CHAPTER 5

Will User Authentication Using Keystroke Dynamics Biometrics Be Interfered by Emotions? – NCTU-15 Affective Keyboard Typing Dataset for Hypothesis Testing

Po-Ming Lee, Liang-Yu Chen, Wei-Hsuan Tsui and Tzu-Chien Hsiao

In this chapter we 1) provide a new dataset collected from real-world for researchers to examine possible influence of emotions on user authentication using keystroke dynamics biometrics, or develop their own systems to recognize emotions using keystroke

Po-Ming Lee
Institute of Computer Science and Engineering
National Chiao Tung University, Hsinchu, Taiwan, ROC

Liang-Yu Chen,Wei-Hsuan Tsui
Institute of Biomedical Engineering
National Chiao Tung University, Hsinchu, Taiwan, ROC

Tzu-Chien Hsiao
Institute of Biomedical Engineering
Department of Computer Science
Biomedical Electronics Translational Research Center and Biomimetic Systems Research Center
National Chiao Tung University,Hsinchu, Taiwan, ROC
e-mail: labview@cs.nctu.edu.tw

Editors: Y. Zhong and Y. Deng, *Recent Advances in User Authentication Using Keystroke Dynamics Biometrics*

DOI: 10.15579/gcsr.vol2.ch5, GCSR Vol. 2, pp. 71-81, 2015
©Science Gate Publishing - Available under CC BY-NC 4.0 International License

dynamics patterns, 2) summarize recent findings in the field of emotion recognition using keystroke dynamics, and 3) provide concrete suggestions to the field of user authentication using keystroke dynamics biometrics based on the empirical findings derived from the proposed dataset.

5.1 Introduction

Emotion plays an essential role in human life because it is the foundation of the motivational system of people. Recent studies in the field of psychology, brain science and communication technology have demonstrated the substantial effects of emotion on human cognition and behavior. Lang et al. (1995) reported an empirical study on the effects of emotional characteristics of stimuli on subjects' cognitive capacity and memory [6]. Bolls et al. (2001) revealed the effect that subjects tend to remember stimulus that elicit negative emotions more, than the stimulus that elicit positive emotions [3]. Later researchers focused on the use of emotional relevant stimulus on attracting the attention of subjects and to make subjects remember more on the presented stimuli. Theories and experimental results of the examinations on the connections between emotion and motivation were also reported [7].

Automatic affect recognition technology has been proposed and has attracted considerable attention since its proposal. The proposed technology aims to help in the Human Computer Interaction (HCI) area. This is because a computer interactive application that does not understand or adapt to a user's context, such as their location, professional, or the emotion states of a user, may lead to usability problems. Such an application could provide annoying feedback, interrupt users in an inappropriate situation, or increase the user's frustration. The main purpose of affect recognition technology is to provide intelligent systems that can provide computer applications the information about the changes of human emotions. As a result, the applications equipped with such a technology can detect and respond to a users' emotion state, and could even give a better user experience as well as provide appropriate feedback in helping users in using the applications. Various methods have been proposed to detect emotions, including the use of facial expressions, voice intonation, and physiological signals collected from human body. All of these proposed methods have high classification rate of success. However, the drawback of these methods is that they can be intrusive to the user and the equipment may be difficult to obtain because of being too expensive.

During the last decade, a novel approach for emotion recognition has been reported, which is by using keystroke dynamics. For keystroke dynamics the time of each key press and release on the keyboard data were collected and keystroke timing features were extracted for analysis. Keystroke dynamic has been used in authentication systems. Just like handwritten signatures, everyone's typing rhythms are unique. For example, a system can authenticate a user based on the provided user name, password, and the typing rhythm. Monrose et al. (2000) suggested that using keystroke dynamics can be interfered by intensive emotions but the relationship between keystroke and emotional states [9] is still unclear. Since proposed by MIT Media Lab in 2003, the technology of emotion recognition using keystroke dynamics [12] has been demonstrated effective in various experimental setups [10] based on various Machine-Learning (ML) techniques [4]. These findings lead to a hypothesis that emotions could influence keystroke dynamics. The hypothesis furthermore leads to several questions. For example, should the performance of the user authentication based on keystroke dynamics biometrics be interfered by emotions? How can a developer eliminate this effect? Previous studies related to the advances in emotion recognition using keystroke dynamics are mostly showcases. Despite the reported performance of the ML classification models that can be built and the provided features used for model building, the source of variance and the effect size of emotions have never been analyzed. Furthermore, the dataset that was used was never provided for open access to other researchers. So, the research questions that may concern the researchers and developers in the field of user authentication currently remain unanswered.

Hence, in this chapter, we 1) provide a new dataset collected from real-world for researchers to examine possible influence of emotions on user authentication using keystroke dynamics biometrics, or develop their own systems to recognize emotions using keystroke dynamics patterns, 2) summarize recent findings in the field of emotion recognition using keystroke dynamics, and 3) provide concrete suggestions to the field of user authentication using keystroke dynamics biometrics based on the empirical findings derived from the proposed dataset.

This chapter is written especially for the researchers and developers in the field of user authentication who concern about the influence of emotions on keystroke dynamics. Due to the reports about recent advances in the technology of emotion recognition using keystroke dynamics, the researchers and developers in the field of user authentication may wonder how these findings are going to fit into their research. Specifically, first, how will such an effect (*i.e.* the influence of emotions) affect their own technology? Second, how to detect such an interference, and third, how to control or eliminate such an interference? Moreover, researchers may be interested to perform their own experiments to gain more information and sense about this phenomenon, but they would need a dataset that contains both keyboard typing information and affective ratings.

The researchers and developers in the field of user authentication can easily gain the information about the latest advances and findings in the development of emotion recognition using keystroke dynamics, can have a glimpse about the possible influence of emotions on their developing authentication technology, and can furthermore derive their own conclusions, test their hypothesis, or develop their own classification models based on the dataset provided. This would satisfy the researchers and developers because the chapter not only provides conclusions and concrete suggestions, but also a dataset collected from real-world for them to test their thoughts and opportunity to improve their own technology.

The following sections and subheadings in this chapter are listed below:

1. Introduction
2. Recent Advances in The Field of Emotion Recognition
3. Description of NCTU-15 Dataset
   - Ethics Statement

  - Subjects
  - Experimental Procedure
  - Stimuli and Self-Report
  - Apparatus

4. Empirical Evidence Derived From The NCTU-15 Dataset

5. On Eliminating The Interference of Emotions in User Authentication

Section 5.1 briefly describes the "why". That is, the goals and the organization of content of this book chapter. Section 5.2 provides a literature study of the field of emotion recognition using keystroke dynamics. This section includes the most important investigation results in the field, in regard to the advances of the proposed method, the findings of the experiments, and the conclusions. The descriptions about the dataset are provided in Section 5.3, in which materials that include ethics statement, subject description, experimental procedure, stimuli and self-report, and apparatus, are provided. Section 5.4 describes the findings derived from the dataset. Both the statistical methods and ML algorithms will be applied to the dataset in order to discuss the nature and the properties of the dataset. The discussion especially focuses on how the findings relate to the readers in the field of user authentication. Moreover, this section will also provide empirical evidence for the conclusions which are to be drawn in Section 5.5. That is, how can user authentication systems based on keystroke dynamics biometrics detect, control, or even remove the effect caused by emotions, in order to improve the robustness of such systems. The questions from the readers, such as which is the size of the effect, what is the source of the effect, how the factors interact, and how to detect the effects, will be answered.

5.2 Recent Advances in The Field of Emotion Recognition

Emotion recognition technology based on keystroke dynamics was not reported in the literature until Zimmermann et al. (2003) first described this approach [12]. These authors proposed an experiment designed to examine the effect of film-induced emotional states (PVHA, PVLA, NVHA, NVLA and nVnA (P = positive, N = negative, H = high, L = low, n = neutral, V = valence, A = arousal)) in subjects, with the keystroke dynamics in regard to keystroke rate per second, average duration of keystroke (from key-down until key-up event). However, they did not actually carry out the work described in their proposal. The use of keystroke dynamics for emotion recognition has two main advantages that make such a technique favorable. The two advantages are that it is non-intrusive and easy-to-obtain because the technique does not require any additional equipment or sensors other than a standard input device, which is the keyboard of a computer. Later, numerous studies in the field of computer science have reported the development of emotion recognition technology based on keystroke dynamics. Vizer et al. (2009) reported the use of ratios between specific keys and all keys to recognize task-induced cognitive and physical stresses from a neutral state [11]. They achieved a classification rate of 62.5% for physical stress and 75% for cognitive stress. The key ratios could represent the frequencies of typing specific keys, which may increase or decrease due to the changes in emotional state. The analysis result was produced based on sophisticated Machine-Learning (ML) algorithms, and hence, the relationship between emotion and these ratios was not identified. Notably, most of the main streams of ML algorithms only produce models that are considered to be a black box, and do not produce readable models. The ML algorithms are usually used for building models from dataset that contains complex relationships which are not able to be identified by a traditional statistical model (*e.g.* t-test, ANOVA). In 2011, Epp et al. (2011) reported a result of building models to recognize experience-sampled emotional states based on keystroke durations and latencies that were extracted from a fixed typing sequence [4]. The accuracy rates of classifying anger, boredom, confidence, distraction, excitement, focus, frustration, happiness, hesitance, nervousness, overwhelmed, relaxation, sadness, stress, and tired, with respect to two-class models, were 75% on average. The study built models by using ML algorithms and a correlation-based feature subset attribute selection method [5]. Although the keystroke features that were used to build the model with the highest accuracy were provided, the relationship between emotion and keystroke dynamics, still, was not reported. Recently, more results related to classification on emotional data using similar feature set have been proposed. Alhothali (2011) reported the use of keystroke features that were extracted from arbitrarily typed keystroke sequences as reaching an 80% accuracy rate of classifying experience-sampled positive and negative emotional states [1]. Bixler and D'Mello (2013) demonstrated a 66.5% accuracy rate on average for two-class models in detecting boredom, engagement, and neutral states. The emotional data used were collected using the experience sampling method [2].

By applying ML methodology for building classification models from various datasets collected from different experimental setups, these studies have suggested that keystroke duration and keystroke latency can be used for model building. One therefore could hypothesize that the keystroke duration and latency may be different when subjects are in different emotional states.

5.3 Description of NCTU-15 Dataset

5.3.1 Ethics Statement

NCTU-15 Dataset is a 15-subject dataset collected under the research project approved by the Institution Review Board (IRB) of the National Taiwan University Hospital Branch. Written informed consents were obtained from all subjects before the experiment. The NCTU-15 dataset is available to all interested researchers upon request to the Institute of Biomedical Engineering, National Chiao-Tung University Institutional Data Access for researchers who meet the criteria for access to confidential data without limitations.

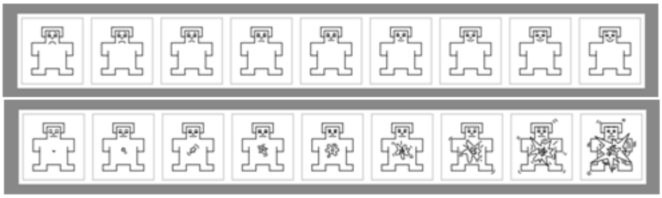

Figure 5.1: Self-Assessment Manikin (SAM) used in this experiment, in which the upper row represents valance and the lower row represents arousal.

5.3.2 Subjects

Fifteen university students participated in the collected data, ranging in age between 20 and 28 (mean = 24.67, standard deviation = 1.91; 11 males, 4 females). All the subjects self-reported that they were non-smoker, healthy, with no history of brain injury, cardiovascular problems, and had normal or corrected-to-normal vision and normal range of finger movement.

5.3.3 Experimental Procedure

An experimental procedure is taken by installing software to subjects' mostly used computer in their daily life to collect data of standard input devices and the affective states of subjects. The software was run in the background and starts every time in the beginning of the startup of the operating system (the software is currently available for Windows because it was written by C#), the keystroke and mouse movement data was collected all the time in a time resolution of 100 nanoseconds, and the software used to collect data will ask subject to self-report the held emotion state every ten minutes when subject have been using the computer for over five minutes. Because it is crucial for producing comparable results, this experiment adopted computer-based Self-Assessment Manikin (SAM) shown in Fig.(5.1), which is provided in [10] for subjects to self-report their emotional states during the experiment. The data collection software stopped and pop-up a message box to the subjects when predefined number of self-report trial is reached (the number was set to fifty for each subject in this study).

5.3.4 Stimuli and Self-Report

In order to elicit users' emotional states in a most natural way and have that way closer to the real-world situations, dataset was not collected in a laboratory with the application of standard emotional stimuli to the subjects. Instead, we adopted the experience-sampling method that lets subjects to use the SAM to report their emotional state periodically in their daily life. Experience-sampling could avoid the influence of memory and time, and collect the accurate description of the feelings of the subjects.

Table 5.1: The distribution of emotional states reported by subjects in regard to valance and arousal.

	Pleasant								Unpleasant
Valance	0	1	2	3	4	5	6	7	8
Count	62	30	209	99	211	70	83	11	35
	Arousal								Calm
Arousal	0	1	2	3	4	5	6	7	8
Count	100	47	171	107	247	74	54	5	5

5.3.5 Apparatus

The computer including the standard input devices, keyboard and mouse that participants mostly used in their daily life. The experimental software is installed in their computers and run in the background. Statistics show those 7 installations on the office desktop computer, 7 installations on their notebook and 1 installation on home desktop computer.

5.4 Empirical Evidence Derived From The NCTU-15 Dataset

The collected data set can be divided into two parts, the self-report feedback provided by the 15 subjects, and the keystroke data. For the keystroke data, the time of each key press on the keyboard was collected but the exact keys pressed by the subjects were not recorded due to the concern of the subjects' privacy issue. The guarantee on the privacy of users is essential for the designing of most emotion recognition technology, and is also the main concern reported by the users in our previous experiments.

For the self-report, totaled 810 feedbacks were collected from the 15 subjects. The distribution of emotional states reported by subjects in regard to valance and arousal is shown in Table 5.1. This table shows that the subjects were aroused accompanied with unpleasant feelings or pleasant feelings during the data collection period.

This section offers the findings about the influence of emotion on keystroke latency. The descriptive statistics of the influence of emotion on keystroke latency are provided in Table 5.2. Figure 5.2 shows the interaction between keystroke latency and emotional dimensions. Figure 5.2 shows that keystroke latency does not change a lot in low arousal and the keystroke latency is shorter for the higher arousal in negative valence. This keystroke latency data was also submitted to a two-way Repeat Measures ANOVA. The ANOVA results are provided in Table 5.3. However, statistically significant difference was not found in the main effect Valence and Arousal in our NCTU-15 dataset. It is worth to note that this result is different from our results obtained in the laboratory [8]. In [8], we conducted a controlled experiment to collect subjects' keystroke data under different emotional states induced by International Affective Picture System (IAPS). Two-way Valence (3) $\times$ Arousal (3) ANOVAs were also used to examine the collected dataset. The findings of that experiment suggest that the effect

Table 5.2: Descriptive statistics of keystroke latency under independent variables Valence × Arousal.

Valence	Arousal	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
negative	low	0.0596	0.0091	0.0414	0.0778
	medium	0.0335	0.0069	0.0197	0.0473
	high	0.0235	0.0078	0.0079	0.0391
neutral	low	0.0627	0.0196	0.0235	0.1019
	medium	0.0957	0.0160	0.0637	0.1277
	high	0.0252	0.0046	0.0160	0.0344
positive	low	0.0579	0.0177	0.0225	0.0933
	medium	0.0551	0.0170	0.0211	0.0891
	high	0.0833	0.0333	0.0167	0.1499

Table 5.3: Repeated measures two-way ANOVA table for keystroke latency. Here shows the result of the 3 (Valence: negative, neutral, positive) × 3 (Arousal: low, medium, high) ANOVA. *p<.05, **p<.01, ***p<.001.

Source of Variance	SS	df	MS	F	P
Subjects	1.6053	14	0.11466		
Valence***	0.0209	2	0.01046	0.33	0.716
Arousal	0.0264	2	0.01320	0.42	0.656
Valence x Arousal***	0.0568	4	0.01421	0.45	0.769
Total	24.4471	740			

of emotion is significant ($p < .001$) in the keystroke duration, keystroke latency, and accuracy rate of the keyboard typing.

The difference in this experiment and our experiment in [8] is the way of eliciting emotions. The controlled experiment avoids the unbalanced induced emotional state and controls the length of time interval of the features extracted. But in real-world application, there should not be any restrictions to the users; for example, when and how they type-in, in their daily life. The length of time interval of extracting features may affect the efficacy of described feature set. If the time interval selected is too long, different emotional state may be included. Conversely, the time interval that is too short may not contain sufficient features for detecting emotion states. To elicit users' emotional states in a natural way, it is common to encounter with inadequate data collection and unbalanced dataset, though the statistical results may not prove the effect of emotion.

The Decision Tree (DT, also "J48" in weka) is also applied to the dataset for building a classification model. Three attributes were used, subjects, keystrokes latency, and the emotional states, to build the model. There are nine emotional states to be classified which are constituted by 3 (Valence: negative, neutral, positive) × 3 (Arousal: low, medium, high). Figure 5.3 shows a skew in the distribution of the nine emotional

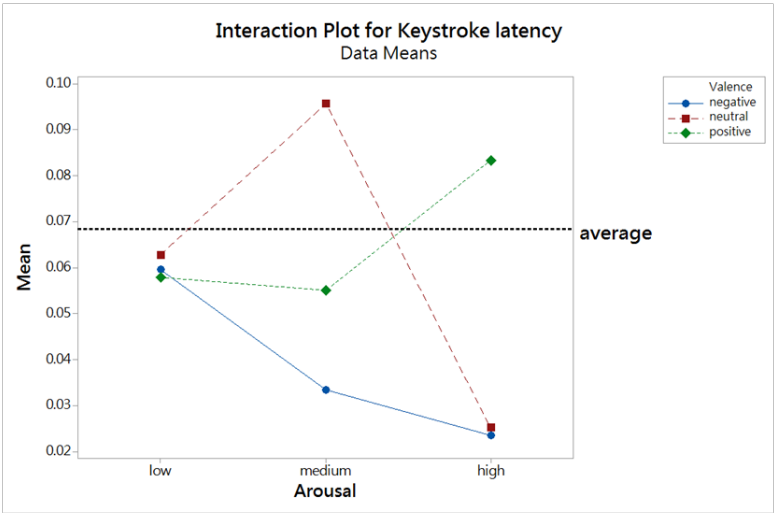

Figure 5.2: The interaction plot for keystroke latency.

states. The adopt method here is oversampling in order to eliminate the problem of class skew and a 10-Fold Cross Validation (10-Fold CV) is used to evaluate the built model. Our result shows that the classification rate is 67.21%.

5.5 On Eliminating The Interference of Emotions in User Authentication

This chapter is written especially for the researchers and developers in the field of user authentication that concern about the influence of emotions on keystroke dynamics. Due to the reports about recent advances in the technology of emotion recognition using keystroke dynamics, the researchers and developers may wonder how these findings are going to fit into their research. Our answers to the questions listed in the Section 5.1 are provided below:

- ***How will such an effect (i.e. the influence of emotions) affects their own technology?*** According to the previous study and the experimental results provided in the chapter, different emotional state may lead to different typing rhythm. However, the variance is small compare to the individual difference in the typing rhythm (see [8] for more details). Moreover, it is worth to note that strong emotions (in terms of high arousal, with positive/negative valence) that may cause larger effects to the keyboard typing, do not continuously happen in users' daily life. So our suggestion is that it would be more cost-efficient to simply neglect the effect of emotions.

- ***How to detect such an interference?*** Keystroke timing features such as duration, latency, and also the accuracy rate of typing can be used to build

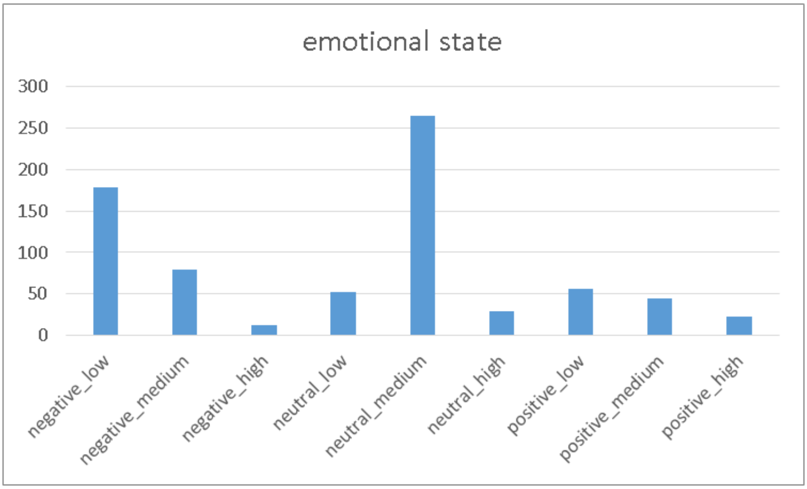

Figure 5.3: The distribution of nine emotional states.

classification models. Machine learning algorithms can be used to build such a classification model and the performance of 70~80% could be reached. By including such a component/function to the original keystroke authentication system, emotional changes can be detected and backup components/mechanisms for authenticating users' identity may be used.

- ***How to control or eliminate such an interference?*** If one wants to control or eliminate the interference of emotions to a keystroke authentication system, we suggest that picture, film or music that allows users to restore neutral and low arousal emotional state can be presented to users' before the authentication process begins. On the other hand, appropriate instructions that are normally used in cognitive experiments that help users to remain calm or return to baseline condition can also be good choices to control the emotional effect if it is required. It is expected that these methods can help in having the users' typing rhythm be maintained in a stable state and improve the accuracy of authentication.

Acknowledgements

This work was fully supported by the Taiwan Ministry of Science and Technology under grant numbers MOST 103-2221-E-009-139. This work was also supported in part by the "Aim for the Top University Plan" of the National Chiao Tung University and Ministry of Education, Taiwan, R.O.C..

References

[1] A. Alhothali. Modeling user affect using interaction events. Master's thesis, University of Waterloo, 2011.

[2] R. Bixler and S. D'Mello. Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits. In *International Conference on Intelligent User Interfaces (IUI)*, pages 225–234, 2013.

[3] P.D. Bolls, A. Lang, and R.F. Potter. The effects of message valence and listener arousal on attention, memory, and facial muscular responses to radio advertisements. *Communication Research*, 28(5):627–651, 2001.

[4] C. Epp, M. Lippold, and R.L. Mandryk. Identifying emotional states using keystroke dynamics. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 715–724, 2011.

[5] M.A. Hall. *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999.

[6] A. Lang, K. Dhillon, and Q. Dong. The effects of emotional arousal and valence on television viewers cognitive capacity and memory. *Journal of Broadcasting & Electronic Media*, 39(3):313–327, 1995.

[7] P.J. Lang. The motivational organization of emotion : Affect-reflex connections. In S.H.M. van Goozen, N.E. Van de Poll, and J.A. Sergeant, editors, *Emotions: Essays on emotion theory*, pages 61–93. Lawrence Erlbaum Associates, 1994.

[8] P.-M. Lee, W.-H. Tsui, and T.-C. Hsiao. The influence of emotion on keyboard typing: an experimental study using visual stimuli. *Biomedical engineering online, 2014. 13(1): p. 81.*, 13:81, 2014.

[9] F. Monrose and A.D. Rubin. Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems*, 16(4):351–359, 2000.

[10] W.-H. Tsui, P. Lee, and T.-C. Hsiao. The effect of emotion on keystroke: an experimental study using facial feedback hypothesis. In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2870–2873, 2013.

[11] L.M. Vizer, L. Zhou, and A. Sears. Automated stress detection using keystroke and linguistic features: an exploratory study. *International Journal of Human-Computer Studies*, 67(10):870–886, 2009.

[12] P. Zimmermann, S. Guttormsen, B. Danuser, and P. Gomez. Affective computing – a rationale for measuring mood with mouse and keyboard. *International Journal of Occupational Safety and Ergonomics*, 9(4):539–551, 2003.