ICROS-SICE International Joint Conference 2009
August 18-21, 2009, Fukuoka International Congress Center, Japan

Keystroke Timing Analysis for Individual Identification in Japanese Free Text Typing

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 : +81-78-367-8618; E-mail: haru@ai.u-hyogo.ac.jp)

Abstract: We developed Web-based typing software in order to collect keystroke data on a LAN, and investigated several characteristics of keystroke dynamics in Japanese free text typing. We performed experiments on 112 subjects, representing three groups according to the number of letters they could type in five minutes. In this experiment, we extracted the feature indices from the keystroke timing for each single letter and for two-letter combinations composed of consonant and vowel pairs in Japanese text. Based on an identification method using a weighted Euclid distance, personal identification for the three groups was evaluated and its high performance was confirmed in proportion to the typing level of the group.

Keywords: Keystroke Dynamics, Biometrics, Free Text, Personal Identification

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 make it distinct from other forms of biometric measures. First, keystroke dynamics can be measured using only a keyboard; special equipment such as that required for fingerprints and retinal scans is not required. Second, this biometric measuring system has uses other than access authorization, such as investigation into identifying individuals attempting illegal access into computer systems.

Most previous research on keystroke dynamics has been related to verification upon user login, using not only knowledge of a series of symbols for password recognition, but also keystroke dynamics as part of the authentication process [1-6]. Short words such as passwords, however, generally do not contain sufficient keystroke information for such dynamics to be comprehended, and so have a lower verification rate as compared with fingerprints or retinal scans. Keystroke dynamics is therefore most often used as a complementary check in password recognition.

Another use of keystroke dynamics, however, is for feature analysis related to human behavior, in a manner similar to handwriting analysis and voice printing. This study focuses on identification from the point of view of using keystroke dynamics related to human behavior. Here, we will consider the use of an analytic method which, rather than using repeated input of a short word for password verification, captures individual characteristics through the input of completely different phrases. By using sentences of a certain length, it is possible to statistically obtain the amount of information required for deriving dynamics. Little research has been performed on the keystroke dynamics of such atypical, free-text typing conditions, and this has only recently become the subject of academic discussion [7-13].

An example of such work is that of Curtin et al. [10], in which profiles of 30 subjects were created based on the inputting of at least 5 documents containing English sentences, and identification was performed by applying the nearest neighbor rule to Euclidean distances calculated for a variety of feature indices. We subsequently expanded on the method of Curtin et al., for example, by applying weightings to the Euclidean distance for each feature and proposing feature amounts related to features peculiar to Japanese [12, 13]. We have performed a detailed analysis on the extraction of feature indices by increasing the number of subjects, who formed three groups based on typing skills. Doing so resulted in a high level of recognition accuracy.

Our previous work [12, 13] concerned the comparison of performances between weighted Euclidean distances and neural networks, and so feature indices on the keystroke of letter pairs were limited to three types (”ka”, ”no”, and ”to”) without missing values. The present study looks only at weighted Euclidean values, and so we are able to examine more than three types of letter pairs, thus further increasing the recognition performance.

2. THE EXPERIMENT AND METHOD OF IDENTIFICATION

2.1 Experimental methodology

In this study, we developed a web-based experimental system that enables a large number of subjects to be studied in a single experiment[14]. As the user interface to collect keystroke data, we adopted a typing learning software that had been familiar to the subjects in their classes so that for them to input documents in a relaxed condition. Fig. 1 shows a snapshot from the interface. Subjects with high typing skills will go on entering the alphabet letters corresponding to the Japanese characters looking at the document displayed on the screen. Sub-

PR0002/09/0000-3166 ¥400 © 2009 SICE

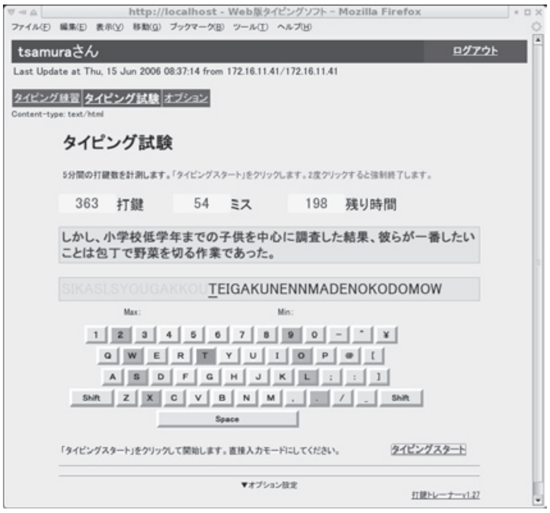

Fig. 1 Snapshot from the interface of keystroke data collecting system

Table 1 Group classification of subjects

Group	G500	G700	G900
Number of input letters (5 min.)	≥ 500	≥ 700	≥ 900
Number of subjects	112	71	52

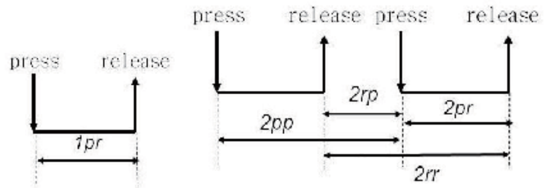

Fig. 2 Keystroke measurements of single letter (left) and letter pair (right)

Table 2 Feature indices of keystroke

notation	explanation
1pr	Average (*1pr.ave*) and standard deviation (*1pr.sd*) of key press durations for 14 single letters (vowels and consonants) of (a, i, u, e, o, k, s, t, n, h, m, y, r, w)
2pr	Average ($2pr.ave$) of key press durations for the second letter of consonant-vowel letter pairs.
2pp	Average ($2pp.ave$) of press to press transition times between consonant-vowel letter pairs.
2rr	Average ($2rr.ave$) of release to release transition times between consonant-vowel letter pairs.
2rp	Average ($2rp.ave$) of release to press transition times between consonant-vowel letter pairs.

jects with low typing skills could refer to the alphabet indication shown below the Japanese texts, if necessary. The number of entered letters, the number of typing errors and remaining time are also displayed.

For each entry, key's character, key pressed time, and key released time are recorded as raw data. Time is measured using UNIX time (millisecond(ms) precision). By using Ajax programming, the recorded data is transferred to the server. Five different documents were used as input documents, and subjects completed 5-minute input sessions separated by at least one week.

Subjects were divided into three groups according to the number of letters they could type in 5 minutes. In this paper, we perform analysis of the results of these three groups: G500 (500 letters or more typed), G700 (700 letters or more), and G900 (900 letters or more). Table 1 lists the number of subjects in each group.

2.2 Extraction of feature indices

In this paper, we refer to previous research by Curtin et al. [10], using feature indices related to single and consecutive double keystrokes for which certain results have been obtained. We used the five measurements shown in Fig. 2 as feature indices for identification of individuals based on Japanese text input, as explained in Table 2.

Table 3 shows the frequencies of 14 single letters and letter pairs (top 14 most frequent ones) per one document typed in 5 minutes, averaged for all input documents from G500 subjects. The frequency of letter pairs overall tends to be lower than that of single letter. In consideration of this result, we set the lower threshold for the appearance of each single letter and letter pairs (N_{TH}) to 3 ($N_{TH} = 3$), and the respective key press durations (T) to be within 300 ms ($T \leq T_{TH} = 300$ ms). Those that were not satisfied with this condition were excepted from the object of evaluation. The average and standard deviation values of the feature indices in Table 2 for all input documents from G500 subjects are shown in Table 4.

Taking raw measurement of a feature index as x, standardization is performed according to the following for-

Table 3 Frequency of each single letter (vowel and cons.) and letter pairs per one document typed in 5 minutes, averaged for all input documents from G500 subjects

a	o	n	i	u	t	e
71.6	60.7	60.3	59.7	47.2	39.5	36.2
k	s	r	m	h	y	w
34.7	27.0	23.4	18.1	14.6	13.0	7.7
no	ka	to	ha	si	ta	ni
8.5	8.2	5.9	5.8	5.7	5.6	5.4
ru	na	ku	ga	ki	te	ko
5.2	4.9	4.9	4.9	4.8	4.6	3.5

Table 4 Average and standard deviation values of the feature indices in Table 2 for all input documents from G500 subjects (ms)

$1pr$		$2pr$	
a	140.4 ± 25.5	no	128.9 ± 21.3
o	127.1 ± 30.1	ka	131.4 ± 19.0
n	107.7 ± 31.2	to	114.2 ± 20.1
i	109.9 ± 24.2	$2pp$	
u	110.0 ± 23.5	no	135.7 ± 29.2
t	124.9 ± 28.2	ka	135.9 ± 31.8
e	129.2 ± 25.3	to	140.8 ± 31.3
k	113.0 ± 28.7	$2rr$	
s	133.7 ± 27.3	no	130.9 ± 30.3
r	133.0 ± 26.6	ka	155.0 ± 34.2
m	121.4 ± 27.9	to	131.9 ± 35.8
h	119.0 ± 24.3	$2rp$	
y	115.2 ± 24.8	no	54.9 ± 31.8
w	139.3 ± 26.6	ka	47.7 ± 33.0
		to	47.5 ± 34.4

mula:

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

Here, x_{min} and x_{max} are the minimum and maximum values obtained from the measurements of feature index x over all subjects.

2.3 Method of identification

The aggregate of feature index patterns (feature vector) after feature extraction is called a profile, and is referred to as data for evaluating and guessing the unknown person inputting a test sample document. A weighted Euclidean distance (WED) is used as a discriminate function for comparison of the reference profiles (known feature vectors) and unknown feature vectors. For example, the squared WED between doc$A1$ (the reference profile of subject A on document 1) and docUK (the feature vector of unknown person on a test sample document), $N^2(\text{doc}A1, \text{docUK})$ is given by

$$N^2(\text{doc}A1, \text{docUK}) \;=\; \sum_{\alpha} \omega_\alpha \left(k_\alpha - r_\alpha\right)^2 \tag{2}$$

Here, α indicates the type of the feature index as shown in Table 2, k_α is the feature quantity in document doc$A1$, and r_α is the feature quantity in document docUK. The weighting ω_α is the reciprocal of the number of letters or letter pairs in the feature index of α.

From the result of Table 3 on the average of all input documents, it is easily expected that considerable number of letter pairs appear less than three times in each document. In this paper, we except such letter pairs from the evaluation of the squared WED as missing data in a profiling document or a sample document. Fig. 3 is an illustration when there exist five letter pairs common to doc$A1$ and docUK. In this case, part of the squared WED for $2pp.ave$ is calculated as follows:

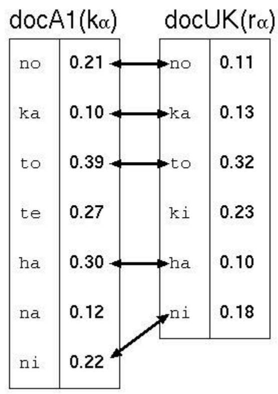

Fig. 3 Illustration of remaining letter pairs and their standardized values of $2pp.ave$ in doc$A1$ and docUK, where there exist five letter pairs common to them. (α=$2pp.ave$)

$$\begin{aligned} N^2_{2pp.ave}(\text{doc}A1, \text{docUK}) = \frac{1}{5}\big[&(0.21-0.11)^2 \\ &+(0.10-0.13)^2 + (0.39-0.32)^2 \\ &+(0.30-0.10)^2 + (0.22-0.18)^2\big] \end{aligned} \tag{3}$$

For identification, we used the nearest neighbor classification rule with the squared WED, and the person whose reference profile had the smallest WED to the test sample docUK was identified as the person with docUK. In order to evaluate recognition accuracy, the leave-one-out cross-validation method was used in the experiment.

In the case that m subjects each key in 5 different documents, the squared WED is computed for $5m$($5m$-1)/2 combinations between each entry of document (as the test sample) and every other entry (as the reference profile). When one of entries by subject A, doc$A1$ is tested, A1 could be attributed to subject A successfully if the WED between doc$A1$ and doc$A*$, (*=2∼5) contains the smallest one. Recognition accuracy is given by the ratio of the succeeded tests to all ones.

3. RESULTS OF THE ANALYSIS

Table 5 shows the recognition accuracies for each feature index by group . One can see that the values for $1pr.ave$, $2pp.ave$, and $2rr.ave$, which are typical feature indices, are larger than those of the other indices. The values for letter pairs, as a whole, were vastly improved (they became three times or more) compared to the accuracy results in our previous work [12, 13]. This is because that the number of the letter pairs contributing to feature indices much increased against the previous case limited to three letter pairs of "ka", "no", and "to".

The recognition accuracies contributed by a combination of $1pr.ave$ and another index are shown in Table 6. Here, "+" means addition of a feature type. The accuracy result on every combined case is higher than that on $1pr.ave$ only, and the combination of $1pr.ave + 2rr.ave$ gives the highest value among them.

Table 5 Recognition accuracies of three groups contributed by individual feature index

Group	G500	G700	G900
$1pr.ave$	67.5%	79.4%	81.2%
$1pr.sd$	31.3	44.5	53.9
$2pr.ave$	20.7	71.0	75.8
$2pp.ave$	38.6	74.6	80.4
$2rr.ave$	52.3	85.1	91.5
$2rp.ave$	18.2	60.0	71.9

Table 6 Recognition accuracies contributed by combination of 1pr.ave and another index.

Group	G500	G700	G900
$1pr.ave$	67.5%	79.4%	81.2%
$1pr.ave + 1pr.sd$	67.5	80.6	88.8
$1pr.ave + 2pr.ave$	68.4	90.1	92.3
$1pr.ave + 2pp.ave$	83.4	97.2	98.5
$1pr.ave + 2rr.ave$	88.5	98.9	99.2
$1pr.ave + 2rp.ave$	80.0	94.1	96.2

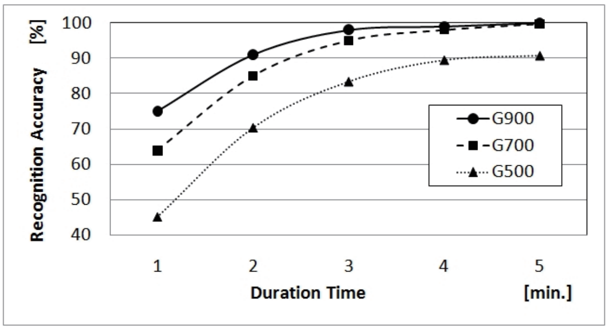

Fig. 4 Dependence of recognition accuracy on input duration of document (1 min. ∼ 5 min.)

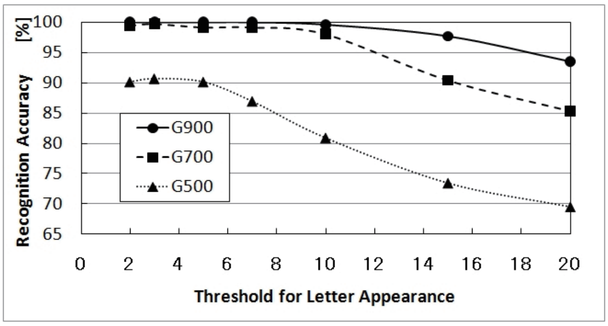

Fig. 5 Dependence of recognition accuracy on threshold for letter appearance ($N_{TH} = 2 \sim 20$)

Furthermore, from the evaluation of multiple combinations of feature indices, we obtain the result shown in Table 7, which indicates the optimal combination (OC): $1pr.ave + 1pr.sd + 2pr.ave + 2pp.ave + 2rr.ave$ that results in the highest recognition accuracy within all groups. The accuracy results on G700 and G900 attain the higher values than 94.7% in the case of English texts by Curtin et. al.[10].

As one can see from Table 5, 6, 7, accuracy results on G500 in which 37% of all are beginner level subjects are fairly 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.

In order to check the adequacy of the input duration of document (5 min.), the threshold for letter appearance ($N_{TH} = 3$) and the threshold for key press duration ($T_{TH} = 300$ ms), we analyzed the dependence of recognition accuracy on these parameters. First, Fig. 4 shows the accuracy dependence on the input duration of document for each group. We found that the choice of 5 minutes for input duration was in a position where the curve just levels off. Based on this finding, we could judge 5 minutes as the necessary and sufficient value in our experiment setting.

Table 7 Recognition accuracies of three groups for optimal combination (OC): 1pr.ave+1pr.sd + 2pr.ave + 2pp.ave +2rr.ave

Group	G500 560 samples	G700 355 samples	G900 260 samples
	90.7% (508)	99.7% (354)	100.0% (260)

Second, the accuracy dependence on the threshold for letter appearance (N_{TH}) and key press duration (T_{TH}) for each group are shown in Fig. 5 and Fig. 6, respectively. The results indicate that $N_{TH} = 3$ and $T_{TH} = 300$ ms are adequate to this application, both being in the region where high accuracy can be obtained.

4. CONCLUSION

In this paper, we studied keystroke dynamics during the input of long free texts in Japanese. Using keystroke times specific to Japanese text input as feature indices, we found their optimal combination yielding the higher recognition accuracy, attaining 100% in G900.

In this experiment, by requesting an interval of at least

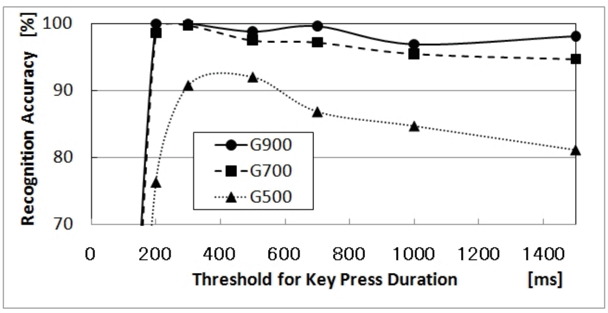

Fig. 6 Dependence of recognition accuracy on threshold for key press duration ($T_{TH} = 100 \sim 1400$ ms)

one week between successive text input sessions and by setting five profiles for each subject, we performed the data taking under the condition similar to an actual application environment, and achieved the high recognition accuracy.

Furthermore, our experiment was conducted on a large scale involving 112 subjects, which allowed for comparisons between groups divided according to typing skill, thus resulting in an investigation that has not been performed in the past.

REFERENCES

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

[2] J. Leggett and G. Williams, "Verifying identity via keystroke characteristics", *International Journal of Man-Machine Studies*, 28(1), pp. 67-76, 1988.

[3] R. Joyce and G. Gupta, "User authorization based on keystroke latencies", *Communications of the ACM*, 33(2), pp. 168-176, 1990.

[4] M. Kasukawa, Y. Mori, K. Komatsu, H. Akaike and H. Kakuda, "An Evaluation and Improvement of User Authentication System Based on Keystroke Timing Data", *Transactions of Information Processing Society of Japan*, Vol.33, No.5, pp. 728-735, 1992 (in Japanese).

[5] M. S. Obaidat and B. Sadoun, "Keystroke Dynamics based Authentication", *Biometrics: Personal Identification in Networked Society*, ed. By A. K. Jain et. al., Springer, pp. 213-230, 1999.

[6] S. Bender and H. Postley, "Key sequence rhythm recognition system and method", *US Patent US7206938*, US Patent Office, 2007.

[7] F. Monrose and A. Rubin," Keystroke Dynamics as a Biometric for Authentication", *Future Generation Computer Systems*, Vol. 16, pp. 351-359, 2000.

[8] F. Bergadano, D. Gunetti and C. Picardi, "User authentication through Keystroke Dynamics", *ACM Transactions on Information and System Security (TISSEC)*, 5(4), pp. 367-397, 2002.

[9] D. Gunetti and C. Picardi,"Keystroke Analysis of Free Text", *ACM Transactions on Information and System Security*, Vol.8, No.3, pp. 312-347, 2005.

[10] M. Curtin, C. Tappert, M. Villani, G. Ngo, J. Simone, H. St. Fort and S.H. Cha, "Keystroke Biometric Recognition on Long-Text Input", *International MultiConference of Engineers and Computer Scientists 2006*, 2006.

[11] M. Villani, C. Tappert, G. Ngo, J. Simone, H. St. Fort and S.H. Cha, "Keystroke Biometric Recognition Studies on Long-Text Input under Ideal and Application-Oriented Conditions", *Proc. Computer Vision and Pattern Recognition Workshop on Biometrics*, 2006.

[12] T. Samura and H. Nishimura, "Performance Analysis of Personal Identification by Keystroke Data in Free Text Typing", *Proceedings of SICE Symposium on Systems and Information 2008*, pp.59-64, 2008 (in Japanese).

[13] 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).

[14] 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).