewine, and A. M. Dale. N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage*, 17(3):1101–1116, Nov. 2002.

[8] C. He and J. Wang. An independent component analysis (ICA) based approach for EEG person authentication. In *3rd Int'l Conf. Bioinformatics and Biomedical Engineering (ICBBE)*, pages 1–4, June 2009.

[9] C. Hema and A. A. Osman. Single trial analysis on EEG signatures to identify individuals. In *6th Int'l Colloquium on Signal Processing and Its Appl. (CSPA)*, pages 1–3, 2010.

Stimuli \ Channel	Pz	O1	O2	Oz
Acronyms	53.33%	58.17%	57.83%	67.83%
Illegal strings	72.00%	71.17%	72.50%	81.17%
Words	68.67%	70.33%	70.17%	78.00%
Pseudowords	57.50%	61.83%	64.17%	68.83%

Table 1. Results of euclidean distance method

Stimuli \ Channel	Pz	O1	O2	Oz
Acronyms	33.83%	45.67%	42.00%	55.67%
Illegal strings	47.00%	43.67%	46.17%	67.17%
Words	49.33%	47.50%	49.17%	62.83%
Pseudowords	36.50%	43.50%	42.67%	49.33%

Table 2. Results of DTW method

[10] C. Hema, M. Paulraj, and H. Kaur. Brain signatures: A modality for biometric authentication. In *Int'l Conf. Electronic Design (ICED)*, pages 1–4, Dec 2008.

[11] H. Jian-feng. Biometric system based on EEG signals by feature combination. In *Int'l Conf. Measuring Technology and Mechatronics Automation (ICMTMA)*, volume 1, pages 752–755, March 2010.

[12] J. Kent. Malaysia car thieves steal finger. BBC News, March 2005. [accessed December 15, 2014].

[13] E. J. Keogh and M. J. Pazzani. Derivative dynamic time warping. In *SDM*, volume 1, pages 5–7. SIAM, 2001.

[14] D. La Rocca, P. Campisi, and G. Scarano. EEG biometrics for individual recognition in resting state with closed eyes. In *Proceedings of the Int'l Conf. the Biometrics Special Interest Group (BIOSIG)*, pages 1–12, Sept 2012.

[15] N.-Y. Liang, P. Saratchandran, G.-B. Huang, and N. Sundararajan. Classification of mental tasks from EEG signals using extreme learning machine. *Int'l J. Neural Systems*, 16(01):29–38, 2006.

[16] E. Maiorana, G. Hine, D. LA ROCCA, and P. Campisi. On the vulnerability of an EEG-based biometric system to hill-climbing attacks algorithms' comparison and possible countermeasures. In *IEEE Sixth Int'l Conf. Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, Sept 2013.

[17] S. Marcel and J. Millan. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Trans Pattern Anal. Mach. Intell.*, 29(4):743–752, April 2007.

[18] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of artificial "gummy" fingers on fingerprint systems. *Datenschutz und Datensicherheit*, 26(8), 2002.

[19] Z. Mu and J. Hu. Research of EEG identification computing based on AR model. In *Int'l Conf. Future BioMedical Information Engineering (FBIE)*, pages 366–368. IEEE, 2009.

[20] M. Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

[21] M. Näpflin, M. Wildi, and J. Sarnthein. Testretest reliability of resting EEG spectra validates a statistical signature of persons. *Clinical Neurophysiology*, 118(11):2519 – 2524, 2007.

[22] P. Nguyen, D. Tran, X. Huang, and D. Sharma. A proposed feature extraction method for eeg-based person identification. In *Int'l Conf. Artificial Intelligence (ICAI)*, 2012.

[23] R. Palaniappan. Method of identifying individuals using VEP signals and neural network. *IEE Proc. Science, Measurement and Technology*, 151(1):16–20, Jan 2004.

[24] R. Palaniappan and D. P. Mandic. Biometrics from brain electrical activity: A machine learning approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):738–742, 2007.

[25] R. Paranjape, J. Mahovsky, L. Benedicenti, and Z. Koles'. The electroencephalogram as a biometric. In *Canadian Conf. Electrical and Computer Engineering*, volume 2, pages 1363–1366 vol.2, 2001.

[26] M. Poulos, M. Rangoussi, N. Alexandris, and A. Evangelou. Person identification from the EEG using nonlinear signal classification. *Methods of information in Medicine*, 41(1):64–75, 2002.

[27] N. A. A. E. M. Poulos, M. Rangoussi. On the use of EEG features towards person identification via neural networks. *Informatics for Health and Social Care*, 26(1):35–48, 2001.

[28] A. Riera, A. Soria-Frisch, M. Caparrini, C. Grau, and G. Ruffini. Unobtrusive biometric system based on electroencephalogram analysis. *EURASIP J. Adv. Sig. Proc.*, 2008, 2008.

[29] M. Ruiz-Blondet, N. Khalifian, B. C. Armstrong, Z. Jin, K. J. Kurtz, and S. Laszlo. Brainprint: Identifying unique features of neural activity with machine learning. In *Proc. 36th Annual Conf. the Cognitive Science Society*, July 2014.

[30] M. Ruiz-Blondet, N. Khalifian, K. J. Kurtz, S. Laszlo, and Z. Jin. Brainwaves as authentication method: Proving feasibility under two different approaches. In *Proc. 40th Annual Northeast Bioeng. Conf. (NEBEC)*, pages 1–2, Apr. 2014.

[31] S. Salvador and P. Chan. FastDTW: Toward accurate dynamic time warping in linear time and space. florida institute of technology, 2004.

[32] H. Shedeed. A new method for person identification in a biometric security system based on brain EEG signal processing. In *World Congress on Information and Communication Technologies (WICT)*, pages 1205–1210, Dec 2011.

[33] D. J. A. Smit, D. Posthuma, D. I. Boomsma, and E. J. C. Geus. Heritability of background EEG across the power spectrum. *Psychophysiology*, 42(6):691 – 697, 2005.

[34] J. Thorpe, P. C. van Oorschot, and A. Somayaji. Pass-thoughts: Authenticating with our minds. In *Workshop on New Security Paradigms (NSPW)*, pages 45–56, New York, NY, USA, 2005. ACM.

[35] S.-K. Yeom, H.-I. Suk, and S.-W. Lee. Person authentication from neural activity of face-specific visual self-representation. *Pattern Recog.*, 46(4):1159 – 1169, 2013.

[36] D. Zhu, X. Zhou, and Q. Guo. An identification system based on portable EEG acquisition equipment. In *Third Int'l Conf. Intelligent System Design and Engineering Applications*, pages 281–284. IEEE Computer Society, 2013.

[37] A. Zquete, B. Quintela, and J. P. da Silva Cunha. Biometric authentication using brain responses to visual stimuli. In A. L. N. Fred, J. Filipe, and H. Gamboa, editors, *BIOSIGNALS*, pages 103–112. INSTICC Press, 2010.