Chapter 10
Personal Identification and Authentication Based on Keystroke Dynamics in Japanese Long-Text Input

Toshiharu Samura
Akashi National College of Technology, Japan

Haruhiko Nishimura
University of Hyogo, Japan

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 a hybrid model for identifying individuals on the basis of keystroke data in Japanese long-text input. By evaluating the identification performance of individuals in the three groups, the effectiveness of the method was found to correspond to the typing skill level of the group.

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 ac-

DOI: 10.4018/978-1-61350-129-0.ch010

cess 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 login, using not only information about a series of input characters for password recognition, but also keystroke dynamics as part of the authentication process (Umphress & Williams, 1985; Leggett & Williams, 1988; Joyce & Gupta, 1990; Kasukawa et. al., 1992; Kotani, Norioka & Horii, 2005; Obaidat & Sadoun, 1999; Bender & Pankanti, 2007; AdmitOneSecurity, 2010). Short words such as passwords, however, generally do not contain sufficient keystroke information for such dynamics to enable user recognition, resulting in a lower verification rate in comparison with fingerprint or retinal scans. Keystroke dynamics is therefore most often used as a complementary check in password recognition.

Similar to handwriting analysis and voice printing, another use of keystroke dynamics is feature analysis related to human behavior. This study focuses on identification and authentication from the point of view of using keystroke dynamics related to human behavior. Here, we consider 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 (Monrose & Rubin, 2000; Bergadano, Gunetti & Picardi, 2002; Gunetti & Picardi, 2005; Curtin et. al., 2006; Villani et.al., 2006; Ahmed & Traoré,2008; Samura & Nishimura, 2008; Samura & Nishimura, 2009a; Samura & Nishimura, 2009b; Tappert, Villani & Cha, 2009; Samura & Nishimura, 2010; Samura & Nishimura, 2011).

In the present study, we propose 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. For identification methods, we use our previously proposed Weighted Euclidean Distance (WED) method (Samura & Nishimura, 2009a; Samura & Nishimura, 2009b), the Array Disorder (AD) method proposed by Gunetti et al. (Bergadano, Gunetti & Picardi, 2002; Gunetti & Picardi, 2005), and a hybrid method that compares given keystroke data using both the WED and AD methods, as well as a combination of the two (Samura & Nishimura, 2010a). While the WED method evaluates the magnitude of differences in feature indices, the AD method evaluates ranking patterns of feature indices. The difference in these approaches leads to significantly higher identification rates for participants with low typing skill when using the WED method in comparison with the AD method. On the other hand, as pointed out by Gunetti et al., relative ranking patterns are resistant to the effects related to differences in input environment (e.g., differing keyboards) and physical or emotional states; thus, in some cases the AD method should be more effective than the WED method. This chapter presents a large-scale study involving 189 participants. We compare the WED and AD methods, and introduce and evaluate a hybrid method that complementarily incorporates the strengths of both.

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. Figure 1 shows example data on Japanese hiragana input by entering combinations of Latin letters. The first field shows the typed let-