Wong, L.-H. et al. (Eds.) (2013). Proceedings of the 21st International Conference on Computers in Education. Indonesia: Asia-Pacific Society for Computers in Education

Analysis of Writing Data for Cheating Detection in e-Testing

Yu YOSHIMURA[a*], Takehiro FURUTA[b], Takahito TOMOTO[c] & Takako AKAKURA[c]
[a]Graduate School of Engineering, Tokyo University of Science, Japan
[b]Nara University of Education, Japan
[c]Faculty of Engineering, Tokyo University of Science, Japan
yoshimura_yuu@ms.kagu.tus.ac.jp[a*]

Abstract: e-Testing is effective in relieving time and space limitations for examinations. However, a drawback is that user authentication employs only login credentials, making cheating easy. We examine variations in personal writing data in e-testing with the aim of detecting cheating. We assume two behaviors: answering questions by oneself and replicating others' answers. We used the sub-stroke method for data analysis, which allowed confirmation of differences in the change over time in writing speed and pen pressure, and differences in average pen angle and direction.

Keywords: e-testing, individual authentication, sub-stroke method, unsteady behavior

1. Introduction

e-Learning is becoming increasingly popular in higher education because it can alleviate time and space restrictions (Suzuki, 2009). However, few universities give credit for lessons that use only e-learning; often students can do e-learning lessons at home, but take regular examinations at the school. e-Testing in the form of web-based exams is effective for reducing the constraints of regular examinations, as it is possible to take exams at home rather than at a testing center. However, most e-testing systems perform user authentication using only a user name and password which are entered at login, making it easy to cheat. For example, persons other than the one enrolled in the course may take the exam, or the Internet or books might be used to retrieve information.

Cheating during exams cannot be prevented using only login authentication. Kikuchi et al. (2008) proposed constant individual authentication using pen pressure data to prevent spoofing in e-testing. Pre-registered writing data (five characters often used in multiple-choice exams) and writing data for characters entered during the exam are used. The method performs individual authentication by comparing these data, and showed the possibility of preventing spoofing by doing so.

However, a method of cheating prevention is also necessary for practical use in e-testing. When cheating during an exam, the examinee may give answers with peculiarities that appear as abnormal values in writing data. We aim to detect these abnormal values at the time of cheating through the use of writing data obtained from a pen tablet.

Writer verification methods using hidden Markov models (Fierrez et al., 2008) and dynamic programming matching (Komiya et al., 2001; Zhao et al., 1996) have been proposed. These studies target signatures, however, and cannot be used for the free-response examination that we are targeting. Moreover, the method of pre-registering the characters used in the exam is to become a burden on the examinees in the free-response examination. A method that does not depend on character shape is desirable.

In this study, we use the sub-stroke method (Tokuno et al., 2002; Shimodaira et al., 2003; Nakai et al., 2003). This method compares similar parts between characters, allowing this method to be applied to various examination styles and a wide variety of languages. Our analysis focuses on cheating detection, so we assume two answering behaviors: usual answering behavior that indicates answering a question on one's own, or answer replication behavior that indicates transcription of the answer of another. We analyzed differences between answer replication and usual answering behavior. Specifically, we compared differences in the change over time and average values of the writing data.

2. Available Writing Data

2.1 Writing Data Obtained from Pen Tablets

Examinees can cheat at any time, making it necessary to constantly authenticate them during e-testing, not only at the time of login. However, methods that are burdensome for examinees cannot be used. We thus propose a method to detect cheating through the use of writing data that can be continuously acquired during the exam. Writing behavior is a natural action while answering problems, and the use of writing data is not burdensome for examinees.

Static and dynamic writing data are obtained from a pen tablet when examinees write answers. Dynamic data consist of time series information including pen tip x- and y-coordinates, pen pressure P, angle between pen and tablet θ, pen direction φ, and writing speed v (Figure 1). Data are saved as a time series with 10 ms sampling intervals. Static data are image data of the answer screen (Figure 2).

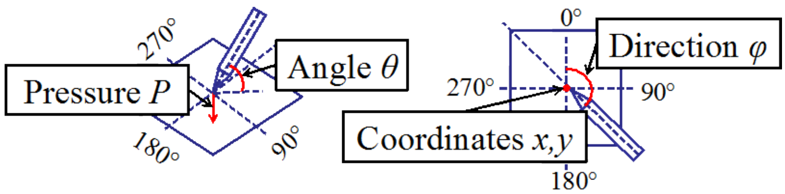

Figure 1. Dynamic Data

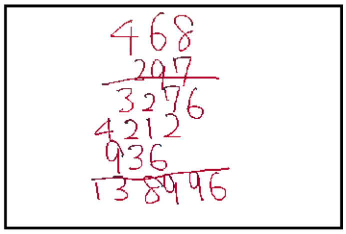

Figure 2. Static Data

2.2 Strokes Division of Dynamic Data

We assume a free-response format in which characters produced are different for each examinee and question. If the test using the language in which the character of a huge variety as Japanese Kanji, pre-registering all characters used in a free-response examination is burdensome for examinees, making it difficult to use matching methods that use character shapes for writer verification. A method that does not depend on character shape could furthermore be applied to various languages in e-testing.

For data analysis and cheating detection we use a sub-stroke method that compares similar character strokes. Such dynamic data are not visible to the naked eye and cannot be intentionally imitated, making it effective as data for cheating detection. The method has two steps: stroke extraction and stroke division. Stroke extraction means extracting dynamic data between the start and end of a stroke. Strokes division means dividing data extracted in the previous step into three parts: “writing-start,” “writing-middle” and “writing-end” (Figure 3). Writing-start is defined as the time from initial pen contact with the tablet until when the pen starts moving. Writing-end is defined as the time from when pen pressure begins to fall until when pen leaves the tablet. Writing-middle is defined as the time between writing-start and writing-end. Each stroke division is assigned a direction number that indicates the direction of a line connecting the start and end points (Figure 4).

In the sub-stroke method, each dynamic data point is classified according to twelve indicators (3 elements × 4 possible direction numbers) (Figure 5). 3 elements “s”, “m”, “e” are as follows; the notation "s" indicates writing-start, "m" indicates writing-middle and "e" indicates writing-end. For example, the notation "s4" indicates writing-start and direction number 4.

By applying sub-stroke method to registered characters and characters written during exams, any character written during the exams allows comparison with similar parts of registered characters. Using by sub-stroke method, the examinee does not have to pre-register all characters used during the exam. From the above, sub-stroke method is a method that does not depend on the form of the character, and it is available in various test formats and languages.

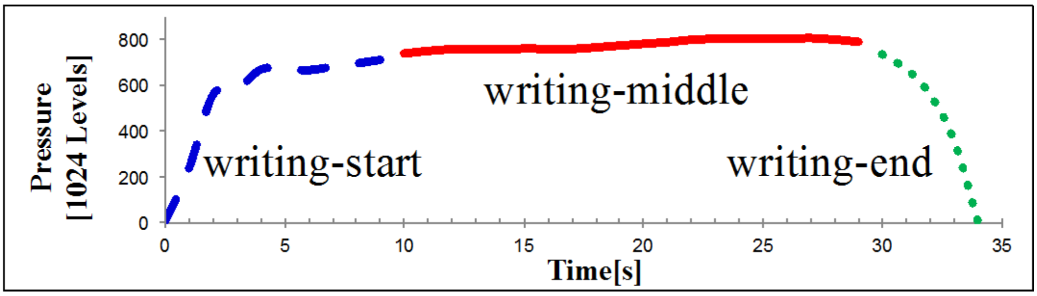

Figure 3. Fragmentation of Stroke

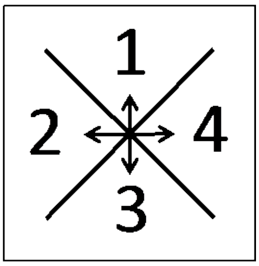

Figure 4. Direction Number

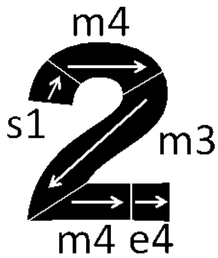

Figure 5. Example of Sub-stroke Method

3. Preliminary Experiment

3.1 Experiment Overview

We performed a comparative analysis of usual answering behavior and answer replication behavior. Usual answering behavior occurs when examinees read a question, think about the answer, and write the answer by oneself. Answer replication behavior occurs when examinees transcribe a model answer from an answer sheet without reading the question sentence.

Experiment subjects were eight students at a Japanese university. Subjects first performed exercises to get used to the pen tablet. Exercises were five simple math questions and transcriptions of five proverbs. Subjects then solved 10 math problems regarding permutations, combinations, and profit and loss calculations to produce analysis data. Six questions were solved using usual answering behavior. Four questions were solved using answer replication behavior (Table 1).

Table 1: Problem Content (usu: usual answering behavior; rep: answer replication behavior)

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
usu	usu	usu	rep	rep	usu	usu	usu	rep	rep
<Profit and loss calculation>					<Combination>				

3.2 Experimental Device

Problems were answered using a pen tablet (Wacom Intous4) and a compatible laptop computer (Fujitsu FMV-A8280) (Figure 6). Writing with the pen on the tablet is output on the computer's monitor (Figure 7). After writing an answer, the user presses a submit button to transfer writing data to the server.

Figure 6. Experimental Devices

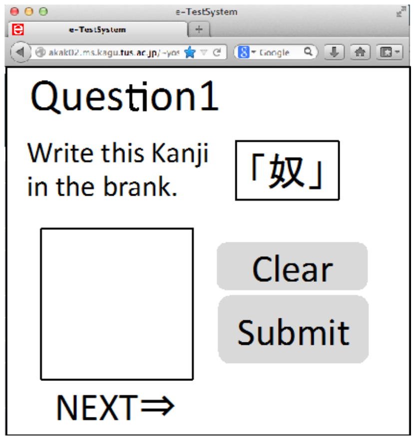

Figure 7. System Screen

3.3 *Data Analysis*

3.3.1 *Analytical Method Using Changes over Time*

Answer replication behavior is abnormal behavior. In the usual answering behavior, the examinee's gaze travels between two places: the monitor and the writing hand. In answer replication behavior, the examinee's gaze moves between three places: the model answers, the monitor, and the hand. There is thus a possibility of differences between the dynamic data during the answer time. We compared changes over time between the dynamic data resulting from answer replication behavior and that from the usual answering behavior. We first applied the sub-stroke method to each question, and then drew an approximate straight line according to the classified indicator among the 48 possible types (4 kinds of dynamic data × 3 elements × 4 directions).

Figure 8 shows and example of the analysis. In that figure, the horizontal axis is the answer time, and the vertical axis is the value of the dynamic information for the characters representing the Japanese vowel sounds (a, i, u, e, o). A "pressure-s4" indicator appears at the four sites in this answer. Data are plotted as a scatter diagram. Approximate straight lines are calculated from among 80 types for each indicator (48 usual answering behavior types and 32 answer replication behavior types). We then calculate the average value of the slope of the approximate straight line for each indicator and each examinee, and check whether the indicators of answer replication behavior are large.

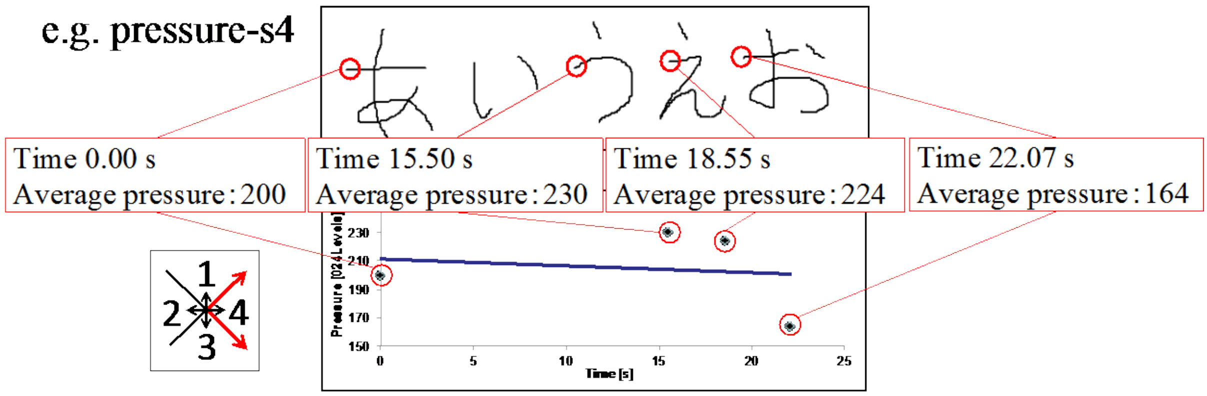

Figure 8. Method of analysis

3.3.2 Results of Analyzing Change over Time

For each indicator and each examinee, we calculated the difference between the average value of the slope of the six answer replications and the four usual answers. We then calculated the mean and variance of the 12 examinees using the calculated difference. Based on the average value, we sort the indicators in descending order. Tables 2 and 3 show the top five indicators of writing speed and pen pressure.

Writing-start and writing-end have higher ranks for pen pressure. Answer replication behavior shows an effect at the start and end of writing strokes. In addition, direction numbers 2 or 4 have a higher rank. We believe that these effects in answer replication behavior appear when writing lines either to the left or the right.

Writing-middle has a higher rank for writing speed. Answer replication behavior affects the middle stroke. In addition, direction numbers 2 or 4 have a higher rank and pen pressure, so we believe that the effects of answer replication behavior appear when writing lines to the left or right. There is no large difference in angle and direction. Differences are seen in intercepts of the straight line approximation. Based on the above, direction and angle are analyzed as the average value of the data.

Attitude when solving a question is different between usual answering behavior and answer replication behavior. Change in pen angle is due to this change in attitude. Therefore, the direction and angle allow comparison of differences between the average values in the usual answering behavior and answer replication behavior.

Table 2: Analysis of Pen Pressure: Average Top Five Indicators of the Difference

Pen pressure	s1	e2	e4	m2	m4
Difference of average value	5.97	5.67	5.08	4.40	3.64
Difference of variance	42.53	82.95	53.61	50.30	24.66

Table 3: Analysis of Writing Speed: Average Top Five Indicators of the Difference

Writing speed	m2	m4	s1	e2	s4
Difference of average value	0.30	0.22	0.21	0.17	0.14
Difference of variance	0.05	0.17	0.09	0.10	0.05

3.3.3 Results of Analyzing Mean Value

Similar to the analysis of change over time, we first apply the sub-stroke method to each question. For each indicator, we calculated the variance and the mean difference for the four usual answer questions and the six answer replication questions for each subject. Based on the average value of the difference, we sorted the indicators in descending order. Tables 4 and 5 show the top five indicators of the angle and direction.

Angle and direction data were confirmed as having differences. The direction numbers 1 and 2 had higher ranks. We believe that the effects of answer replication behavior appear when writing lines to the right or left. These results confirm that there is a change in pen angle when performing abnormal behavior such as transcribing answers from others. Little difference in writing speed and pen pressure was indicated. We believe that this is a result inherent to angle and direction.

Table 4: Analysis of Direction: Average Top Five Indicators of the Difference

Direction	s1	m1	e2	s4	m3
Difference of average value	1.71	0.69	0.65	0.58	0.54
Difference of variance	168.26	9.95	13.80	16.43	11.08

Table 5: Analysis of Angle: Average Top Five Indicators of the Difference

Angle	**s1**	**s2**	**e2**	**s1**	**s3**
Difference of average value	**5.39**	**4.12**	**3.91**	**3.72**	**3.43**
Difference of variance	**23.78**	**6.70**	**8.76**	**4.90**	**3.37**

4. Conclusion and Future Work

We analyzed differences in writing data between usual answering behavior and answer replication behavior for the purpose of detecting cheating in e-testing. Analysis of change over time for same questions between examinees indicated differences in writing speed and pen pressure. Effects of differences in attitude were observed in some indicators of direction and angle. In conclusion, we found two types of differences in change over time and average value.

Future research will be directed at increasing the versatility of the method, for example, by extending the sub-stroke method to curves. We also intend to carry out experiments regarding other cheating patterns, as well as further data analysis.

Acknowledgements

This study was supported in part by a Grant-in-Aid for Challenging Exploratory Research (No. 25560121) from the Japan Society for the Promotion of Science.

References

K. Suzuki. (2009). E-Learning in Japan: Past, Present, and Future. *KAEM and the 4th BK21 GGRTE International Conference: Technology and Future Learning Space proceedings*, pp.9-17.

S. Kikuchi, T. Furuta, and T. Akakura. (2008). Periodical Examinees Identification in e-Test Systems Using the Localized Arc Pattern Method. *Association of Pacific Rim Universities 9th Distance Learning and the Internet Conference 2008 proceedings*, pp.213-220.

Y. Komiya, T. Ohishi and T. Matsumoto. (2001). A pen input on-line signature verifier integrating position, pressure and inclination trajectories. *IEICE Trans. Inf. & Syst., 84(7)*, pp. 833-83.

P. Zhao, A. Higashi, and Y. Sato (1996). On-Line Signature Verification by Adaptively Weighted DP Matching. *IEICE Trans. Inf. & Syst., 79(5)*, pp.535-541.

J. Fierrez, J. Ortega-Garcia, D. Ramos, and J. Gonzalez-Rodriguez. (2007). HMM on-line signature verification: Feature extraction and signature modeling. *Pattern Recognition Letters, 28(16)*, pp.2325-2334.

J. Tokuno, N. Inami, A. Matsuda, M. Nakai, H. Shimodaira, and S. Sagayama. (2002). Context-dependent substroke model for HMM-based on-line handwriting recognition. *Proc. IWFHR-8*, pp78-83.

H. Shimodaira, T. Sudo, M. Nakai, and S. Sagayama. (2003). On-line overlaid-handwriting recognition based on substroke HMM. *Proc. ICDAR 2003*, pp.1043-1047.

M. Nakai, H. Shimodaira, and S. Sagayama. (2003). Generation of hierarchical dictionary for stroke-order free kanji handwriting recognition based on substroke HMM. *Proc. ICDAR 2003*, pp.514-518.