SonicPrint: A Generally Adoptable and Secure Fingerprint Biometrics in Smart Devices

Aditya Singh Rathore[1], Weijin Zhu[1], Afee Daiyan[1], Chenhan Xu[1], Kun Wang[2], Feng Lin[3], Kui Ren[3], Wenyao Xu[1]
[1]University at Buffalo, the State University of New York, Buffalo, New York, USA
[2]University of California, Los Angeles, California, USA
[3]Zhejiang University, Zhejiang, China
{asrathor,weijinzh,afeedaiy,chenhanx,wenyaoxu}@buffalo.edu,wangk@ucla.edu,{flin,kuiren}@zju.edu.cn

ABSTRACT

The advent of smart devices has caused unprecedented security and privacy concerns to its users. Although the fingerprint technology is a go-to biometric solution in high-impact applications (e.g., smartphone security, monetary transactions and international-border verification), the existing fingerprint scanners are vulnerable to spoofing attacks via fake-finger and cannot be employed across smart devices (e.g., wearables) due to hardware constraints. We propose *SonicPrint* that extends fingerprint identification beyond smartphones to any smart device without the need for traditional fingerprint scanners. *SonicPrint* builds on the fingerprint-induced sonic effect (FiSe) caused by a user swiping his fingertip on smart devices and the resulting property, i.e., different users' fingerprint would result in distinct FiSe. As the first exploratory study, extensive experiments verify the above property with 31 participants over four different swipe actions on five different types of smart devices with even partial fingerprints. *SonicPrint* achieves up to a 98% identification accuracy on smartphone and an equal-error-rate (EER) less than 3% for smartwatch and headphones. We also examine and demonstrate the resilience of *SonicPrint* against fingerprint phantoms and replay attacks. A key advantage of *SonicPrint* is that it leverages the already existing microphones in smart devices, requiring no hardware modifications. Compared with other biometrics including physiological patterns and passive sensing, *SonicPrint* is a low-cost, privacy-oriented and secure approach to identify users across smart devices of unique form-factors.

CCS CONCEPTS

• **Security and privacy → Biometrics**; • **Human-centered computing → Ubiquitous and mobile devices**; • **Hardware →** *Signal processing systems.*

ACM Reference Format:
Aditya Singh Rathore[1], Weijin Zhu[1], Afee Daiyan[1], Chenhan Xu[1], Kun Wang[2], Feng Lin[3], Kui Ren[3], Wenyao Xu[1]. 2020. SonicPrint: A Generally Adoptable and Secure Fingerprint Biometrics in Smart Devices . In *The 18th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '20), June 15–19, 2020, Toronto, ON, Canada.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3386901.3388939

1 INTRODUCTION

Fingerprint technology has become highly ubiquitous with wide-scale adoption in smartphones and smart devices for user identification. Unlike the privacy concerns raised by face biometrics (San Francisco face ban [1]), low degree-of-freedom in iris detection [2] and inferior robustness of voice authentication [3], fingerprint possesses high social acceptance due to its uniqueness and usability. In current digital era, the necessity to track vital signs, automate day-to-day tasks and improve the quality of life fuels the growth of diverse smart devices. A report by Gartner describes a typical family home to contain 500 smart objects by 2022 [4]. For protecting the smart environment, it is evident that smart devices will continue to rely on biometrics, with fingerprint being the first choice as key to user's confidential data.

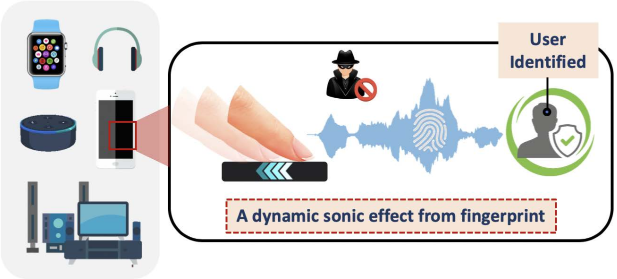


Figure 1: A new dimension of fingerprint sensing adoptable across diverse smart devices and resilient to fake-finger spoofing.

Even after decades of development, the existing fingerprint modalities suffer from two limitations. For accurate user identification, a high-resolution fingerprint needs to be acquired through dedicated hardware scanners (e.g., optical, capacitive or thermal [5]), which are expensive and cumbersome. The diversity in terms of embedded sensors, shape and size makes the adoption of existing fingerprint biometrics infeasible in upcoming smart devices. Secondly, fake-fingers can be exploited to spoof fingerprint traits [6–8]. Even the in-display ultrasound sensors, targeted towards enhancing usability, are susceptible to 3D finger models [9].

It is a known fact that when two objects slide against each other, kinetic energy is released in the form of sonic wave and heat. The harmonics of this friction-excited sonic wave are dependent on the

MobiSys '20, June 15–19, 2020, Toronto, ON, Canada

ACM ISBN 978-1-4503-7954-0/20/06…$15.00
https://doi.org/10.1145/3386901.3388939

surface characteristics of objects and their internal composition. Our key contribution is the observation that the sonic waves from a user swiping his fingertip on a surface can serve as biometric traits. Since every person has a unique fingerprint, we hypothesize that two users swiping their fingertips on a common surface should result in distinct fingerprint-induced sonic effect (FiSe). Although the statistical properties of FiSe may change depending on the user's swiping speed, pressure or surface roughness, the inherent uniqueness is dependent on the surface texture (i.e., fingerprint ridge patterns) and the finger's constitution. If this hypothesis holds, FiSe can be observed across smart devices and measured using inbuilt microphones. The goal of this work is to explore the knowledge and validation of a new fingerprint sensing modality and open discussions for emerging mobile security research.

As a natural communication interface, microphones are actively employed in IoT-enabled devices, virtual-reality (VR) headsets as well as smart city initiatives [10]. To this end, our novel biometric provides two distinct advantages over existing fingerprint technologies as illustrated in Figure 1:

• **Adoptability:** our method requires no specific hardware and utilize low-cost off-the-shelf sensors in smart devices. The biometric trait is available across devices with diverse flexibility, geometry and composition.

• **Anti-Spoofing:** unlike the traditional fingerprint methods, our proposed approach leverages the advantages of both fingerprint and audio domain to prevent against fake-fingers, replay or side-channels attacks.

Building on this, we aim to transform everyday smart devices into fingerprint scanners. To achieve this, three challenges need to be addressed: (1) FiSe is typically of low power and submerged in dynamic background noises. How to acquire the target FiSe without any information loss? (2) To enable high accessibility and acceptance, it is important to provide freedom to the users while swiping the surface. In the case where a user's swiping speed and pressure are not controlled, how to select appropriate features that closely resemble the fingerprint? (3) For real-world applications, it is critical that the FiSe cannot be compromised. How to evaluate the vulnerability of our system which relies on characteristics of both fingerprint and audio domain?

In this work, we propose the first systematic framework that leverages FiSe from a user swiping on smartphone and smart devices as a new biometric. We first validate the uniqueness of fingerprint-induced sonic patterns by comparing the resulting spectrum of fingerprints with different textures. Then, we leverage the underlying microphone in a smartphone to acquire FiSe and investigate a sequence of spectral and wavelet denoising approaches for background isolation. An adaptive segmentation method is designed to remove the tap noise and other entities which can be easily misinterpreted as the target signal. Afterward, we propose a novel taxonomy that highlights the semantic relationship between the fingerprint and audio domain and identifies multi-level features that fundamentally share the same concept as a fingerprint. Based on these insights, we design and implement our system, *SonicPrint*, to facilitate secure sensing of FiSe for user identification. Finally, a comprehensive evaluation is performed with 31 participants on five smart devices across six sessions over two months to validate the effectiveness and inclusiveness of *SonicPrint* under real-world scenarios.

Summary: Our contribution in this work is three-fold:

- We explore a novel fingerprint-based biometric approach for user identification. We find that when a user swipes his fingertip on a surface, the FiSe contains intrinsic fingerprint information.
- We design and implement *SonicPrint*, an end-to-end biometric system to facilitate secure, accessible and user-friendly fingerprint sensing on everyday smart devices in practice.
- We validate the effectiveness and inclusiveness of *SonicPrint* through extensive experiments with results showing up to 98% accuracy. We conduct comprehensive studies to show the resilience of *SonicPrint* against fake-finger and replay attacks.

2 BACKGROUND AND PRELIMINARIES

In this section, we provide a background on friction-excited sonic waves and the rationale behind its uniqueness in terms of human-to-material interaction. We also perform a feasibility study to prove this concept.

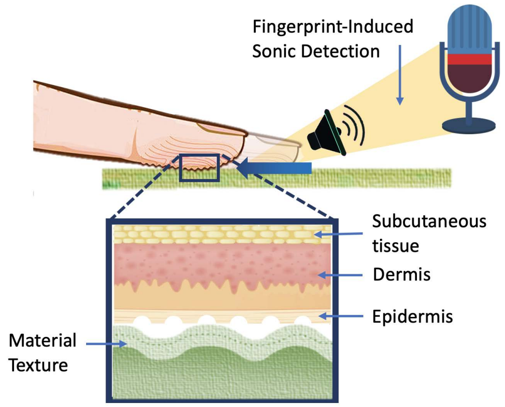


Figure 2: FiSe arises from the friction between fingerprint and surface and can be sensed by a conventional microphone.

2.1 Fingerprint-Induced Sonic Effect

Friction develops from two surfaces sliding against one another irrespective of the intensity of their relative motion. This friction leads to distinct waves and oscillations within the interacting mediums resulting in the emission of sonic waves to the ambient environment [11]. In daily life, there are several instances of friction-excited sonic waves from an interaction between sneakers on the floor or chalk on the blackboard. In this paper, the context of sonic wave differs from the roughness noise, which is generally random (e.g., rubbing of two sandpapers). Under strong contact conditions, the sliding surfaces become a coupled system and generate an intricate and often nonlinear response. Previous studies have shown that physical parameters, including speed and pressure, only affect the magnitude of power spectral density to a certain extent, but not the overall distribution [12]. The roughness of the sliding surfaces impacts the sound pressure level (SPL) as:

$$\Delta SPL = 20\log_{10}(\frac{R_2}{R_1})^m, \quad (1)$$

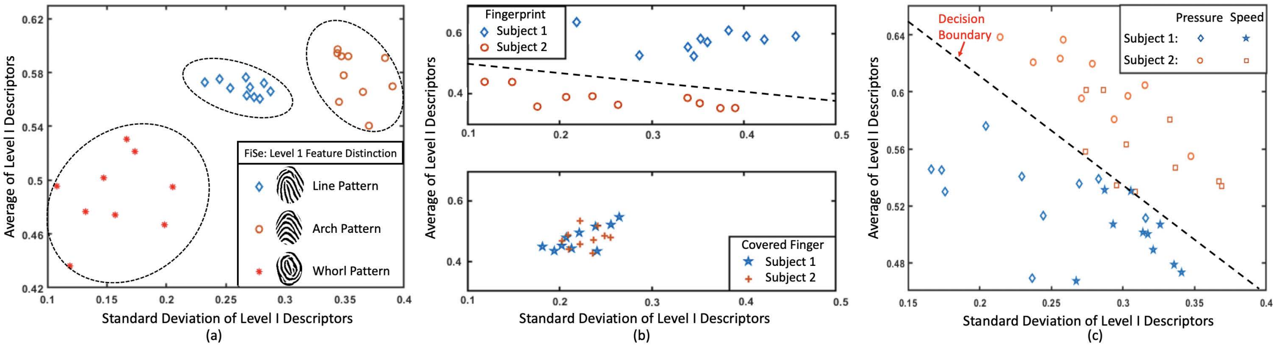


Figure 3: A proof-of-concept (three subjects) for FiSe-based identification under the impact of (a) different fingerprint patterns; (b) fingerprint and covered finger interaction with surface; (c) human dynamics (i.e., swiping speed and pressure).

where R_2 and R_1 correspond to the roughness of friction pair and m is an empirical factor varying based on the surface texture. The SPL of sonic waves can be similar between different friction pairs and thus impacts its sensing rather than uniqueness. A person with rough fingertip would produce a more audible sonic wave when rubbing a surface, in contrast to a soft skin fingertip. More importantly, for different friction pairs (e.g., finger against metal vs. finger against plastic), the uniqueness of sonic waves arise from the interface properties (i.e., texture) and the constitution of objects (e.g., weight distribution). The surface deformation during contact is highly minute [13] and its intensity is inconsequential to surface roughness.

Hypothesis: When a user swipes his fingertip on any surface (refer to Figure 2), the resulting friction-excited sonic wave depends on the intrinsic fingerprint patterns, underlying structure of finger and opposing material. Since every user has a unique fingerprint, the FiSe from two users swiping on the same surface should be different. Moreover, the low SPL of FiSe provides a strong resilience against spoofing attacks.

2.2 A Feasibility Study

Proof-of-concept Setup: To validate the uniqueness of FiSe from different fingerprints, we conduct a preliminary study (n = 3) to perform straight-downward swipes with the right index finger (dry state), 20 times each, on the back surface (aluminum) of a commodity smartphone. The subjects are told to swipe naturally without exerting intense pressure or speed. During the second trial, we cover the subject's fingertip with a scotch tape and repeat the swipe actions. In another experiment, we ask two subjects to repeat 15 swipes with gradually increasing pressure and speed in each trial. The resulting FiSe is recorded by the inbuilt microphone (sampling rate-44.1KHz) of a smartphone. For the sake of isolating environmental dependency, this study is performed in a conference room (21°C) with low ambient noise. After processing the fingerprint-induced sonic waves, we aim to extract features that can provide a clue towards the inherent fingerprint.

Feature Distinction Analysis: Level I characteristics of the fingerprint depend on its macro details, i.e., the pattern and ridge flow and can be visually perceived through naked eye [14]. Similarly, in the audio domain, power-based temporal features highlight the changes in signal over time and perceptual features (e.g., pitch, loudness) have semantic meaning to a human listener. Therefore, we select features, including temporal centroid, log attack time, harmonicity, pitch and spectral features (i.e., centroid, crest, decrease, entropy, flatness, rolloffpoint and spread) as Level I friction descriptors. For ease of the comparison, Figure 3 illustrates the variations against average and standard deviation of descriptors after normalization. Each FiSe yields a data point on the graph and the points from multiple FiSe by the same fingerprint exhibit a cluster.

Insights: Our preliminary analysis reveals that (1) every user has a unique fingerprint pattern (e.g., line, arch and whorl pattern in Figure 3(a)) which generates a unique FiSe during the swipe action; (2) Figure 3(b) proves that distinctiveness of FiSe is dependent on the fingerprint rather than the overall geometry of the fingertip; (3) variation in pressure and speed has a limited effect on the identifiability of FiSe (see Figure 3(c)). However, considering different swipe dynamics, the decision boundary might overlap when only using Level I friction descriptors.

Summary: We prove that FiSe depends on the underlying fingerprint. To improve accuracy, we continue to recruit appropriate features highlighting the intrinsic fingerprint information (Level II and Level III) from the sonic waves. In the following paper, we discuss the application of FiSe for smartphone security and provide insights to consider before real-world deployment. We also examine other types of smart devices in Section 9.1.

3 THREAT MODEL

We consider an attacker, namely Alice, who intends to steal private information from the victim's smartphone. We assume that the smartphone comprises of a standalone defense mechanism, i.e., *SonicPrint*. Upon realizing the fundamental operation of *SonicPrint*, Alice explores the existing literature for methods to compromise fingerprint and audio security. Specifically, we consider the following attack scenarios:

• **Fingerprint phantom attack:** Typically, Alice can either exploit social media of the victim or remotely capture the desired fingerprint through high-resolution cameras. Afterward, the fingerprint and overall finger geometry can be utilized to create a fingerprint phantom (i.e., fake-finger). This fake-finger is highly identical to the victim's live-finger and can be used to spoof the system. During the entire process, the victim has no idea of the ongoing threats. It

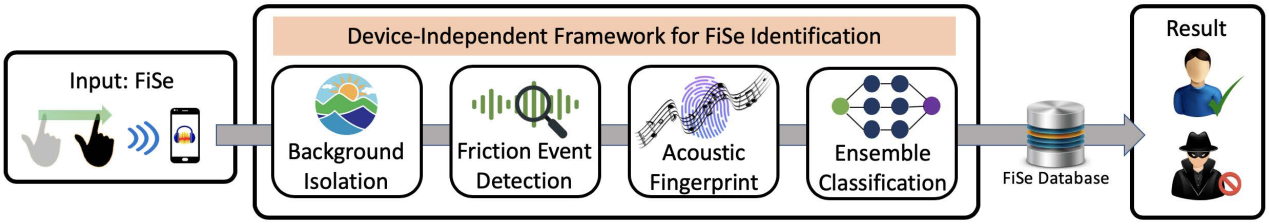


Figure 4: The overview of *SonicPrint*, a fingerprint-biometric based user identification system.

is worth mentioning that conventional fingerprint scanners can be compromised using this stealthy attack [15].

- **Replay and Side-Channel attack:** Without the victim's knowledge, Alice places a high-sensitive microphone near the smartphone and records the FiSe during an access attempt. This recording is replayed to the target device through direct FiSe matching or vibration injections by leveraging sophisticated hardware. Studies show that this attack can compromise the security of traditional voice authentications within five trials [16].

In our work, we assume that Alice is not able to place the recording device very close to the victim's smartphone (i.e., < $20cm$) during the access attempt. This assumption is practical since the malicious hardware will be at line-of-sight to the victim, raising his suspicion. Moreover, even prominent biometrics (e.g., voice) are inapplicable under the identical scenario. We assume that Alice cannot create a biological replica of victim's finger using organic 3D printers. These printers are economically unfeasible, costing millions of dollars and require advanced knowledge of printing. *SonicPrint* can leverage FiSe for secure user identification.

4 SONICPRINT SYSTEM OVERVIEW

By analyzing the FiSe caused by fingertip and surface interaction, *SonicPrint* can reveal fingerprint dependent characteristics in the received signal. Figure 4 illustrates four primary modules of *SonicPrint*: (1) Background isolation; (2) Friction event detection; (3) Acoustic fingerprint analysis; (4) Ensemble classification. First, when a user swipes his fingertip on the smartphone surface, the inbuilt microphone is used to capture the FiSe. A series of pre-processing techniques including clutter suppression, target enhancement and ambient denoising are applied to acquire the precise sonic wave. Once its position is verified, a multi-level representation of acoustic fingerprint is obtained from specific features of the target signal. Finally, the representation is input to an ensemble classifier to precisely identify the legitimate user.

5 FISE PROCESSING SCHEMES

In this section, we discuss the nature of friction excited sonic waves from a coupled system consisting of fingertip and material. When a user swipes his fingertip on the smartphone surface, a FiSe is generated, which can be captured by the inbuilt microphone and can span the entire frequency band (0-22KHz).

5.1 Pre-processing

The sonic wave is typically submerged in the dynamic ambient noises (e.g., human talking, music) due to its low power. Considering the diverse and known frequency bands in the noise spectrum, it is effective to use high-order cutoff in one-pass filters. However, this also eliminates the intrinsic fingerprint information in the lower frequency bands. To remove the low frequency noise from human speech and music, we employ a high-pass filter with cutoff 2.2KHz to remove the arbitrary clutter and recover the signal with a frequency range from 2.2KHz to 22KHz.

5.2 Sonic Effect Enhancement

Although the human voice and background clutter can be separated based on information content, the FiSe might be perceived as generic noise due to its low power. Multi-band spectral subtraction [17] is a widely used method to enhance the target signal that is degraded by additive noise without introducing any distortions. Given that noise does not affect the entire frequency band of FiSe uniformly, we need to ideally subtract the appropriate noise spectrum from each frequency bin. This would restrict any excessive subtraction of intrinsic fingerprint information. We acquire the clean and enhanced spectrum of FiSe in the ith frequency band by:

$$|\hat{S}_i(k)|^2 = |Y_i(k)|^2 - \alpha_i \delta_i |\hat{D}_i(k)|^2 \quad b_i < k < e_i, \tag{2}$$

where Y_i is the power spectrum of noisy FiSe signal, $\hat{D}_i$ is the noise estimate, b_i and e_i are starting and ending frequency bins. α_i is an over-subtraction factor and δ_i is empirically chosen for each frequency band. For calculating δ_i, we leverage a pre-recorded two second audio sample in daily environment with human voices as noise estimate. We update over-subtraction factor α_i as:

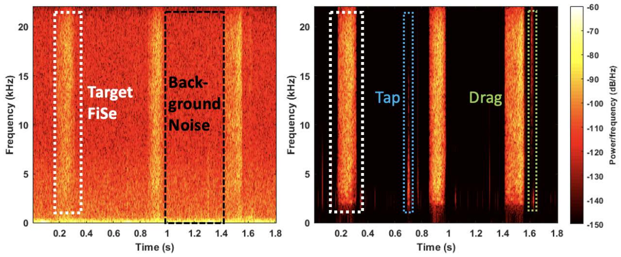
$$\alpha_i = c_1 \cdot log_{10}\left(\frac{\sum_{k=b_i}^{e_i} |Y_i(k)|^2}{\sum_{k=b_i}^{e_i} |\hat{D}_i(k)|^2}\right) + c_2, \tag{3}$$

where c_1, c_2 are empirically chosen values. After nonlinear power spectrum subtraction, the enhanced FiSe is derived from its spectrogram. However, there still exists residual clutter between the intervals of FiSe.

5.3 Denoising-Aware Wavelet Reconstruction

Given the advantages of (1) multi-scale analysis [18] and (2) optimal resolution in frequency and time domain, we employ wavelet denoising to eliminate the residual noise from the FiSe that remains even after sonic effect enhancement. Using maximal overlap discrete wavelet transform (MODWT) [19], the signal is first subjected to decomposition to acquire detail coefficients (α_k) and approximation coefficients (β_k):

$$\alpha_k^{(J)} = \sum_{n \in Z} x_n \bar{g}_{n-2^J k}^{(J)} \quad \beta_k^{(l)} = \sum_{n \in Z} x_n \bar{h}_{n-2^l k}^{(l)}, \tag{4}$$



(a) (b)

Figure 5: The spectrogram of (a) original and (b) denoised FiSe from three swipe actions.

where the levels $J \in Z$ and $l \in \{1, 2, 3, ..J\}$. We choose the Daubechies 3 wavelet (dB3) and reduce the FiSe to 6 levels. Afterward, we apply the detail coefficient threshold for each level to discard the ambient clutter. Finally, a level-dependent reconstruction is employed using all the coefficients as:

$$x_n = \sum_{k \in Z} \alpha_k^{(J)} \overline{g}^{(J)}_{n-2^J k} + \sum_{l=1}^{J} \sum_{k \in Z} \beta_k^{(l)} \overline{h}^{(l)}_{n-2^l k}, \qquad (5)$$

where $\overline{g}$ and $\overline{h}$ are rescaled discrete orthogonal functions. The spectrogram of FiSe before and after the processing stage is shown in Figure 5. It is worth mentioning that the signal-to-noise ratio (SNR) is significantly improved from -3 to 23 decibels. In the next subsection, we discuss the challenges of localizing the FiSe in the overall signal and our proposed solution.

5.4 Friction Event Detection

Considering the FiSe is caused by a user swiping his fingertip on the smartphone surface, there are three challenges in tracing the target's precise location in the measured signal:

- The length of FiSe would vary among different swipes and different users. Generally, the FiSe from a swipe action ranges from 0.05 to 0.3 seconds.
- Due to the variations in SPL of FiSe from human dynamics during the swipe action (see Section 2.1), the traditional threshold-based separation [20] algorithms are inadequate without optimization.
- During a swipe action, there may be an initial tap sound (i.e., finger colliding with device surface) or closing drag sound (i.e., finger slipping when lifting) enclosing the FiSe. Since the amplitude of tap and drag sound are arbitrary, peak detection methods are ineffective.

To this end, we specially design our segmentation process (see Algorithm 1), to address the above challenges and isolate the starting and ending periods of each FiSe.

i) Adaptive Detection via HMM model: The hidden Markov Model (HMM) has proven to be an effective method for acoustic event detection [21]. It computes the probability of an occurrence of FiSe in every segment of the recorded signal and only consider those with high probability as friction events. Specifically, we first divide the recorded sample in non-overlapping frames, where each frame is 0.01 second period. A discrete fourier transform (DFT) is applied to each frame, after which an unbiased noise variance is calculated based on the optimally smoothed power spectral density estimate and spectral minima from each frequency band [22]. Finally, a widely used log-likelihood ratio test and HMM-based hang-over scheme [23] is used to determine the probability of friction event. To regulate the prior SNR [24] in log-likelihood, we define two additional parameters, i.e., TargetToSilence (TTS) probability and SilenceToTarget (STT) probability.

Algorithm 1 Roughness-aware FiSe detection.

```
Input: $x(k), W$: k frames from signal with size W
       $STT, TTS$: Probability thresholds for HMM
Output: $R$: Target fingerprint-induced sonic
1:  $P_i, L_i, T \leftarrow 0$                                   ▹ Initialize parameters.
2:  for $i \in \{1, .., k\}$ do $P_i = HMM(x(k), STT, TTS)$;     ▹ Compute
    likelihood $P_i$ that frames contain FiSe.
3:  end for
4:  $T = segments(P_i \geq 0.9)$;                                ▹ Extract segments.
5:  $\{dry, balanced, soft\} = compare(count(T), k \times W)$;   ▹ Predict roughness by comparing number of detected vs expected
    segments based on window size.
6:  if $!dry$ then
7:      Repeat lines 2-4 with optimized thresholds;
8:  end if
9:  $L = Phase(x(k))$;                                           ▹ Compute linearity index.
10: $minTime, maxTime = duration(\Delta max(P_i, L_i))$;
    ▹ Determine the period range of target FiSe based on consecutive high likelihood and linearity index.
11: if $minTime \leq len(T) \leq maxTime$ then
12:     $R = T$;                                                 ▹ Verify duration of segments.
13: end if
14: return $R$
```

In a scenario where the swipe action of a user is controlled, the TTS and STT thresholds can be set to a fixed value, similar to speech recognition applications. However, considering the dynamic nature of FiSe, a fixed cutoff would lead to a reduction in the identified target segments. Therefore, for ensuring the identification of FiSe with even low audibility, we design an adaptive technique that ranks the roughness of user's fingertip based on the statistical analysis of the signal. In particular, the roughness can be categorized as dry, balanced or soft by comparing the number of detected FiSe vs. the expected FiSe based on overall period. Depending on the predicted roughness, the TTS and STT probabilities are optimized to retrace the optimal friction events. In scenarios where the SPL of FiSe is very low, our adaptive detection can raise the number of identified events by more than 84% (counted manually).

ii) Phase-based Detection: The tap sound and drag sound are of arbitrary characteristics and challenging to remove by conventional statistical methods (e.g., maximum amplitude, mean, standard deviation). Previously, phase-based detection schemes have been proposed to suppress the impact noise [25]. The acoustic signal is first divided into non-overlapping frames of $0.01s$. Considering that there is only one dominant pulse of magnitude a at n_0 in the current frame, the signal $x(n) = 0$ except at $n = n_0$. Afterward, a DFT is applied to individual frames with the kth frequency bin and the phase slope as:

$$X(k) = |X(k)| e^{j\theta(k)} = a e^{-j2\pi kn/N}, \qquad (6)$$

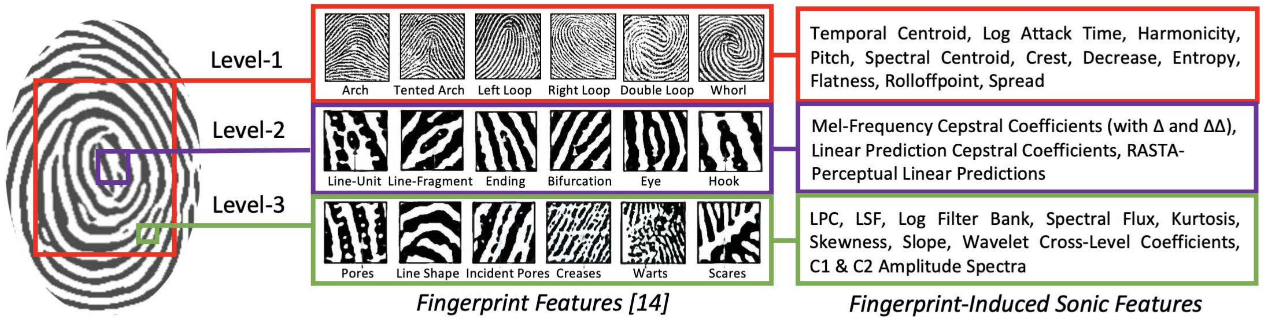

Level-1
Level-2
Level-3
Arch
Tented Arch
Left Loop
Right Loop
Double Loop
Whorl
Line-Unit
Line-Fragment
Ending
Bifurcation
Eye
Hook
Pores
Line Shape
Incident Pores
Creases
Warts
Scares
Temporal Centroid, Log Attack Time, Harmonicity, Pitch, Spectral Centroid, Crest, Decrease, Entropy, Flatness, Rolloffpoint, Spread
Mel-Frequency Cepstral Coefficients (with Δ and ΔΔ), Linear Prediction Cepstral Coefficients, RASTA-Perceptual Linear Predictions
LPC, LSF, Log Filter Bank, Spectral Flux, Kurtosis, Skewness, Slope, Wavelet Cross-Level Coefficients, C1 & C2 Amplitude Spectra
Fingerprint Features [14]
Fingerprint-Induced Sonic Features

Figure 6: A taxonomy of multi-level friction descriptors corresponding to intrinsic fingerprint.

$$\Delta\theta(k) = \tan^{-1}\frac{Im(\overline{X}(k)\cdot\overline{X}^{*}(k-1))}{Re(\overline{X}(k)\cdot\overline{X}^{*}(k-1))}\quad \overline{X}(k) = \frac{X(k)}{|X(k)|}, \tag{7}$$

where $*$ represents the complex conjugate. Lastly, based on the phase slope and the n_0 position in current frame, a linearity index is defined as:

$$LI_{\theta}(k) = \Delta\theta(k) - \frac{-2\pi n_0}{N}. \tag{8}$$

The linearity index varies significantly between the FiSe and residual noise. However, its magnitude for tap/drag sound is similar to the FiSe, implying that they are of similar phase. Therefore, we employ the last processing step to select optimal FiSe events.

iii) Duration Verification: The sequence of occurrences with a high magnitude linearity index differs between the tap/drag sound and the FiSe. Even in cases where a user has a soft fingertip with low SPL, the duration of FiSe would still be distinct. Based on the insights from HMM model and the linearity index, we conduct a final check by removing the segments whose duration does not lie from 0.05 to 0.3 seconds.

It is worth noting that the proposed event detection is applicable for acquiring FiSe across different smart devices and surfaces (see Section 9.1) since it does not assume the swiping behavior of users.

6 TAXONOMY OF ACOUSTIC FINGERPRINT

The uniqueness of friction-excited sonic wave is dependent on the texture of contact surface, i.e., the fingerprint. As shown in Section 2.2, Level I friction descriptors are not sufficient since they can only relate to Level I optical fingerprint patterns. To this end, we propose a novel taxonomy (see Figure 6) that bridges the gap between Level II and III fingerprint patterns and acoustics to select valid features for FiSe classification.

6.1 Level II Friction Descriptors

In the fingerprint domain, Level II features involve Galton characteristics, also known as minutiae points (e.g., hooks and bifurcations). These features possess a high variance between fingerprints of different users and are actively used in classification models. For the discrimination of audio sources, features such as the mel-frequency cepstral coefficients (MFCC) are essential since they can capture the timbral characteristics. Other cepstral features generally employ the perceptual filter bank and autoregression model to approximate the spectral envelope. Based on this semantic relationship, for the Level II friction descriptors, we select 14 MFCC (with Δ and ΔΔ), 12 linear prediction cepstral coefficients (LPCC) and 27 perceptual linear predictions (RASTA-PLP [26]). These descriptors can provide insights into the minutiae features of the fingerprint.

6.2 Level III Friction Descriptors

Although being unique, Level II fingerprint features are prone to spoofing since they could be visually perceived through the naked eye or even in low-resolution images. Thus, Level III fingerprint features are proposed based on the dimensional ridge information, including width, pores and edge contour. Similarly, short-time fourier transform and adaptive time-frequency decomposition can reveal various physical attributes of FiSe. These features have inferior meaning to human perception [27] and thus are difficult to spoof. To reveal the intrinsic fingerprint from FiSe, we select 12 linear prediction coefficients (LPC), 12 linear spectral frequencies (LSF), 26 log filter bank and spectral statistics (i.e., flux, kurtosis, skewness and slope) as Level III friction descriptors. Besides, we also employ 16 wavelet cross-level coefficients and 32x20 C1 & C2 amplitude spectra [28] relating to the texture of FiSe.

6.3 SonicPrint Identification

Two-check Feature Selection: Majority of the feature selection methods [29, 30] focus on finding *minimal-optimal* subset based on the classification accuracy. Yet, the limited accuracy for a specific model is not sufficient to confirm a feature as irrelevant. Therefore, we employ Boruta algorithm [31] to determine the *all-relevant* features for FiSe classification. It relies on the computationally efficient Random Forest classifier to iteratively discard the less relevant features. We utilize a two-step correction, i.e., Benjamini Hochberg FDR for evaluating features against random and Bonferroni correction [32] for testing identical features repeatedly. After applying the feature selection on our multi-level friction descriptors, the majority of features are chosen, except C1 & C2 amplitude spectra. The feature vector, initially of 802 features, is reduced to ***162 friction descriptors*** and fed to our classification model.

Ensemble Classifiers: As the first exploratory study using FiSe for biometrics, we employ the following prediction models which have shown superior performance in user identification [33–35]:

- Logistic Regression (LR): It models the outcome through logistic sigmoid function to deliver a probability measure that is mapped to a specific class. We set the maximum iterations as 1000 and a cross-entropy loss for multi-class problem.

- Support Vector Machine (SVM): It is a statistical learning method with linear kernel that determines an optimal hyperplane to divide classes by maximizing the margin between the closest points.
- Random Forest (RF): It fits specific decision tree classifiers on the sub-samples and employs averaging to reduce overfitting. We set the estimators as 200 and use an entropy criterion for prediction.
- Linear Discriminant Analysis (LDA): By utilizing the Bayes' rule and approximating class conditional densities to samples, it creates a linear decision boundary to separate the classes. We select singular value decomposition as the solver.
- Gaussian Mixture Model (GMM): It provides a parametric probability distribution of audio signal and related features and characterizes the weighted sum of Gaussian components as a density function. We assume 5 components in our model.

From our empirical analysis, LDA is most suited for FiSe classification, followed by RF and SVM. Therefore, we assign a weight to each classifier (LR, SVM, RF, LDA, GMM) as 1, 2, 2, 3, 1, respectively. Finally, we perform hard voting on the observations generated from the classifiers to decide the legitimate user.

7 EVALUATION SETUP

7.1 Experimental Settings

We conduct a pilot study to validate the uniqueness of FiSe caused by the swipe motion on a smartphone. From reviewing the recent development in touch-based biometrics [36], we observe that two swipe actions are the most convenient and acceptable among users, as shown in Figure 7. **1Hand Swipe:** a user holds his phone naturally in right-hand and uses the index finger of the same hand to swipe on the surface. **2Hand Swipe:** left-hand firmly holds the phone while the other is used to perform the swipe. The 2Hand swipe is more robust to artifacts and allows for precise stroke capture. To provide a better understanding of the experimental process, we create a code to describe the performed swipe action. The code comprises of three parts, i.e., **Swipe-Sensing Distance-Surface**. The swipe could vary between 1Hand and 2Hand; sensing distance differs among 1cm, 7cm or 11cm from inbuilt microphone; and surface could be aluminum, glass or others. Our experimental setup for the pilot study involves the participants to sit on a chair in a conference room with low ambient noise. The participants are asked to perform *1Hand-7cm-aluminum* swipes in a straight-downward direction on the back of the smartphone. Afterward, they are required to complete *2Hand-1cm-glass* swipes at the front of the smartphone. *To ensure that the obtained insights are applicable in real-world scenarios, physical attributes (i.e., speed, pressure or roughness) of the finger are not controlled during the swipe action, throughout the remainder of this paper.* We employ the Google Pixel 2 smartphone with a 0-22KHz range microphone to record the FiSe caused by the swipe action. It is 14.4cm(5.7$inch$) x 6.8cm(2.7$inch$) x 1.5cm(0.6$inch$) in size and weighs only 161.5g, which is lightweight for easy use in daily life. It works on a Qualcomm Snapdragon 835 with an Octa-Core processor. The recorded signal is fed to *SonicPrint* for further analysis.

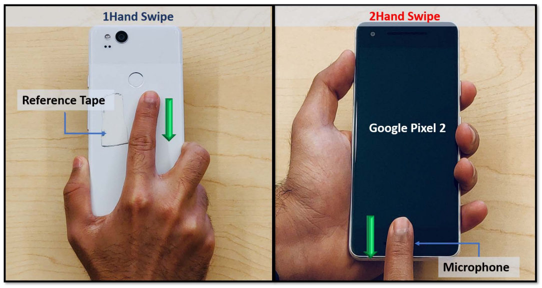


Figure 7: The evaluation setup with subject performing 1Hand and 2Hand swipe on the smartphone surface with right index finger.

7.2 FiSe Collection and Partition

As the first exploration of utilizing FiSe for user identification, we recruit 31 subjects (25 males and 6 females) within the age-group of 18-50 years in our study. None of the subjects have any damage to their fingerprint. For both the experiments involving 1Hand and 2Hand swipes, every subject performs six trials each. In each trial, the subject swipes at the specific position 30 times continuously. A 15min break separates every two consecutive trials to ensure non-uniform speed and pressure during swipes. Furthermore, the six trials for each experiment are spread across three weeks. A trial consists of 1min recording for each person. In total, every subject performs 180 *1Hand-7cm-aluminum* and 180 *2Hand-1cm-glass* swipe actions. The generated FiSe is recorded by the inbuilt microphone (sampling rate of 44.1KHz) and later fed to *SonicPrint*. After denoising and segmentation, a total of 4099 1Hand swipes (~130 per participant) and 4405 2Hand swipes (~140 per participant) are selected for training and testing. A 10-fold stratified cross-validation approach is applied to normalized features during user identification. The reason behind choosing stratified approach relates to the bias in classification models. During prediction, every instance is weighted equally, implying that a few over-represented classes can dominate the evaluation metrics. Thus, a stratified model ensures that each fold in cross-validation is representative of the whole dataset, thereby optimizing the bias and variance [37]. We employ other cross-validation and direct matching algorithms, in Section 9, to evaluate the inclusiveness of *SonicPrint* in real-world scenarios. **Evaluation Metrics:** We introduce balanced accuracy (BAC), F-score, equal error rate (EER) and receiver operating characteristics (ROC) curve as metrics in our evaluation model. They are insensitive to class distribution which is critical for identification schemes.

7.3 SonicPrint Usability & Social Acceptance

SonicPrint requires the users to naturally swipe on their smartphone cover to acquire the unique FiSe. To assess the practicality and acceptance of *SonicPrint* in the real-world, we surveyed the 31 participants recruited in our pilot study. Of all the 31 participants, 80% are male and 20% are female. After completing the experiments, we ask the participants a few questions regarding their experience with our system. 71% of them preferred to perform 2Hand swipes on the front surface of the smartphone, while 29% preferred 1Hand swipes on the back cover. On a scale of 1 to 10, all the participants are requested to rate the comfortability while performing multiple

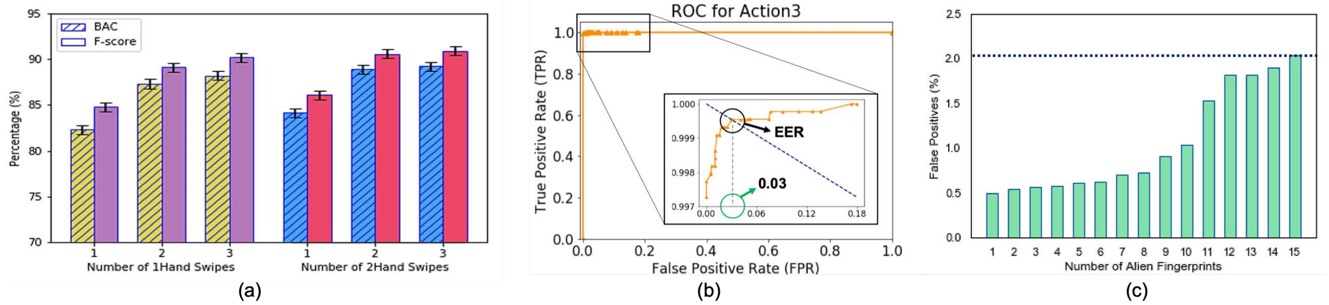


Figure 8: The performance between (a) *Action1, Action2*; (b) *Action3*; (c) *Action3* (unsupervised).

swipe actions. We record an average score of 9.35, validating the ease-of-use of *SonicPrint*. Furthermore, we employ a 4-point Likert scale (ranging from Strongly Disagree to Strong Agree) [38]. This scale determines the participant's willingness to adopt *SonicPrint* in daily life for unlocking a smartphone or accessing protected information. 80% of the participants answered with a score of 4 points, while the rest gave a score of 3 points. These results show high acceptance of *SonicPrint* among subjects, especially when made aware of the threats in traditional fingerprint scanners.

8 ACCURACY & RELIABILITY STUDY

As a potential breakthrough technology, it is critical to evaluate the performance and reliability of *SonicPrint*. Our smartphone-based pilot study comprises user identification using FiSe obtained from two actions: (1) *Action1: 1Hand-7cm-aluminum* swipes; (2) *Action2: 2Hand-1cm-glass* swipes. For each action, we make a comparison of evaluation metrics by increasing the number of swipes per sample performed by the user.

i) Action1 performance: After performing 10-fold stratified cross-validation on 4099 samples, the observed BAC and F-score are shown in Figure 8(a). The number of inputs, i.e., swipes per sample is increased from one to three and the variation in performance is recorded. The BAC for 1, 2 and 3 inputs is 82.3%, 87.3% and 88.2% while the F-score is 84.8%, 89.1% and 90.2% respectively. From ROC curve, the area-under-curve (AUC) is observed to be 85.3%, 89% and 88.6% as swipes per sample increases.

ii) Action2 performance: We report the BAC and F-score for 10-fold stratified cross-validation on 4405 samples in Figure 8(a). The BAC for 1, 2 and 3 inputs is 84.15%, 88.9% and 89.2% while the F-score is observed as 86.1%, 90.6% and 90.9% respectively. We compute AUC as 85.8%, 88.2% and 88.7% for increasing inputs. For *Action1* and *Action2*, the performance improves by augmenting more swipes per access attempt.

Performance Reliability: To ensure that the observed performance is not dependent on the size of training and testing dataset, we vary the number the splits in K-fold (from 3 to 10) and note the results. For both *Action1* and *Action2*, the BAC and F-score remain stable, within a margin of ±2%, exhibiting the reliability of *SonicPrint* even under less amount of training samples.

Insights: While the previous results demonstrate the uniqueness of FiSe as a biometric trait, they also provide vital clues to improve *SonicPrint*. One reason for the lower performance of 1Hand (*Action1*) to 2Hand (*Action2*) swipes is due to its sensing distance from the microphone. A close proximity of swipe action with microphone ensures high SNR and allows for more precise capture of the FiSe. The 2Hand swipes provide a superior control to the users to ensure that their fingerprint properly interacts with the opposing surface. A rich textural material facilitates strong coupling between the fingerprint and surface to produce a more distinct FiSe. Since the glass material in *Action2* is a smooth surface, the performance can be further enhanced by selecting a more suitable material to interact with the fingerprint.

iii) Action3 performance: Based on these insights, we conduct another experiment, *Action3*, to analyze the *SonicPrint* performance under ideal conditions. We place the smartphone in a common protective case made from synthetic leather and ask the 31 subjects to perform *2Hand-1cm-leather* swipes. We collect 4572 FiSe during swipe events and perform 10-fold stratified cross-validation. The BAC and F-score for one swipe per sample is 98.3% and 98.4%, respectively. Figure 8(b) shows the ROC curve where the observed EER and AUC are 0.03 and 97.5%. We examine the performance reliability by changing the splits in K-fold from 3 to 10, with results showing a ±1% variation in scores.

Alien Fingerprint: To examine the vulnerability of *SonicPrint* against alien fingerprints (i.e., samples not trained in advance), we randomly choose 16 subjects and train the model using their *2Hand-1cm-leather* swipes. The remaining 15 subjects are used for testing in Figure 8(c). Our system can successfully reject the alien fingerprints using the threshold value of classification score. The results prove our insights and confirm that the users can be precisely recognized by *SonicPrint*.

Identification vs Authentication: A conventional fingerprint scanner in smart devices grant access to a user by matching his input to a pre-trained template. This task is similar to binary classification in authentication problems [39]. Our previous results show the capability of *SonicPrint* to perform a more challenging task of user identification (in other words, multi-class classification) which is desirable in the IoT environment (e.g., smarthome). Nevertheless, we also evaluated *Action1* and *Action2* performance for user authentication (i.e., each subject is compared against others, in a one-against-one fashion) to observe comparable evaluation metrics (+2%). Furthermore, we vary the number of randomly selected subjects from 2 to 30 and note the BAC score in Figure 9. As the number of subjects increase, the performance decreases. An interesting observation is that after 15 subjects, our model learns to effectively determine features that can accurately differentiate

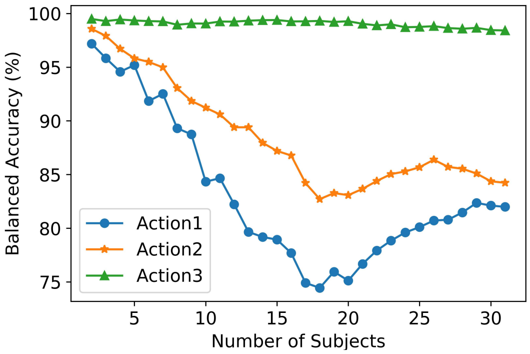


Figure 9: The trend of balanced accuracy with increasing number of subjects.

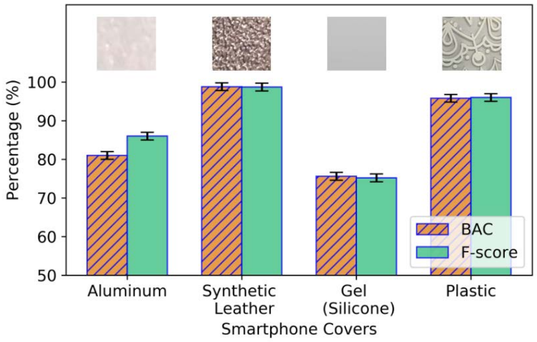


Figure 10: Evaluation among diverse surfaces.

the subject-specific FiSe. A comprehensive evaluation of relative entropy in FiSe can be a lucrative venue for future work.

9 INCLUSIVENESS STUDY

In this section, we consider several factors that could affect the ability of *SonicPrint* to identify the users. *The existing biometric technologies are inapplicable under many of these conditions, e.g., smartwatch or smart assistant cannot support faceID while traditional fingerprint does not work under moisture.* For the following evaluations, we highlight the base performance by using one swipe per sample. While our results are applicable across all fingers, we ask subjects to use their right index finger during experiments.

9.1 Surface Exploration

Impact of Surface Texture: The uniqueness of FiSe relies on the user's fingerprint and the contact surface. Although fingerprint possesses a fixed composition, the opposing surface may vary from a smooth to a coarse texture. To analyze the effect of surface texture, we recruit 10 subjects and ask each of them to perform 100 *1Hand-7cm* swipes on four common smartphone covers, i.e., aluminum, synthetic leather, gel (silicone) and plastic. The BAC and F-score are shown in Figure 10.

Insights: The results confirm that current implementation of *SonicPrint* is more suitable for rough surfaces (e.g., synthetic leather) than smooth surfaces (e.g., gel). Although plastic has a rigid but smooth surface, its superior performance is due to the engravings in material on which the swipe action is performed. The material texture influences the SPL of FiSe; a high sensitive microphone is required to increase accessibility of our system across smooth surfaces. *SonicPrint* can drive a new form of user identification using surfaces with satisfactory texture.

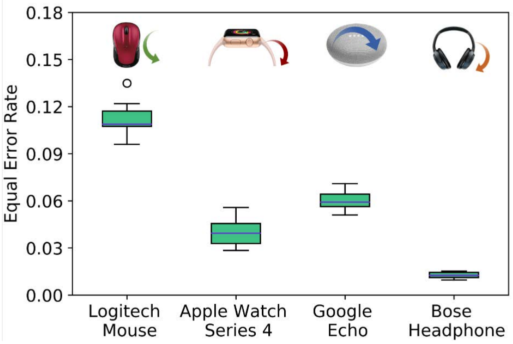


Figure 11: The EER for curved smart devices.

Impact of Surface Geometry: *SonicPrint* would be highly valuable if it can be deployed across all smart devices, regardless of their geometric structure. We investigate the FiSe on four popular smart devices with an increasing level of curvature: Bose Headphones, Google Echo, Apple Watch Series 4 (leather strap) and Logitech mouse. The Logitech mouse comprises an inward surface while the rest are outward. As a first exploratory study, we position the microphone in Google Pixel 2 near the surfaces of considered smart devices and record the FiSe during swipe action. From 5 subjects, an overall of 1981 FiSe are collected. For each device, the K-Fold splits during the cross-validation is varied from 4 to 10 and the EER is illustrated in Figure 11.

Insights: The performance of *SonicPrint* depends on the curvature of smart devices, with lower EER for smaller diameter. The higher EER of a mouse is due to its larger curvature since it is challenging to maintain the entire fingerprint in contact with the surface during the swipe action. The influence of curved geometry can be reduced by controlling the swiping speed to generate a stronger coupling between the fingerprint and devices.

9.2 Fingerprint Sensitivity

Impact of Partial Fingerprint: In practical scenarios, a higher degree of freedom during swipe action would result in different portions of a fingertip to interact with the contact surface. To examine the system performance with prominent partial fingerprints, we ask 10 subjects to individually perform 400 *2Hand-1cm-glass* swipes on the smartphone. For every 100 swipes, the subject interacts with different regions (i.e., full, right, tip and left) of their fingertip.

Insights: Figure 12 shows that the accuracy of identifying users depend on the coverage area of their fingertip. Due to the challenge in executing a consistent tip or left swipe with the right index finger, subjects unconsciously face more area of their fingerprints towards the contact surface. Furthermore, subjects reveal that their fingernails collide with the surface during the tip swipes, thereby causing distortions in the FiSe. These signals are typically disregarded by our system, thus requiring the users to perform multiple attempts in case the tip is utilized. Nevertheless, *SonicPrint* is sensitive to partial fingerprints and its performance can be further improved by ensuring suitable contact surfaces.

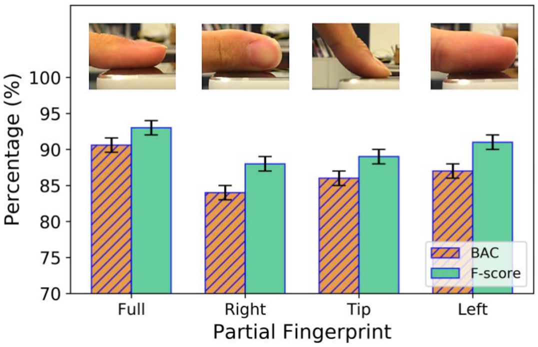


Figure 12: Evaluation of partial fingerprint.

Impact of Moisture: It is a known fact that the presence of moisture on fingertip adversely affects the sensing capability of traditional fingerprint scanners. Considering FiSe relies on the fingerprint, it is necessary to evaluate its sensitivity to moisture. We experiment with 5 subjects where each of them performs 300 *2Hand-1cm-aluminum* swipes on the smartphone. For first 100 swipes, the fingertip of subjects is in dry state. In next 100 swipes, the finger is placed inside a glass of water and then the moist fingertip is used to swipe. Finally, excessive lotion is rubbed on the subject's fingertip.
Insights: Table 1 demonstrates that excessive moisture lowers the performance of *SonicPrint*. However, we achieve satisfactory results if the moisture level is equivalent to the water. The reason is that a few consecutive swipes allow the water to be removed from the fingertip surface, thereby having minimal difference against a dry fingertip. In comparison, an expensive and high resolution sensor (e.g., Cross Match Technologies PartolID) demonstrates a 39∼56% false matching rate due to moisture [40]. *SonicPrint* shows a better tolerance to moisture compared to existing fingerprint scanners.

9.3 Swipe Dynamics

Impact of Swipe Pattern: Although our system employs the uniqueness of FiSe for user identification, it would be ideal if one can integrate a secondary dimension of soft characteristics, i.e., the swipe pattern (similar to pattern password) to the biometric trait. We investigate the performance under diverse swipe patterns in Figure 13. For this experiment, each of 10 subjects is asked to complete 400 *2Hand-7cm-aluminum* swipes on the smartphone. For every 100 swipes, four widely used motion patterns are performed, including line, zig-zag, circle and star, with increasing level of complexity.
Insights: In contrast to the expected scenario, the performance of *SonicPrint* is proportional to the complexity of swipe pattern. Raising the complexity also increases the length of recorded FiSe, ensuring that sufficient user-specific information is present in each sample. Among other variables, the system performance can be easily improved by reducing the distance between the swipe action and the microphone.

Table 1: *SonicPrint* resilience to moisture.

	Dry	*Water*	*Lotion*
BAC	92.7%	90%	60.1%
F-score	93.8%	87%	68%

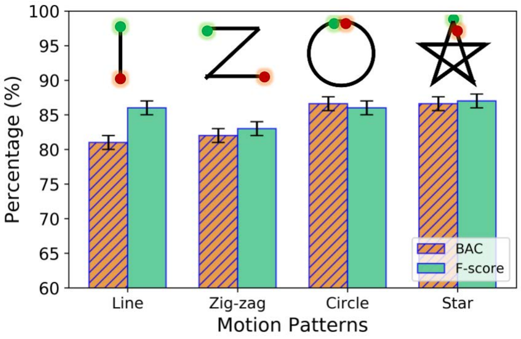


Figure 13: Evaluation among swipe patterns.

Impact of Sensing Location: As the size and dimensionality of smart device increases, the distance of swipe action might vary with the inbuilt microphone. Besides, different users would prefer different locations based on the structure of their palm while holding the smartphone. Therefore, we investigate the performance of *SonicPrint* at three locations, i.e., high, middle and low at a distance of 11*cm*, 7*cm* and 1*cm* from the inbuilt microphone, respectively. The smartphone (Google Pixel 2) is placed inside the synthetic leather cover to ensure consistency of material across locations. Each of 5 subjects performs 100 1Hand swipes at every location.
Insights: In Figure 14, we observe the identifiability of users depends on the location of swipe action with respect to the inbuilt microphone. It is worth mentioning that if the training dataset comprises of swipes from high or low locations, it is still possible to identify users with newer swipes at the middle location. The lower cross-performance is due to the distinct surface texture of synthetic leather at a specific location.

9.4 Acoustic Noise Resilience

The low sound pressure level (SPL) of FiSe brings a trade-off between high security and precise sensing. In an environment with dynamic noises, a high-order cutoff filter can reduce the background clutter, but also leads to information loss. To determine the capability of *SonicPrint* in various acoustic noises, we inform 5 subjects to do 150 *2Hand-1cm-glass* swipes on the smartphone at three noise levels: (1) **Stationary noise (23.6dB):** inside a running car; (2) **Human motion noise (29.4dB):** subject walking around with a phone, within 3*m* distance; (3) **Voice, music and environment noise (40.9dB):** windy outdoors with environmental temperature

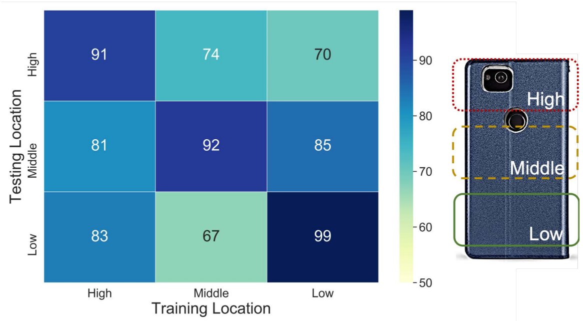


Figure 14: The BAC for sensing location.

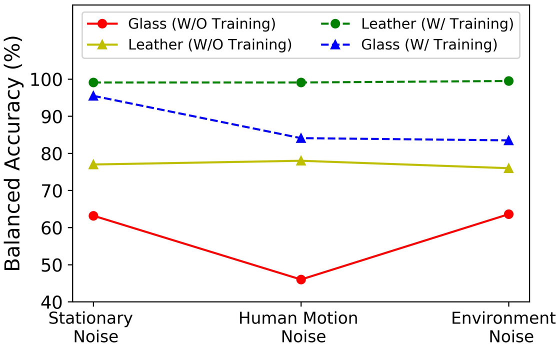


Figure 15: Evaluation under acoustic noises.

at 23°C, humans talking and music playing in the background. Although 2.2KHz high-pass cutoff is suitable for indoor ambience, it is not sufficient to eliminate loud noises. From empirical observations, we fix the cutoff to 4.4KHz for stationary engine noise and human motion noise and 11KHz for loud noise primarily from winds. We repeat the experiment on synthetic leather cover and illustrate the outcomes for with/without noise training in Figure 15.

Insights: We observe that the surface texture plays a vital role during sensing in presence of acoustic noises. Yet, accuracy is adversely affected when the model is only trained in control environment. *SonicPrint* can precisely identify users with diverse training samples. We discuss the possible enhancements in Section 11.

10 VULNERABILITY STUDY

10.1 Fingerprint Phantom Attack

We assume that Alice has access to the fingerprint and other geometrical characteristics (e.g., width, thickness) of left index finger of a legitimate user. Using this information, she aims to build a replica of the target's finger and breach the biometric security. Among accessible spoofing materials [41], Alice utilizes gelatin which can most closely relate the texture of live finger [6] and can even spoof capacitive fingerprint scanners [8]. To explore this, we recruit 5 subjects having fingers of various sizes and execute these steps:

- We ensure that the entire finger of each subject is covered by multiple layers (5 to 8) of latex material.
- Between each successive layer, we wait for 10 minutes to lose the moisture; the finger is kept still so that no pressure marks or creases occur on the coating.
- Once the latex coating becomes firm, we gently enclose it with baking powder as we remove the latex.
- We prepare a mixture of one part gelatin, glycerin and water and use a conventional microwave to heat the mixture. Finally, we pour the mixture inside the recovered latex coating and leave it to dry for 24 hours. The latex coating is then discarded to obtain the gelatin fake-finger, as illustrated in Figure 16.

We ask each subject to use their live left index finger and perform 100 *2Hand-7cm-aluminum* and 100 *2Hand-1cm-glass* swipes on the smartphone. Afterward, we repeat the process by informing subjects to utilize their fake-fingers to complete swipe actions. We train the *SonicPrint* on recordings from live fingers and test fake-fingers during identification. For the fake-finger recordings, we observe that our pre-processing module discards 300 (out of 500) aluminum and 450 (out of 500) glass FiSe. Out of the remaining, only 32 (6.4%) aluminum and 21 (4.2%) glass FiSe are misclassified as live fingers.

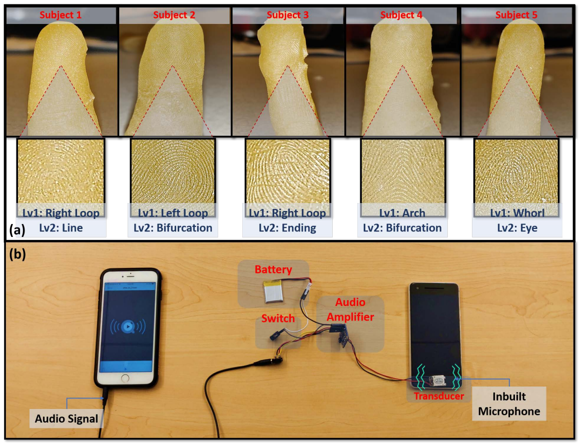


Figure 16: (a) Gelatin fake fingers with multi-level fingerprint textures; (b) vibration injections via audio transducer.

10.2 Replay and Side-Channel Attack

We assume that Alice knows the underlying mechanism of *SonicPrint* to sense the sonic waves for user identification. Through a high-resolution camera, Alice can acquire the victim's fingerprint from a distance of 2*m* [42]; however, no FiSe can be obtained from a similar distance due to its low SPL. Therefore, we envision an unrealistic scenario where she leverages a high-sensitive microphone (i.e., Fifine-K670) and positions it at very close proximity of 20*cm* and 30*cm* facing the target smart device. The microphone captures the FiSe during an access attempt by a legitimate user.

Attack via Microphone: the recording is replayed to the inbuilt microphone of target smartphone by direct FiSe replay. Overall, 4 subjects conduct 500 *2Hand-7cm-aluminum* swipes on Google Pixel 2 and the inbuilt and secondary microphone concurrently records the FiSe. For attack through a direct transfer, merely 4.8% and 3.2% of replayed FiSe match with the original recording, even at a close distance of 20*cm* and 30*cm* respectively.

Attack via Vibration Channel: we consider a scenario where Alice attempts to forge the swipe action of legitimate user as vibration signals for identification. The pre-recorded audio signal is passed through the coil of transducer and a dynamic electromagnetic field is generated which makes the actuator vibrate the smartphone (see Figure 16). Although these vibrations are propagated from a very close distance (i.e., top of smartphone), all are rejected by *SonicPrint*, making side-channels attacks via hidden transmitters ineffective.

10.3 Potential Vulnerabilities

Section 10 examines the anti-spoofing capability of *SonicPrint* against intricate presentation attacks. We discuss other physical attacks and conditions that may affect the biometric capability below:

Denial-of-service: Observing the low SPL of FiSe, Alice can leverage additional speakers to project white noise towards the target microphone while Bob is performing the swipe action. Unlike additive noise, a multiplicative signal can be utilized which raises the challenge for *SonicPrint* pre-processing module. However, any audible noise would be noticeable to Bob, decreasing the stealthiness of attack. To this end, Alice could employ inaudible ultrasound signals [43] which exploits the loophole in hardware non-linearity; the traces left after the attack are challenging to remove in the recorded FiSe signal. We consider this as a potential scope for future research.

Surface Texture: In this paper, we have tested seven widely used materials (i.e., aluminum, synthetic leather, glass, silicone, plastic, fiber (Google echo) and rubber (Mouse)) on smart devices. Learned from the rationale, every material should have a unique coupling with the user's fingerprint. Although *SonicPrint* can be applied to other materials including paper, wood and textiles, the trade-off between usability and security needs to be explicitly considered (for instance, a high audible FiSe is more usable but less secure).

11 DISCUSSION

Aging Effects: For every biometric trait, a different degree of variation occurs over time. We ask 5 subjects to each complete 100 *2Hand-1cm-glass* swipes on the smartphone. These trials are performed every week for two months. We train the *SonicPrint* on the recordings from the first week and test the remaining dataset. We observe a sharp drop in the performance (by 38%) after three weeks, which continues to reduce over time. The reason is due to employing the same smartphone in all performance studies, causing accelerated aging. In real practice, every user has their own personalized devices. Moreover, specialized materials that are more resistant to aging can be used for superior longevity.

Microphone Sensitivity: *SonicPrint* leverages the low-cost microphone of smartphone for FiSe acquisition. Although our system shows a satisfactory performance under ideal conditions, the overall results can be significantly improved by adopting high sensitive microphones. These microphones can precisely detect FiSe from even swipe actions on smooth surfaces in a noisy environment. Users would not be required to perform the swipe as close to the microphone, increasing the level of freedom and user acceptance.

Accuracy and Improvements: *SonicPrint* achieves 84% and 98% identification rates with a single trial on standard and high-texture smartphone surface, respectively. This is comparable to recent low-cost solutions using vibrations [44, 45], gait patterns [46] and passive sensing [47] for authentication. Yet, the most significant contribution of *SonicPrint* is its adoptability across smart devices (refer to Section 9.1) which is not supported by existing solutions. Our proposed approach can also be used as secondary biometrics; improvements in microphone frequency response and deep learning approaches can be considered for our future exploration.

System Considerations: As a starting point, *SonicPrint* is a promising biometric with high adoptability and anti-spoofing capabilities. However, a practical deployment in the real-world requires reflection on following criteria: (1) *Privacy:* The audible nature of FiSe makes it prone to theft via a conventional recording device. For a countermeasure, the user can be asked to perform a specialized gesture (e.g., zig-zag or star pattern in Section 9.3) during the training process. These gestures are uncommon in normal user behavior, thereby increasing the difficulty for an attacker to acquire the target FiSe outside the recognition period. (2) *Power consumption:* The current power consumption relies on the microphone (<100mW); we envision that a touch trigger can be employed to activate FiSe recording, thereby limiting battery usage in smart devices. (3) *Recognition time:* By employing computationally inexpensive algorithms, *SonicPrint* can identify a user in less than 1 second period, further facilitating its deployment in smart devices.

12 RELATED WORK

Touch-based Biometrics: Touch-based implicit authentication relies on the unconstrained movement patterns of users when they interact with their smartphone. The location of finger taps could be inferred from the motion sensors [48, 49]. Based on this insight, the touch dynamics was explored as a soft biometric trait for user authentication [50, 51]. Different parameters such as the rhythm, strength, angle of applied force [52] or the size and axis length of finger touch area [53] can depict the user's individuality. Despite the enhancements in security [54–56], it was shown that mimicry attacks have a bypass rate of 86%, even with partial knowledge of the underlying features of touch biometric [57]. Recently, researchers have employed induced body electric potentials (iBEP) or body guided communications as a new biometric [58, 59]. However, it requires the user to continuously wear a token device and can be spoofed through injection attacks. Our method relies on the uniqueness of fingerprint and cannot be spoofed via mimicry or side-channel attacks.

Acoustic Sensing: In 2011, researchers proposed that the acoustic signatures caused by an object impacting with a screen surface could identify its type (i.e., fingernail, knuckle, tip) [60]. Afterward, the domain of acoustics-based touch interaction was enhanced by monitoring continuous sound via structure-borne sound propagation [61] for inferring the finger tapping and movements of the user [62]. When a vibration motor excites a surface, the presence of devices [63] or user-specific gestures [44] can be sensed by the inertial sensors. However, these approaches have limited accessibility due to the requirement of additional vibration transmitters and receivers and more importantly, are vulnerable to the Denial-of-Service (DoS) attacks. A recent study captures the finger sound caused by thumb rubbing the finger for gesture recognition [64], yet requires the user to wear a ring during the sensing process. To the best of our knowledge, we provide the first study on exploring the intrinsic fingerprint information in friction-excited sonic waves for secure user identification.

13 CONCLUSION

Existing fingerprint biometric is vulnerable to spoofing attacks (e.g., fake-fingers) and cannot be adopted in upcoming smart devices due to hardware constraints. In this paper, we introduce a new dimension of fingerprint sensing using the friction-excited sonic wave caused by a fingerprint to surface interaction. We develop *SonicPrint* that utilizes the FiSe from a user swiping his fingertip on everyday smart devices for identification. The system is adoptable, user-friendly and difficult to counterfeit with an identification accuracy up to 98%. We also show the inclusiveness of *SonicPrint* under partial fingerprints, motion patterns and surface geometry. In the future, we aim to consider users having damaged fingerprints while exploring high-sensitive microphones with ultrasonic range to improve the system accuracy.

ACKNOWLEDGEMENTS

We thank our shepherd, Dr. Philip Levis, and all anonymous reviewers for their insightful comments on this paper.

REFERENCES

[1] K. Conger, R. Fausset, and S. F. Kovaleski, "San francisco bans facial recognition technology," May 2019. [Online]. Available: https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html

[2] S. Bharadwaj, H. S. Bhatt, M. Vatsa, and R. Singh, "Periocular biometrics: When iris recognition fails," in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, 2010, pp. 1–6.

[3] L. Zhang, S. Tan, and J. Yang, "Hearing your voice is not enough: An articulatory gesture based liveness detection for voice authentication," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017, pp. 57–71.

[4] Gartner, "The future smart home: 500 smart objects will enable new business opportunities." [Online]. Available: https://www.gartner.com/en/documents/2793317

[5] A. Ross and A. Jain, "Biometric sensor interoperability: A case study in fingerprints," in *International Workshop on Biometric Authentication*. Springer, 2004, pp. 134–145.

[6] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of artificial" gummy" fingers on fingerprint systems," in *Optical Security and Counterfeit Deterrence Techniques IV*, vol. 4677. International Society for Optics and Photonics, 2002, pp. 275–289.

[7] S. S. Arora, K. Cao, A. K. Jain, and N. G. Paulter, "3d fingerprint phantoms," in *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 684–689.

[8] H. Kang, B. Lee, H. Kim, D. Shin, and J. Kim, "A study on performance evaluation of the liveness detection for various fingerprint sensor modules," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, 2003, pp. 1245–1253.

[9] D. Winder, "Samsung galaxy s10 fingerprint scanner hacked - here's what you need to know," Apr 2019. [Online]. Available: https://www.forbes.com/sites/daveywinder/2019/04/06/samsung-galaxy-s10-fingerprint-scanner-hacked-heres-what-you-need-to-know/#10c88305d423

[10] "Mems microphones market size, share: Industry trends report, 2025." [Online]. Available: https://www.grandviewresearch.com/industry-analysis/mems-microphones-market

[11] A. Akay, "Acoustics of friction," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1525–1548, 2002.

[12] B. L. Stoimenov, S. Maruyama, K. Adachi, and K. Kato, "The roughness effect on the frequency of frictional sound," *Tribology international*, vol. 40, no. 4, pp. 659–664, 2007.

[13] H. B. Abdelounis, A. Le Bot, J. Perret-Liaudet, and H. Zahouani, "An experimental study on roughness noise of dry rough flat surfaces," *Wear*, vol. 268, no. 1-2, pp. 335–345, 2010.

[14] A. K. Jain, Y. Chen, and M. Demirkus, "Pores and ridges: High-resolution fingerprint matching using level 3 features," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 1, pp. 15–27, 2006.

[15] C. Barral and A. Tria, "Fake fingers in fingerprint recognition: Glycerin supersedes gelatin," in *Formal to Practical Security*. Springer, 2009, pp. 57–69.

[16] H. Feng, K. Fawaz, and K. G. Shin, "Continuous authentication for voice assistants," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. ACM, 2017, pp. 343–355.

[17] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise." in *ICASSP*, vol. 4. Citeseer, 2002, pp. 44 164–44 164.

[18] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2015, pp. 1472–1480.

[19] D. B. Percival and A. T. Walden, *Wavelet methods for time series analysis*. Cambridge university press, 2006, vol. 4.

[20] Q. Li, J. Zheng, A. Tsai, and Q. Zhou, "Robust endpoint detection and energy normalization for real-time speech and speaker recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 3, pp. 146–157, 2002.

[21] Y. Bi, M. Lv, C. Song, W. Xu, N. Guan, and W. Yi, "Autodietary: A wearable acoustic sensor system for food intake recognition in daily life," *IEEE Sensors Journal*, vol. 16, no. 3, pp. 806–816, 2015.

[22] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on speech and audio processing*, vol. 9, no. 5, pp. 504–512, 2001.

[23] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE signal processing letters*, vol. 6, no. 1, pp. 1–3, 1999.

[24] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

[25] A. Sugiyama, R. Miyahara, and K. Park, "Impact-noise suppression with phase-based detection," in *21st European Signal Processing Conference (EUSIPCO 2013)*. IEEE, 2013, pp. 1–5.

[26] H. Hermansky and N. Morgan, "Rasta processing of speech," *IEEE transactions on speech and audio processing*, vol. 2, no. 4, pp. 578–589, 1994.

[27] D. Mitrović, M. Zeppelzauer, and C. Breiteneder, "Features for content-based audio retrieval," in *Advances in computers*. Elsevier, 2010, vol. 78, pp. 71–150.

[28] H.-S. Kim and J. Smith, "Synthesis of sound textures with tonal components using summary statistics and all-pole residual modeling," in *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx-16)*, 2016, pp. 129–136.

[29] J. Neumann, C. Schnörr, and G. Steidl, "Combined svm-based feature selection and classification," *Machine learning*, vol. 61, no. 1-3, pp. 129–150, 2005.

[30] A. Janecek, W. Gansterer, M. Demel, and G. Ecker, "On the relationship between feature selection and classification accuracy," in *New challenges for feature selection in data mining and knowledge discovery*, 2008, pp. 90–105.

[31] M. B. Kursa, W. R. Rudnicki, *et al.*, "Feature selection with the boruta package," *J Stat Softw*, vol. 36, no. 11, pp. 1–13, 2010.

[32] S. R. Narum, "Beyond bonferroni: less conservative analyses for conservation genetics," *Conservation genetics*, vol. 7, no. 5, pp. 783–787, 2006.

[33] C. Bo, L. Zhang, X.-Y. Li, Q. Huang, and Y. Wang, "Silentsense: silent user identification via touch and movement behavioral biometrics," in *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 2013, pp. 187–190.

[34] J. Angulo and E. Wästlund, "Exploring touch-screen biometrics for user identification on smart phones," in *IFIP PrimeLife International Summer School on Privacy and Identity Management for Life*. Springer, 2011, pp. 130–143.

[35] S. Kwon and S. Narayanan, "Robust speaker identification based on selective use of feature vectors," *Pattern Recognition Letters*, vol. 28, no. 1, pp. 85–89, 2007.

[36] Z. Ali, J. Payton, and V. Sritapan, "At your fingertips: Considering finger distinctness in continuous touch-based authentication for mobile devices," in *2016 IEEE Security and Privacy Workshops (SPW)*. IEEE, 2016, pp. 272–275.

[37] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, no. 2. Montreal, Canada, 1995, pp. 1137–1145.

[38] S.-O. Leung, "A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point likert scales," *Journal of Social Service Research*, vol. 37, no. 4, pp. 412–421, 2011.

[39] N. K. Ratha, R. M. Bolle, V. D. Pandit, and V. Vaish, "Robust fingerprint authentication using local structural similarity," in *Proceedings Fifth IEEE Workshop on Applications of Computer Vision*. IEEE, 2000, pp. 29–34.

[40] M. A. Olsen, M. Dusio, and C. Busch, "Fingerprint skin moisture impact on biometric performance," in *3rd International Workshop on Biometrics and Forensics (IWBF 2015)*. IEEE, 2015, pp. 1–6.

[41] T. Chugh, K. Cao, and A. K. Jain, "Fingerprint spoof buster: Use of minutiae-centered patches," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2190–2202, 2018.

[42] S. Swanson and S. Swanson, "Fingerprints go the distance," Oct 2012. [Online]. Available: https://www.technologyreview.com/s/422400/fingerprints-go-the-distance/

[43] N. Roy, S. Shen, H. Hassanieh, and R. R. Choudhury, "Inaudible voice commands: The long-range attack and defense," in *15th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 18)*, 2018, pp. 547–560.

[44] J. Liu, C. Wang, Y. Chen, and N. Saxena, "Vibwrite: Towards finger-input authentication on ubiquitous surfaces via physical vibration," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017, pp. 73–87.

[45] J. Li, K. Fawaz, and Y. Kim, "Velody: Nonlinear vibration challenge-response for resilient user authentication," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2019, pp. 1201–1213.

[46] Y. Ren, Y. Chen, M. C. Chuah, and J. Yang, "Smartphone based user verification leveraging gait recognition for mobile healthcare systems," in *2013 IEEE International Conference on Sensing, Communications and Networking (SECON)*. IEEE, 2013, pp. 149–157.

[47] W.-H. Lee and R. B. Lee, "Multi-sensor authentication to improve smartphone security," in *2015 International Conference on Information Systems Security and Privacy (ICISSP)*. IEEE, 2015, pp. 1–11.

[48] L. Cai and H. Chen, "Touchlogger: Inferring keystrokes on touch screen from smartphone motion." *HotSec*, vol. 11, no. 2011, p. 9, 2011.

[49] E. Miluzzo, A. Varshavsky, S. Balakrishnan, and R. R. Choudhury, "Tapprints: your finger taps have fingerprints," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACm, 2012, pp. 323–336.

[50] T. Vu, A. Baid, S. Gao, M. Gruteser, R. Howard, J. Lindqvist, P. Spasojevic, and J. Walling, "Distinguishing users with capacitive touch communication," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 197–208.

[51] Y. Meng, D. S. Wong, R. Schlegel, *et al.*, "Touch gestures based biometric authentication scheme for touchscreen mobile phones," in *International Conference on Information Security and Cryptology*. Springer, 2012, pp. 331–350.

[52] N. Zheng, K. Bai, H. Huang, and H. Wang, "You are how you touch: User verification on smartphones via tapping behaviors," in *2014 IEEE 22nd International Conference on Network Protocols*. IEEE, 2014, pp. 221–232.

[53] H. Yang, L. Chen, K. Bian, Y. Tian, F. Ye, W. Yan, T. Zhao, and X. Li, "Taplock: Exploit finger tap events for enhancing attack resilience of smartphone passwords,"

in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 7139–7144.

[54] M. Shahzad, A. X. Liu, and A. Samuel, "Secure unlocking of mobile touch screen devices by simple gestures: you can see it but you can not do it," in *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 2013, pp. 39–50.

[55] M. Sherman, G. Clark, Y. Yang, S. Sugrim, A. Modig, J. Lindqvist, A. Oulasvirta, and T. Roos, "User-generated free-form gestures for authentication: Security and memorability," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. ACM, 2014, pp. 176–189.

[56] Y. Chen, J. Sun, R. Zhang, and Y. Zhang, "Your song your way: Rhythm-based two-factor authentication for multi-touch mobile devices," in *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2015, pp. 2686–2694.

[57] H. Khan, U. Hengartner, and D. Vogel, "Targeted mimicry attacks on touch input based implicit authentication schemes," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2016, pp. 387–398.

[58] Z. Yan, Q. Song, R. Tan, Y. Li, and A. W. K. Kong, "Towards touch-to-access device authentication using induced body electric potentials," *arXiv preprint arXiv:1902.07057*, 2019.

[59] V. Nguyen, M. Ibrahim, H. Truong, P. Nguyen, M. Gruteser, R. Howard, and T. Vu, "Body-guided communications: A low-power, highly-confined primitive to track and secure every touch," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 2018, pp. 353–368.

[60] C. Harrison, J. Schwarz, and S. E. Hudson, "Tapsense: enhancing finger interaction on touch surfaces," in *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 2011, pp. 627–636.

[61] Y.-C. Tung and K. G. Shin, "Expansion of human-phone interface by sensing structure-borne sound propagation," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2016, pp. 277–289.

[62] K. Sun, T. Zhao, W. Wang, and L. Xie, "Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 2018, pp. 591–605.

[63] M. Goel, B. Lee, M. T. Islam Aumi, S. Patel, G. Borriello, S. Hibino, and B. Begole, "Surfacelink: using inertial and acoustic sensing to enable multi-device interaction on a surface," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2014, pp. 1387–1396.

[64] C. Zhang, A. Waghmare, P. Kundra, Y. Pu, S. Gilliland, T. Ploetz, T. E. Starner, O. T. Inan, and G. D. Abowd, "Fingersound: Recognizing unistroke thumb gestures using a ring," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, p. 120, 2017.