Person Authentication Using Finger Snapping — A New Biometric Trait

Conference Paper *in* Lecture Notes in Computer Science · October 2016

DOI: 10.1007/978-3-319-46654-5_84

CITATIONS
2

READS
344

4 authors:

Yanni Yang
The Hong Kong Polytechnic University
11 PUBLICATIONS **57** CITATIONS
SEE PROFILE

Feng Hong
Ocean University of China
91 PUBLICATIONS **856** CITATIONS
SEE PROFILE

Yongtuo Zhang
UNSW Sydney
6 PUBLICATIONS **72** CITATIONS
SEE PROFILE

Zhongwen Guo
Ocean University of China
112 PUBLICATIONS **921** CITATIONS
SEE PROFILE

Person Authentication Using Finger Snapping — A New Biometric Trait

Yanni Yang[1], Feng Hong[1(✉)], Yongtuo Zhang[2], and Zhongwen Guo[1]

[1] Department of Computer Science and Technology, Ocean University of China, Qingdao, China
hongfeng@ouc.edu.cn

[2] School of Computer Science and Engineering, University of New South Wales, Sydney, Australia

Abstract. This paper presents a new biometric trait, finger snapping, which can be applied for person authentication. We extract a set of features from finger snapping traces according to time and frequency domain analysis. A prototype is developed on Android smartphones to realize authentication for users. We collect 6160 snapping traces from 22 subjects for continuous 7 days and 324 traces from 54 volunteers across three weeks. Extensive experiments confirm the measurability, permanence, uniqueness, circumvention, universality and acceptability of the finger snapping to realize biometrics based authentication. It shows that the system achieves 6.1 % average False Rejection Rate (FRR) and 5.9 % average False Acceptance Rate (FAR).

Keywords: Finger snapping · Biometric trait · DTW · Smart device

1 Introduction

The security of smart devices has been a major concern for people nowadays. For example, a range of methods have been applied for user authentication on smartphones and smart watches, such as password, PIN and fingerprint [1]. They can be either easily stolen by attackers or need extra sensors for input. In this paper, a new biometric trait, finger snapping, is applied for person authentication. The sound of finger snapping is easy to capture with the microphone embedded in the smart devices. Besides, it is easy to perform and do not require explicit remembrance for the reason that finger snapping only depends on muscle memory.

Finger snapping is an act of making an impulsive sound with one's fingers and palm [2]. It is often done by connecting the thumb with another (middle, index or ring) finger, and then moving the other finger immediately downward to hit the palm. Such act of finger snapping involves physiological characteristics which refer to inherited traits that are related to human body, as the sound of finger snapping is differentiated by the size of palm and skin texture. In addition, it also involves behavioral characteristics which refer to learned pattern of a person, as it is the movement of the finger creates the sound.

Z. You et al. (Eds.): CCBR 2016, LNCS 9967, pp. 765–774, 2016.
DOI: 10.1007/978-3-319-46654-5_84

Intuitively, the original time series of finger snapping sound is applied to authenticate users. However, through experiments in Sect. 2, the average FAR of this method is 45 %, which is relatively high for person authentication purpose. So we concentrate our research to the direction of *finding the unique features contained in the finger snapping sound for user authentication*. This problem is solved by investigating features which represent the physiological and behavioral characteristics of finger snapping from the original snapping sounds. On the one hand, time domain features of finger snapping sound are explored, since it is a motion during which fingers hit the palm. On the other hand, we regard the person's palm as a musical instrument and try to locate features in frequency domain. Overall, we extract zero-crossing rate, and root mean square as time domain features, and the Mel-Frequency Cepstral Coefficiets (MFCC), spectrum power and spectral centroid as frequency domain features. Finally, Dynamic Time Warping (DTW) [3] is applied to realize authentication. Through the evaluation phase, we verify that the finger snapping, as a biometric trait, can meet the requirements of a combination of several factors including measurability, permanence, uniqueness, circumvention, universality and acceptability [4].

2 Related Work

Biometric traits have been widely used in user authentication on smart devices, like fingerprint which requires an extra scanner to input the trait. In terms of the biometric trait, it generally falls into two categories: physiological and behavioral characteristics. Physiological characteristics are related to the shape of the body, such as face [5], fingerprint [1], iris [6], etc. Physiological characteristics rely on biological features that users uniquely have, but often require special sensors like extra sensors or user engagement. For iris and retina based identification, bespoke special sensors and making eyes close to certain scanner are both required, which increase the cost and inconvenience for users.

Behavioral characteristics take the implicit patterns of user behavior for authentication, including but not limited to typing rhythm, gait [7], voice [8] and motion [9]. For example, motion based authentication utilize the way of user waving smartphones [9] which calls for the user to remember the fixed gesture and try to perform the same gestures all the time, restricting accuracy and convenience. Another type of behavioral characteristics is voice, which is also an authentication method using sound traces like finger snapping. Many acoustic features and recognition algorithms have been put forward, giving us clues on finger snapping sound processing.

Compared with the above biometrics, finger snapping contains both physiological and behavioral characteristics, and it requires no special sensors and explicit remembrance as the finger snapping depends only on muscle memory.

3 Constitution and Detection of Finger Snapping

This section first introduces the constitution of finger snapping sound and explains the motivation why it can be taken as a biometric trait. Then we describe our data

collection process and bring forward the method to detect and extract the raw snapping traces from the whole snapping sound files. At last, we give the authentication results on raw time series comparison to illustrate the importance of feature extraction.

3.1 Constitution of Finger Snapping

The finger snapping sound is made of three parts [2]: (1) The friction sound between the middle finger and the thumb which is weak and unnoticeable. (2) The impact sound made by the middle finger colliding with the cavity formed by contacting the ring finger with the palm. (3) The pop sound created by the fast compression and the pursuant decompression of air. The pop sound is the clearest sound among the three parts as it is caused by a compression of air between the fast moving middle finger and the cavity formed by the palm and middle finger.

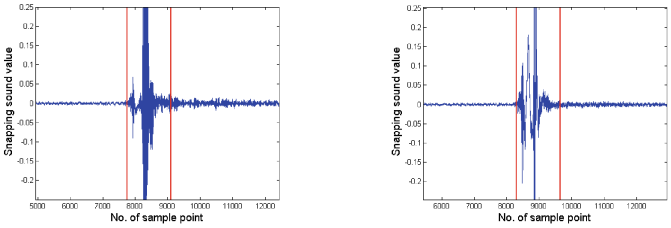
The snapping sound is differentiated by the skin texture and the cavity created by palm and finger which can be categorized as physiological characteristics. In addition, the finger movement also has an impact on the snapping sound, with different strength and speed, which can be related to behavioral characteristics. So finger snapping is a combination of both physiological and behavioral characteristics.

3.2 Collection of Finger Snapping

We collect the finger snapping traces on a commercial Android based smartphone, HUAWEI Honor 7. An application is developed on the smartphone to collect snapping sounds and interact with the user to tell whether he is the owner or not. Two datasets are collected: Training Set and Testing Set. Training Set consists of 22 subjects' 6160 snapping sounds across continuous 7 days. These subjects' ages are from 19 to 39, including 14 males and 8 females. Testing Set consists of 54 volunteers' 324 snapping traces, which are collected when they are doing the attacking experiment on the smartphone. We collect each volunteer's finger snapping traces every Tuesday and Thursday for three weeks. Collection is carried out in two common laboratories.

3.3 Finger Snapping Detection

Finger snapping detection is to locate the exact proportion of snapping sound in the whole file. Through our empirical observation, it is found that the time span of the exact finger snapping is generally below 3 ms. Hence, with the sampling rate of 44.1 kHz, the exact snapping sound is within 1350 sample points. Therefore, a sliding window is applied to detect the largest power difference of the whole sound samples and cut out 1350 sample points as the exact finger snapping sound, called snapping trace in the following. Figure 1 shows examples of finger snapping detection for two subjects, and the part between the two vertical solid lines is the detected snapping sound.



(a) (b)

Fig. 1. Finger snapping detection of (a) subject1 (b) subject2

3.4 Raw Data Analysis

Intuitively, the whole raw time series of snapping trace (i.e., the whole 1350 samples) is applied to authenticate users. Here, Hidden Markov Model (HMM) [10] is applied to the raw time series of the snapping traces, as the process of finger snapping can be considered as a Markov process with hidden states. The HMM model is trained with 40 traces of each subject to get an authentication threshold. All the other subjects' traces are used to test the model for each subject. However, the average FAR among all the subjects of the direct comparison is 45 %, which is not acceptable for person authentication purpose. So we concentrate our research to the direction of *finding the unique features underlying in the finger snapping sound for user authentication.*

4 Feature Extraction

We extract features from the finger snapping traces for person authentication based on Training Set. Here several features referring to acoustic analysis on time and frequency domain are investigated step by step.

4.1 Time Domain Feature

We explore time domain features because they reflect the behavioral characteristics. Two widely used time domain features are selected for finger snapping authentication, including zero-crossing rate (ZCR) and root mean square (RMS). Figure 2(a) shows the ZCR and RMS values of 40 snapping traces from each of the 4 subjects. It can be seen that the point sets are distributed without too much intersection. We only show 4 subjects' values for clear visualization, and all the traces of other 18 subjects also receive the similar results.

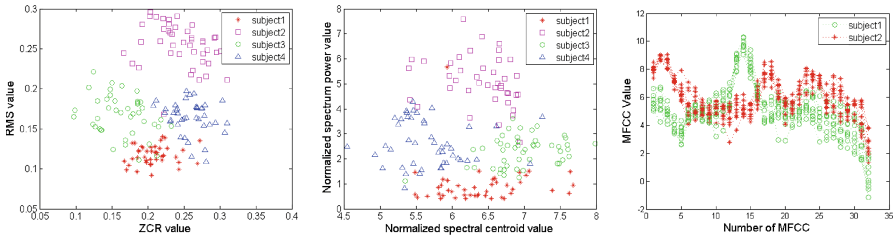
4.2 Frequency Domain Feature

We regard the hand as a musical instrument for snapping fingers. Considering that musical instruments may have their own particular frequency, we study frequency domain features of the snapping sounds. We select the MFCC, spectrum

power and spectral centroid features because they can reflect the uniqueness of different individuals' finger snapping.

Spectrum Power and Spectral Centroid: After taking the FFT of the snapping trace, we calculate the total spectrum power and spectral centroid of the snapping trace. Spectral centroid is the gravity center of the power spectrum [11]. The values of spectrum power and spectral centroid are normalized between 0 and 10. Figure 2(b) shows the spectrum power and spectral centroid of 40 snapping traces from each of the 4 subject. The point set boundaries are clear among different individuals, so spectrum power and spectral centroid will be useful for user authentication. We only show 4 subjects' values for clear visualization in Fig. 2(b), and the traces of other 18 subjects also receive the similar results.

MFCC: Mel-Frequency Cepstral Coefficients (MFCC) are proposed to realize speech and speaker recognition [5]. It is an approximation of the human auditory system's response. Figure 2(c) shows two subjects MFCC value distributions of 15 snapping traces respectively. There are obvious differences between these two subjects which indicate the uniqueness for different individuals and the MFCC sets of the same individual remain a similar trend. In order to facilitate observation, we only give 2 subjects MFCC sets in Fig. 2(c), and the other 20 subjects present the observable distinctiveness as well.



(a) (b) (c)

Fig. 2. Visualization of (a) ZCR and RMS (b) Spectrum power and spectral centroid (c) MFCC

5 System Design

The design of our user authentication system based on finger snapping is shown in Fig. 3. The system leverages snapping sound sensed by the microphones in smart devices and does not rely on the priori that knowing any ambient environment information. The result of the system is a decision: owner or attacker. The system contains five steps:

Sensing: To record the finger snapping sound using the embedded microphone in smart devices. Then the exact sound of the finger snapping is detected.

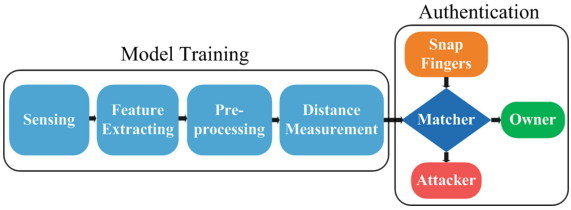


Fig. 3. System Design

Feature Extracting: A set of acoustic features from both the time and frequency domain are extracted from the snapping sound traces.

Preprocessing: Since noises can overwhelm the snapping sound, there may be a few low quality snapping inputs. Enrollment threshold is defined to remove low quality traces for model training. First, we calculate the feature vector's Manhattan distance between one trace and all the other traces from the owner. Each trace gets its own average Manhattan distance d_i, which represents the relevance with the owner's normal snapping state. The enrollment threshold δ is represented in Eq. (1), where m and var are the mean and standard variance of all d_i. Those traces with d_i values exceeding δ are the invalid enrollments. It is emphasized that preprocessing is only applied on traces used to train the model.

$$\delta = m + 2 \times var \tag{1}$$

Distance Measurement: DTW distance measurement is applied for authentication. DTW finds a non-linear warping path between two feature vectors to obtain an optimized similarity distance. Here, DTW is not applied to cope with length diversity, as the feature vector of finger snapping is length-fixed. It is the special property of MFCC that makes us use DTW algorithm. In the process of calculating MFCC, the spectrum power is mapped to mel-scale using overlapping windows which share some common areas with each other. This indicates that adjacent MFCC values can be similar, so number 1 MFCC in one feature set can be aligned with number 2 MFCC in another set.

Matcher: It decides whether the snapping sound is made by the owner or the attacker with the DTW output. As only owner's traces are available for authentication situation, DTW needs a template and a threshold to get the result. The template is selected to be the trace from the owner which achieves the highest pairwise similarity with all the other training traces. All the training traces will be compared with the template and the outputs are sorted to get a certain value as the authentication threshold. The procedure of determining the template and threshold can be regarded as the training process. A test trace will be compared with the template to get the distance output. If the output is smaller than the threshold, the test trace will be accepted as the owner's.

6 Evaluation

In this section, the performance of finger snapping authentication is evaluated. Based on Training Set of 6160 snapping traces from 22 subjects across 7 days, we evaluate the FER, select the training parameters and analyze the permanence of finger snapping. Combining Training Set with Testing Set, we verify the uniqueness of finger snapping. Circumvention results are evaluated by recording owner's snapping sound with attacker's smartphone. We further do a survey of 74 people on universality and acceptance issue of finger snapping authentication. Finally, the cost of finger snapping authentication is analyzed, including the computational time and power consumption on smartphones.

6.1 Failure to Enroll Rate

FER is the rate at which low quality inputs occur and they are regarded as invalid enrollment. The preprocessing phase eliminates the low quality inputs before training models. In order to give a comprehensive result, all the collected traces of 22 subjects are used to obtain FER. The average FER among all subjects is 4.4 %, which shows that finger snapping can be effectively collected.

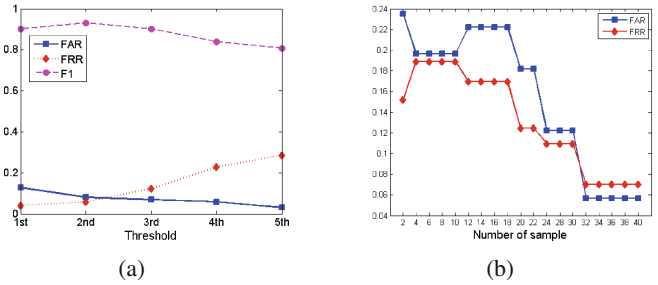


Fig. 4. Performance of different (a) threshold (b) trace capacity

6.2 Threshold and Trace Capacity

We investigate the effect of the choice of authentication threshold and the number of training traces on authentication performance with all the snapping traces in Training Set. The template is selected from the Day0' traces for each subject. Then all the training traces are compared with the template by DTW and the outputs are sorted to get the $1st$, $2nd$, $3rd$, $4th$ and $5th$ highest values. They are respectively taken as the authentication threshold. Figure 4(a) depicts the average FRR, FAR and F1 score of different thresholds for all subjects. It shows that the $2nd$ highest value is proper to be set as the authentication threshold as it achieves the highest F1 score with low FRR and FAR.

For the capacity of training traces, we vary the number from 1 to 40 to obtain the template and threshold. The performance on average FRR and FAR for all

subjects with different trace capacity is shown in Fig. 4(b). When the number is higher than 32, the FRR and FAR will stay at the lowest level, so we choose 32 as the capacity length.

6.3 Permanence and Uniqueness

The snapping traces in Training Set and Testing traces are used to evaluate the permanence and uniqueness of finger snapping.

FRR: To obtain FRR, we use each subject Day0's traces for training and use Day1-Day6's traces for testing. The average FRRs among all 22 subjects from Day1 to Day6 are shown in Table 1. The average FRR (6.1 %) across 7 days and the standard variance (0.9 %) show that the FRRs stay low and relatively stable across days. Besides, the testing traces are not preprocessed. Among the FRR traces, some owner's traces are rejected because of low quality inputs (FER traces).

Table 1. Average FRR of all subjects on Day0

Day	1	2	3	4	5	6	Mean	std
FRR	6.4 %	4.8 %	5.9 %	7.2 %	5.4 %	6.7 %	6.1 %	0.9 %

FAR: To obtain FAR, 22 subjects in Training Set are regarded as owners. All the snapping traces in Testing Set which is collected from 54 volunteers and the other 21 subjects' traces are used to attack the owner's finger snapping. The FAR is illustrated in Table 2. It shows the FAR is kind of subject dependent. The average FAR is 5.9 %, and all the FARs are lower than 10 % for all subjects.

Table 2. FAR of all subjects on Day0

Subject	1	2	3	4	5	6	7	8	9	10	11
FAR	3.8 %	6.5 %	4.5 %	6.2 %	4.7 %	5.8 %	7.6 %	3.5 %	7.9 %	8.5 %	5.5 %
Subject	12	13	14	15	16	17	18	19	20	21	22
FAR	6.4 %	5.6 %	7.4 %	6.8 %	5.2 %	4.8 %	7.5 %	4.4 %	5.7 %	5.8 %	4.9 %

6.4 Circumvention

In terms of circumvention, one question might be brought up: What if the attacker records the owner's snapping sound to attack the owner's device? We imitate this situation by recording the owner's snapping with the attacker's smartphone side by side to the owner's phone. Then the attacker replays the recorded file to attack owner's model. The owner's phone type is HUAWEI Honor

7, and the attacker's phone types are HUAWEI Honor 7, Samsung Galaxy 3 and 4, Google Nexus 5, iPhone 5 and 6. All the 22 subjects play the role of the owner in turn, and attack attempts all failed. This is because the frequency response of the microphone and speaker of the off-the-shelf devices, like smartphones, does not match with each other. So the snapping sounds recorded by the microphone are thereby badly distorted through the speaker when being replayed.

6.5 Universality and Acceptability

A survey is carried out on 74 people about whether they can snap their fingers and accept the finger snapping authentication. Results show that 86.5 % of the respondents can snap fingers, of which 89.2 % would like to authenticate themselves using a simple finger snap. Besides, through our finger snapping collecting phase, we come to find out that people who could not snap their fingers can learn to do it after understanding the method of finger snapping.

6.6 Computational Time and Power Consumption

For each of the 22 subjects, we further measure the time and energy cost with 32 snapping traces. The average time on calculating DTW distances and deciding the template and threshold is $7.144s$ on the smartphone of HUAWEI Honor 7. The average time to decide whether a test snapping trace is the owner's or not is only $0.118s$ on the smartphone. In terms of the power consumption on the smartphone, it costs totally $4.36J$ for the training process, and $0.12J$ for authenticating a test snapping trace.

7 Conclusion

This paper proposes a new biometric trait, finger snapping, to realize person authentication. Finger snapping is recorded and detected with the microphone embedded in the off-the-shelf smart devices. The acoustic features are extracted from both time and frequency domain, including zero-crossing rate, root mean square, time average energy, MFCC, spectrum power and spectral centroid. We apply DTW to evaluate the similarity between the feature vectors of the snapping traces. Experiments have confirmed the measurability, uniqueness, permanence, circumvention, universality and acceptability of the finger snapping authentication.

Acknowledgments. We show thanks to the volunteers who participated in the process of finger snapping collection. This research is partially supported by the National Science Foundation of China (NSFC) under Grant Number 61379128 and 61379127.

References

1. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of Fingerprint Recognition, 2nd edn. Springer, Heidelberg (2009)
2. "Wikipedia" "Finger snapping". https://en.wikipedia.org/w/index.php?title=Finger_snapping&oldid=691142888
3. Müller, M.: Dynamic time warping. In: Müller, M. (ed.) Information Retrieval for Music and Motion, pp. 69–84. Springer, Heidelberg (2007)
4. Jain, A.K., Ross, A., Nandakumar, K.: Introduction to Biometrics. Springer, Heidelberg (2011). ISBN 978-0-387-77325-4
5. Li, S.Z., Jain, A.K. (eds.): Handbook of Face Recognition, 2nd edn. Springer, Heidelberg (2011). ISBN 978-0-85729-931-4
6. Kevin, W.B., Karen, H., Patrick, J.F.: Image understanding for iris biometrics: a survey. Comput. Vis. Image Underst. **110**(2), 281–307 (2008)
7. Yu, G., Li, C.-T.: A robust speed-invariant gait recognition system for walker and runner identification. In: 2013 International Conference on Biometrics (ICB) (2013)
8. Shannon, R.V., and Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M.: Speech recognition with primarily temporal cues. In: American Association for the Advancement of Science (1995)
9. Hong, F., Wei, M., You, S., Feng, Y., Guo, Z.: Waving authentication: your smartphone authenticate you on motion gesture. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (2015)
10. Nadeu, C., Macho, D., Hernando, J.: Time and frequency filtering of filter-bank energies for robust HMM speech recognition. In: Speech Communication. Elsevier (2001)
11. Wang, J., Lee, H., Wang, J., Lin, C.: Robust environmental sound recognition for home automation. IEEE Trans. Autom. Sci. Eng. **5**(1), 25–31 (2008)