Shared Multi-Keyboard and Bilingual Datasets to Support Keystroke Dynamics Research

Ahmed Anu Wahab, Daqing Hou,
Mahesh Banavar, Stephanie Schuckers
ECE, Clarkson University
Potsdam, NY, USA
{wahabaa,dhou,mbanavar,sschucke}@clarkson.edu

Kenneth Eaton, Jacob Baldwin, Robert Wright
Assured Information Security, Inc
Rome, NY, USA
{eatonk,baldwinj,wrightr}@ainfosec.com

ABSTRACT

Keystroke dynamics has been shown to be a promising method for user authentication based on a user's typing rhythms. Over the years, it has seen increasing applications such as in preventing transaction fraud, account takeovers, and identity theft. However, due to the variable nature of keystroke dynamics, a user's typing patterns may vary on a different keyboard or in a different keyboard language setting, which may affect the system accuracy. In other words, an algorithm modeled with data collected using a mechanical keyboard may perform significantly differently when tested with an ergonomic keyboard. Similarly, an algorithm modeled with data collected in one language may perform significantly differently when tested with another language. Hence, there is a need to study the impact of multiple keyboards and multiple languages on keystroke dynamics performance. This motivated us to develop two free-text keystroke dynamics datasets. The first is a multi-keyboard keystroke dataset comprising of four (4) physical keyboards - mechanical, ergonomic, membrane, and laptop keyboards - and the second is a bilingual keystroke dataset in both English and Chinese languages. Data were collected from a total of 86 participants using a non-intrusive web-based keylogger in a semi-controlled setting. To the best of our knowledge, these are the first multi-keyboard and bilingual keystroke datasets, as well as the data collection software, to be made publicly available for research purposes. The usefulness of our datasets was demonstrated by evaluating the performance of two state-of-the-art free-text algorithms.

CCS CONCEPTS

• **Security and privacy → Biometrics**.

KEYWORDS

keystroke dynamics, datasets, free-text, multi-keyboard, bilingual

ACM Reference Format:
Ahmed Anu Wahab, Daqing Hou,, Mahesh Banavar, Stephanie Schuckers and Kenneth Eaton, Jacob Baldwin, Robert Wright. 2022. Shared Multi-Keyboard and Bilingual Datasets to Support Keystroke Dynamics Research. In *Proceedings of the Twelveth ACM Conference on Data and Application Security and Privacy (CODASPY '22), April 24–27, 2022, Baltimore, MD, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3508398.3511516

CODASPY '22, April 24–27, 2022, Baltimore, MD, USA

ACM ISBN 978-1-4503-9220-4/22/04…$15.00
https://doi.org/10.1145/3508398.3511516

1 INTRODUCTION

As more people rely on their computers and mobile devices to trade, learn, and interact with the rest of the world, the demand for identity verification grows and the need to provide trust, prevent fraud and secure accounts becomes critical. Keystroke dynamics is a promising solution to these needs because it identifies a user based on their typing rhythms, but requires no extra hardware other than the keyboard readily available on a computer. Furthermore, it is passive and non-intrusive. That is, it can run in the background in a frictionless manner without interfering with the user's activities.

Works on keystroke dynamics can be categorized into fixed (static-text) or free (dynamic-text) based on typed content. Sometimes, keystroke dynamics can also be neither completely fixed-text nor free-text but somewhat in the middle, which can be called semi-fixed-text [13].

The continuous monitoring of a user's activities to verify that the user is who they say they are throughout the session is referred to as continuous authentication [8]. Continuous authentication has seen competitive results in terms of high accuracy or low equal error rate (EER) [3, 4, 8, 9], but its ultimate acceptance as a means of authentication in practical applications requires it to be sufficiently robust under various conditions. A particular constraint to its robustness is the lack of publicly shared free-text datasets to study the impact of multiple keyboards and multiple languages in keystroke dynamics. Keystroke dynamics algorithms have been tested with data collected from multiple users but in a keyboard-agnostic way, leaving out the possibility of keyboard and language variety untested. As a result, when deployed, algorithms known to have achieved high accuracy using a particular keyboard type may perform poorly after it is tested on a different keyboard type. The same concern applies to languages as well.

This motivated us to develop two free-text keystroke datasets. The first is a multi-keyboard dataset comprising of four (4) physical keyboards - mechanical, ergonomic, membrane and laptop keyboards - and the second is a bilingual dataset involving both English and Chinese. Data were collected from a total of 86 participants (60 participants for the multi-keyboard dataset and 26 participants for the bilingual dataset) using a web-based keylogger system in a laboratory. Furthermore, to demonstrate the usefulness of our datasets, we evaluated the performance of two state-of-the-art free-text algorithms (Instance-based Tail Area Density (ITAD) Metric [3] and D-Vectors [4]). To the best of our knowledge, our free-text datasets are the first multi-keyboard and bilingual keystroke datasets (with the data collection software) that are publicly available for research.

This paper is organized as follows. Section 2 surveys commonly used free-text datasets. In Section 3, we describe the collection process for the multi-keyboard and bilingual keystroke datasets. Section 4 demonstrates the usefulness and quality of our datasets on two state-of-the-art keystroke dynamics algorithms. The conclusion is given in Section 5.

2 RELATED WORK

There are a few publicly available free-text keystroke dynamics datasets. These include the Clarkson I dataset [12], the Buffalo dataset [11], the Clarkson II dataset [9], the Aalto University dataset [6], the ACB [2], and the Torino dataset [8].

The Clarkson I dataset [12] was collected in a lab setting with the same desktop computer, keyboard and the English language. The data is mostly free text while subjects answered survey questions that were carefully designed around the interest areas of the subject's population to ensure a fluent response. The dataset contains 840K keystrokes and a total of 39 participants. However, all data were collected using a single keyboard type.

The Buffalo dataset [11] is collected by SUNY Buffalo and contains both long fixed text and free text keystrokes data. The keystrokes are collected in the lab from 157 participants who are skillful at using the keyboard. There is a total keystroke count of 2.14M and the average user has 1.44K keystrokes while the max one has 18.3K. Most users have 3 sessions. A subset of the 157 participants used keyboards with different key layouts to input text information.

Unlike other datasets, the Clarkson II dataset [9] is completely uncontrolled. A logger was loaded on the user's own computer and passively recorded participants' keystrokes during their normal behavior without requiring or restricting them to do anything. Hence, users had the choice to use any application of their choice. There are 103 participants in the dataset, with a total keystroke count of 12.9M. The dataset was collected over a span of three years or more. The average user has 125K keystrokes, while the max user has 625K keystrokes. A single keyboard was used per user throughout the collection.

A more recent and large free-text keystrokes dataset is the Aalto University dataset [6]. It was collected from 168,000 participants within three months. Subjects were required to memorize some English sentences and then asked to type them as quickly and accurately as they could. These sentences were randomly selected from a set of 1,525 examples with 3 words and 70 characters maximum. Subjects could type more than 70 characters because there is a chance that participants would forget words or add new characters when typing. The data was collected in an uncontrolled environment. All participants in the database completed each session on either a desktop or laptop keyboard but not both.

There are some works in the literature that study the effect of multi-keyboard on keystroke dynamics performance. Giot et al. [7] had a publicly shared multi-keyboard dataset called the "GREYC" where 100 participants contributed at least 5 sessions of data by typing the passphrase "greyc laboratory". Participants typed the passphrase six times each on a laptop and on a USB keyboard.

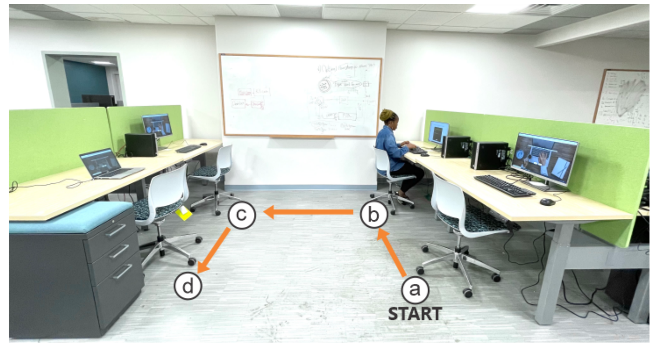

Figure 1: A participant is contributing multi-keyboard data in our laboratory.

The dataset is fixed-text and the password used was imposed unrealistically on participants, whereas in a practical scenario where everyone freely chooses their own unique password.

Bours and Ellingsen [5] had participants perform a task on a laptop keyboard and repeat the same task using a regular desktop keyboard. The task was divided into 3 parts - first, each participant typed his own name 20 times; second, each user typed other participants' names twice; and third, each user typed two texts per keyboard, based on an image that was shown to them. Only 15 participants have contributed data to the study and there are fewer samples contributed per participant. Unfortunately, the dataset is private, so it can not foster research findings on the impact of multi-keyboard on keystroke dynamics.

Our datasets are quite different and unique because they hold data collected from participants who have entered content using four (4) different physical keyboards each, and also in two languages (English and Chinese). Unlike others, our datasets can be used to study the impact of the variation in users' typing patterns across multiple keyboards and two languages. There are a total of 86 participants and 2.35M keystrokes in our datasets - 60 participants contributed 1.66M keystrokes for the multi-keyboard and 26 participants contributed 694K keystrokes for the bilingual dataset.

3 DATA COLLECTION PROCEDURE

As shown in Figure 1, three desktop computers and one laptop were set up in our laboratory for the multi-keyboard study. Connected to the first (a), second (b) and third (c) desktop computers are a mechanical, ergonomic and membrane keyboard, respectively. The fourth (d) is a laptop keyboard.

3.1 The Four Physical Keyboards

Each of the four QWERTY physical keyboards used in this study (mechanical, ergonomic, membrane and laptop as shown in Figure 2), has its own pros and cons in terms of key travel (the total distance from the key at rest to full depression until it hits bottom), force applied, and the noise that the keyboard produces. As demonstrated later in this paper, the size, shape, and position of the keys on a keyboard affect the performance of keystroke dynamics-based authentication systems.

a) Mechanical b) Ergonomics

c) Membrane d) Laptop

Figure 2: The four physical keyboards used in the study a) Mechanical keyboard b) Ergonomic keyboard c) Membrane keyboard d) Laptop keyboard.

3.1.1 Mechanical Keyboard. Mechanical keyboards use individual physical springs and switches to deploy each key. They have a longer key travel, heavier keys that feel sturdier, and provide much more direct feedback to the user with a physical and satisfying sensation when keys are depressed. As a result, they could be noisy to use. They are mostly preferred by typists and gamers. The mechanical keyboard used in this study is a Das keyboard (DASK4MKPROSIL-3G7-r1.5) with the MX-Brown switch (Figure 2a).

3.1.2 Ergonomic Keyboard. Ergonomic keyboards (Figure 2b) are designed to reduce typing strain on the fingers and hands. The keys are laid out at angles that make them comfortably reachable without strain on the fingers, wrist, hands or arms. Most ergonomic keyboards split the letter keys into two halves, rotating the keys to point down toward the lower corners of the keyboard. The rotation allows the arms to approach the keyboard from a more natural angle. They are more expensive than the regular typing keyboards and could take a while to get used to because of the shape and key positioning. We have used the PERIBOARD-512 ergonomic keyboard in this study.

3.1.3 Membrane Keyboard. Membrane keyboards (Figure 2c) are designed with little to no space between the individual keys, and with shorter key travel. Unlike mechanical keyboards which use separate keys that are each attached to individual switches, a membrane keyboard uses a single pressure pad/membrane to register keystrokes. The keys are pressure-sensitive, with each character outlined on a flat surface. In general, membrane keyboards are cheaper and quieter than traditional keyboards and we have used the Logitech MK270.

3.1.4 Laptop Keyboard. Laptops use either the traditional, chiclet or mechanical keyboards, although the chiclet style is becoming the keyboard of choice for thin-and-light laptop makers because it is thinner and more versatile in design. The 14-inch laptop computer (DELL Latitude 5410) used in this study comes with a chicklet style keyboard (Figure 2d). The chiclet keyboard (also known as the island style) is an evolution of the membrane keyboard using the same principle of a single rubber sheet with keys that appear to pop out of the laptop's body through separate cut-outs for each key. The keys usually have flat tops and do not slope off around the edges, creating an impression that there is more space between the keys. This also makes it easier for cleaning the keyboard.

3.2 The Web System Design

A web system was designed for the data collection using Python and the Django and hosted on an Amazon cloud server. The user interface (UI) and user experience (UX) of the system design could affect the way participants type during data collection. Therefore, care was taken to ensure that the pages, texts and buttons are visually appealing, engaging and easy to use - requiring little to no external human guide during the study. The UI of the web system comprises a Signup/Sign-in screen, a keyboards selection screen, a Q/A tasks selection screen and a Q/A screen.

3.2.1 Signup/Sign-in Screen. This screen is for new and returning users to sign up/sign in to the system. A participant's first and last names are required to signup/sign-in. To protect the participants' personal identifiable information, a seven-digit random number known as the "uniqueID" is assigned to every participant at signup.

3.2.2 Keyboard Selection Screen. After a participant signs up or signs in to the web system, they will see the keyboards screen. To reduce human error and interference, this screen was design to enforce the data collection order shown in Figure 1. That is, every participant answers four questions for each of the mechanical, ergonomic, membrane, and laptop keyboards, in this order. Once the four questions are completed for a keyboard, the participant will be automatically logged out the system and reminded to move to the next computer. The participant is not allowed to select a completed keyboard any more. The keyboard completion status is tracked by the system so the participant can resume from where they leave off when switching between the four computers.

3.2.3 Tasks Selection Screen. The tasks screen displays to the participant four (4) question-answer (Q/A) tasks associated with the selected keyboard. Unlike the keyboard selection screen, the participant can choose the next Q/A task for completion in any order. Each task requires the participant to type at least 50 words as an answer. Table 1 shows the list of 16 open-ended Q/A tasks and their assignments per keyboard.

3.2.4 Q/A Screen. This screen takes a user's input to the task questions and logs the keystrokes in the background. The screen has a word count showing the number of words typed. The total typed words must be 50 or more before the the submit button is clickable. To ensure that responses are typed and keystrokes are captured at all times, users are prevented from copying or pasting content into the text area. This is also monitored and enforced by the system.

3.3 Data Collection

Data was collected from a total of 86 participants after signing an IRB-approved consent form - 60 participants for the multi-keyboard in a laboratory environment and 26 participants for the bilingual (English-Chinese), remotely. Both datasets were collected over a time span of 8 months and all participants were compensated for participating in the study.

Table 1: Question prompts per keyboard. Participants answer the same questions per keyboard in both visits.

Keyboards	Question prompts
Mechanical	1. If you are to share your happiest moment, what would it be and why? Ensure to type at least 50 words and reference the date of such moment. 2. Do you prefer traveling by car or plane and where do you prefer to stay when you go on a vacation? Also, Describe the most interesting person you met on one of your travels. 3. Movies are fun, interesting, captivating and entertaining. Which movie would be your best of all time? Write about why you love the movie, if you don't mind watching it all over again and if you could recommend it to someone. 4. Outdoor recreation encompasses a variety of outdoor activities: Backpacking, bicycling, birdwatching, boardsailing/windsurfing, camping, canoeing, rock climbing, fishing, hiking, hunting, kayaking, rafting, running/jogging, sailing, scuba diving, skateboarding, skiing, snowboarding, stand-up paddling, surfing, trail running, wakeboarding and wildlife viewing. Do you like outdoor sports and which outdoor sport are your favorites?
Ergonomic	1. Write about your experience on any of your first days at the Kindergarten, Middle School, High School, or University. How did you feel (exciting, thrilled etc.)? Who did you meet like a friend or a professor or an advisor? What did you enjoy about it (playing games with friends or meeting with advisors etc)? 2. Write about your positive experience with one of your professors, teachers, advisors, counselors. How did they help shape your life? 3. A student new to the Clarkson campus wants to walk to the Sergi's Restaurant on Market Street for dinner, which is near Clarkson Inn. Describe in detail how they can walk from the Cheel parking lot to the restaurant and the things they will see along the way. 4. What is your favorite sports team? Why do you root for this team? What feelings are evoked in you when the team wins? What does it mean to be part of the fandom?
Membrane	1. On a gloomy and rainy day, what do you do to raise your spirits? Do you cook? Do you talk to someone? Do you read a book or watch a movie? 2. What is your go-to comfort food? Describe what aspects of it bring you comfort. Is it the texture, the taste, or the memories the food evokes? Who would you like to share the meal with to make the experience better? 3. Describe a close family member like a mother, father, or sibling. What is their personality? What are their favorite activities? What are your memories of them? 4. Describe a course you have taken in college or high school. What were the major topics that were covered? How has this course helped you in your life or career?
Laptop	1. On a daily basis, you need to go to the store, your job, school or to visit friends. How do you prefer to get around? Do you use multiple methods? What are the advantages and disadvantages of each? Have you lived somewhere where you have used public transportation and what methods did you take? How does this compare to transportation in a college town? 2. How has technology changed over your lifetime? What technology do you spend the most time on? Are there times that you purposefully limit your use of technology? why? 3. Due to Covid-19, our lifestyles have been restricted in many ways. What is the first thing that you will want to do and enjoy, when we go back to our previous normal beings. It can be anything that you wish or desire post covid-19 life. And how will you do it and why would you like to do it? 4. Write about your favorite animal. Why is it your favorite? Do you still own it? How long have you owned it? What was your experience with it?

3.3.1 Multi-Keyboard Data Collection. Each multi-keyboard participant visited our lab twice with each visit estimated to last about an hour, although most participants completed the tasks much earlier. The visits are about a week apart for a large percentage of the participants but only 2 to 3 days apart for some. For the first visit, the participant walked through the four computers starting with the desktop computer hooked up with the mechanical keyboard all the way to the last computer (laptop) (in order shown in Figure 1). Signup is required for all new users. After signup, the user is shown the list of all keyboards with the mechanical keyboard being the only selectable option on the first computer. Upon clicking on the mechanical keyboard option, four Q/A tasks are displayed and the user can choose in any order of his choice. A keylogger captures the user's inputs: key presses, key releases, key code and timestamps with a precision in milliseconds. A prompt to move to the next computer is displayed on the screen when the user completes all the 4 Q/A tasks for a given keyboard. The user is also notified when all Visit 1 tasks are completed. Since the system tracks the user's visit and the completed keyboards, the user can resume from where they left off after switching computers. A total of 16 tasks are required to complete Visit 1. In the second visit, the user repeats the same process and the same set of questions as the first visit.

3.3.2 Bilingual Data Collection. The bilingual participants accessed the data collection web system remotely via a web link from anywhere and on any computer of their choice. Since the goal for the bilingual dataset was to study the effect of language variation on typing patterns, only one computer keyboard is required (one owned by the participant). Participants were screened to ensure their fluency in both English and Chinese language. The same physical keyboard was used to input data for both languages. However, users were required to change the keyboard input language to Chinese in the computer settings before contributing data in Chinese. Participants are required to complete at least 8 tasks in English and

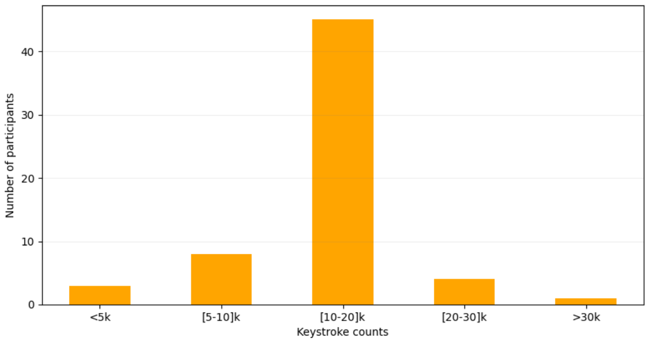

Figure 3: Distribution of total keystrokes contributed per user across all 60 participants for the multi-keyboards dataset.

another 8 tasks in Chinese from Table 1. Most users had completed 16 tasks for both languages.

3.4 Keystroke Logger and Database

The keystroke logger, written in JavaScript, runs continuously in the background and is completely passive and non-intrusive. Captured keystrokes are stored asynchronously to a remote database without interrupting the user's activity. This was achieved with Django Background Task, a database-backed work queue for Django, loosely based around Ruby's DelayedJob library. The keystroke logger captures the user's 7 digits unique ID, key presses, key releases and their respective timestamps in milliseconds precision, task prompt number (referred to as question), selected keyboard and visit.

A remote MySQL database was used to store data. Tables were created for storing users and keystrokes data respectively. Columns in the users table include unique ID (the randomly generated 7 digit number), first name, last name, keyboard sessions (the number of completed keyboards) and visit (the number of completed visits). Note that the shared data have been anonymized and contains no personally identifiable information. The columns in the keystrokes table for the multi-keyboard data includes unique ID, key name, release (0 for press and 1 for release), timestamp, keyboard, visit and question ID. The same keystrokes table was used for the bilingual data, except that the "visit" column was discarded.

4 INITIAL EVALUATION

To demonstrate the usefulness of our datasets for studying the effects of multi-keyboard and multi-language in keystroke dynamics, we replicated two free-text algorithms - ITAD metric [3] and D-Vectors model [4] - on the datasets.

4.1 Instance-based Tail Area Density (ITAD) Metric

The ITAD metric [3] is an instance-based metric that relies solely on the tail area under the probability density function (PDF) of each keystroke dynamics feature (Equation 1).

$$\text{similarity score} = \begin{cases} CDF_{\text{train}}(x), & \text{if } x \leq M_{\text{train}} \\ 1 - CDF_{\text{train}}(x), & \text{if } x > M_{\text{train}} \end{cases} \quad (1)$$

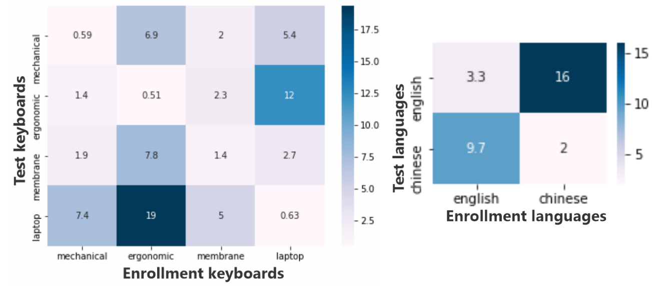

Figure 4: The EERs (%) for same keyboard and cross-keyboard on our multi-keyboard dataset (left) and the EERs (%) for same language and cross-language on our multi-keyboard dataset (right) using ITAD metric.

where CDF_{train} is the empirical cumulative distribution function of each feature in the profile, and M_{train} is the median. The average ITAD metric of all feature instances in a test sample is thresholded to make an authentication decision. The ITAD metric is always between 0 and 0.5 and can be interpreted as a similarity score; the larger it is, the closer the test sample is to the profile.

4.1.1 *Multi-keyboard Experiment and Results:*

For the multi-keyboard dataset, the average participant contributes about 14,000 keystrokes in both visits and across all four keyboards. Figure 3 shows the distribution of total keystrokes contributed per user. Five (5) kinds of features including monograph (the time interval between the press and release of a key) and digraphs (the latency between successive keys such as down-down, up-down, down-up and up-up) were extracted from the raw data. For each participant per keyboard type, all keystrokes (average of 1,800) from the first visit are used for enrollment. Data from the second visit was used for testing. More specifically, each test sample comprises 200 keystrokes. As a result, if a user has an average of K test samples and there are a total of N users, then there will be K genuine test samples and (N-1)*K impostor samples. These scores are averaged across all feature instances. Performance is measured in terms of EER. The EER is calculated per user and then averaged across all users. Figure 4 (left) depicts a heatmap of the EERs across all four keyboards. As shown by each column in Figure 4 (left), cross-keyboard - enrolling and testing with different keyboards - affects keystroke dynamics performance where keyboard size and layout appear to have the biggest impact on users' performance. Second, when the laptop keyboard is used to test against all other keyboard enrollments, it gives the worst performance of a 19% EER. Therefore, we hypothesize that the keyboard size is the major factor at play here. Furthermore, when the ergonomic keyboard is used for enrollment, it performs the worst among all keyboard enrollments. Therefore, we hypothesize that the keyboard layout is the major factor.

4.1.2 *Bilingual Experiment and Results:*

The goal of the bilingual experiment is to study the effect of cross-language in keystroke dynamics. The average participant contributes about 10,000 keystrokes for both English and Chinese language. We extracted monograph and digraphs features from the dataset. For each participant per language, the first two questions (known as prompt 1 and prompt 2 in the dataset) were used for enrollment (average of 900 keystrokes). Data from the remaining questions was used for testing and each test sample comprises 200 keystrokes. As a result, if a user has an average of K test samples and there are a total of N users, then there will be K genuine test samples and (N-1)*K impostor samples. These scores are averaged across all feature instances.

The bilingual experiments results are shown in Figure 4 (right). Our results show that there is a deterioration in performance for cross-language in keystroke dynamics. Enrolling with Chinese data and testing with English results in a performance loss of (16-2)=14%, while enrolling with English data and testing with Chinese results in a performance loss of (9.7-3.3)=6.4%. While further investigation is required, we hypothesized that users could be more familiar with typing in one language than the other even though they are fluent in both languages.

4.2 D-Vectors

The Deep Vectors (D-Vectors) [4] keystroke authentication algorithm is unconventional. Unlike most methods, D-Vectors does not use digraphs that are traditionally used in keystroke dynamics. Instead, the D-Vectors method converts N sequential keystroke events into an image that is fed into a convolutional neural network to featurize the data and learn the projection. The authors argue that this approach enables the model to extract non-intuitive features and reason over arbitrary length sequences of keystrokes which contributes to its increased accuracy. As a deep learning approach, this method is data hungry and takes hours to train. However, once trained this method is computationally inexpensive [4].

For our experiments we divided the datasets by subjects in a 70/30 train test split. 70% of a user's data is used to train the D-Vectors model and then the remaining 30% of the data is used for evaluation purposes. In testing, for the multi-keyboard experiment, we used a profile with 6 samples of 150 keystrokes each (a total of 900 keystrokes) for enrollment while for the bilingual experiment, we used a profile of size 5 (750 keystrokes) for enrollment.

Figure 5 (left) shows the EERs across all four keyboards. Similar to the results from the ITAD metric, performance drops for cross-keyboard compared to using the same keyboard. More so, enrolling with the ergonomic keyboard and testing with the laptop keyboard gave the worst performance (13.6%), an observation that is consistent with the results from the ITAD metric. Figure 5 (right) shows the bilingual experiment results, which is also consistent with the ITAD metric results that the performance of keystroke dynamics deteriorates when enrolling and testing with different languages (cross-language).

Overall, the statistical instance-based ITAD metric results outperformed the D-vectors model because the D-vectors model, like any other deep learning models [1], is data hungry. D-vectors may produce more competitive results when fed with a larger amount of keystroke data.

5 CONCLUSION

The robustness of keystroke dynamics is contingent upon studying the effect of multi-keyboard and multi-language on its performance, so that algorithms or strategies can be developed to deliver stable verification accuracy regardless of the keyboard type or language

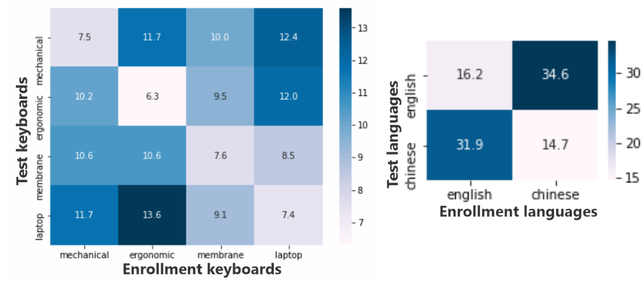

Figure 5: The EERs (%) for same keyboard and cross-keyboard on our multi-keyboard dataset (left) and the EERs (%) for same language and cross-language on our multi-keyboard dataset (right) using the D-Vectors.

used for enrollment and testing. To this end, we developed two novel datasets - multi-keyboard and bilingual datasets - to be used for research purposes. In demonstrating the usefulness of the datasets, we replicated two state-of-the-art algorithms on them. The results of our initial evaluation indicates that cross-keyboard and cross-language do significantly affect keystroke dynamics performance.

Maxion and Commuri [10] found that there are discrepancies between timestamps recorded from a USB keyboard and a laptop keyboard (PS/2). In addition to the keyboard size as the major factor, we intend to further investigate why the laptop, when used to test against all other keyboard enrollments, gave worst performances.

REFERENCES

[1] Alejandro Acien, Aythami Morales, John V Monaco, Ruben Vera-Rodriguez, and Julian Fierrez. 2021. TypeNet: Deep Learning Keystroke Biometrics. *arXiv preprint arXiv:2101.05570* (2021).

[2] Abdullah Alshehri, Frans Coenen, and Danushka Bollegala. 2017. Accurate continuous and non-intrusive user authentication with multivariate keystroke streaming. In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. SCITEPRESS-Science and Technology Publications.

[3] Blaine Ayotte, Mahesh Banavar, Daqing Hou, and Stephanie Schuckers. 2020. Fast free-text authentication via instance-based keystroke dynamics. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 2, 4 (2020), 377–387.

[4] Jacob Baldwin, Ryan Burnham, Andrew Meyer, Robert Dora, and Robert Wright. 2019. Beyond speech: Generalizing d-vectors for biometric verification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 842–849.

[5] Patrick Bours and JØrgen Ellingsen. 2018. Cross keyboard keystroke dynamics. In *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*. IEEE, 1–6.

[6] Vivek Dhakal, Anna Maria Feit, Per Ola Kristensson, and Antti Oulasvirta. 2018. Observations on typing from 136 million keystrokes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[7] Romain Giot, Mohamad El-Abed, Baptiste Hemery, and Christophe Rosenberger. 2011. Unconstrained keystroke dynamics authentication with shared secret. *Computers & security* 30, 6-7 (2011), 427–445.

[8] Daniele Gunetti and Claudia Picardi. 2005. Keystroke analysis of free text. *ACM Transactions on Information and System Security (TISSEC)* 8, 3 (2005), 312–347.

[9] Jiaju Huang, Daqing Hou, Stephanie Schuckers, Timothy Law, and Adam Sherwin. 2017. Benchmarking keystroke authentication algorithms. In *2017 IEEE Workshop on Information Forensics and Security (WIFS)*. IEEE, 1–6.

[10] Roy Maxion and Vrishab Commuri. 2020. This Is Your Behavioral Keystroke Biometric on Rubbish Data. (2020).

[11] Yan Sun, Hayreddin Ceker, and Shambhu Upadhyaya. 2016. Shared keystroke dataset for continuous authentication. In *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 1–6.

[12] Esra Vural, Jiaju Huang, Daqing Hou, and Stephanie Schuckers. 2014. Shared research dataset to support development of keystroke authentication. In *IEEE International joint conference on biometrics*. IEEE, 1–8.

[13] Ahmed Anu Wahab, Daqing Hou, Stephanie Schuckers, and Abbie Barbir. 2021. Utilizing Keystroke Dynamics as Additional Security Measure to Protect Account Recovery Mechanism.. In *ICISSP*. 33–42.