# A Preliminary Study of a New Soft Biometric Finger Recognition for Keystroke Dynamics

Syed Zulkarnain Syed Idrus, Estelle Cherrier, Christophe Rosenberger, Patrick Bours

# A Preliminary Study of a New Soft Biometric: Finger Recognition for Keystroke Dynamics

Syed Zulkarnain Syed Idrus, Estelle Cherrier and Christophe Rosenberger
*Université de Caen Basse-Normandie, UMR 6072 GREYC, F-14032 Caen, France*
*ENSICAEN, UMR 6072 GREYC, F-14050 Caen, France*
*CNRS, UMR 6072 GREYC, F-14032 Caen, France*
{*syed-zulkarnain.syed-idrus,estelle.cherrier,christophe.rosenberger*}*@ensicaen.fr*

Patrick Bours
*NISLab, Gjøvik University College, Gjøvik, Norway*
*patrick.bours@hig.no*

June 9, 2012

## Abstract

*Keystroke dynamics is an interesting biometric modality as a user can be authenticated while typing a passphrase or a password on a keyboard. In order to improve the accuracy of biometric systems, it is possible to exploit some prior information that can be known or extracted from the biometric raw data. This process is known as "soft biometrics". In this paper, we propose a new soft biometric approach for keystroke dynamics consisting in extracting from keystroke dynamics templates if the user types with one or two hands. Preliminary experimental results show a correct accuracy recognition equal to 80%.*

## 1 Introduction

Keystroke dynamics is a low cost biometric modality as it enables to authenticate or identify an individual based on its way of typing a passphrase or a password [1]. This biometric modality does not provide as good recognition results as for iris or fingerprint but is very familiar to users (everybody is used to type a password on a computer) [2; 3]. In order to improve performance, many solutions are possible such as multibiometrics (fuse different algorithms or features) or soft biometrics (consisting in using *a prior* information).

In this paper, we address the latter approach. There are few works concerning soft biometrics for keystroke dynamics. Epp *et al.* show it is possible to detect the emotional state of an individual through its way of typing [4]. In this case, detecting anger and excitation is possible in 84% cases. Recently, Giot *et al.* shows it is possible to detect the gender of an individual through the typing of a fixed text [5]. The gender recognition rate is superior to 90% and the use of this information, in association to the keystroke dynamics authentication, reduces the EER of 20%.

The purpose of this research is to conduct a preliminary study to predict if the user uses one finger, two fingers or all fingers upon login during an authentication process. Subsequently, the objective is to propose and generate novel models of authentication that can be utilised as references and also enhance the keystroke dynamics authentication recognition *vis-à-vis* biometric security applications or systems.

This study is related to soft biometrics, in the sense that this information is not sufficient to authenticate nor identify a user. Predicting if a user types in a usual way (i.e. with 1, 2 or more fingers) can reinforce the authentication/identification performed by another biometric system.

The contributions of this paper are threefold. Firstly, we first present a comparison of the average typing time depending on the way of typing (with one finger, two fingers or all fingers). Then, we introduce a definition of the complexity of typing a password and we look for a possible link between the complexity and the total typing time. Finally, a study is conducted on the accuracy rate of the proposed method.

This paper is organised as follows. The issues within this domain i.e. biometrics technology are addressed in Section 2. Section 3 is devoted to the description of the proposed methodology. The database is described, together with the enrolment process, and the tools that will be used for analysis purposes. In Section 4, we present the results obtained, showing that this preliminary study deserves to be pursued and deepen. Section 5 presents the conclusions and the future works to be addressed.

# 2 Problem Statement

No single biometric is expected to effectively satisfy the needs of all identification (authentication) applications. Subsequently, a number of biometrics have been proposed, studied and evaluated. Thus, each biometric has its strengths and limitations, and each biometric appeals to a particular identification (authentication) application [6; 7]. The acceptability of a biometric as an application is often a trade-off between the sensitivity of a community to various perceptions or taboos and value or convenience offered by a biometric-based identification [6].

The general problem of personal identification raises a number of important research issues: which identification technologies are the most effective to achieve accurate and reliable identification of individuals? Some of these problems are well-known open problems in the allied areas, for example, pattern recognition and computer vision, while the others need a systematic cross-disciplinary effort. Authors [6] believed that biometrics technology alone may not be sufficient in order to resolve these issues effectively, thus the solutions to the outstanding open problems may lie in the innovative engineering designs exploiting constraints, and otherwise, it would be unavailable to the applications and in harnessing the biometrics technology in combination with other allied technologies.

# 3 Proposed Methodology

Biometric recognition is said to be a part of the solution, it is however, not a solution *per se* [8]. Therefore, this preliminary experiment aims at seeking a portion of the solution and revealing if there are some interesting research tracks in this protocol. Subsequently, should there be any, we intend to pursue further with this research.

## 3.1 Database

Since this is an initial stage of the study, the population will not be a vast amount of numbers. Therefore, five people were selected and had volunteered to participate in this experiment. However, we will see that the number of sessions, the number of passwords and the different considered cases can compensate for the small number of users. To create the database, some experimentation tools are required, which does not incur an additional cost as all are already available in the laboratory: a laptop, an external keyboard, a software (GREYC Keystroke software) to perform the keystroke process and to store the keystroke data. The location and position of the hardware are to be in a stagnant position and immoveable throughout the session for the authenticity of the outcomes. Any adjustments or changes either in the system or externally are to be done only by the operator(s) of this experiment. Now, we describe the considered cases: three ways of typing; three passwords; and one keyboard.

We define three classes depending on the number of fingers used to type, denoted $C_1$ to $C_3$:

| | |
|---|---|
| $C_1$ | only one finger is used |
| $C_2$ | one finger per hand is used |
| $C_3$ | all fingers of both hands are used |

We define three passwords of eight characters that are randomly chosen from a (French or English) dictionary, denoted $P_1$ to $P_3$:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $P_1$ | a | n | n | u | a | i | r | e |
| $P_2$ | s | e | a | s | o | n | a | l |
| $P_3$ | d | i | a | l | o | g | u | e |

We define three keyboards identification (ID), denoted $K_1$ to $K_3$:

| | |
|---|---|
| $K_1$ | for typing using one finger |
| $K_2$ | for typing using two fingers |
| $K_3$ | for typing using all fingers |

## 3.2 Enrolment process

For the enrolment of each password $P_j$, $j = 1, 3$ and for each finger class $C_i$, $i = 1, 3$, each user will have to key-in five times. By using $C_1$ with only one hand, each user is expected to key-in five entries for each password character without any error. If there are errors, the current entry has to resume and the user will have to proceed until five successful entries have been recorded into the system. The first, second and third character passwords entries are to be typed in a normal typing pace. Then, the same finger must be used to press the 'Enter' key, and hence the enrolment is then captured for $C_1$.

For the $C_2$ stage is realized by using two hands (one finger of each hand), and the same protocol as $C_1$ is adopted. In the last and final stage, i.e. $C_3$, the user is required to use more than two or all of his or her fingers, which can be used freely and without any constraints.

Therefore, at the end of the enrolment, we have gathered 225 data (= 3 passwords x 3 classes of finger x 5 users x 5 entries) in the database.

## 3.3 Data analysis

We present in this part the main contributions of this preliminary study. We look for criteria which can discriminate the way of typing, i.e. with one finger, two fingers, or more than two fingers. The first criterion we propose to analyze is the total time necessary to type the password. This information is given for any entry by the GREYC Keystroke software. We denote $t_{i,j,k,l}$ the time for class $C_i$, password $P_j$, user $k$ and entry $l$. So, we can compute for each class $C_i$:

$$\mathrm{T}_{\mathrm{total,i}} = \sum_{j=1}^{3} \sum_{k,l=1}^{5} t_{i,j,k,l} \tag{1}$$

The second criterion relies on the definition of the complexity of the passwords. We propose to evaluate the typing complexity of each password $P_j$, denoted $CP_j$, by calculating the distance of the location between each key on the keyboard namely the alphabets 'a' to 'z' (we use a French keyboard). The keyboard is represented as a grid, whereby each key has its own location number identified as the key codes. More precisely, the letter 'a' is represented as (0.00 0), 'z' as (1.00 0), 'q' as (0.25 1), 's' as (1.25 1) and so on. All those values are an estimate of determining the unit, where the unit is the size of the button for both horizontally (*x*-axis) and vertically (*y*-axis). Figure 1 illustrates the graphical notion of the location of keys namely the key codes on the keyboard. Then, the correlation between the complexity $CP_j$ of password $P_j$ and the time needed for typing $P_j$ can be evaluated through the Pearson Correlation factor:

$$\mathrm{corr_j} = \frac{Cov(CP_j, \sum_{i,k,l} t_{i,j,k,l})}{\sigma(CP_j).\sigma(\sum_{i,k,l} t_{i,j,k,l})} \tag{2}$$

Where $Cov$ represents the covariance matrix and $\sigma()$ is the standard deviation.

The last criterion we propose relies on the recognition rate using Support Vector Machine learning (SVM) introduced by [9]. It consists of an evaluation of the class ($C_i$) recognition rate in function of the ratio of data kept for the learning stage. We use LIBSVM [10] with default values.

The computation on SVM is done for 100 iterations for each percentage of the learning ratio (we selected between 1% and 90% of total data to define the training set

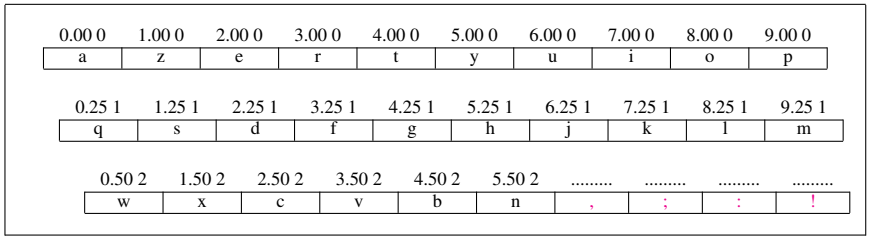


Figure 1: Location of keys on a *AZERTY* keyboard

of data) and calculating the average to produce the recognition rate. The expectation of this is to obtain higher recognition rates that are consistent. This computation takes approximately five hours and it is run separately for the two sets of test namely three classes ($C_1$ to $C_3$) and two classes ($C_1$ & $C_2$_$C_3$) of finger.

# 4 Results

The expected outcome of this study is to foresee if the proposed model is able to predict if the user uses one, two or more fingers to key-in the password during the authentication process. Figure 2 illustrates the average total typing time $\mathrm{T}_{\mathrm{total,i}}$, $cf$ equation (1) of five users for the three scenarios ($C_1$ to $C_3$) to type three passwords $P_1$ to $P_3$. We can see that the curves corresponding to $C_1$, $C_2$ and $C_3$ are at three different timing levels. This may be explained as $C_1$ uses one finger (with one hand), $C_2$ uses two fingers (with two hands) and $C_3$ uses all fingers (with two hands), and hence having different flight times. Thus, by looking at the curves, surely the more fingers we use to type, the lower the flight time it will be and the less the time it takes to type the password.

By performing the computation on all three passwords and their complexity difficulty, the objective is to find out if there is a significant difference between passwords with finger classes. The results of the matrix generated on the total time (in millisecond) for the three classes of finger ($C_1$ to $C_3$) are shown in Table 1, which later will be used to compute the correlation coefficient.

Table 1 shows the average time $\mathrm{T}_{\mathrm{total,i}}$ taken for typing each password $P_j$ for all five users in different

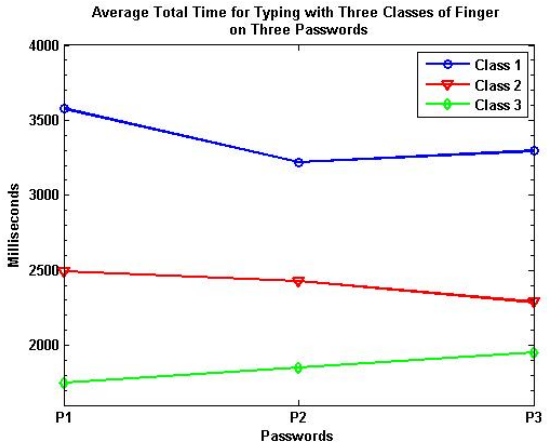


Figure 2: Average Total Time of Typing

scenarios $C_i$. The results illustrate that the more fingers the user used to type the passwords, obviously the less the typing time is. For example, for password ($P_1$), by using $C_3$, the value is 1748 ms, which is less than any other finger classes. Subsequently, it is the same for passwords ($P_2$) and ($P_3$). Somehow or rather, it is strange to see that the less complex password takes the most time using one finger, which is $C_1$_$P_1$. We have not performed any analysis on this, thus at this point in time it is undetermined.

Table 2 shows the results in terms of correlation coefficient computation $\mathrm{corr_j}$, $cf$ equation (2) from the results in Table 1. Judging by the column on the right, it states that there is some similarity between all the classes of finger and passwords complexity. We are able to see that there is a strong correlation between the two classes of

Table 1: Average Total Time and Complexity Difficulty of Typing

| Passwords | $P_1$ | $P_2$ | $P_3$ |
|---|---|---|---|
| Classes of Finger | Average Total Time | | |
| $C_1$ | 3577 ms | 3219 ms | 3294 ms |
| $C_2$ | 2493 ms | 2428 ms | 2286 ms |
| $C_3$ | 1748 ms | 1852 ms | 1957 ms |
| Complexity Difficulty | 30.2182 | 34.2133 | 51.2893 |

finger namely two and all fingers as opposed to one finger, for which the time taken to type the passwords. The results show that two and all fingers produced the value of $-0.9914$ and $0.9369$, respectively, where the values are closer to the absolute value i.e. 1, unlike the one finger with the value of $-0.4807$. Thus, there is a high correlation respectively between $C_2$ and $C_3$, and password complexity.

Table 2: Correlation Coefficients

| | | | | |
|---|---|---|---|---|
| 1.0000 | 0.5916 | -0.7570 | -0.4807 | $C_1$ |
| 0.5916 | 1.0000 | -0.9747 | -0.9914 | $C_2$ |
| -0.7570 | -0.9747 | 1.0000 | 0.9369 | $C_3$ |
| -0.4807 | -0.9914 | 0.9369 | 1.0000 | |
| $C_1$ | $C_2$ | $C_3$ | | |

Figure 3 illustrates the results of the recognition rates on different learning ratios with all classes $C_1$, $C_2$ and $C_3$. It appears that the results are not very good. From the learning ratios of 1% and up to 45%, it starts of with a very low recognition rate and gradually increasing. But, it sustains its performance at 45% to 90% with an average of 65% for the recognition rate. Thus, the plotted data are not very impressive and does not produce significant impact of the experiment.

Figure 4 illustrates the results of the recognition rates on different learning ratios with one finger ($C_1$) vs. more than one finger (i.e. $C_2$ and $C_3$). The results are much better, where the learning ratios of 1% and up to 25% have a steep increase in the corresponding recognition rates. Furthermore, the performance between 25% and 90% learning ratios sustain a much better percentage at 85% on average for the recognition rate. Hence, the curve projected is finer and produces a significant impact towards the results of this experiment.

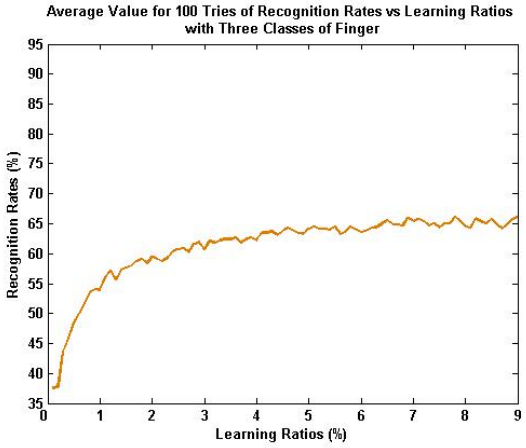


Figure 3: Recognition Rates (3 classes)

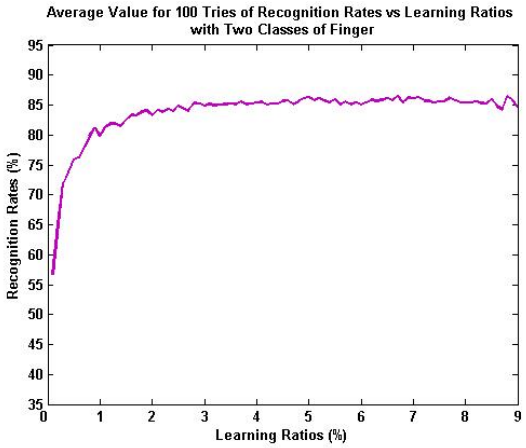


Figure 4: Recognition Rates (2 classes)

# 5 Conclusions and Future Work

We proposed a new soft biometric approach for keystroke dynamics. It consists of predicting the way of typing of the user, by defining three classes, depending on the number of fingers used to type a password. In this preliminary study, we focused on different criteria to classify the users, namely the average total time of typing, the correlation between this time and the complexity of the chosen password. After a study of the recognition rate using SVM, we noticed that better results are obtained with only two classes: one finger vs. more than one finger.

Since the presented results reveal promising, further study will be conducted on a larger data set. We also plan to exploit the video capture by the integrated webcam, to enhance the performances by a fusion method.

# References


[1] R. Giot, M. El-Abed, and C. Rosenberger, "Keystroke dynamics overview," in *Biometrics / Book 1* (D. J. Yang, ed.), vol. 1, ch. 8, pp. 157–182, InTech, July 2011.

[2] A. Messerman, T. Mustafic, S. Camtepe, and S. Albayrak, "Continuous and non-intrusive identity verification in real-time environments based on free-text keystroke dynamics," in *Biometrics (IJCB), 2011 International Joint Conference on*, pp. 1–8, IEEE, 2011.

[3] P. Bours, "Continuous keystroke dynamics: A different perspective towards biometric evaluation," *Information Security Technical Report*, no. 0, pp. –, 2012. In Press, Corrected Proof.

[4] C. Epp, M. Lippold, and R. Mandryk, "Identifying emotional states using keystroke dynamics," in *Proceedings of the 2011 annual conference on Human factors in computing systems*, pp. 715–724, 2011.

[5] R. Giot and C. Rosenberger, "A new soft biometric approach for keystroke dynamics based on gender recognition," *Int. J. Info. Tech. and Manag., Special Issue on "Advances and Trends in Biometrics by Dr Lidong Wang*, vol. 11, no. 1/2, pp. 35–49, 2012.

[6] A. Jain, R. Bolle, and S. Pankanti, *Introduction to Biometrics: Personal Identification in Networked Society*, ch. 1 - Introduction, pp. 1–41. Boston, MA: Kluwer Academic, 1999.

[7] W. Stallings and L. Brown, *Computer Security: Principles and Practice*. United States of America: Pearson Prentice Hall, 2008.

[8] A. K. Jain, "Next generation biometrics," December 2009. Department of Computer Science and Engineering, Michigan State University / Department of Brain and Cognitive Engineering, Korea University.

[9] V. Vapnik, *Statistical learning theory*. Wiley, 1998.

[10] C.-C. Chang and C.-J. Lin, "Libsvm - a library for support vector machines."