Accurate Combined Keystrokes Detection Using Acoustic Signals

Jian Wang, Rukhsana Ruby, Lu Wang, Kaishun Wu
Guangdong Province Key Laboratory of Popular High Performance Computer
College of Computer Science and Software Engineering, Shenzhen University
wangjiantyrion@gmail.com, rukhsana@ece.ubc.ca, {wanglu,wu}@szu.edu.cn

***Abstract*—With the development of acoustic localization schemes using mobile devices, keystroke detection has received tremendous attention from the academia and industry. Currently, most of the existing systems focus on the recognition of a single keystroke and they are limited by several restrictions. In this paper, we consider the idea of signal variation caused by the combination of two combined keystrokes, and propose an acoustic-based scheme that can detect the combined keystrokes effectively. Our system exploits the blind signal separation technique to deal with the mixed signals, resultant from typing two separate keys simultaneously. Then, we apply feature extraction and pattern recognition algorithms to recognize the combined keystrokes. Extensive experiments have been conducted in a laboratory environment with mobile phones equipped with two microphones. Our results show that for several combinations of two keystrokes, on average, we can achieve 78.4% recognition accuracy.**

***Keywords*-Combined keystrokes detection, acoustic recognition, blind signal separation, smartphone.**

I. Introduction

Over the last few decades, keystrokes detection has become a very active topic due to the emergence of acoustic-based localization and eavesdrop-based techniques. These techniques enable users to obtain the keystrokes content with the help of mobile devices. With the improvement of hardware quality, mobile devices have been widely applied in localization problems to achieve centimeter-level accuracy, which are useful to solve keystrokes detection problems. Although eavesdropping helps in designing keystrokes detection techniques, according to the principle of information security, a proper keystrokes detection mechanism can be applied to prevent eavesdropping further effectively.

Consequently, in response to this research trend, many keystrokes detection methods have been proposed that take advantages of different technologies. In general, all existing approaches can be classified into WiFi-based, vision-based and acoustic-based categories. WiFi-based methods [1],[2] use commercial Access Point (AP) and Network Interface Controllers (NICs) to detect gesture when one person keeps pressing on the keyboard. These methods mainly detect the keystrokes contact through the change of wireless channels. Moreover, these methods require the system to have stabilized experimentation scenario, and the layout of the laboratory can affect the results further. Vision-based methods adopt cameras and computer vision techniques to recognize keystrokes. The shortcoming of these methods is, they are limited by light conditions.

Acoustic-based methods in [3],[4] and [5] provide a new direction to the keystroke detection problem, which collect the sound signals of keystrokes with high sensitivity. Some features and methods, such as Amplitude Spectrum Density (ASD) [6], Mel Frequency Cepstral Coefficient (MFCC) and Time Difference of Arrival (TDOA) [7] are used to distinguish different keystrokes. However, the aforementioned existing methods can only detect a single keystroke, and none of them focused on the combined keystrokes. In view of these limitations, initially, we conducted some experiments to verify the suitability of the existing methods in the detection of combined keystrokes. Basically, these experiments prompted us towards the direction of this research.

Now, in this context, one may ask a crucial question, i.e., *can we detect the combined keystroke using acoustic signals?* In a Blind Signal Separation (BSS)-based system, the separated signals must have non-Gaussian distribution, and must be relatively independent. Consequently, we propose an effective strategy to obtain independent signals separately. Thanks to the recognizable features in sound signals, we can transform the localization problem to the classification one, and the resultant classifier can detect the combined keystrokes effectively.

To recognize the combined keystrokes efficiently, there are still two challenges. *First, it is hard to extract the raw acoustic signals of keystrokes when two keys are pressed simultaneously.* The microphones receive mixed signals resultant from the typing of two keys, and most signal fragments are overlapped with each other. *Second, it is difficult to classify different combined keystrokes based on some representative feature(s).* The core of this problem is a localization problem in the near field. Traditional angle-based localization methods cannot solve this problem efficiently as some captured signal waveforms are similar when different keys are typed. Moreover, it is also challenging to select suitable features for classification.

To the best of our knowledge, this is the first signal separation-based approach to recover the combined keystrokes. The main contributions of this paper are listed as follows.

- We have utilized acoustic signals from mobile phones to detect the combined keystrokes when two keys are pressed simultaneously. It is the first known attempt to achieve a fine-gained combined keystrokes recognition.
- We have employed BSS and Independent Component Analysis (ICA) techniques to separate mixed signals. Feature matching and machine learning algorithms are used for classifying the combined keystrokes.

978-1-5090-5696-5/16 $31.00 © 2016 IEEE
DOI 10.1109/MSN.2016.9

- We have conducted a comprehensive experimentation in a laboratory environment to evaluate our proposed system. The outcome of the experimentation has proved that smartphones equipped with our proposed scheme can detect combined keystrokes effectively. Furthermore, the average recognition accuracy of the proposed scheme is 78.4%.

The rest of the paper is organized as follows. A brief overview of the related work is provided in Section 2. We describe the detailed methodology of the proposed scheme in Section 3. Followed by the experimentation setup, in Section 4, we evaluate the performance of our system. Finally, Section 5 concludes the paper with some future direction.

II. Related Work

In this section, we briefly review the related work on keystroke detection and acoustic localization techniques, and discuss their shortcomings in the context of this work.

In [3], Zhuang et al. showed that it is possible to recognize users' keystrokes by differentiating the sound emanation. They used unsupervised learning technique while taking cepstrum features into account to solve the problem. Berger et al. [8] proposed a dictionary-based method to reconstruct every single word. Ubik [4] leveraged Amplitude Spectrum Sensitivity (ASS) to localize the keystrokes on the conventional surface. Every solid surface has different attribute. Consequently, to realize multi-scenario localization problem, cumbersome training is required. Zhu et al. [5] utilized smartphones to record sound signal from keystrokes. The TDOA method was applied to calculate the relative positions. This approach was based on geometry. Comparing to the aforementioned works, snooping keystroke [7] is an example of a keystroke eavesdropping system without training, linguistic model and multiple devices. The keys on keyboard are divide into several groups by the TDOA method. Then, the acoustic feature, MFCC distinguish them further in groups. Unlike all these works, our research focuses on locating keystrokes when a user clicks two keys simultaneously.

There are some recent works on recognizing keystrokes with the aid of wireless signals [9],[10],[11]. These works basically are based on Channel State Information (CSI) of wireless signals. Chen et al. [1] made the case for a keystroke detection mechanism by combining changes in the wireless channel and interference cancellation. [2] has exploited WiFi devices to find the connection between keystrokes and CSI. These solutions require an experimentation environment which is not interfered by other factors. Our design exploits the sound signal captured by mobilephones, the detection of which is more sensitive.

Clearshot et al. [12] presented a novel approach to recover text typed on keyboards in a video. Computer vision technique and language model are fused to eavesdrop keystrokes. The ways how to compromise electromagnetic emanations of keystrokes have been discussed by [13]. The authors in this paper analyzed the entire spectrum and computed

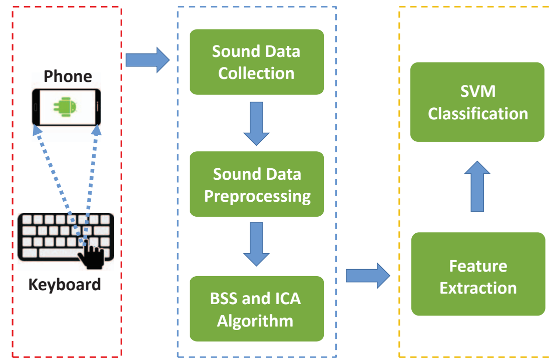


Figure 1: System block diagram.

Short Time Fourier Transform (STFT) of the signal. Then, a trigger model was defined to scan continuous radiation. Marquardt et al. [14] demonstrated that the accelerometer of mobilephones can be used to reconstruct the keystrokes. They also developed a classier to classify keystroke events using a neural network. Wang et al. [15] devised a system that can reveal a user's private PIN sequence by utilizing embedded sensor in wearable devices.

Relevant to our proposed system, there are some existing works on acoustic localization, which perform accurate location sensing using off-the-shelf audio hardware. BeepBeep [16] employed time-of-arrival (TOA) metric of the acoustic signal to calculate the distance. Using sound waves, and then employing TOA and ASS techniques, Phone-to-Phone ranging system [17] has a granular localization accuracy. Other acoustic localization-based works include [18] and [19]. These works achieve indoor localization with smartphones without taking any infrastructure support into account.

III. Methodology

In this section, we first provide an overview of the mechanism for detecting combined keystrokes along with design challenges. Detailed components, including data collection, blind source separation, localization and classification algorithms are discussed to see how to address these challenges.

A. Overview and design challenges

To begin with, we introduce two necessary assumptions which are as follows. First, near-field model of microphone arrays is different from the far-field model. In our experimentation setup, the sound signals are regarded as spherical wave and the microphone arrays are almost in a straight line. Second, sound is a baseband and non-stationary signal. Hence, the traditional direction of arrival (DOA)-based algorithms do not work in this case.

The basic idea is straightforward although there remain several challenges in the case of implementation. First, when two keys are typed simultaneously, the microphones collect the mixed signals. The independent components of the mixed signals need to be separated without distortion. Second, how to reveal the signal of each key accurately

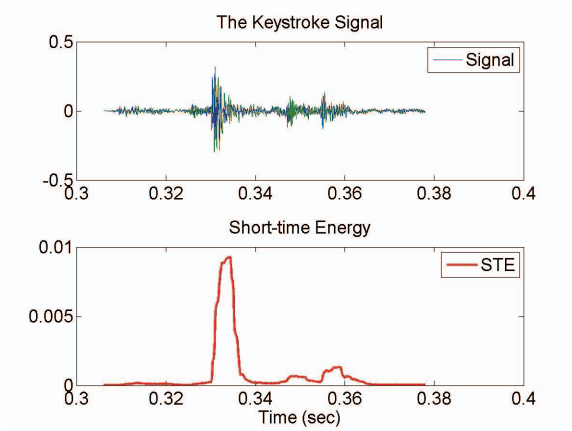


Figure 2: An illustration of the signal segmentation method, that includes received signals by two microphones (top) and short-time energy of one microphone (bottom).

after the separation. Third, with the assistance of positional information, we need to design an efficient location tracking algorithm to improve the accuracy of the sound source localization problem in the near filed.

Fig. 1 demonstrates the block diagram of our system. There are five functional blocks in this arena for sound data collection, sound data processing, BSS and ICA algorithms, features extraction and SVM classification. Among these, two components are crucial to address the aforementioned challenges. BSS and ICA algorithms separate the mixed signals. Features extraction scheme exploits location information from the independent signal components.

B. Sound signal preprocessing

1) Segmentation and filter: We conduct an experiment accompanied with five combined keystrokes. As shown in [3], a keystroke signal has three peaks: *touch peak*, *hit peak* and *release peak*. The *touch peak* and *hit peak* are near and cannot be separated clearly. In order to extract the whole keystroke segment, we introduce sound energy to find the start and end points of the keystroke.

Particularly, we calculate the energy $E_i(t)$ of the signal sequence $x_i(t)$ and the corresponding accumulated energy $A_i(t)$ is given by

$$E_i(t) = kx_i(t)^2, \tag{1}$$

$$A_i(t) = \sum_{n=t}^{t+w} E_i(n), \tag{2}$$

where w is the length of the sliding time window. We set the threshold γ to determine the start and end points. The bottom part of Fig. 2 shows the energy of keystrokes received from one microphone.

To remove some high frequency noise and ensure high quality feature, we adopt *Butterworth* bandpass filter to cutoff the signal sequence. We observed the frequencies of the cherry keyboard keystrokes, and found that the range lies approximately between $100Hz$ and 1000Hz. The sample rate is 48kHz and we set the cutoff frequency as $w_c = [\frac{2\pi*100}{48000}, \frac{2\pi*1000}{48000}] \approx [0.013, 0.13]$rad/s.

2) Noise removal: Since the signals are affected by the hardware and the variability of the indoor environment, the sound signal can be the raw data with some random noise. We adopt a wavelet-based denoising scheme to smooth the raw data. It has better performance compared to the fourier-based denoising method in time and frequency domains. In the process, we choose soft thresholding mechanism and *sym8* wavelet basic. Besides, we can get the threshold from the SureShrink strategy, which is a technique of selecting a threshold by minimizing Stein's unbiased risk estimator and it is given by

$$\lambda_j^S = \arg\min_{\lambda}[SURE^S(\lambda, d_j)], \tag{3}$$

where $SURE^S(\lambda, d_j)$ is the threshold function for the Stein's unbiased estimator risk and d_j represents the detail coefficient at level j of the decomposed signal.

C. BSS and ICA algorithms

Once the sound data is collected, the signals in each microphone are mixed. The signals should be separated, which plays an important role in the overall detection process. BSS is the technique that separates a set of original source signals from a set of mixed signals with the aid of some information. Fig. 3 shows an illustration of the BSS technique.

The original source signal values $[s_1, s_2, ..., s_n]^T$, e.g., the individual sound in a cocktail party, are some samples of these random variables. If we observe *n* linear mixtures x_1,...,x_n of *n* independent components, mixture x_j can be given by

$$x_j = a_{j1}s_1 + a_{j2}s_2 + ... + a_{jn}s_n. \tag{4}$$

In the ICA model, we assume that each mixture x_j as well as each independent component s_k is a random variable. In the subsequent discussion, we use vector and matrix notations to represent multiple random variables. Consequently, we denote the random vector variable whose elements are the mixture variables x_1,...,x_n by **X**, and likewise the random vector with elements $s_1, s_2, ..., s_n$ by **S**. Furthermore, let us denote a matrix with elements a_{ij} by **A**. Hence, the mixture model can be briefly expressed by

$$\mathbf{X} = \mathbf{AS}. \tag{5}$$

In the above statistical model, the independent components and the mixture matrix are assumed to be unknown. After estimating matrix **A**, we can compute its inverse form, and it is denoted by **W**. Then, each independent component is simply given by

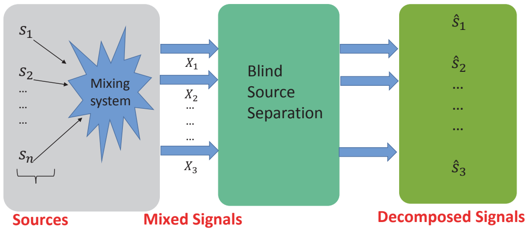


Figure 3: An illustration of the BSS method, including sources, mixed signals and decomposed signals.

$$s_j^{(i)} = w_j^T x_i. \tag{6}$$

The basic idea of ICA algorithm is, for the observed $[x_1, x_2, ..., x_n]^T$ mixtures, the optimal values of the parameters $\mathbf{W}$ are computed in an iterative manner. The detailed breakdown of this algorithm is as follows.

Step 1: To make the values of $\mathbf{x}$ centered, the mean value of the vector $\mathbf{x}$ is subtracted from its elements.

Step 2: In order to transform the resultant observed vector linearly, the whitening transformation is necessary. The method of this transformation is to use the eigen value decomposition (EVD) of the covariance matrix $E\{\mathbf{xx^T}\} = \mathbf{EDE^T}$. The whitening process can be denoted by

$$\tilde{\mathbf{x}} = \mathbf{ED}^{-\frac{1}{2}}\mathbf{E^T x}, \tag{7}$$

where $\mathbf{D}^{-\frac{1}{2}} = \text{diag}(d_1^{-\frac{1}{2}}, ..., d_n^{-\frac{1}{2}})$. Besides, we need to ensure that the covariance matrix of $\tilde{\mathbf{x}}$ satisfies the relation in $E\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\right] = \mathbf{I}$.

Step 3: According to the Karush-Kuhn-Tucker (KKT) conditions, the optimal value of $E\{G(\mathbf{W}^T\mathbf{x})\}$ obeys the following formula.

$$E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - \beta\mathbf{w} = 0, \tag{8}$$

where $E\{(\mathbf{w^T x})^2\} = \|\mathbf{w}\|^2 = 1$. $G(\mathbf{W}^T\mathbf{x})$ is some non-quadratic function with parameter $\mathbf{W}^T\mathbf{x}$. Let denote the derivative of nonquadratic function $G(.)$ by $g(.)$. We use the Newton method to solve this equation. The result is a Jacobian matrix $JF(\mathbf{w})$, and is given by

$$JF(\mathbf{w}) = E\{\mathbf{xx^T}g'(\mathbf{w}^T\mathbf{x})\} - \beta\mathbf{I}. \tag{9}$$

We transform the Jacobian matrix to the diagonal one, and then invert it. After algebraic simplification, we obtain the following equation

$$\mathbf{w}^+ = E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - E\{g'(\mathbf{w}^T\mathbf{x})\}\mathbf{w}. \tag{10}$$

Step 4: Before initiating the iterations, let $\mathbf{w} = \mathbf{w}^+/\|\mathbf{w}^+\|$. Then, give an initial random weight vector $\mathbf{w}$, if the resultant solution does not converge, we repeat *Step 4*.

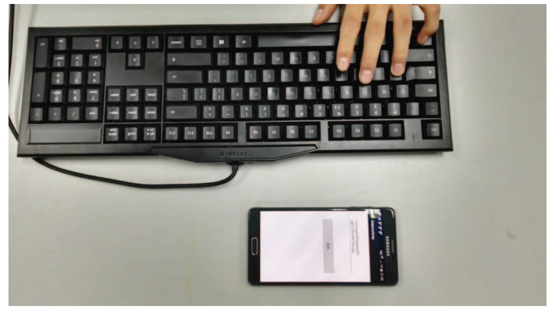


Figure 4: A sample experimental setup in a laboratory environment.

In our system, we adopt the above algorithm to obtain the approximate signal information of the keystrokes. This transformation does not affect the quality of the feature, which is verified in Section IV.

D. Feature extraction and matching

Following the procedure in the previous subsection, we obtain the separated signal data, which are generated from different combined keystrokes. Now, we want to extract the feature information in these data that can distinguish different keystrokes. After a thorough experimentation, we choose MFCC feature, that is crucial in distinguishing different keystrokes. MFCC is a feature, which is widely used in automatic speech and speaker recognition systems. Fig. 5 shows the outcome of the feature extraction method for one sample combined keystrokes. In our implementation, we consider the number of filterbank channels as 20, and the cepstral coefficient is set to 12.

We have extracted several group features of different keys. Then, we need to choose an effective classifier to recognize keystrokes. In particular, we use the Support Vector Machine (SVM) to classify different keystrokes, which is a supervised learning model. In supervised learning, one is given a set of training examples $X_1 ... X_N$ with labels $y_1 ... y_N$ to predict y_{N+1} from the given X_{N+1}. The non-linear SVM model can formally be described as

$$\min_{w} \frac{\|\mathbf{w}\|^2}{2}, \tag{11}$$

subject to $y_i(\mathbf{w}.(\Phi(\mathbf{x}_i) + b) \geq 1, i = 1, 2, ..., N$. Here, $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{id})^T$ is the corresponding attributes of the ith sample and $\mathbf{w}$ is the parameter of the SVM model. In order to obtain the optimal clustering results, we adopt the dual-based optimization technique. And, the corresponding Lagrangian function is given by

$$L_D = \sum_{n=1}^{N} \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j), \tag{12}$$

where λ_i is the Lagrange multiplier. After solving the above equation, we can obtain optimal λ_i $\mathbf{w}$ and b. The radial basic function (RBF) $\Phi(\mathbf{x}_i)$ is chosen to transform

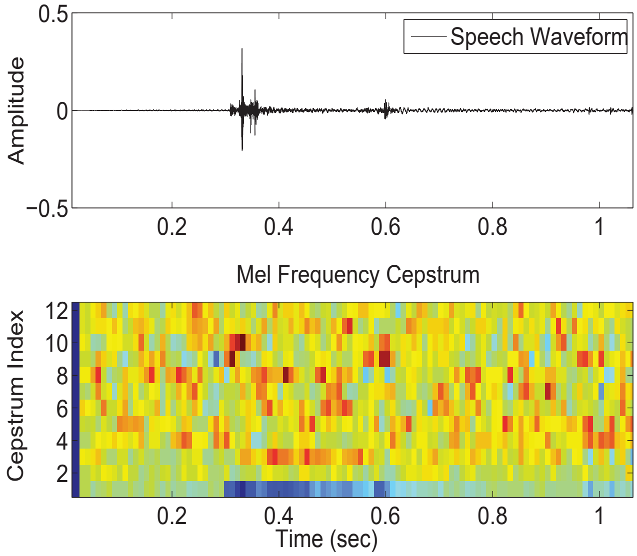
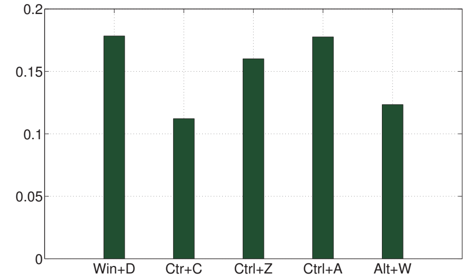
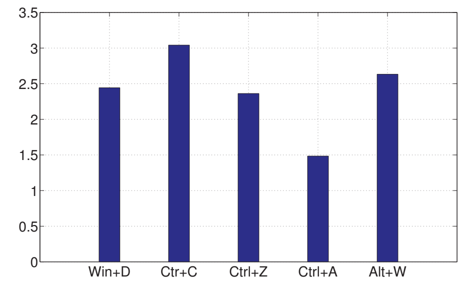


Figure 5: A sample outcome of the feature extraction procedure.



(a) Average RMSE



(b) Average SNR

Figure 6: Performance of the signal separation method. (a) the RMSE between the raw signals and the separated signals, (b) the SNR of the separated signals.

the samples from lower dimensional space to the higher dimensional one. Due to the space constraint, the detailed feature extraction mechanism including the way to obtain the optimal parameters are omitted in this paper.

IV. Experiments and Performance Evaluation

In this section, we first describe the details of the experimental settings and the implementation of our system. Then, to justify the effectivity of our system, we show the results of our experimentation.

A. Experimental setup

1) Software and hardware: In our experiment, sound signals are collected by Samsung Galaxy Note 4. The operating system of this phone is Android 5.1.1. There are three microphones on the edge of this phone. We have developed a Android application to drive the microphones and record the keystroke signals. The experimenters are taken with the keyboard Cherry G80-3800, the dimension of which is 7×8 square centimeter. The keystroke sound is much louder than the random noise in the lab. Besides, the interval of the keystroke sound is shorter than that of regular keyboards. After collecting data, we have utilized MATLAB (R2014a) to analyze raw data.

2) Experimental scenario: We have conducted thorough experiments on computer desks in a quiet laboratory room with some random noise. The smartphone uses 48kHz for audio recoding, and it is placed in front of the keyboard. As Fig. 4 shows, the distance is 10 centimeter between the bottom edge of phone and the keyboard.

B. Experimental Results

In this section, we evaluate the overall performance of our proposed combined keystrokes recognition system. Specifically, we have used a matlab toolbox to measure the performance of blind source separation algorithms. The results are shown in two bar charts. Furthermore, we use a form to evaluate the performance of keystrokes recognition. In the form, each column represents the instances of a predicted class, and each row represents the instances of an actual class.

1) Separation performance: We evaluate the performance of the signal separation algorithm in terms of Root Mean Squared Error (RMSE) and Signal-to-Noise Ratio (SNR). The SNR metric measures the quality of the estimated signal compared to the quality of the original signal. And, the RMSE is a frequently used measure, which is the difference between a value predicted by a model or an estimator and the actually observed value. The average RMSE and SNR of the collected data are presented in Fig. 6 for different combinations of keystrokes. The results verify the validity of the signal separation algorithm.

2) Recognition performance: Fig. 7 shows the form of the combined keystrokes recognition results. The average recognition accuracy of five combination of keystrokes is about 78.4%. The presented results verify that we can obtain an acceptable level of accuracy in recognizing the combined keystrokes by using our proposed system.

V. Conclusions and Future Work

In this paper, we proposed an acoustic-based system that can recognize combined keystrokes. The expected function-

	Win+D	Ctr+A	Ctr+C	Ctr+Z	Alt+W
Win+D	67.0%	1.5%	4.0%	7.0%	20.5%
Ctr+A	11.0%	86.0%	1.0%	1.5%	0.5%
Ctr+C	9.0%	2.0%	83.0%	3.0%	3.0%
Ctr+Z	8.0%	4.0%	2.0%	83.0%	3.0%
Alt+W	5.0%	6.5%	10.0%	5.5%	73.0%

Figure 7: The final results of different combination of keystrokes.

ality of the system is achieved by analyzing mixed acoustic signals which are collected by smartphones equipped with two microphones. Followed by the verification of the feasibility of the signal separation mechanism, our work outperforms other existing works with a key property that it has relatively higher accuracy in recognizing combined keystrokes. During our test experiments, on average, we achieved 78.4% recognition accuracy.

Current system can only recognize a limited number of combined keystrokes. In our future work, we intend to recognize more different combination of keystrokes. We also plan to bring the microphone array in the near field to improve the recognition accuracy of our current system. Moreover, since the experimentation scenario is a classic near-field model, it can employ the improved time and angle estimation procedure to locate the source position. However, before that, we need to verify its feasibility further. Besides, based on our previous works [20],[21],[22], we plan to exploit the data rate of WiFi networks in detecting combined keystrokes. We would like to leave these more interesting and challenging works as our future study.

VI. Acknowledgment

This current research is supported in part by the Shenzhen Science and Technology Foundation KQCX20150324160536457,JCYJ20150324141711621,China NSFC Grant 61472259, 61502313, Program for New Century Excellent Talents in University (NCET-13-0908), Guangdong Talent Project 2014TQ01X238, 2015TX01X111 and GDUPS (2015), Tencent "Rhinoceros Birds"-Scientific Research Foundation for Young Teachers of Shenzhen University.

References

[1] B. Chen, V. Yenamandra, and K. Srinivasan, "Tracking Keystrokes Using Wireless Signals," *Proc. ACM Mobisys*, pp. 31–44, 2015.

[2] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke Recognition Using WiFi Signals," *Proc. ACM Mobicom*, pp. 90–102, 2015.

[3] L. Zhuang, F. Zhou, and J. D. Tygar, "Keyboard acoustic emanations revisited," *Proc. ACM CCS*, p. 373, 2005.

[4] J. Wang, K. Zhao, X. Zhang, and C. Peng, "Ubiquitous keyboard for small mobile devices: harnessing multipath fading for fine-grained keystroke localization," *Proc. ACM Mobisys*, pp. 14–27, 2014.

[5] T. Zhu, Q. Ma, S. Zhang, and Y. Liu, "Context-free Attacks Using Keyboard Acoustic Emanations," *Proc. ACM CCS*, pp. 453–464, 2014.

[6] S. P. Tarzia and R. P. Dick, "Indoor Localization without Infrastructure using the Acoustic Background Spectrum," *Proc. ACM Mobisys*, pp. 155–168, 2011.

[7] J. Liu, Y. Wang, G. Kar, Y. Chen, J. Yang, and M. Gruteser, "Snooping Keystrokes with mm-level Audio Ranging on a Single Phone," *Proc. ACM Mobicom*, pp. 142–154, 2015.

[8] Y. Berger, A. Wool, and A. Yeredor, "Dictionary attacks using keyboard acoustic emanations," *Proc. ACM CCS*, pp. 245–254, 2006.

[9] K. Wu, H. Tan, Y. Liu, J. Zhang, Q. Zhang, and L. M. Ni, "Side channel: Bits over interference," *IEEE Transactions on Mobile Computing*, vol. 11, no. 8, pp. 13–24, 2010.

[10] K. Wu, H. Tan, H. Ngan, Y. Liu, and L. M. Ni, "Chip error pattern analysis in ieee 802.15. 4," *IEEE Transactions on Mobile Computing*, vol. 11, no. 4, pp. 543–552, 2012.

[11] K. Wu, H. Li, L. Wang, Y. Yi, Y. Liu, Q. Zhang, and L. M. Ni, "Hjam: Attachment transmission in wlans," *IEEE Transactions on Mobile Computing*, vol. 12, no. 12, pp. 2334–2345, 2013.

[12] D. Balzarotti, M. Cova, and G. Vigna, "ClearShot: Eavesdropping on keyboard input from video," *Proc. IEEE Symposium on Security and Privacy*, pp. 170–183, 2008.

[13] M. Vuagnoux and S. Pasini, "Compromising electromagnetic emanations of wired and wireless keyboards," *Proc. ACM USENIX*, no. Lcd, pp. 1–16, 2009.

[14] P. Marquardt, A. Verma, H. Carter, and P. Traynor, "(sp) iPhone: decoding vibrations from nearby keyboards using mobile phone accelerometers," *Proc. ACM CCS*, pp. 551–562, 2011.

[15] C. Wang, X. Guo, Y. Wang, Y. Chen, and B. Liu, "Friend or Foe ? Your Wearable Devices Reveal Your Personal PIN," *Proc. of the 11th ACM on Asia Conference on Computer and Communications Security*, pp. 189–200, 2016.

[16] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "Beepbeep: a high accuracy acoustic ranging system using cots mobile devices," *Proc.of the international conference on Embedded networked sensor systems*, pp. 1–14, 2007.

[17] U. C. S. Barbara, D. Chu, X. Chen, and T. Moscibroda, "SwordFight : Enabling a New Class of Phone-to-Phone Action Games on Commodity Phones Categories and Subject Descriptors," *Proc.of the 10th international conference on Mobile systems, applications, and services*, pp. 1–14, 2012.

[18] J. Scott and B. Dragovic, "Audio location: Accurate low-cost location sensing," *Proc.of International Conference on Pervasive Computing*, pp. 1–18, 2005.

[19] Y.-c. Tung and K. G. Shin, "EchoTag : Accurate Infrastructure-Free Indoor Location Tagging with Smartphones," *Proc. ACM Mobicom*, pp. 525–536, 2015.

[20] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni, "Csi-based indoor localization," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 7, pp. 1300–1309, July 2013.

[21] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with wi-fi!" *IEEE Transactions on Mobile Computing*, vol. 15, no. 11, pp. 2907–2920, 2016.

[22] L. Wang, X. Qi, J. Xiao, and K. Wu, "Exploring smart pilot for wireless rate adaptation," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4571–4582, 2016.