

PAPER • OPEN ACCESS

Preliminary study on EEG based typing biometrics for user authentication using nonlinear features

To cite this article: Intan Amalina *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **557** 012035

View the [article online](#) for updates and enhancements.

Preliminary study on EEG based typing biometrics for user authentication using nonlinear features

Intan Amalina¹, A Saidatul¹, CY Fook¹ and Z Ibrahim²

¹ Biosignal Processing Research Group (BioSIM), School of Mechatronic Engineering, Universiti Malaysia Perlis, Arau, Perlis, Malaysia

² Faculty of Technology, University of Sunderland, St Peter's Campus, Sunderland, SR6 0DD, United Kingdom

Abstract. The application of password usage in databased user authentication method is unsafe, due to higher chance of off-line guessing attacks. Hackers are able to save our personal information and can detect them until the perfect match is acquired. Thus, the aim of this research is to examine EEG signals during typing task from three engineering students (right handed) with age of between 19 to 23 years old to identify the regular human typing pattern to differentiate between one user to another. All the subjects were asked to perform three times typing tasks for 5 minutes with rest in between for 30 seconds. The subjects were required to rest before and after performing the typing tasks for one minute. Truscan EEG device (Deymed Diagnostic, Alien Technic, Czech Republic) with frequency sampling of 1024 Hz and 19 channels was used. This study applied Infinite Impulse Response (IIR) notch filter to remove 50Hz powerline artefacts and proposed the implementation of nonlinear features such as Distribution Entropy (DistEn) and Approximate Entropy (ApEn). The features were extracted from channels F3 and F4. The extracted entropies features vector are used as an input to k-Nearest Neighbour (kNN) classifier. As a conclusion, the kNN and LDA classifier giving promising accuracy which are 82.22% (F3), 88.52% (F4), 94.81% (F3 and F4) and 80.37% (F3), 81.48% (F4), 89.63% (F3 and F4) respectively.

1. Introduction

Biometric is used to identify the identity of an input sample when compare to a template to identify specific user by certain characteristics[1]. Those individual unique characteristics of the users can be in a form of DNA and fingerprint. Authentication is a procedure to confirm the identity of user and validate the reachability of the service to that user.

Authentication based on passwords involves username or ID that is rare to the user[2]. The passwords must be unrepeated, long, complex, consist of both letters and numerals[3]. A secure password is necessary to have lowercase and uppercase letters, special characters, as well as symbols. The full length of the password must be up to 8 characters or more.

Nowadays, passwords are unlimited to databased user authentication. It requires a great measure of performance security. The password is not secured because when users enter their password in public area and observed by spy cameras, this can cause password theft. In order to overcome this problem, the brain wave patterns are recorded during typing task to identify the regular human typing pattern to differentiate between one user and another.



2. Previous Study

Research was done by C. Lin and C. Wu to chance the possibility of using multi-channel electroencephalography (EEG) being collected before movements in order to predict human errors. The task done was hear-and-type task which imitated everyday numerical data entry tasks performed by phone operators. A computer program read out 30 nine-digit, randomly generated numbers without decimals in each trial. Those subjects typed out the numbers. For signal processing, the raw data at 1000 Hz sampling rate were filtered by to 30 Hz bandpass filter. Then 3 temporal features were extracted in a 150-ms time window. For classification, LDA classifier detected 74.34% of numerical typing errors [4].

Research done in [5] was to investigate human-fatigue using smartphone. Each of the subjects involved would have eight writing parts to be done at separate time. In each parts, 1193 taps would take 11.6 minutes. They are asked to type pre-specified writing using their smartphones. The session is separated into 10 segments. 4 features were extracted from the EEG signals. Support vector machine (SVM) and artificial neural network (ANN), are used detect the tiredness level of users. The accuracy obtained were 88.8% and 77.5% respectively.

3. Data Acquisition

Three male participants aged of between 19 and 23 years old have involved in this experiment. Right handed with no health issues are chosen. Furthermore, left-handed subjects, those with high blood pressure problem, those with exposure to general anaesthesia and smokers are excluded.

EEG signals were taken by using Truscan EEG device (Deymed Diagnostic, Alien Technic, Czech Republic) with 19 channels to track the brain signals. EEG signals were measured at 19 locations which are Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1 and O2. The position of electrodes follows the international 10-20 system as in Figure 1. EEG signals were taken by EEG machine consists of 16 bits A/D resolution. The participants took 5-10 minutes to remain calm before starting the trials. The sampling frequency was 1024 Hz. The impedances were remained lower than 5 k Ω . The cutoff range signal was 1Hz to 80Hz.

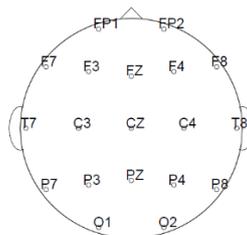


Figure 1. The location of electrodes based on 10-20 International System[6]

At first, the subjects filled in the demographic form and questionnaires. Then, they were given 5 minutes to rest. After that, the subjects were told to take a seat in front of LCD computer screen. All subjects were required to perform three times typing tasks for 5 minutes per each trial with rest in between for 30 seconds. The subjects were given a text for them to copy. The subjects were asked to rest before and after performing the typing tasks for a minute.

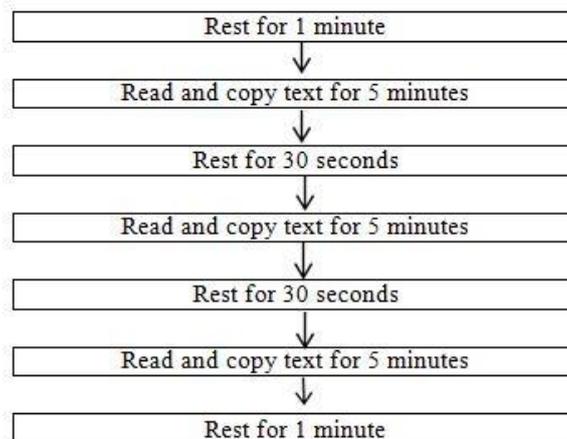


Figure 2. Experimental procedure

4. Signal Processing

In EEG signal processing, the raw signals were processed to identify the human regular typing patterns. The brain signals went through a few stages in order to be classified. To filter out the noise in the signal, the pre-processing method is applied. Then the informations in filtered signal are removed to describe their behaviours.

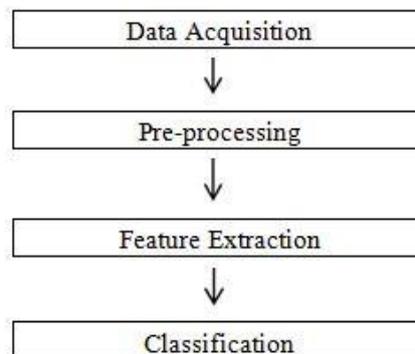


Figure 3. Flowchart of signal processing

4.1. Pre-processing

Raw EEG signals that have been taken have amplitudes in microvolts and consists of frequency up to 80 Hz. Thus, pre-processing is needed which involves filtering, and removal of noise. Notch filter of 50 Hz cutoff frequency were applied to eliminate power line interference[7]. The artefacts are particularly caused by wires, light fluorescence and electrode that involves in the data acquisition[8].

4.2. Feature Extraction

In this study, 2 types of nonlinear features are used due to the nature of EEG signal that is non-stationary signal. EEG signals change over time. The nonlinear features are tested separately and by combination to achieve a high accuracy.

4.2.1. *Approximate Entropy (ApEn)*. It was applied to measure nonlinear difference contained in filtered signal. It identifies any alterations in the behaviour and differentiate the sample resemblances by pattern length and similarity coefficient[9]. Time series with high irregular pattern results in higher ApEn value whereas time series with highly same patterns results in less ApEn value.

4.2.2. *Distribution Entropy (DistEn)*. DistEn evaluates the complexity of time-series by approximating empirical probability density function (ePDF) in the state space[10]. It has been revealed, that time-series with disorganized regime contains dispersive ePDF. This will make ePDF become concentrative for both random and periodic information. Furthermore, ePDF periodic data can concentrate in some separate distances. Therefore, it can be concluded that chaotic series will have higher values of DistEn whereas in periodic series, lower values of DistEn will be obtained.

4.3. Classification

4.3.1 *k Nearest Neighbour (kNN)*. The kNN classifier applies a feature distance in a data set to assign the group of the data belongings[11]. A group is formed when the distance within the data is near to each other whereas many groups are created when the distance within the data is far. kNN differentiate a testing data with a training data. The objects are arranged by greatest vote of its neighbours. k values vary to detect the match class between training and testing data.

4.3.2 *Linear Discriminant Analysis (LDA)*. LDA is easy in computational procedure and produces best results. Training feature vectors are divided by linear hyper planes to have a smaller variance inside the clusters. The data supposed to have normal distribution. The classes also need to have same covariance matrices. Furthermore, maximization in Fisher's criterion provides the greatest value of the distance among the classes means and the lowest value of variance[12].

5. Results and Discussions

Both extracted entropies are used as an input to classifiers such as k Nearest Neighbour (kNN) and Linear Discriminant Analysis (LDA). The entropies are extracted from F3 channel, F4 channel and both F3 and F4 channels.

Table 1. kNN and LDA classifier results based on nonlinear features and EEG channels

Classifiers	Nonlinear Features	Channels	Percentage (%)
kNN	ApEn & DistEn	F3	82.22
		F4	88.52
		F3 and F4	94.81
LDA	ApEn & DistEn	F3	80.37
		F4	81.48
		F3 and F4	89.63

Based on the Table 1, the combination of approximate entropy (ApEn) and distribution entropy (DistEn) shows a high accuracy. This means that the nonlinear features need to be combined in order to reach a high accuracy. By comparing both classifiers, it can be said that the kNN classifier shows the higher accuracy than LDA classifier. This is due to its nature which is effective to be used when the training data is large enough. Lastly, the results of this study display that the frontal channels F3 and F4 produce unique brainwave signals in individuals while performing the typing tasks.

6. Conclusion

In this research, it can be concluded that channels F3 and F4 give significant and unique brainwave signals in individuals. In order to ensure each individual's identity can be attained, repeating typing tasks are performed. Overall, by comparing both classifiers, the kNN and LDA classifier, the highest performance accuracy can be achieved by kNN which is 94.81%.

Acknowledgments

We would like to acknowledge the support and encouragement given by the PPK Mechatronic School, University of Malaysia Perlis (UniMAP). We also thank those colleagues from UniMAP who provided guidance in assisting the research.

References

- [1] R. Malathi and R. Jeberson Retna Raj, "An Integrated Approach of Physical Biometric Authentication System," *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 820–826, 2016.
- [2] Y. Zhao, S. Li, and L. Jiang, "Secure and Efficient User Authentication Scheme Based on Password and Smart Card for Multiserver Environment," *Secur. Commun. Networks*, vol. 2018, 2018.
- [3] R. Wash *et al.*, "Understanding Password Choices : How Frequently Entered Passwords Are Re-used across Websites This paper is included in the Proceedings of the Understanding Password Choices : How Frequently Entered Passwords are Re-used Across Websites," *Symp. Usable Priv. Secur.*, no. Soups, pp. 175–188, 2016.
- [4] C. J. Lin and C. Wu, "Predicting numerical data entry errors by classifying EEG signals with linear discriminant analysis," *Behav. Inf. Technol.*, vol. 34, no. 8, pp. 787–798, 2015.
- [5] H. Al-Libawy, A. Al-Ataby, W. Al-Nuaimy, M. A. Al-Tae, and Q. Al-Jubouri, "Fatigue Detection Method Based on Smartphone Text Entry Performance Metrics," *Proc. - 2016 9th Int. Conf. Dev. eSystems Eng. DeSE 2016*, pp. 40–44, 2017.
- [6] "EEGcomplexityAD_Kliman."
- [7] B. Harender and R. K. Sharma, "EEG signal denoising based on wavelet transform," *Proc. Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2017*, vol. 2017–Janua, pp. 758–761, 2017.
- [8] Hallmarq, "MRI artifacts - and how to avoid them," pp. 1–10, 2017.
- [9] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. Bhat, and J. E. W. Koh, "Application of entropies for automated diagnosis of epilepsy using EEG signals: A review," *Knowledge-Based Syst.*, vol. 88, pp. 85–96, 2015.
- [10] P. Li, C. Yan, C. Karmakar, and C. Liu, "Distribution entropy analysis of epileptic EEG signals," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015–Novem, pp. 4170–4173, 2015.
- [11] M. Pozzi and A. Der Kiureghian, "Response Spectrum Compatible PSD for High-Frequency Range," *22nd Conf. Struct. Mech. React. Technol.*, no. 1, pp. 5–8, 2013.
- [12] K. E. C. Vidyasagar, M. Moghavvemi, and T. S. S. T. Prabhat, "Performance evaluation of contemporary classifiers for automatic detection of epileptic EEG," *2015 Int. Conf. Ind. Instrum. Control. ICIC 2015*, no. Icic, pp. 372–377, 2015.