Scientific
Research
Publishing

Keyboard Dependency of Personal Identification Performance by Keystroke Dynamics in Free Text Typing

Yoshitomo Matsubara[1], Toshiharu Samura[2], Haruhiko Nishimura[1]

[1]Graduate School of Applied Informatics, University of Hyogo, Kobe, Japan
[2]Department of Electrical and Computer Engineering, National Institute of Technology, Akashi College, Akashi, Japan
Email: me1213@ob.akashi.ac.jp, samura@akashi.ac.jp, haru@ai.u-hyogo.ac.jp

Abstract

We have proposed some methods for feature extraction and identification that enable identification of individuals through long-text input as an important topic in keystroke dynamics research. As to the robustness in practical circumstances, there exists a question on the keystroke dynamics how much the recognition accuracy is influenced by the change of keyboard. By comparing the performance in the cases of using the same keyboard and different keyboards, the dependencies on keyboards are evaluated through three implemented experiments for subjects. As a result, it is found that we do not need to worry about the keyboard difference for users whose typing skills reach high level with about 900 or more letters in 5 minutes, and only for the remaining users it would be necessary to register their profile data with respect to each keyboard they use in order to avoid recognition accuracy degradation.

Keywords

Keystroke Dynamics, Biometrics, Information Security, Keyboard Dependency

1. Introduction

Timing data for keystrokes follow a fixed pattern, and the biometric approach using such data is called keystroke dynamics. There has been considerable research on keystroke dynamics owing to the advantage of using only a keyboard without requiring special equipment such as a fingerprint scanner [1]-[4]. Most of them have focused on user authentication at the time of log-in, using not only information about a series of input characters

for password verification, but also keystroke matching as a part of the authentication process [5] [6]. On the other hand, it is possible to obtain sufficient information for deriving keystroke dynamics statistically by using sentences of a certain length. The keystroke authentication with such long-text input has recently become the subject of academic discussion [7]-[16] and can be used to detect spoofing after log-in process by observing keystroke data when using a system.

In our previous studies [11]-[16], we have proposed methods for feature extraction and identification that enable identification of individuals through long-text input as an important topic in keystroke dynamics research. We use keystroke timing for single character and paired character sequences when the user is inputting Latin characters. For identification methods, we adopt our previously proposed the Weighted Euclidean Distance (WED) method [11] [12], the Array Disorder (AD) method proposed by Gunetti *et al.* [8] [9], and the WED + AD method which integrates the WED and AD methods [13]-[15]. Through a large-scale study involving 189 participants, it has been confirmed that the WED + AD method stably provides the best recognition accuracy among three methods [15]. Furthermore, we have constructed new profile generation methods, the profile-updating and profile-combining methods, to reinforce the robustness of keystroke dynamics and demonstrated the effectiveness of them through the examinations with subject's practical data [17]-[19].

As to the robustness in practical circumstances, there exists a question on the keystroke dynamics how much the recognition accuracy is influenced by the change of keyboard. If the influence is significant, users would need to register their profile data keyboard by keyboard they use. Pointed out by many researchers, this point has not been examined yet. In this paper, we investigate dependencies on keyboards for the personal identification by keystroke dynamics in Japanese free text typing. By comparing the performance in the cases of using the same keyboard and different keyboards [20], the dependencies on keyboards are evaluated through three implemented experiments for subjects.

In Section 2, we briefly introduce the data collection and the feature extraction we have dealt with so far. Next, we present three identification methods in free text typing in Section 3. In Section 4, evaluation procedure for keyboard difference is given, and experimental results are shown in Section 5. Finally, conclusion is given in Section 6.

2. Data Collection and Feature Extraction

This section describes keystroke data collection and how to extract features from the data. Participants input a different text each time for 5 minutes. For each entry, typed key, key pressed time and key released time are recorded as raw data. **Figure 1** shows an example of collected keystroke data. The first, second and third fields show typed key, whether the key was pressed (*p*) or released (*r*) and UNIX time of the event respectively. For collecting the data, we use a web-based system, and it has typing support software that is familiar to the participants, thereby lowering effects related to unfamiliarity and nervousness. **Figure 2** shows a screenshot of the software interface used in this process. Since this experiment focuses only on Latin alphabet input keystroke, conversion into Japanese kanji characters is not designed. Detailed explanation about the interface of this system is described in [20] [21].

Next, we describe feature extraction from keystroke data. The notation 1*pr* in **Figure 3** indicates the time from press to release of a single key and is referred to below as key press duration. The notation 2*rr* indicates

```
b, p, 1197417770648
a, p, 1197417770733
b, r, 1197417770791
i, p, 1197417770816
a, r, 1197417770823
i, r, 1197417770872
k, p, 1197417770972
k, r, 1197417771039
i, p, 1197417771112
i, r, 1197417771167
```

Figure 1. Example of collected keystroke data (typed key, pressed (p) or released (r), and UNIX time in order from left).

Figure 2. Snapshot of keystroke data collection system.
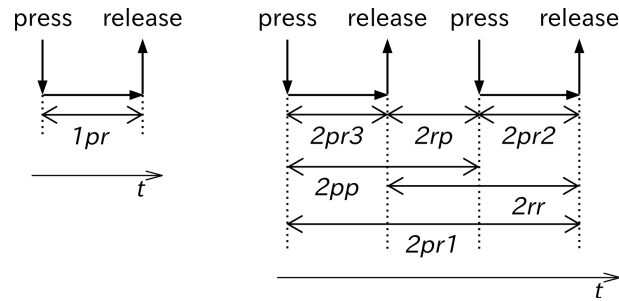

Figure 3. Keystroke measurements of single letter (left) and letter pair (right).

the time from the release of one key to the release of the following key when typing a consonant-vowel pair. The time from release of the first key to the time of pressing the following key (2*rp*) and the time from pressing the first key to pressing the second key (2*pp*) are also considered. Furthermore, 2*pr*1 indicates the time from pressing the first key to releasing the next key, 2*pr*2 indicates the key press duration when typing the second (vowel) key, and 2*pr*3 indicates the key press duration when typing the first (consonant) key.

The average and standard deviation of each of the seven measures described above are used as the feature indices for identification of individuals. For these feature indices, standardization is performed according to the following equation:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

(1)

Here, $x_{\min}$ and $x_{\max}$ respectively refer to the minimum and maximum values obtained from the feature indices of all subjects. It has been confirmed in our previous works [11] [13]-[15] that the standard deviations for six letter pairs (2**) in feature indices have little contribution to improve recognition accuracy. Therefore, we do not adopt them in this study.

3. Identification Methods

This section describes identification methods using the Weighted Euclidean Distance (WED) proposed by Samura and Nishimura [11] [12], the Array Disorder (AD) proposed by Gunetti and Picardi [8] [9], and the WED + AD proposed by Samura and Nishimura [13]-[15].

3.1. Weighted Euclidean Distance (WED)

Taking the first profiling document of Typist A as *docA*1, the profiling document of each participant can be represented as *docA*1, *docA*2, $\cdots$, *docAN*, *docB*1, *docB*2, $\cdots$, *docBN*, *docC*1, $\cdots$. An unknown document is represented as *docUK*. The *WED*(*docUK*,*docA*1) used as the identification function is given by the following equation:

$$WED\left(docUK,docA1\right) = \sqrt{\frac{1}{m}\sum_{\alpha=1}^{m}\frac{1}{n_\alpha}\sum_{i=1}^{n_\alpha}\left(k_{\alpha(i)} - r_{\alpha(i)}\right)^2} \tag{2}$$

The index of the feature indices α (=1, 2, $\cdots$) is then 1*pr.av*, 1*pr.sd*, $\cdots$. Furthermore, m is the number of contributing feature indices, $\alpha(i)$ indicates the feature index α for the i-th character (single letter or letter pair), and n_α is the number of characters therein. n_α will vary greatly with respect to the number of characters compared when, for example, taking the keystroke feature indices for single letters and those for two-letter combinations. $k_{\alpha(i)}$ is the $\alpha(i)$ feature index standardized according to Equation (1) for a profiling document (e.g., *docA*1), and $r_{\alpha(i)}$ is that for $\alpha(i)$ of an unknown document (*docUK*). The *WED* is normalized to 0 - 1 using the weightings $\frac{1}{m}$ and $\frac{1}{n_\alpha}$. **Figure 4** illustrates the 2*pp.av* feature indices for two documents, *docUK* and *docA*1. For five 2*pp.av* letter pairs common to the two documents, the *WED* becomes as follows:

$$WED_{2pp.av}\left(docUK,docA1\right) = \sqrt{\frac{1}{5}\left(0.33^2 + 0.06^2 + 0.12^2 + 0.22^2 + 0.25^2\right)}. \tag{3}$$

3.2. Array Disorder (AD)

The AD, which is called the R-measure in Gunetti and Picardi [9], but referred to as the Array Disorder in the present study, ranks characters according to their feature index values, and evaluates the disorder of the rankings. Standardized feature indices are sorted in increasing order, the difference in rankings of each is extracted, and the total of each is taken as distance.

$$AD\left(docUK,docA1\right) = \frac{1}{m}\sum_{\alpha=1}^{m}\frac{1}{\omega\left(n_\alpha\right)}\sum_{i=1}^{n_\alpha}\left|rk_{\alpha(i)} - rr_{\alpha(i)}\right| \tag{4}$$

$$\omega\left(n_\alpha\right) = \begin{cases} \dfrac{n_\alpha^2}{2} & \left(n_\alpha : \text{even}\right) \\[3mm] \dfrac{n_\alpha^2 - 1}{2} & \left(n_\alpha : \text{odd}\right) \end{cases} \tag{5}$$

When n_α characters are used to compare a feature index α, if n_α is even then the distance is divided by $\frac{n_\alpha^2}{2}$; if n_α is odd then the distance is divided by $\frac{n_\alpha^2 - 1}{2}$. Finally, the value is normalized to the range 0 - 1 by

docUK				docA1	
ki	0.02	← 0.33 →		ki	0.35
sa	0.18	← 0.06 →		sa	0.24
tu	0.25	← 0.12 →		tu	0.13
no	0.10	← 0.22 →		no	0.32
hu	0.34	← 0.25 →		hu	0.09
wo	0.41				

Figure 4. Illustration of the Weighted Euclidean Distance (WED) processing for existing letter pairs and their standardized values of 2*pp.av* in *docUK* and *docA*1, where there are five letter pairs in common between them.

dividing the value by the number of contributing feature indices *m*. **Figure 5** shows the feature index values rearranged in increasing order from **Figure 4**. The sum of the difference in rankings of each character is normalized, resulting in the following calculation of the *AD* value:

$$AD_{2pp.av}\left(docUK,docA1\right) = \frac{2}{5^2-1}\left(4+2+0+2+4\right). \tag{6}$$

3.3. WED + AD

In contrast to the WED, which evaluates the magnitude of differences in feature index values between documents, the AD focuses on differences between documents in ranking patterns of the feature indices. The WED + AD complementarity incorporates the features of the WED and AD. The distance of the WED + AD is given by the following equation:

$$\left[WED+AD\right]\left(docUK,docA1\right) = WED\left(docUK,docA1\right) + AD\left(docUK,docA1\right) \tag{7}$$

In this measure, neither *WED* nor *AD* dominates because they are normalized to the range 0 - 1. Detailed explanation for the above method is given in [15].

Identification is performed using the nearest-neighbor rule. In other words, comparisons are performed between the unknown document and each of the profile documents, and the typist of the profile document that gives the lowest value is taken to be the typist of the unknown document. For example, given five profile documents (*A1*-*A5*) for a subject (Typist A), we can expect that if Typist A also typed the unknown document then its distance will be close to one of the five states of *A1*-*A5*. In the case where the typist of the closest profile document is matched with the typist of the test document, the identification is classified as a success; otherwise, the identification is classified as a failure. Such validation is performed in turn on each document used in this study, and the recognition accuracy is calculated by (the number of successful identifications)/(the number of test documents) × 100%.

4. Evaluation Procedure for Keyboard Difference

To investigate the influence of keyboard difference, we calculate the recognition accuracy in the following two cases. As shown in **Figure 6**, one is the case where an extracted unknown document (*docUK*) typed with keyboard X is identified with profile documents typed with the same keyboard X, and the same for Y. After this
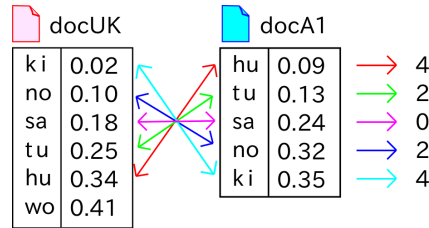

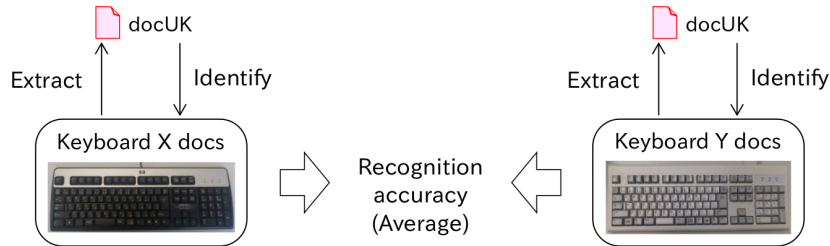**Figure 5.** Illustration of the Array Disorder (AD) processing for the example in **Figure 4**.

Figure 6. Evaluation procedure in the case where an extracted unknown document (*docUK*) typed with keyboard X is identified with profiled documents typed with the same keyboard X, and the same for Y (intra-keyboard evaluation).

process, each recognition accuracy from keyboard X and Y is averaged. We call this procedure "intra-keyboard evaluation." The other is the case where an extracted unknown document (*docUK*) typed with keyboard X is identified with profiled documents typed with the other keyboard Y, and *vice versa* as illustrated in **Figure 7**. Each recognition accuracy from keyboard X and Y is averaged as well as the intra-keyboard evaluation. We call this procedure "inter-keyboard evaluation."

We perform experiments for subjects who have 5 or more documents, each of which contains 500 or more letters. Since this criterion has been used as a standard condition in our previous works [11]-[17], we adopt it in this paper too. As shown in **Figure 8**, the number of documents from each user is set to 5 by choosing 5 documents at random from those created by a subject who typed an excess number of documents. Analysis is performed taking, as a single set, the number of documents equal to 5× (the number of subjects). Considering possible bias in document selection, analysis is performed five times with 5 different document sets.

5. Experimental Results

We implemented three experiments exchanging two types of keyboards.

5.1. Experiment Using Desktop Keyboards with Low Typing Skill Subjects (Experiment I)

In this experiment, two keyboards were prepared as shown in **Figure 9**. One is Hewlett-Packard BK-0316 and
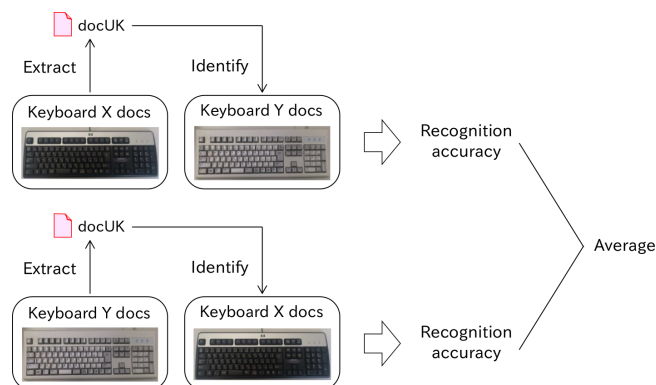

Figure 7. Evaluation procedure in the case where an extracted unknown document (*docUK*) typed with keyboard X is identified with profiled documents typed with the other keyboard Y, and *vice versa* (inter-keyboard evaluation).
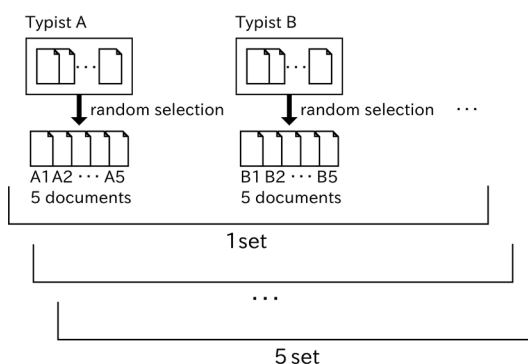

Figure 8. Procedure of preparing 5 document sets for evaluation.

Figure 9. Keyboard 1 [Hewlett-Packard BK-0316] (left) and keyboard 2 [NEC MT-109PSXP] (right) used in Experiment I.

the other is NEC MT109PSXP. We call these keyboards keyboard 1 and keyboard 2, respectively. Both keyboards belong to the standard desktop keyboards with the keystroke depth of 3.5 - 4.0 mm and the key pitch of 19 mm. The subjects of this experiment were 35 participants who could type 500 or more letters per 5 minutes, and we call this group G500. **Figure 10** shows the histogram of percentage of documents on the number of input letters, which indicates the variety of subject's typing skills in G500. The average of input letters per 5 minutes is 866 ± 267 with keyboard 1 and 969 ± 280 with keyboard 2.

The recognition accuracy for G500 is shown in **Figure 11**. The method using the WED + AD, which is expected to provide the highest recognition accuracy, provides about 96% in the intra-keyboard evaluation and about 86% in the inter-keyboard evaluation. It degrades severely because of changing keyboards and is not adequate level even in the case of using the same keyboard. This seemed to be caused by the existence of low typing skill subjects in G500, and then we left 23 subjects who could type 700 or more letters per 5 minutes with both keyboards (G700) and re-examined. **Figure 12** shows the histogram in G700 as well as **Figure 10**. The average of input letters per 5 minutes becomes 995 ± 237 with keyboard 1 and 1090 ± 238 with keyboard 2, which increase by about 120 from those in G500.

Figure 13 shows the recognition accuracy for G700. Comparing the result with G500 in **Figure 11**, the recognition accuracy for G700 in the inter-keyboard evaluation increases only 1% - 2%. However, the accuracy for G700 in the inter-keyboard evaluation improves about 8%, and it is found that G500 is more affected by the keyboard difference than G700. Although the numbers of subjects in G500 and G700 are different, it is suggested that low typing skill users are easier to be influenced by the change of keyboards than high typing skill ones. Thus, we can expect that the higher the typing skill is getting, the lesser the influence by the keyboard difference becomes.
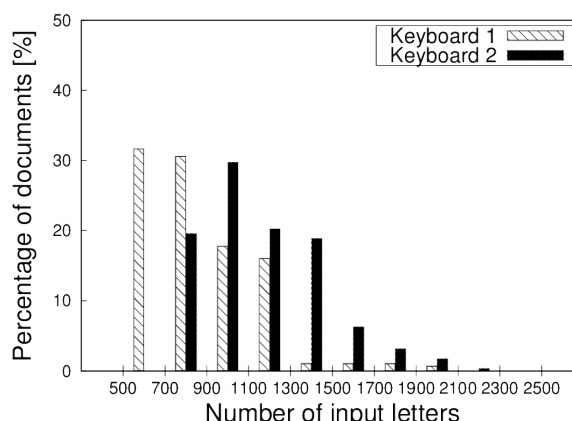

Figure 10. Histogram of percentage of documents on the number of input letters for 35 subjects (G500) in Experiment I.
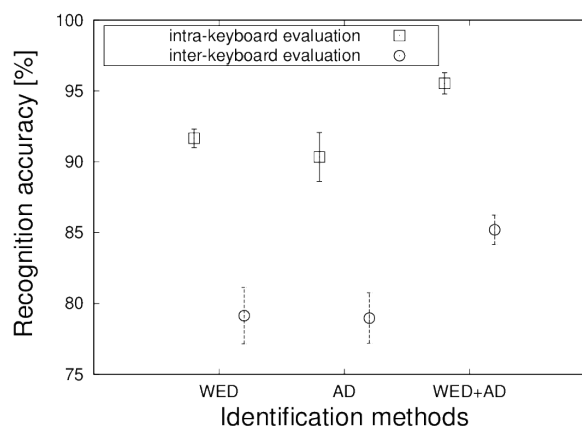

Figure 11. Recognition accuracy in intra-keyboard and inter-keyboard evaluations for G500 in Experiment I, obtained by three identification methods.
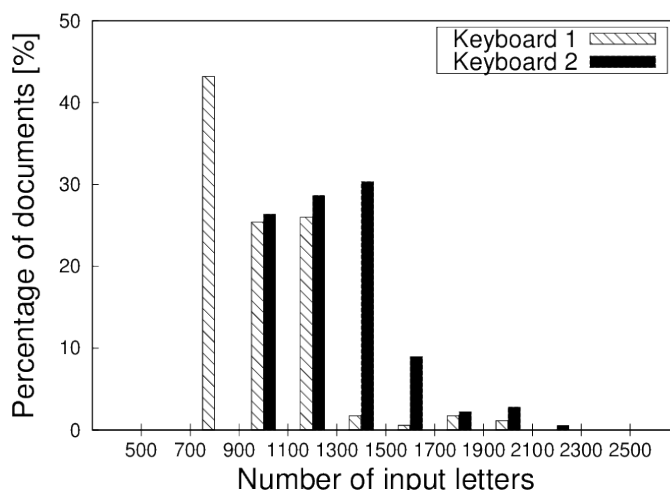
Figure 12. Histogram of percentage of documents on the number of input letters for 23 subjects (G700) in Experiment I.
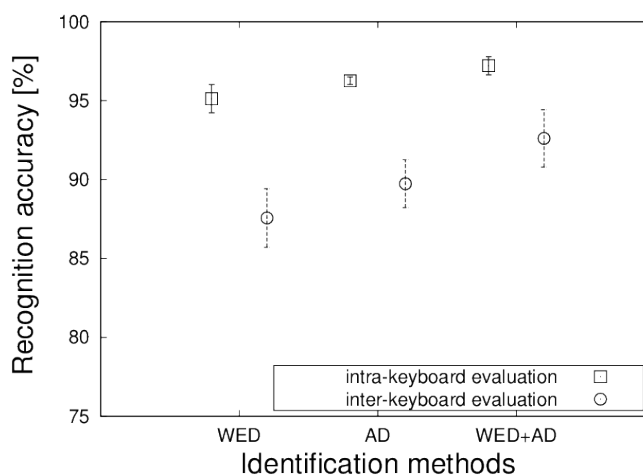

Figure 13. Recognition accuracy in intra-keyboard and inter-keyboard evaluations for G700 in Experiment I.

5.2. Experiment Using Desktop Keyboards without Low Typing Skill Subjects (Experiment II)

This experiment was performed on 27 subjects, and they were much more skilled-typists than those in Experiment I. In this experiment, we used two desktop keyboards as shown in **Figure 14**. One is BTC 5313 classic keyboard of which keystroke depth and key pitch are 3.9 ± 0.2 mm and 19 mm respectively. We call it keyboard 3. The other is Emprex 5211AU keyboard (USB keyboard) of which keystroke depth and key pitch are 3.5 ± 0.5 mm and 19 mm respectively. We call it keyboard 4. Both keyboards belong to the standard desktop keyboards with the keystroke depth and the key pitch are 3.5 - 4.0 mm and 19 mm. **Figure 15** shows the histogram of percentage of documents on the number of input letters, which indicates the variety of subject's typing skills in this experiment. The average of input letters per 5 minutes is 1371 ± 313 with keyboard 3 and 1411 ± 301 with keyboard 4. These amounts have no significant difference each other and are about 350 more than those of G700 in Experiment I.

The recognition accuracy in this experiment is shown in **Figure 16**. High recognition accuracy around 99% is obtained in both cases of intra-keyboard and inter-keyboard evaluations with the WED + AD method. Any influence of the keyboard difference is not seen in the result. It is confirmed that high typing skill users in such a case are not almost influenced by the change of keyboards.

Figure 14. Keyboard 3 [BTC 5313 classic keyboard] (left) and keyboard 4 [Emprex 5211AU keyboard] (right) used in Experiment II.
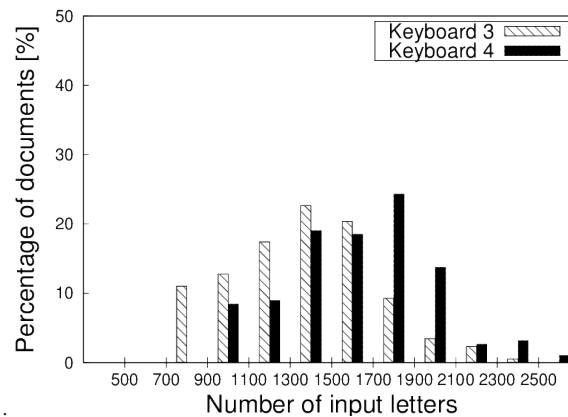

Figure 15. Histogram of percentage of documents on the number of input letters for 27 subjects in Experiment II.
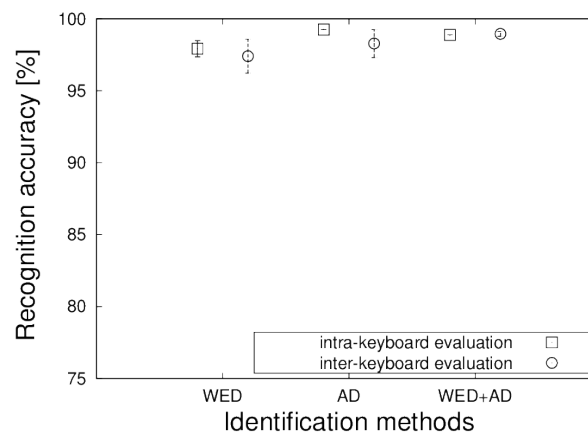

Figure 16. Recognition accuracy in intra-keyboard and inter-keyboard evaluations in Experiment II.

5.3. Experiment Using Desktop and Laptop Keyboards without Low Typing Skill Subjects (Experiment III)

This experiment was performed on 21 subjects, and they were also much more skilled-typists than those in Experiment I. **Figure 17** shows the desktop and laptop keyboards used in this experiment. As a desktop keyboard, the keyboard 4 adopted in Experiment II was used again. As a laptop keyboard, Hewlett-Packard ProBook 4710s was used, of which keystroke depth and the key pitch are 3.5 ± 0.5 mm and 19 mm respectively. We call it keyboard 5. **Figure 18** shows the histogram of percentage of documents on the number of input letters, which indicates the variety of subject's typing skills in this experiment. The average of input letters per 5 minutes is 1464 ± 246 with keyboard 4 and 1427 ± 297 with keyboard 5. There is no significant difference between these amounts as well as in Experiment II, and they are about 400 more than those of G700 in Experiment I.

The recognition accuracy in this experiment is given in **Figure 19**. The accuracy with the WED + AD method is around 99% and does not suffer the influence of keyboard difference in spite of using a laptop keyboard. Although feels of typing are different between desktop and laptop keyboards, it seems not to be a factor affecting the performance.

Figure 17. Keyboard 4 [Emprex 5211AU keyboard] (left) and keyboard 5 [Hewlett-Packard ProBook 4710s] (right) used in Experiment III.
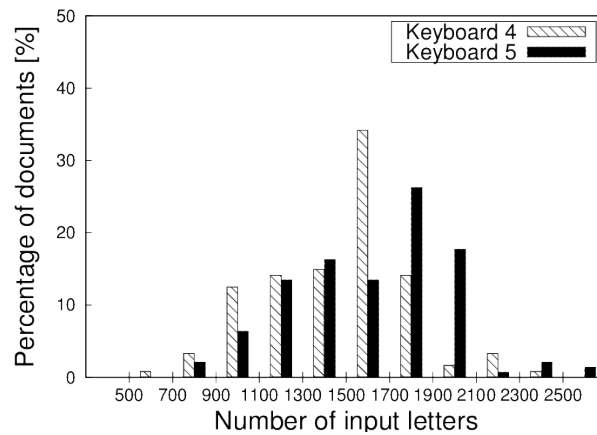

Figure 18. Histogram of percentage of documents on the number of input letters for 21 subjects in Experiment III.
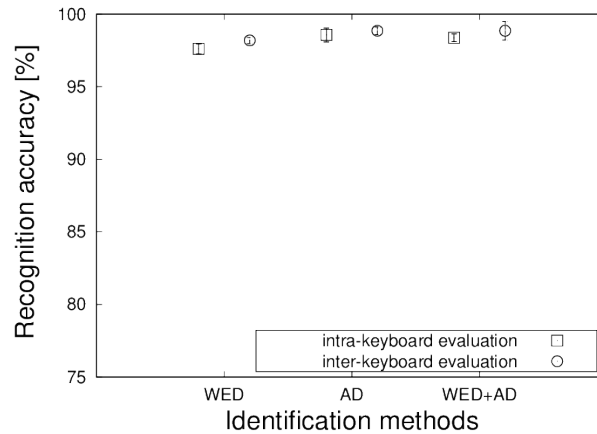

Figure 19. Recognition accuracy in intra-keyboard and inter-keyboard evaluations in Experiment III.

6. Conclusion

In this paper, we have investigated keyboard dependency of the personal identification by keystroke dynamics in Japanese free text typing. In Experiment I, it was confirmed that low typing skill users were easy to be influenced by changing keyboards, and then the recognition accuracy considerably decreased. Through the cases in Experiments II and III for high typing skill users, however, it was found that three identification methods kept high recognition accuracy against the keyboard difference. Especially, even in the situation using desktop and laptop keyboards, the recognition accuracy with the WED + AD method did not become worse. In practice, considering the results in Experiments I, II and III, we do not need to worry about the keyboard difference for users whose typing skills reach high level with about 900 or more letters in 5 minutes, and only for the remaining us-

ers it would be necessary to register their profile data with respect to keyboard they use in order to avoid recognition accuracy degradation. Finally, similar tendencies of the keyboard dependency would also be obtained in the cases of other languages, although we have focused only on Japanese text typing.

Acknowledgements

References

[1] Banerjee, S. and Woodard, D. (2012) Biometric Authentication and Identification Using Keystroke Dynamics: A Survey. *Journal of Pattern Recognition Research*, **7**, 116-139. http://dx.doi.org/10.13176/11.427

[2] Killourhy, K.S. (2012) A Scientific Understanding of Keystroke Dynamics. No. CMU-CS-12-100. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh.

[3] Alsultan, A. and Warwick, K. (2013) Keystroke Dynamics Authentication: A Survey of Free-Text Methods. *International Journal of Computer Science Issues*, **10**, 1-10.

[4] The, P.S., Teoh, A.B.J. and Yue, S. (2013) A Survey of Keystroke Dynamics Biometrics. *The Scientific World Journal*, **2013**, Article ID: 408280.

[5] Umphress, D. and Williams, G. (1985) Identity Verification through Keyboard Characteristics. *International Journal of Man-Machine Studies*, **23**, 263-273. http://dx.doi.org/10.1016/S0020-7373(85)80036-5

[6] Joyce, R. and Gupta, G. (1990) Identity Authentication Based on Keystroke Latencies. *Communications of the ACM*, **33**, 168-176. http://dx.doi.org/10.1145/75577.75582

[7] Monrose, F. and Rubin, A. (2000) Keystroke Dynamics as a Biometric for Authentication. *Future Generation Computer Systems*, **16**, 351-359. http://dx.doi.org/10.1016/S0167-739X(99)00059-X

[8] Bergadano, F., Gunetti, D. and Picardi, C. (2002) User Authentication through Keystroke Dynamics. *ACM Transactions of Information and Systems Security*, **5**, 367-397. http://dx.doi.org/10.1145/581271.581272

[9] Gunetti, D. and Picardi, C. (2005) Keystroke Analysis of Free Text. *ACM Transactions of Information and System Security*, **8**, 312-347. http://dx.doi.org/10.1145/1085126.1085129

[10] Tappert, C.C., Villani, M. and Cha, S. (2009) Keystroke Biometric Identification and Authentication on Long-Text Input. *Behavioral Biometrics for Human Identification*: *Intelligent Applications*, IGI Global, 342-367.

[11] Samura, T. and Nishimura, H. (2009) Personal Identification by Keystroke Dynamics in Japanese Free Text Typing. *Transactions of the Institute of Systems*, *Control and Information Engineers*, **22**, 145-153. http://dx.doi.org/10.5687/iscie.22.145

[12] Samura, T. and Nishimura, H. (2009) Keystroke Timing Analysis for Individual Identification in Japanese Free Text Typing. *Proceedings of ICROS-SICE International Joint Conference* 2009, 3166-3170.

[13] Samura, T. and Nishimura, H. (2010) A Hybrid Model for Individual Identification Based on Keystroke Data in Japanese Free Text Typing. *Transactions of the Society of Instrument and Control Engineers*, **46**, 676-684. http://dx.doi.org/10.9746/sicetr.46.676

[14] Samura, T. and Nishimura, H. (2011) Keystroke Dynamics for Individual Identification in Japanese Free Text Typing. *SICE Journal of Control*, *Measurement*, *and System Integration*, **4**, 172-176. http://dx.doi.org/10.9746/jcmsi.4.172

[15] Samura, T. and Nishimura, H. (2011) Personal Identification and Authentication Based on Keystroke Dynamics in Japanese Long-Text. *Continuous Authentication based on Biometrics*, *Data*, *Models*, *and Metrics*, IGI Global, 212-231.

[16] Samura, T. and Nishimura, H. (2011) Keystroke Timing Analysis for Personal Authentication in Japanese Long Text Input. *Proceedings of the SICE Annual Conference* 2011, 2121-2126.

[17] Matsubara, Y., Samura, T. and Nishimura, H. (2012) Robustness of Keystroke Dynamics against Changing in Input and Recognition Conditions for Japanese Atypical Text. *Proceedings of* 2012 *International Symposium on Technology for Sustainability*, Bangkok, November 2012, 103-106.

[18] Samura, T., Matsubara, Y. and Nishimura, H. (2013) Performance Assessment in Keystroke Dynamics by Combined Profile Documents for Free Text Typing. *Proceedings of SICE Annual Conference* 2013, 265-270.

[19] Matsubara, Y., Samura, T. and Nishimura, H. (2015) Profile Generation Methods for Reinforcing Robustness of Keystroke Authentication in Free Text Typing. *Journal of Information Security*, **6**, 131-141. http://dx.doi.org/10.4236/jis.2015.62014

[20] Samura, T. and Nishimura, H. (2011) Influence of Keyboard Difference on Personal Identification by Keystroke Dy-

namics in Japanese Free Text Typing. *Proceedings of 6th International Conference on Emerging Trends in Engineering & Technology* (*ICETET*-12), 30-35.

[21] Samura, T. (2008) A Web-Based Typing Learning System That Supports Japanese Text. *Journal of Education in the College of Technology*, **31**, 471-476.