Keystroke Timing Analysis for Personal Authentication in Japanese Long Text Input

Toshiharu Samura[1] and Haruhiko Nishimura[2]

[1]Department of Electrical and Computer Engineering, Akashi National College of Technology, Hyogo, Japan
(Tel : +81-78-946-6079; E-mail: samura@akashi.ac.jp)
[2]Graduate School of Applied Informatics, University of Hyogo, Hyogo, Japan
(Tel : +81-78-303-1934; E-mail: haru@ai.u-hyogo.ac.jp)

Abstract: We have investigated several characteristics of keystroke dynamics in Japanese long-text input. We performed experiments with 189 participants, classified into three groups according to the number of letters they could type in five minutes. In this experimental study, we extracted feature indices from the keystroke timing for each alphabet letter and for each two-letter combination composed of a consonant and vowel in Japanese text. Taking into account two identification methods using Weighted Euclidean Distance (WED) and Array Disorder (AD), we proposed WED+AD method for authentication on the basis of keystroke data in Japanese long-text input. By evaluating the authentication performance of individuals in the three groups, the effectiveness of the method was found to correspond to the typing skill level of the group.

Keywords: Computer Security, Human Interface, Keystroke Dynamics, Japanese Long-Text, Biometric Measures

1. INTRODUCTION

Timing data for keystrokes follows a fixed pattern, and biometric measures that use such data are called keystroke dynamics. Keystroke dynamics has two features that differentiate it from other forms of biometric measures. First, keystroke dynamics can be measured using only a keyboard; special equipment, such as fingerprint and retinal scanners, is not required. Second, this biometric measuring system has applications other than access authorization, such as investigation into the identity of malicious users who attempt to gain unauthorized access to computer systems.

Most previous research on keystroke dynamics has focused on user authentication during log-in, using not only information about a series of input characters for password recognition, but also keystroke dynamics as part of the authentication process [1, 2]. On the other hand, we considered the use of an analytic method that captures individual characteristics through the input of completely different phrases, rather than using repeated input of a short word for password verification. By using sentences of a certain length, it is possible to obtain sufficient information for deriving dynamics statistically. Little research has been performed on the keystroke dynamics of such long-text input, and this has only recently become the subject of academic discussion[3-8].

We have proposed a method for feature index extraction and identification that enables identification of individuals through long-text input as a fundamental topic in keystroke dynamics research. Here, we use keystroke timing for single character and paired character sequences when the user is inputting Latin characters. In this paper, we further consider the personal authentication case where test documents contain a typing sample from a person completely unknown to the system. For identification methods, we use our previously proposed Weighted Euclidean Distance (WED) method[5, 6, 8], the Array Disorder (AD) method proposed by Gunetti et al.[3], and a WED+AD method that add WED distance to AD distance. By performing a large-scale experiment involving 189 participants, we are able to obtain new results that compare authentication performance between groups of users with differing typing skill levels.

```
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
```

Fig. 1 Illustration of keystroke data

2. KEYSTROKE DATA AND EXTRACTION OF FEATURE INDICES

Consider a situation in which a user inputs text of a given length. During keyboard entry, the system performs background measurements of key press and release times. Fig. 1 shows example data on Japanese hiragana input by entering combinations of Latin letters. The first field shows the typed letter, the second field shows the key press (p) or release (r), and the third field shows the UNIX system time of the event (millisecond precision). Measurements can be performed to acquire such keystroke data by using keystrokes for either single alphabet letters or letter pairs. While a relatively large amount of previous research has focused on letter

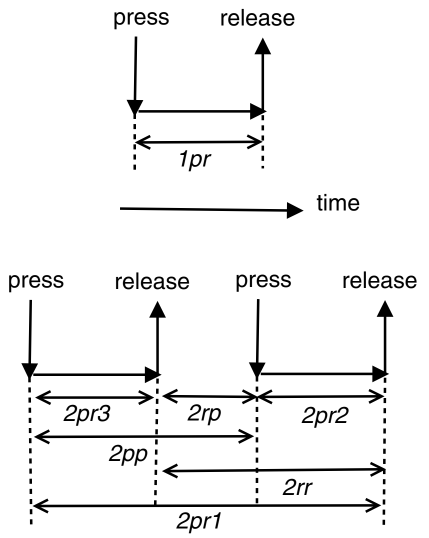

Fig. 2 Keystroke measurements for a single letter (above) and letter pair(below)

pairs[3], there has been relatively little research examining both single letter and letter pairs [4-8]. When typing in English, certain combinations of three letters or more frequently occur, for example, the and tion; however, because Japanese typing involves the input of hiragana characters (formed from either a single vowel letter or a consonant-vowel letter pair), particular combinations of three or more letters occur infrequently. This study therefore does not address feature indices for combinations of three or more letters. The notation 1pr in Fig. 2 indicates the time from press to release of a single key, and is referred to below as key press duration. The notation $2rr$ in Fig. 2 indicates 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 ($2rp$) and the time from pressing the first key to pressing the second key ($2pp$) are also considered. Furthermore, $2pr1$ indicates the time from pressing the first key to releasing the next key, $2pr2$ indicates the key press duration when typing the second (vowel) key, and $2pr3$ indicates the key press duration when typing the first (consonant) key. As shown in Table 1, the average and standard deviation of each of the seven measures described above are used as the feature indices for identification of individuals. These measurements are, however, not independent of each other; for example, $2pp = 2pr3 + 2rp$. Therefore, $2pp$ might seem to be unnecessary for evaluation, but the evaluation of $2pp$ corresponds to the evaluation of the correlation between $2pr3$ and $2rp$ because in general $(2pp.sd)^2 = (2pr3.sd)^2 + (2rp.sd)^2 - 2((2pr3 \cdot 2pr).av - 2pr3.av \cdot 2rp.av)$. Furthermore, the use of multiple feature indices ensures highly robust data processing.

Taking these feature indices as x, standardization is performed according to the following equation:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (0 \leq x' \leq 1) \qquad (1)$$

Here, x_{min} and x_{max} respectively refer to the minimum and maximum values obtained from the feature indices of all subjects.

Table 1 Feature indices of keystroke

notation	explanation
1pr	Average (*1pr.av*) and standard deviation (*1pr.sd*) of key press durations for alphabet single letters of `a` to `z`
2rr	Average ($2rr.av$) and standard deviation ($2rr.sd$)of release to release transition times between consonant-vowel letter pairs.
2rp	Average ($2rp.av$) and standard deviation ($2rp.sd$)of release to press transition times between consonant-vowel letter pairs.
2pp	Average ($2pp.av$) and standard deviation ($2pp.sd$) of press to press transition times between consonant-vowel letter pairs.
2pr1	Average ($2pr1.av$) and standard deviation ($2pr1.sd$)of press to release transition times between consonant-vowel letter pairs.
2pr2	Average ($2pr2.av$) and standard deviation ($2pr2.sd$)of key press durations for the second letter of consonant-vowel letter pairs.
2pr3	Average ($2pr3.av$) and standard deviation ($2pr3.sd$)of key press durations for the first letter of consonant-vowel letter pairs.

3. METHOD OF IDENTIFICATION

Profiles are created through extraction of feature indices for each typist (A, B, C, $\cdots$) using multiple (N) documents. The profile collection is then used to determine the typist who input an unknown document.

In this section, we describe the WED method, the AD method and WED+AD method for comparison of the feature indices for the unknown document and subsequent comparison with each profile.

3.1 Weighted Euclidean Distance (WED) method

Taking the first profiling document of Typist A as doc$A1$, the profiling document of each participant can be represented as doc$A1$, doc$A2$,$\cdots$,docAN, doc$B1$, doc$B2$,$\cdots$, docBN, doc$C1$, $\cdots$. An unknown document is represented as docUK. The squared WED $\mathrm{WED}^2(\mathrm{doc}A1, \mathrm{docUK})$ used as the identification func-

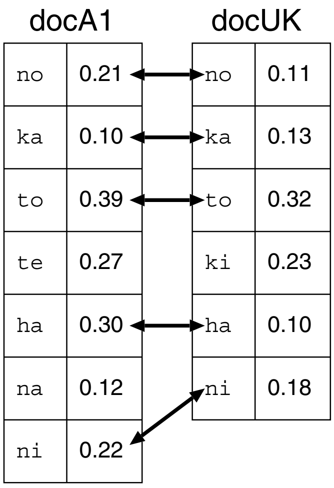

Fig. 3 Illustration of the Weighted Euclidean Distance (WED) method for remaining letter pairs and their standardized values of $2pp.av$ in $docA1$ and $docUK$, where there are five letter pairs in common between them.

tion is given by the following equation:

$$\mathrm{WED}^2(\mathrm{doc}A1, \mathrm{docUK}) = \frac{1}{m}\sum_{\alpha=1}^{m}\left(\frac{1}{n_\alpha}\right)\sum_{i=1}^{n_\alpha}\left(k_{\alpha(i)} - r_{\alpha(i)}\right)^2 \quad (2)$$

The index of the feature indices $\alpha(= 1, 2, \cdots)$ is then $1pr.av$, $1pr.sd$, $\cdots$. 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 eq.(1) for a profiling document (e.g., doc$A1$), and $r_{\alpha(i)}$ is that for $\alpha(i)$ of an unknown document (docUK). The WED is normalized to $0 \sim 1$ using the weightings $\frac{1}{m}$ and $\frac{1}{n_\alpha}$.

During actual feature index extraction, a lower threshold N_{TH} for the number of occurrences of each character type (single letter and letter pair) and an upper threshold N_{TH} for key press duration other than $2pr1$ are introduced. As a result, when comparing $k_{\alpha(i)}$ and $r_{\alpha(i)}$, there can be loss of k-r pair feature indices for a given input document in accordance with the setting of the lower threshold N_{TH} for that document. In cases where the keystroke feature indices for all k-r pairs in a given document are missing, then the document is removed from consideration for the participant.

Fig. 3 illustrates the $2pp.av$ feature indices for two documents, namely, $docA1$ and $docUK$. For the five $2pp.av$ letter pairs common to the two documents, the squared WED is given by the following equation:

$$\begin{aligned} WED^2_{2pp.av}(docA1, docUK) = [&(0.21-0.11)^2 \\ &+(0.10-0.13)^2+(0.39-0.32)^2 \\ &+(0.30-0.10)^2+(0.22-0.18)^2]/5 \end{aligned} \quad (3)$$

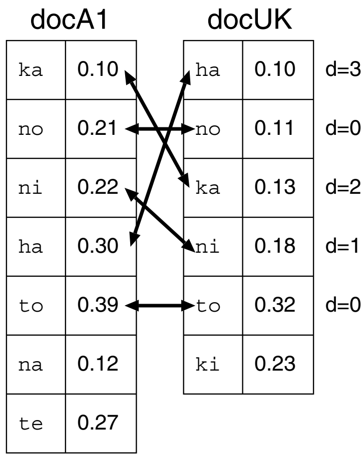

Fig. 4 Illustration of Array Disorder (AD) method for the example in Fig. 3.

3.2 Array Disorder (AD) method

The AD method, which is called the R-measure in Gunetti and Picardi[3], 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 are compared, and the totals of each are taken as the distance.

When n_α characters are used to compare a feature index α, if n_α is even then the distance is divided by $n_\alpha^2/2$; if n_α is odd then the distance is divided by$(n_\alpha^2 - 1)/2$. Finally, the value is normalized to the range $0 \sim 1$ by dividing the value by the number of contributing feature indices m.

Fig. 4 shows the feature index values from Fig. 3 rearranged in increasing order. The sum of the difference in rankings of each character (the d value in the figure) is normalized, resulting in the following calculation of the AD value:

$$AD_{2pp.av}(docA1, docUK) = \frac{1}{(5^2-1)^2}[3+0+2+0] \quad (4)$$

3.3 WED+AD method

In contrast to the WED method, which evaluates the magnitude of differences in feature index values between documents, the AD method focuses on differences between documents in ranking patterns of the feature indices. This study introduces WED+AD method that complementarity incorporate the features of the WED and AD methods. In this method, neither WED nor AD dominates because they are normalized to the range $0 \sim 1$.

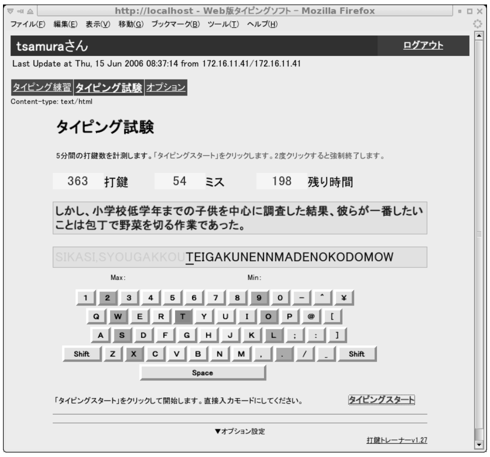

Fig. 5 Screenshot of interface of keystroke data collecting system.

4. EXPERIMENTAL METHODOLOGY

This section describes the keystroke data collection system [9] and the classification of participants into groups. In this study, we used a web-based system that is able to collect keystroke data from a large number of participants in a single experiment. The system uses typing support software that was familiar to the participants, thereby lowering effects related to unfamiliarity and nervousness. Fig. 5 shows a screenshot of the software interface used in this study.

The document display screen allows participants who are skilled typists to input text while viewing the Japanese text displayed in the upper row. Less skilled typists can type while viewing the Latin alphabet text displayed in the middle row. Latin alphabet text is removed from the screen as it is typed, allowing confirmation of mistyped characters. The top row displays the number of keystrokes, the number of errors, and the amount of time remaining. Because this experiment focuses only on Latin alphabet input keystroke, by design Latin character input is not converted into Japanese kanji characters. While participants are typing, browser-embedded JavaScript code records character input, key press times, and key release times. Times are recorded using UNIX times (millisecond precision). Recorded data were sent to a server using Ajax. Participants input data during 5-min sessions separated by at least one week. Lower thresholds for single letters and letter pairs, N_{TH}, were set at 3 occurrences ($N_{TH} = 3$), and key press duration T for keystrokes other than $2pr1$ were set at 300 ms or less ($T < T_{TH} = 300$).

Analysis was performed after dividing participants into three groups according to the number of letters that they could type in 5 min: G500 for those who could type at least 500 letters, G700 for those who could type at least 700 letters, and G900 for those who could type at least 900 letters. Table 2 shows the number of participants in each category.

The number of documents obtained from each participant varied from 5 to 18. As shown in Fig. 6, the number of documents from each user was set to 5 by choosing 5 documents at random from those created by participants who typed an excess number of documents. Analysis was performed taking, as a single set, the number of documents equal to 5 times the number of participants. To account for possible bias in document selection, analysis was performed five times on 5 similar document sets.

Table 2 Group classification of subjects.

Group	G500	G700	G900
Number of input letters (5 min.)	≥ 500	≥ 700	≥ 900
Number of subjects	189	127	52
Number of all documents (1 set)	945	635	440

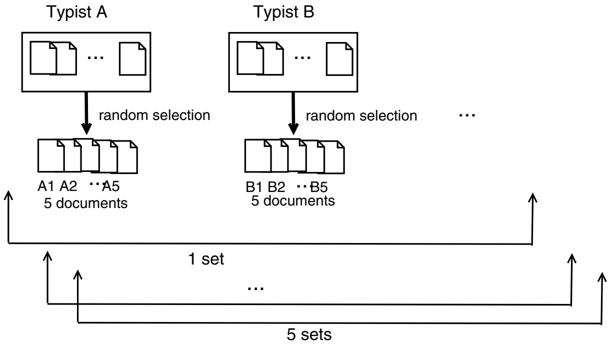

Fig. 6 Procedure for preparing document sets.

We next describe the basic features (statistics) of the documents examined. The documents were divided into three groups according to the number of letters they contained: 500 to 699 letters, 700 to 899 letters, and 900 or more letters. Fig. 7 and 8, respectively, show the average number of occurrences of single letters (vowel or consonant) and letter pair combinations (consonant plus vowel) for the most common twenty occurrences, ranked in decreasing order of the occurrences in those documents containing more than 900 letters.

The number of letter pair occurrences is much lower than that of the single letter occurrences, and the probability of a missing pair increases with lower ranking. As can be seen from the graphs, for those typists with input rates of 500-699 letters there will be many letter pairs for which the occurrence count is below the lower threshold $N_{TH} = 3$.

5. ANALYSIS RESULTS

For the personal authentication case, we use the method of Gunetti and Picardi[3] to evaluate user authentication. Taking the profiling documents of Typist A as doc$A1$, doc$A2$, $\cdots$, doc$A5$, in eq.(5), we express the median distance (md) from the identification function for an unknown document docUK in terms of the distance of each profile document d as follows:

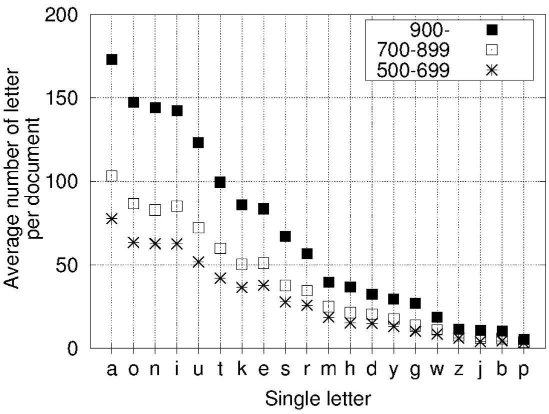

Fig. 7 Frequency of each single letter (vowel or consonant) per document typed in 5 min, averaged for all input documents for input numbers of 500–699, 700–899, and 900 or higher.

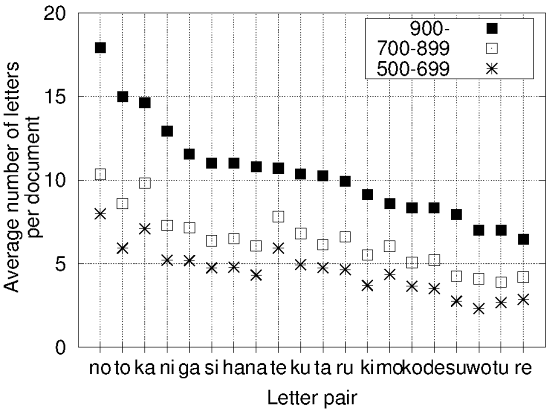

Fig. 8 Frequency of each letter pair (consonant plus vowel) per document typed in 5 min, averaged for all input documents for input numbers of 500–699, 700–899, and 900 or higher.

$$md(UK, A) = [d(docUK, docA1) + d(docUK, docA2) + \cdots + d(docUK, docA5)]/5 \quad (5)$$

he median distance m for all documents from typist A is given by eq.(6):

$$m(A) = [d(docA1, docA2) + d(docA1, docA3) + \cdots + d(docA4, docA5)]/{}_5\mathrm{C}_2 \quad (6)$$

When the following conditions are fulfilled, an unknown document docUK is taken to belong to A:

1. The median distance $md(UK, A)$ between Typist A and the unknown document docUK is shorter than the median distance between all other typists.
2. (2-i) $md(UK, A) < m(A)$
   or
   (2-ii) Taking Typist B to have the next shortest value of md after Typist A, the following condition is satisfied:
   $md(UK, A) - m(A) \leq md(UK, X) - md(UK, A)$

False rejection rates (FRR) and false acceptance rates (FAR) are used as an evaluation index of user authentication. In this study, as shown in Fig. 6, each set contained M participants (G500, G700, and G900 in Table 2), each having $N_{doc} = 5$ profiling documents. FRR evaluation is performed as follows. As in the leave-one-out cross validation method, a document is removed from the test documents of a participant, and is taken as the unknown document $docUK$. If the conditions 1 and 2 above match with the typist of the unknown document then the identification is considered a success; otherwise, it is considered a failure (a false rejection). This operation is repeated $M \times N_{doc}$ times, and FRR $=$ (the number of false rejections)$/(M \times N_{doc}) \times 100$.

FAR evaluation is performed as follows. An unknown document assumed to have been prepared by an external attacker is used as the unknown document $docUK$. If a document prepared by one of the test participants is used as the unknown document $docUK$, that participant will be treated as a non-registrant, so all documents prepared by the participant are removed. Next, the remaining documents and the unknown document are compared against $(M-1)$ participants to see if a participant meeting criteria 1 and 2 above is found. If the corresponding participant is found, then the authentication is considered a failure (false acceptance), and if no participants fulfilling the criteria are found then the document is classified as an external attack and identification is considered successful. This operation is repeated $M \times N_{doc}$ times, FRA $=$ (the number of false acceptance$/(M \times N_{doc}) \times (M-1)) \times 100$.

Considering the results of our papers[7], we excluded the combinations of the feature indices of standard deviations from letter pair keystrokes, and we examined combination of all remaining feature indices,$1pr.av + 1pr.sd + 2rr.av + 2rp.av + 2pp.av + 2pr1.av + 2pr2.av + 2pr3.av$.

The results of performing user authentication using the WED, AD, and WED+AD methods are shown in Table 3. Thus, the present authentication method shows FRR$\sim$ 3% and FAR$\sim 0.01\%$ when using WED+AD method for G700 and G900.

The performance on G500 in which 37% of all are beginner level subjects is lower than those on G700 and G900 which include normal and advanced level subjects. We guess this is due to the dispersive formation of the temporal patterns in beginners keying when looking at the screen and keyboard one after the other.

These accuracies are lower than those of physiological biometrics, e.g., fingerprints and iris, however, are relatively higher than those of behavioral biometrics, e.g., signature and voice.

6. CONCLUSION

We investigated a method for identification of individuals using keystroke data from long-text input. Specifically, we proposed 14 feature indices related to the typing of single letters and consecutive letter pairs using Latin

Table 3 Results of authentication for three groups when using the feature combination $1pr.av + 1pr.sd + 2rr.av + 2rp.av + 2pp.av + 2pr1.av + 2pr2.av + 2pr3.av$ for WED, AD and WED+AD.

Method	WED	AD	WED+AD
G900			
FRR(%)	5.82 ± 1.00	2.68 ± 0.41	2.14 ± 0.34
FAR(%)	0.021 ± 0.006	0.015 ± 0.004	0.016 ± 0.008
G700			
FRR(%)	7.18 ± 0.48	5.13 ± 0.31	3.06 ± 0.74
FAR(%)	0.024 ± 0.006	0.011 ± 0.005	0.011 ± 0.004
G500			
FRR(%)	10.37 ± 0.24	9.78 ± 0.85	7.98 ± 0.38
FAR(%)	0.032 ± 0.004	0.019 ± 0.004	0.017 ± 0.007

alphabet input, confirmed improvements in recognition related to the use of 8 of those feature indices, and constructed an associated evaluation index. For the authentication method, we introduced a WED+AD method that incorporates the AD method proposed by Gunetti et al. into the WED method.

By performing a large-scale experiment involving 189 participants, we were able to obtain new results that compare identification performance between groups of users with differing typing skill levels. We found that WED+AD method had recognition accuracy rates of FRR $\sim$ 3% and FAR $\sim$ 0.01 % for participants able to type at least 140 letters per minute (G700 and G900). While this study used Japanese text input in its investigation, the feature indices and authentication methods used can also be applied to other languages.

Finally, we would like to close with a discussion of possible applications of keystroke dynamics. In previous research, keystroke dynamics have generally been consider to be a supplementary security measure used in conjunction with password authentication during system logins to prevent unauthorized access. Additional incorporation of keystroke dynamics for long-text input can be used for continuous system monitoring for unauthorized users after login occurs. Specific examples include the introduction of an keystroke dynamics system to e-Learning materials and online tests to detect impersonation and unauthorized access [4, 10]. The inclusion of encoded keystroke data in e-mails could also be used in place of digital signatures.

We used WED+AD method as hybrid identification method, but we plan to examine other methods in future studies.Furthermore, there are many areas for further research related to factors that must be considered for practical use, such as using different keyboards, and the unfamiliarity and nervousness of users. We hope to address these topics in future studies.

REFERENCES

[1] D. Umphress and G. Williams, "Identity verification through keyboard characteristics", *International Journal of Man-Machine Studies*, Vol. 23, pp. 263–273,(1985).

[2] R. Joyce and G. Gupta, "User authorization based on keystroke latencies", *Communications of the ACM*, Vol. 33, No. 2, pp. 168–176 (1990).

[3] D. Gunetti and C. Picardi, "Keystroke Analysis of Free Text", *ACM Trans. Inf. Syst. Secur.*, Vol. 8, No. 3, pp. 312–347 (2005)

[4] C.C. Tappert, M. Villani, and S. Cha, "Keystroke biometric identification and authentication on long-text input", *In L. Wang and X. Geng (Eds.), Behavioral Biometrics for Human Identification: Intelligent Applications, Medical Information Science Reference*, pp. 342–367 (2009).

[5] T. Samura and H. Nishimura, "Personal Identification by Keystroke Dynamics in Japanese Free Text Typing", *Transactions of the Institute of Systems, Control and Information Engineers*, Vol. 22, No. 4, pp. 145–153 (2009) (in Japanese).

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

[7] T. Samura and H. Nishimura, "A Hybrid Model for Individual Identification Based on Keystroke Data in Japanese Free Text Typing", *Transactions of the Society of Instrument and Control Engineers*, Vol. 46, No. 11, pp. 676–684 (2010) (in Japanese).

[8] T. Samura and H. Nishimura, "Keystroke Dynamics for Individual Identification in Japanese Free Text Typing", *SICE Journal of Control, Measurement and System Integration*, Vol. 4, No. 2, pp. 172–178 (2011).

[9] T. Samura, "A Web-based Typing Learning System That Supports Japanese Text", *Journal of Education in the College of Technology*, Vol.31, pp. 471–476 (2008) (in Japanese).

[10] A. A. E. Ahmed and T. Traore, "Employee surveillance based on free text detection of keystroke dynamics", *In M. Gupta and R. Sharman (Ed.), Handbook of research on social and organizational liabilities in information security, Idea Group Publishing*, pp. 47–63 (2008).