Acoustical User Identification Based on MFCC Analysis of Keystrokes

Matus PLEVA[1], Eva KIKTOVA[1], Jozef JUHAR[1], Patrick BOURS[2]

[1]Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Letna 9, 040 01 Kosice, Slovakia

[2]Norwegian Information Security Lab, Gjovik University College, Teknologivegen 22, N-2815 Gjovik, Norway

matus.pleva@tuke.sk, eva.kiktova@tuke.sk, jozef.juhar@tuke.sk, patrick.bours@hig.no

Abstract. *This paper introduces a novel approach of person identification using acoustical monitoring of typing the required word on the monitored keyboard. This experiment was motivated by the idea of COST IC1106 (Integrating Biometrics and Forensics for the Digital Age) partners to acoustically analyse the captured keystroke dynamics database using widely used time-invariant mathematical models tools. The MFCC (Mel-Frequency Cepstral Coefficients) and HMM (Hidden Markov Models) was introduced in this experiment, which gives promising results of 99.33 % accuracy, when testing 25 % of realizations (randomly selected from 100) identifying between 50 users/models. The experiment was repeated for different training/ testing configurations and cross-validated, so this first approach could be a good starting point for next research including feature selection algorithms, biometric authentication score normalization, different audio & keyboard setup tests, etc.*

Keywords

Acoustical analysis, biometrics, HMM, keystroke, MFCC, user identification.

1. Introduction

The problem of user identification using different biometric features is widely discussed not only for forensic purposes, but also for improving the usability of current information technologies and gadgets. Some people are looking for systems which will recognize them without giving any password, ID card, NFC (near field communication), biometric features (gait, face, fingerprint, handwriting, iris, voice, etc.) or verification questions. This identification could provide new dimension in human-computer interaction, when passing by your digital home assistant you could be informed about your next appointments or possible free time activities around or introduced by your friends' posts from social networks. Of course some people are afraid of identification on public places where Orwell's Big Brother scenarios could be recognized.

The keystroke dynamics (KD) analysis started years ago with analysis of timing information gathered from keyboard driver as key up/down, hold, flight, release, press etc. Using this approach is cheap, because no additional hardware needs to be used; only the software needs to be installed to capture this information. A significant number of researches reported interesting results based on this timing analysis and a deep survey by Banerjee & Woodard could be found in [1] or different classification approaches review by Karnan et al. in [2]. Using the surveillance camera the typed keys could be also recognized using vision analysis techniques described in [3]. There is also a so called static and dynamic approach, where for the dynamic one the keystrokes are monitored continuously during the work on the PC and when the system recognizes that the authenticated user is no longer at the workstation the locking of the session is executed and additional authentication process required [4].

The problem of acoustical keystrokes analysis for user identification purposes is not so widely covered nowadays and for example the acoustical analysis was used for keys identification or authentication purposes from free typed text which is a very challenging task [5] or recognizing the key pressed acoustically investigated by IBM research labs [6] and also revisited by current technologies (language model analysis included) by Berkeley University [7]. It could be said that the user identification based on acoustical monitoring of

the keyboard could be a very interesting research area and after tuning the system and the acoustic score normalization it could also be used for authentication purposes in the future.

The paper is organized as follows. In Section 2. the acoustical keystroke database collection is described. Section 3. provides an overview of the features and algorithms used for acoustical analysis of the keystrokes. Section 3. presents the results obtained during the different training and testing scenarios, and in the last Section the discussion about the results and future work will be concluded.

2. Database Collection

The data collection was prepared after analysis of previously collected acoustical keystroke data described in the papers [8] and [9], where the same word "kirakira" was typed and inspired by [10] we used four sessions for every user with short pause in-between them. The database was recorded on NIS Lab (Norwegian Information Security) in Gjovik where every of 50 participants typed 100 times the same word "password" in four consecutive sessions, and he has no display for visual control of the typed characters. The supervisor stops the participant after successfully typed 25 correct session passwords. According to available IT volunteers gender distribution there were 40 male participants comparing to 10 females of average age around 26 years.

The database capturing was done using cheap widely used webcam microphone (Logitech model QuickCam Pro 9000) in 10 cm distance from the desktop keyboard (DELL model SK-8135) in semi-controlled en-
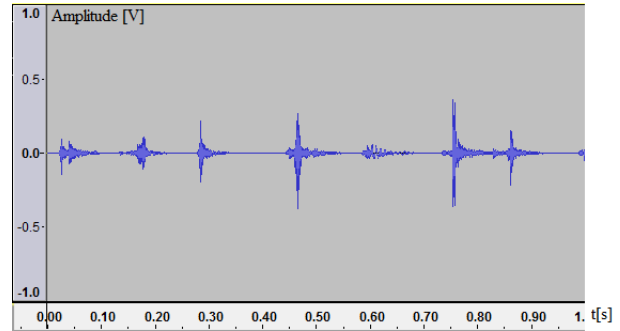

Fig. 2: Example of the captured audio file waveform.

vironment (quiet room, but with no sound insulation) as shown depicted in Fig. 1. Besides the audio data (depicted in Fig. 2) the timing information was also captured in the attached workstation PC.

Accuracy was used as the main evaluation metric in this case. The accuracy is defined as ratio of all tested recordings N and all tested recordings N decreased by the substitution error S according to Eq. (1).

$$ACC(Accuracy) = \frac{N - S}{N} \cdot 100. \qquad (1)$$

From the captured audio (44.1 kHz 16 bit PCM mono converted to 16 kHz for MFCC computation) and timing information files a database was compiled using the session number information in the filename. This information was later used to try to test how the number of the session relates with the discriminative potential of the recordings. In other words if it is better to use the first, middle or last session for training purposes or if the session number influence the way we type on the keyboard.

3. Acoustical Analysis of the Keystroke Recordings

First of all the feature extraction algorithm needs to be decided for particular purpose of acoustical analysis of keystroke sounds. From previous research of currently used techniques presented in [5] we decided to start with MFCC coefficients (keystroke MFCC coefficients depicted in Fig. 3), which seems to have better
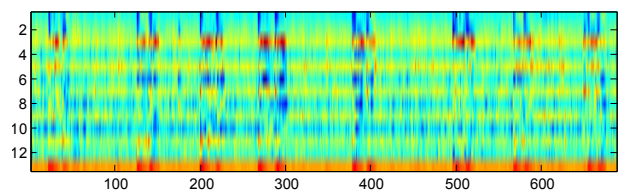

Fig. 1: Database collection setup.

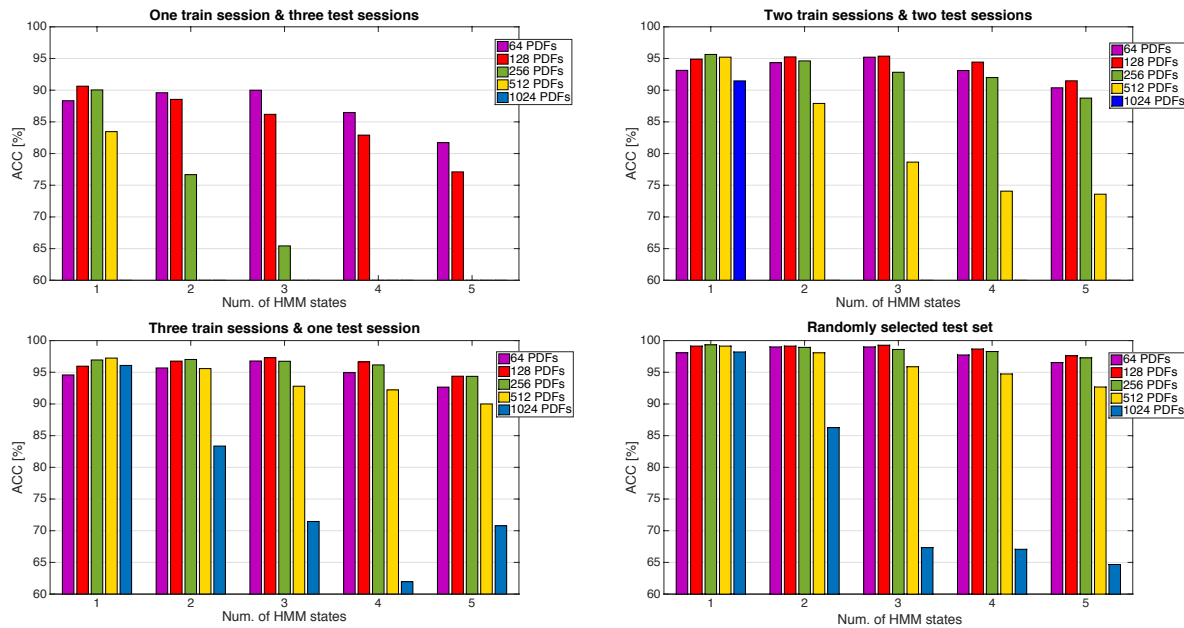**Fig. 3:** 13 MFCC coefficients of keystroke sound recording.

Fig. 4: Cross-validated accuracy of the user identification divided according to test set amount and randomly selected 25 % test set.

discrimination results comparing to FFT [7]. Also our previous experiences from acoustical events detection as gunshots or breaking glass reviewed in [11] shows that MFCC could be successfully used.

The acoustical analysis of time-invariant sounds could be done using widely used HMM models. The different timing of keystrokes sound and the well know techniques for training the HMM models could help to easy setup construction. The HTK tools were chosen for training and testing purposes [12].

The MFCC coefficients were extracted using 25 ms Hamming window and 10 ms frame shift. Mel-filter bank was created by 26 filters. The log energy was also computed together with first (velocity) and second (acceleration) time derivatives of the basic 12 cepstral coefficients. The resulted 39 dimension MFCC vector was used for training and testing the fully ergodic HMM models (with possibility to jump over next states or backwards) using Viterbi decoding algorithm.

The first approach was to select randomly one fourth of every user recordings to testing set. After discussions about possible real life applications we decide to try also other setups, for example only one session was used for training (we tried all 4 ones sequentially) etc.

For example in real life the user is not patient enough in the training phase of the system, so he needs to learn quickly, and then it is usually tested the trained models much more times. But of course having bigger training data increase the accuracy so finally we tried all combinations of training and testing sessions setup.

For HMM models we also decided to try not only one state transition matrix (excluding the input and output state) but also increasing the number of states, but then we have problems to train the models because of small amount of training data in some scenarios. The increasing number of PDF (Probability Density Functions) should also help the system accuracy, but there was also a problem with small amount of training data depending on number of training sessions used in particular scenario. The models with number of PDF smaller then 64 was outperformed so we do not present them in the paper.

4. Conclusion and Results Discussion

As you can see in Fig. 4 the accuracy (number of correct identifications divided by number of all test recordings) increased if more sessions were used for training and if we randomly select the test recordings, but this could be a problem in real life application. The best results were achieved for one state 256 PDF model trained on 75 % randomly selected recordings 99.33 %, which was 97.03 % after cross validation (where whole sessions were tests sets). The accuracy decreased when using two sessions for training to 95.64 % after cross validation. When using only one session for training the system achieved 90.62 % (cross-validated) and in this case the best result was 92.91 % for second session used for training, and worse 88.93 % for the fourth training session.

In future work we want to examine other possible features for user identification task, different transformation and discriminant analysis algorithms developed in our lab [13], and work on resulted score normalization for improving the biometric authentication potential of the acoustical analysis of KD.

Acknowledgment

References

[1] BANERJEE, S. P. and D. L. WOODARD. Biometric Authentication and Identification using Keystroke Dynamics: A Survey. *Journal of Pattern Recognition Research.* 2012, vol. 7, iss. 1, pp. 116–139. ISSN 1558-884X. DOI: 10.13176/11.427.

[2] KARNAN, M., M. AKILA and N. KRISHNARAJ. Biometric personal authentication using keystroke dynamics: A review. *Applied Soft Computing.* 2011, vol. 11, iss. 2, pp. 1565–1573. ISSN 1568-4946. DOI: 10.1016/j.asoc.2010.08.003.

[3] ROTH, J., X. LIU, A. ROSS and D. METAXAS. On continuous user authentication via typing behavior. *IEEE Transactions on Image Processing.* 2014, vol. 23, iss. 10, pp. 4611–4624. ISSN 1057-7149. DOI: 10.1109/TIP.2014.2348802.

[4] BOURS, P. Continuous keystroke dynamics: A different perspective towards biometric evaluation. *Information Security Technical Report.* 2012, vol. 17, iss. 1, pp. 36–43. ISSN 1363-4127. DOI: 10.1016/j.istr.2012.02.001.

[5] ROTH, J., X. LIU, A. ROSS and D. METAXAS. Investigating the discriminative power of keystroke sound. *IEEE Transactions on Information Forensics and Security.* 2015, vol. 10, iss. 2, pp. 333–345. ISSN 1556-6013. DOI: 10.1109/TIFS.2014.2374424.

[6] ASONOV, D. and A. RAKESH. Keyboard Acoustic Emanations. In: *IEEE Symposium on Security and Privacy.* Oakland: IEEE, 2004, pp. 3–11. ISBN 0-7695-2136-3. DOI: 10.1109/SECPRI.2004.1301311.

[7] ZHUANG, L., F. ZHOU and J. D. TYGAR. Keyboard acoustic emanations revisited. *ACM Transactions on Information and System Security (TISSEC).* 2009, vol. 13, iss. 1, pp. 1–26. ISSN 1094-9224. DOI: 10.1145/1609956.1609959.

[8] DOZONO, H., S. ITOU and M. NAKAKUNI. Comparison of the adaptive authentication systems for behavior biometrics using the variations of self organizing maps. *International Journal of Computers and Communications.* 2007, vol. 1, iss. 4, pp. 108–116. ISSN 2074-1294.

[9] NAKAKUNI, M., H. DOZONO and S. ITOU. Adaptive authentication system for behavior biometrics using supervised pareto self organizing maps. In: *Proceedings of WSEAS International Conference on Mathematics and Computers in Science and Engineering MAMECTIS.* Corfu: WSEAS, 2008, pp. 277–282. ISBN 978-960-474-012-3.

[10] ROTH, J., X. LIU, A. ROSS and D. METAXAS. Biometric authentication via keystroke sound. In: *2013 International Conference on Biometrics (ICB).* Madrid: IEEE, 2013, pp. 1–8. ISBN 978-1-4799-0310-8. DOI: 10.1109/ICB.2013.6613015.

[11] KIKTOVA, E., M. LOJKA, M. PLEVA, J. JUHAR and A. CIZMAR. Comparison of different feature types for acoustic event detection system. In: *Multimedia Communications, Services and Security.* Krakow: Springer, 2013, pp. 288–297. ISBN 978-3-642-38558-2. DOI: 10.1007/978-3-642-38559-9_25.

[12] YOUNG, S., G. EVERMANN, M. J. F. GALES, T. HAIN, D. KERSHAW, G. MOORE, J. ODELL, D. OLLASON, D. POVEY, V. VALTCHEV and P. C. WOODLAND. The HTK Book. Htk^3 [online]. 2006. Available at: http://htk.eng.cam.ac.uk/docs/docs.shtml.

[13] VISZLAY, P. Unsupervised Linear Discriminant Subspace Training Based on Heuristic Eigenspectrum Analysis of Speech. In: *13th Scientific Conference of Young Researchers.* Herlany: IEEE, 2013, pp. 261–264. ISBN 978-80-553-1422-8.

About Authors

Matus PLEVA was born in Kosice, Slovakia in 1977. He received his Ph.D. from the Department

of Electronics and Multimedia Communications of the Faculty of Electrical Engineering and Informatics at the Technical University of Kosice in 2010. He is the MC member in IC1106 COST action (Integrating Biometrics and Forensics for the Digital Age). He has published over 80 technical papers in journals and conference proceedings (35 in Scopus). His research interests include speech processing, human-machine interaction, mobile applications, networking, security, biometrics, etc.

Eva KIKTOVA was born in Liptovsky Mikulas, Slovakia in 1984. In 2009 she graduated M.Sc. (Ing.) at the Department of Electronics and Multimedia Communications of the Faculty of Electrical Engineering and Informatics at the Technical University of Kosice (with maiden name Vozarikova). In 2013 she recieved Ph.D. at the same department in the field of Telecommunications, where she works as a researcher. She has published over 40 technical papers in journals and conference proceedings (17 in Scopus). Her research is oriented on the field of the acoustic event detection and classification, speaker recognition and speaker diarization.

Jozef JUHAR was born in Poproc, Slovakia in 1956. He graduated from the Technical University of Kosice in 1980. He received Ph.D. degree in Radioelectronics from Technical University of Kosice in 1991, where he works as a Full Professor at the Department of Electronics and Multimedia Communications. He works as a head of the same department. He is author and co-author of more than 200 scientific papers (85 in Scopus). His research interests include digital speech and audio processing, speech/speaker identification, speech synthesis, development in spoken dialogue and speech recognition systems in telecommunication networks.

Patrick BOURS was born in Sittard in the Netherlands. He got his M.Sc. and Ph.D. in Discrete Mathematics from Eindhoven University of Technology in the Netherlands with a specialization in Coding Theory. He worked for 10 years for the Dutch Government in the area of asymmetric crypto and since 2005 he is working at Gjovik University College, first as a PostDoc, since 2008 as an Associate Professor and since 2012 as Professor. His research focus is on behavioral biometrics with a special interest in keystroke dynamics and continuous authentication (with 50 Scopus documents). His research also includes finding innovative manners to identify persons based on their daily behavior.