EarSlide: a Secure Ear Wearables Biometric Authentication Based on Acoustic Fingerprint

ZI WANG, Augusta University, USA
YILIN WANG, Florida State University, USA
JIE YANG*, University of Electronic Science and Technology of China, China

Ear wearables (earables) are emerging platforms that are broadly adopted in various applications. There is an increasing demand for robust earables authentication because of the growing amount of sensitive information and the IoT devices that the earable could access. Traditional authentication methods become less feasible due to the limited input interface of earables. Nevertheless, the rich head-related sensing capabilities of earables can be exploited to capture human biometrics. In this paper, we propose EarSlide, an earable biometric authentication system utilizing the advanced sensing capacities of earables and the distinctive features of acoustic fingerprints when users slide their fingers on the face. It utilizes the inward-facing microphone of the earables and the face-ear channel of the ear canal to reliably capture the acoustic fingerprint. In particular, we study the theory of friction sound and categorize the characteristics of the acoustic fingerprints into three representative classes, pattern-class, ridge-groove-class, and coupling-class. Different from traditional fingerprint authentication only utilizes 2D patterns, we incorporate the 3D information in acoustic fingerprint and indirectly sense the fingerprint for authentication. We then design representative sliding gestures that carry rich information about the acoustic fingerprint while being easy to perform. It then extracts multi-class acoustic fingerprint features to reflect the inherent acoustic fingerprint characteristic for authentication. We also adopt an adaptable authentication model and a user behavior mitigation strategy to effectively authenticate legit users from adversaries. The key advantages of EarSlide are that it is resistant to spoofing attacks and its wide acceptability. Our evaluation of EarSlide in diverse real-world environments with intervals over one year shows that EarSlide achieves an average balanced accuracy rate of 98.37% with only one sliding gesture.

CCS Concepts: • **Security and privacy → Biometrics**.

Additional Key Words and Phrases: Biometrics, Fingerprint, Friction, User Authentication, Earable

ACM Reference Format:
Zi Wang, Yilin Wang, and Jie Yang. 2024. EarSlide: a Secure Ear Wearables Biometric Authentication Based on Acoustic Fingerprint. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 1, Article 24 (March 2024), 29 pages. https://doi.org/10.1145/3643515

1 INTRODUCTION

Capitalizing on their emerging ability to sense and interpret head-related features, ear wearables (earables) have become an emerging platform for various personal applications including communication, entertainment and

*This author is the corresponding author.

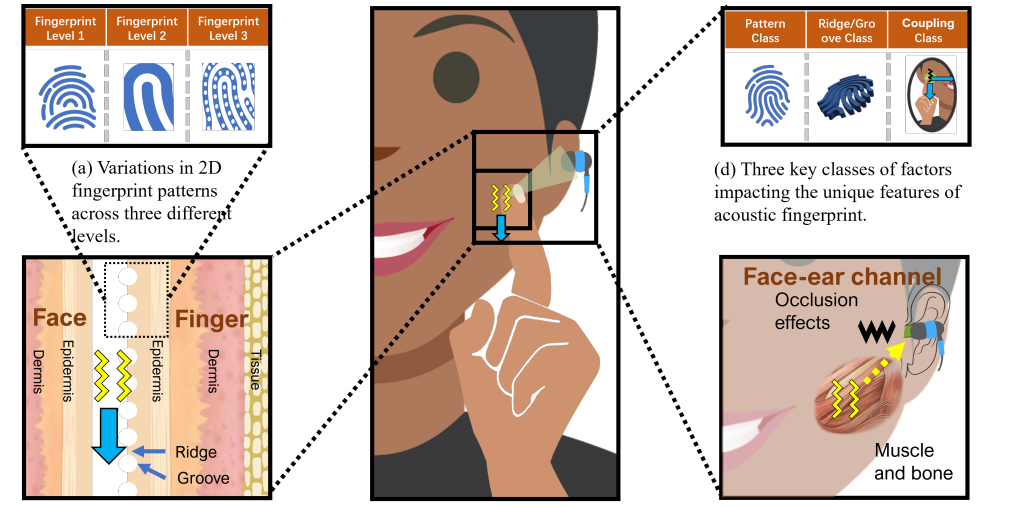
Authors' addresses: Zi Wang, Augusta University, 1120 15th Street, Augusta, GA, 30912, USA, zwang1@augusta.edu; Yilin Wang, Florida State University, 1017 Academic Way, Tallahassee, FL, 32306, USA, yilwang@cs.fsu.edu; Jie Yang, University of Electronic Science and Technology of China, No.4, Section 2, North Jianshe Road, Chengdu, Sichuan, 610054, China, jie.yang@uestc.edu.cn.

2474-9567/2024/3-ART24
https://doi.org/10.1145/3643515

healthcare [42]. The rapidly growing market is evidenced by the widespread adoption of new-generation devices like true wireless earbuds [58]. In fact, earbuds like Apple's Airpods have outperformed other wearable technology in terms of revenue, including smartwatches and glasses. Projections indicate that, in a decade, the earables market will surpass the 5 billion dollar mark [18]. Many researches are being directed towards leveraging earables for various applications, such as monitoring daily activities, tracking body gestures, and fitness assistance [12].

The widespread adoption of earables presents new security challenges [12], as these devices increasingly gain access to sensitive, multi-source information, often through AI assistants. Furthermore, earables have significant promise as tokens that mediate access to online accounts and a wide range of devices in IoT environments. Therefore, there is a growing need for secure authentication methods to prevent unauthorized access to sensitive information and resources. However, directly implementing traditional authentication methods on earables can be challenging due to the limited hardware space in earables. These devices are already densely integrated, resulting in a lack of suitable user interfaces to support traditional authentication methods such as fingerprint. Additionally, voice authentication, a current solution for earables' user interface, has been shown to be vulnerable to spoofing attacks [25, 32, 47] in many scenarios.

On the other hand, the extensive adoption of earables paves the way for new sensing possibilities. New-generation earables equipped with advanced sensors can capture unique human biometrics, such as ear canal geometry and ear canal deformation [16, 54]. Furthermore, these advanced sensing capabilities present researchers with the opportunity to reimagine or repurpose traditional authentication physiological information and biometrics, including in-ear EEG, heart rate, and blood pressure monitoring [9, 34, 49].



(a) Variations in 2D fingerprint patterns across three different levels.

(b) Microscopic view of the finger-face sliding interaction.

(c) Acoustic fingerprint generated by finger-face sliding captured by inward-facing microphone.

(d) Three key classes of factors impacting the unique features of acoustic fingerprint.

(e) The role of the face-ear channel and occlusion effects.

Fig. 1. Core concepts underpinning the proposed EarSlide system.

Fingerprint, a traditional form of identification, have long been integral to forensic science and personal identification processes due to its uniqueness in three different levels of features exists in the 2D patterns, as shown in Fig. 1 (a). As technology has advanced, fingerprint recognition has become increasingly pervasive, seen in its adoption for authentication purposes in smartphones and IoT devices. However, two major challenges hinder

the direct implementation of fingerprint authentication in earables. Firstly, to acquire a reliable fingerprint, high-resolution optical scanners or specialized equipment such as thermal scanners are typically required, which are not practical for earables due to their size and cost. Secondly, fingerprint can be easily compromised and are vulnerable to spoofing attacks. For example, latent fingerprint left on surfaces can be lifted and used to create counterfeit fingerprint. Fake fingerprint can be fabricated using common materials like silicone or gelatin for replay attacks. Meanwhile, traditional fingerprint authentication systems primarily focus on the 2D pattern information captured from static images of fingerprints. These systems are typically based on matching minutiae points or ridge patterns, such as bifurcations and ridge endings. While this approach has been quite effective, it does not take full advantage of all the information available in a fingerprint. In particular, it neglects the rich 3D structure of the ridges and grooves in a fingerprint.

In this paper, we propose EarSlide, a secure earable authentication system that utilizes the advanced sensing opportunities on new-generation earables to capture the acoustic fingerprint for user authentication. In particular, we repurpose the face and earables as a natural scanner to indirectly sense the fingerprint biometrics. We utilize the inward-facing microphone to capture the acoustic fingerprint generated when the users are sliding their finger on the face. When the finger comes into contact with the face, the complex fingerprint patterns creates acoustic fingerprint, transforming energy into friction sound waves, this process is illustrated in Fig. 1 (b). The harmonics of these acoustic fingerprint vary depending on the unique 2D patterns and 3D ridge/valley structure of individual fingerprint. Different from traditional fingerprint authentication only utilize 2D patterns, we incorporate the 3D information in acoustic fingerprint and indirectly sense the ridges and grooves to enhance the robustness and accuracy of the authentication process.

The key insight is that the acoustic fingerprint carry the distinct and unique biometric information of fingerprint. Two different participants performing the same sliding gestures will generate distinct acoustic fingerprint. As shown in Fig. 1 (c), as the acoustic fingerprint are transferred through the face-ear channel and captured by the inward-facing microphones, the unique face-ear channel provide additional encryption to the acoustic fingerprint. Also, the occlusion effect helped to enhanced the signal-to-noise ratio (SNR) of captured acoustic fingerprint. They both contribute to securely capturing fingerprint features and guard our system against replay and mimic attacks as in Fig. 1 (e). Compared with traditional biometrics, EarSlide is secure against spoofing attacks and widely acceptable.

Spoofing Attack Resistance. The acoustic fingerprint originate from friction excited by the fingerprint's 2D patterns and 3D ridge/valley structure, which are transformed into acoustic signals. These acoustic signals further transfer through the unique face-ear channel as shown in Fig. 1 (e), possessing individual distinctiveness. This exclusive channel serves as a private and secure medium that modulates and encrypts the sound waves. As a result, this transmission of acoustic fingerprint in face-ear channel, concealed within the human skull, offers increased resilience against spoofing attacks compared to traditional biometrics (e.g., face and voice) that are more susceptible to mimic and replay attacks [11].

Enhanced Acceptability and Compatibility. EarSlide offers an eye-free and require minimal user operation, which is no more than a simple slide, making it convenient and socially acceptable. On the existing commercial devices that are already equiped with inward-facing microphones, EarSlide can be deployed through software updates, providing new authentication possibilities. In addition, when finger sliding gestures are employed as an authentication and an human-computer interface simultaneously, it offers fast, convenient, and unobtrusive access to device unlocking, command issuance, and notification handling.

In our work, we first study the theory of friction, and its conterpart in the acoustic fingerprint domain. Through our analysis, we categorized three classes of impact factors, i.e., pattern class, ridge groove class, and coupling class, that influence the finger-face sliding acoustic fingerprint. Secondly, we design a set of representative finger sliding gestures. We choose five sliding gestures to represent acoustic fingerprint' characteristics as well as to balance the easy to perform. Thirdly, building on the insights from our impact factors analysis, we then designed

a feature extraction mechanism that captures the corresponding features from the three classes of impact factors. These features provide a unique characteritics of acoustic fingerprint for each user, which forms the basis of our authentication system.

Meanwhile, to make the authentication process efficient and adaptable, we implemented a Siamese Neural Network (SiameseNN). The network leverages the extracted features and compares the authentication attempts against the enrolled user profile. This approach allows EarSlide to adapt to the unique characteristics of each user's acoustic fingerprint, enhancing the reliability of the authentication process. In addition, we designed a user behavior mitigation strategy. This strategy is designed to minimize the effects of variability in user behavior, i.e., changes in sliding speed between different instances. Lastly, we conducted a feasibility study to validate the effectiveness of EarSlide. We also performed evaluations in real-world settings with 20 different subjects. These evaluations further demonstrate the robustness and applicability of EarSlide in various scenarios, underscoring its potential as a reliable and secure earable authentication system.

The contributions of our work are summarized as follows:

- We investigate the theory of friction and the acoustic fingerprint. The distinctive acoustic fingerprint's characteristics are categorized into three representative classes: pattern-class, ridge-groove-class, and coupling-class. These classes encapsulate the inherent attributes of acoustic fingerprint, offering a unique set of biometric features for robust authentication.
- We design and implement EarSlide, an acoustic fingerprint based earable authentication system. EarSlide presents several advantages, including its resilience against spoofing attacks, enhanced user acceptability and compatibility, and an adaptable user authentication framework. Leveraging the secure face-ear channel, and the inward-facing microphone on earables, EarSlide not only captures sound waves reliably but also enhance the security and privacy of users.
- We conduct experiments in diverse real-world environments to assess the performance of EarSlide. Our results indicate that EarSlide achieves an average balanced accuracy rate of 98.37% in differentiating fingerprint. This high accuracy rate shows EarSlide's potential as a robust, reliable, and secure earable authentication system.

2 RELATED WORKS

2.1 Biometric Based Authentication

2.1.1 Traditional Biometric Based Authentication. Biometric authentication systems generally rely on physiological and behavioral characteristics. These systems primarily fall into two categories: those based on physiological traits like fingerprints, palm prints, and facial features, and those based on behavioral patterns.

Fingerprint-based authentication, due to its uniqueness and ease of use, has seen widespread adoption in various applications, quickly finding its way into mobile devices, wearables, and IoT devices [14, 33]. In addition to consumer electronics, it is extensively used in forensic science, government, and municipal administration [14, 33]. However, fingerprint-based systems are vulnerable to various malicious attacks, including fingerprint obfuscation, impersonation [35, 36], and even physical spoofing by using artificially crafted fingerprints [1, 46]. Face authentication is another popular biometric method, integrated into most modern smartphones [7, 50]. Leading smartphone manufacturers like Apple and Samsung incorporate face recognition technology for unlocking devices and other security-related applications. Nevertheless, facial recognition systems are also susceptible to various spoofing attacks, including face morphing [15, 44]. Moreover, the recent COVID-19 pandemic highlighted a practical limitation of face-based authentication, where it struggles to function effectively when users are wearing masks. Another category of user authentication utilizes the behavioral characteristics, such as voice [60], gait [40] , signature [41] , vital signs [28] , finger gestures [48], and human activities [52].

Due to the wide acceptance of smart speaker in the IoT environment and on mobiles, voice-based user authentication has become most beloved and integrated with many devices. However, recent studies found that the voice biometric is vulnerable to spoofing attacks [19], such as replay attacks [13, 17] and speech-synthesize attacks [55].

2.1.2 Functional Biometrics Based Authentication. Numerous recent studies have been concentrating on Functional Biometrics [27], which use the human body as a transfer function to generate a distinctive response to a stimulus, subsequently utilized for user authentication. Recent biometric proposals include those that focus on the ear canal and its deformation. For instance, Arakawa et al. [4] suggest using earphones to capture the static geometry of the ear canal. They extract Mel-frequency cepstral coefficients (MFCCs) from the reflected acoustic signals within the ear canal to distinguish between users. Similarly, EarEcho [16] utilizes the acoustic characteristics of the static geometry of the ear canal for user authentication. Meanwhile, Wang et al. [54] have proposed using in-ear wearables to capture ear canal deformation for continuous authentication. However, systems like EarEcho require the emission of sound to probe the ear canal, which could be intrusive for individuals sensitive to the probe sound.

2.2 Earable Interaction and Sensing

Earables, with their unique sensing positions, advanced sensing capabilities, and computing capacities, have garnered research interest as a new type of ubiquitous computing platform [42]. They can sense a wide variety of human-centered information, including body motion, facial expressions [37], gait [5, 39], ear canal deformations [20], and lung function [56]. Additionally, earables can be employed in healthcare monitoring, such as respiratory [43] and cardiovascular monitoring [8], and even emotion estimation [6]. Furthermore, earables are actively studied for enhancing user authentication security by leveraging captured biometrics, such as static ear canals [16, 31], dynamic ear canal deformations [54], cochlear responses [30], and tooth gesture-generated sonic waves [53].

One major category of earables-based interaction is touch-based controls, which are integrated into many commercial earbud products. However, touch sensors can occupy additional hardware space and sometimes suffer from limited interaction areas and unintended touch activation [12, 21]. Voice-based controls, such as Apple AirPods' Siri voice assistance [3, 59], offer another popular category, although they may not be ideal in all scenarios.

Face-based interaction methods have also been explored, enabling users to sense and utilize hand movements or manipulate the earable and ear. Examples include mid-ear hand or finger gesture recognition [10, 38] and facial movement sensing [26]. Previous research on on-face finger gestures has identified a few recognizable gestures [57], but our system extends this capability by allowing the input of all alphabets and numbers. This added complexity provides a desirable input method in scenarios where traditional interactions may be less feasible.

3 PRELIMINARIES

3.1 Fingerprint Biometrics

Fingerprints have a long history of being used as a method of identification. Human fingerprint usually refer to the 2D pattern of the epidermal ridges of human fingers. The fingerprint 2D patterns are commonly described in a taxonomy of three different levels. The 1st level refers to the general global ridge flow pattern, the 2nd level contains minutiae points, and the 3rd level is pores, the local shape of ridge edges, etc [33], as shown in Fig. 2.

At the 1st level, the fingerprint are characterized by nearly parallel ridges that form distinguishable configuration regions. The regions are usually called singularities or singular regions and are usually comprised of three major pattern categories, i.e., whorls, loops, and arches, which are typically characterized by circle shapes, bow

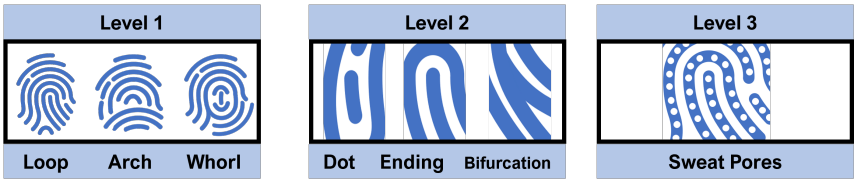


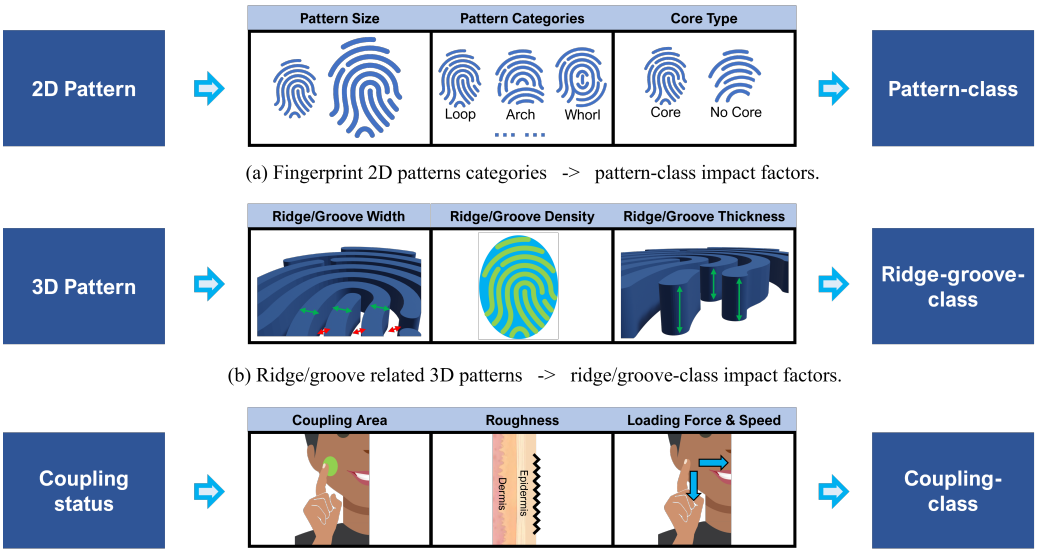
Fig. 2. Taxonomy of three levels of fingerprint pattern features.

shapes, and triangle shapes, respectively. The center point of fingerprint patterns is often called the core. The core point corresponds to the center of the north most loop type singularity or is associated with the point of maximum ridge line curvature.

The 2nd level features of the fingerprint 2D patterns focus more on small details, which are called minutiae. Minutiae refer to different types of how the ridges are ended or discontinued. In traditional fingerprint identification algorithms, the minutiae are the most commonly used features. The 2nd level also includes ridges' dimensional characteristics such as width, shape, edge contour, breaks, creases, and scars.

More fine-grained details can be extracted from the pattern at the 3rd level. Under this level, it could be observed that each epidermis ridge contains pores along its trajectory. These pores are dotted on the outer skin of the finger and anchored to the inner skin. The size of the pores could range from 60 to 250 pm. Besides the pore shape and the relative pore positions, the 3rd level features also include other inherent ridge appearances, such as the alignment and form of each ridge unit.

3.2 Friction Acoustics



(a) Fingerprint 2D patterns categories -> pattern-class impact factors.

(b) Ridge/groove related 3D patterns -> ridge/groove-class impact factors.

(c) Coupling status of sliding surface -> coupling-class impact factors.

Fig. 3. Impact factors of acoustic fingerprint.

Friction arises from the interaction of two sliding surfaces. As these surfaces move relative to each other, energy is generated and transferred from one to the other. A portion of this kinetic energy is converted into thermal energy during this process, while another portion is dispersed as acoustic energy [2]. The complex friction sound

produced by interaction of two objects are influenced by the characteristics of the sliding surfaces themselves and a multitude of other factors including factors of the ***2D Pattern***, the ***3D Pattern*** and the ***Coupling Status***.

In accordance with Fig. 3 (a), a major factor influencing the acoustic characteristics of finger-face sliding is the ***2D Pattern*** of the sliding surfaces, a central concept in friction acoustics. The foundation of these sounds lies in the interaction between two surfaces in motion relative to one another. Here, the primary acoustic impactors encompass the size, categories, and core types of the fingerprint patterns. It is acknowledged that surface texture significantly affects the friction sound and friction resulting from a sliding motion [29]. Larger patterns, such as those found on bigger tires, often generate louder friction sound due to a higher amount of material contact with the surface. Conversely, smaller, low-resistance patterns tend to produce quieter sounds. This leads us to the ***Pattern-Class*** impact factors mainly include the fingerprint patterns size, categories, and core type in acoustic fingerprint.

Moving to Fig. 3 (b), the ***3D Pattern***, specifically the ridges and grooves, play a substantial role in friction sound. The inherent characteristic of ridges and grooves of acoustic fingerprint bring us the additonal 3D information, while traditional 2D Pattern based fingerprint authentication only utilize the 2D information which lacks using of 3D information. The 3D Pattern information includes the ridge's width, density, and thickness, as well as the groove's characteristics. In essence, textures with a larger mean texture depth are linked to higher friction and friction sound. Aggressive, symmetrical tread patterns tend to create more friction sound, whereas continuous, straight grooves cause weaker friction sound. This concept is similar to the friction sound variations observed in different types of tires due to their tread pattern designs. The sound differences are largely attributed to the air trapped and subsequently released from the grooves. Applying this understanding to the domain of acoustic fingerprint, ***Ridge/Groove-Class*** impact factor like the ridge orientation and ridge-groove thickness ratio become significant.

As depicted in Fig. 3 (c), the ***Coupling Status*** also has a notable influence on the friction sound. Coupling Status, referring to the conditions at the interface where the sliding surfaces meet, is a critical concept in understanding the physics of friction. This status largely depends on the coupling area, roughness, loading forces, and speed. There are two main types of Coupling Status: decoupling (or weak coupling) and coupling (or strong coupling). In decoupling, the friction impact is confined to the interface area, creating responses in each sliding piece nearly independently of the other. However, in strong coupling, the force impact extends beyond the interfaces, turning the sliding surface pair into a coupled system that elicits a more complex and frequently non-linear response. An everyday example of this can be seen when a car travels over an unpaved road made of sand and gravel; the tire and road decouple, creating more vibration and friction sound. This understanding interprets the ***Coupling-Class*** impact factors in the study of acoustic fingerprint.

3.3 Impact Factors to Acoustic Fingerprint

In this section, we analysis the factors that influence the characteristics of acoustic fingerprint produced when a fingertip slides against the face. This exploration serves as a foundation for understanding and characterizing the unique acoustics generated by acoustic fingerprint, as well as to extract the most representative features. We propose a taxonomy that emphasizes the impact of three classes of factors: ***Pattern-Class*** factors, ***Ridge-Groove-Class*** factors, and ***Coupling-Class*** factors, as illustrated in Fig. 3.

Pattern-Class, as shown in Fig. 3 (a), factors are primarily tied to the 2D structure of the fingerprint pattern. The sounds produced by these patterns result from the contact between the face surface and the fingerprint's unique pattern block. Larger pattern sizes correspond to greater amplitudes and an increased presence of low-frequency components. Additionally, the 1st level fingerprint patterns, including circles, bows, and triangles, correspond to the common patterns known as whorls, loops, and arches, respectively, each generating a different frequency component to build up the majority of spectrum. The 2nd level patterns involve Galton characteristics

(minutiae points), which vary greatly between individuals and play a significant role in the sliding process, influencing the pitch and cepstral of the acoustic fingerprint. The 3rd level patterns, including pores and edge contours, contribute to the fine-grained components in each harmonic portion of the sliding acoustic fingerprint. The type of core, associated with the point of maximum ridge line curvature, generates rich acoustic signals, while individuals without cores produce less pronounced acoustics in terms of both amplitude and frequency.

Ridge-Groove-Class, as illustrated in Fig. 3 (b), factors largely originate from the width, density, and thickness of the ridge and groove, as well as the ratio between them. Ridge-groove-induced sound primarily results from the acoustic fingerprint between air and the groove, with a significant proportion being correlated to air-pumping noise, which is generated by the compression and expansion of air between ridges and grooves [24]. The unique features of ridge-groove structures contribute to distinctive acoustic fingerprint. For instance, females typically have higher ridge density, while males have wider ridges, and the average distance between ridges is larger in males compared to females [51]. The width, density, and thickness of the ridges and grooves determine the number and viscosity of acoustic fingerprint during each slide. The orientation of the ridges also affects the direction of the sliding, and the ridge/groove thickness ratio impacts the harmonic ratio of the acoustic fingerprint.

The ridges and grooves of acoustic fingerprint form unique 3D information, and this is an additonal aspect of the acoustic fingerprint's identity comapred with traditional fingerprint authentication. The arrangement, width, depth, and slope of these ridges and grooves are distinctive for each individual, which result in differences in air pumping in grooves, thus generate distinct acoustic fingerprint when sliding, providing an additional layer of information that can be used for authentication. By utilizing the 3D information, we can enhance the robustness and accuracy of the authentication process.

Coupling-Class, as depicted in Fig. 3 (c), factors refer to the coupling status of the two sliding surfaces, specifically whether they are coupled or decoupled. This class includes factors such as coupling area, loading force, roughness, and speed, as shown in the last row of Fig. 3. The degree of coupling or decoupling influences the frequency response and eigenfrequencies of the acoustic fingerprint. A system with a lower loading force, smaller coupling area, higher roughness, and faster sliding speed tends to be decoupled and unpredictable, resulting in a frequency response and eigenfrequencies that align with an isolated system. Conversely, a system with a higher loading force, larger coupling area, lower roughness, and slower sliding speed tends to be more coupled, causing the frequency response and eigenfrequencies to shift to a higher frequency band. Meanwhile, variations of sliding speed could have negative impact on the extracted features which need to be further handeled by mitigating the user behavior.

3.4 Sensing Acoustic Fingerprint

Our work reutilizes the face and inward-facing microphone as one nature scanner and uses the earable to capture the acoustic fingerprint. There are several noteworthy advantages of our sensing approach. First, the sensed acoustic fingerprint are modulated by the face-to-ear propagation channel of the user. The channel is unique to each person due to the differences in face bone, muscle, and tissue. It is unlikely for an adversary to obtain such a unique private channel to simulate the sensed acoustic fingerprint at earable. Therefore, the acoustic fingerprint travels through the face tissues and skull, which act as an encrypted security channel that modulates and encrypts the acoustic fingerprint. The face-ear channel encrypts and seriously weakens the transferred signals at the same time.

Second, to reliably capture the acoustic fingerprint, we choose the inward-facing microphone to capture the attenuated signals to take advantage of the occlusion effect. The design of inward-facing microphone insert into the ear canal effectively filters out most environmental sounds. And the occlusion effect of the ear canal chamber could be leveraged to heighten the low-frequency signals that carry the information of acoustic fingerprint.

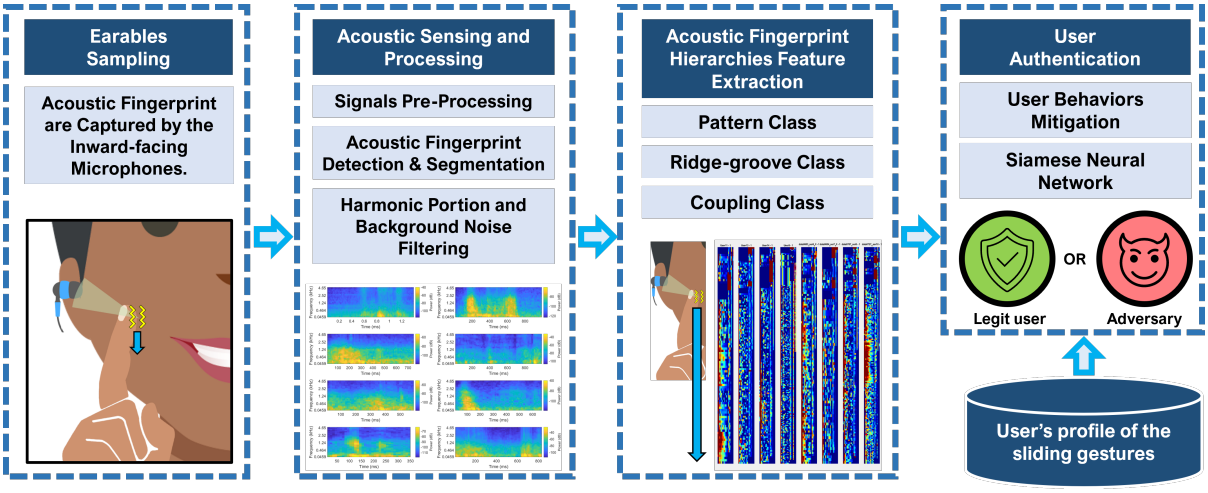


Fig. 4. System flow of EarSlide.

Specifically, a chamber forms as the ear canal is obstructed by the earbuds. And the chamber will rebound the signals with the ear canal wall and the eardrum, thus amplifying the low-frequency component of the signals to achieve better SNR.

4 SYSTEM DESIGN

4.1 System and Attack Models

Our system utilizes earables equiped with an inward-facing microphone to detect the acoustic fingerprint produced by the unique fingerprint during finger-face sliding gestures. The authentication can be initiated on-demand or employed as an unobtrusive authentication method when finger-face sliding actions serve as a eyes-free user interface. To authenticate a user, they must wear the earable and execute one or more finger-face sliding gestures on either side of the face when wearing the earables. The user's unique acoustic fingerprint profile should be pre-enrolled.

For attack models, we primarily address two types of spoofing attacks: *mimic attacks* and *replay attacks*. Mimic attacks involve adversaries wearing the victim's earable and replicating the same gestures as the victim to attempt bypassing authentication. Replay attacks are also considered, where adversaries initially eavesdrop on the acoustic fingerprint generated by the victim's fingerprint, possibly by recording the sound in close proximity to the victim. Subsequently, they replay the recorded sound to the authentication system to launch spoofing attacks. Evaluation of the system peformance under differnet attacks are detailed in section 5.3.

4.2 System Overview

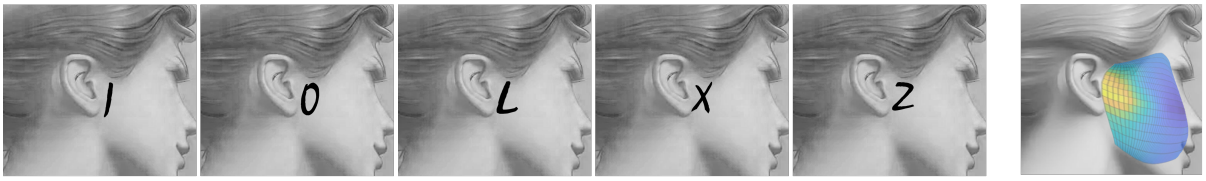
Our proposed system, EarSlide, leverages the inward-facing microphones embedded in earables to accurately capture acoustic fingerprint for earable authentication. As illustrated in Fig. 4, the operation of EarSlide can be dissected into four primary stages. Firstly, the system captures time-series acoustic fingerprint generated by finger sliding gestures through an earable device. This data is then transfered via the user's unique face-ear channel, which serves as a secure and private channel for transmission.

In the second stage, the raw acoustic data undergoes signals preprocessing. This involves the identification and segmentation of finger gestures, along with the elimination of noise and outliers. We achieve this through the implementation of harmonic ratio and background noise filters.

The third phase is the detection and extraction of representative features from three distinct classes of the acoustic fingerprint. Specifically, EarSlide extracts multi-classes of acoustic features, such as Mel-frequency cepstral coefficients (MFCC), Linear prediction cepstral coefficients (LPCC), Spectral Centroid, Spectral Spread, pitch etc. These features are from three distinct classes: pattern-class, ridge-groove-class, and coupling-class. These classes encapsulate the 2D patterns, the 3D patterns of ridge and groove, and the coupling status of the sliding surfaces, respectively. Then further indirectly sense the fingerprint of the users. These critical processes and features will also be discussed further in Section 4.4.

Lastly, the extracted and assembled features are introduced into a Siamese neural network for user authentication. To accommodate the potential variations in user behavior, EarSlide incorporates a user behavior mitigation module in this stage leveraging dynamic time warping. EarSlide offers the flexibility to authenticate a user based on a singular finger gesture or a composite of multiple gestures, providing a robust and adaptable framework for user authentication.

4.3 Gesture Design and Face Area Sensitivity Analysis



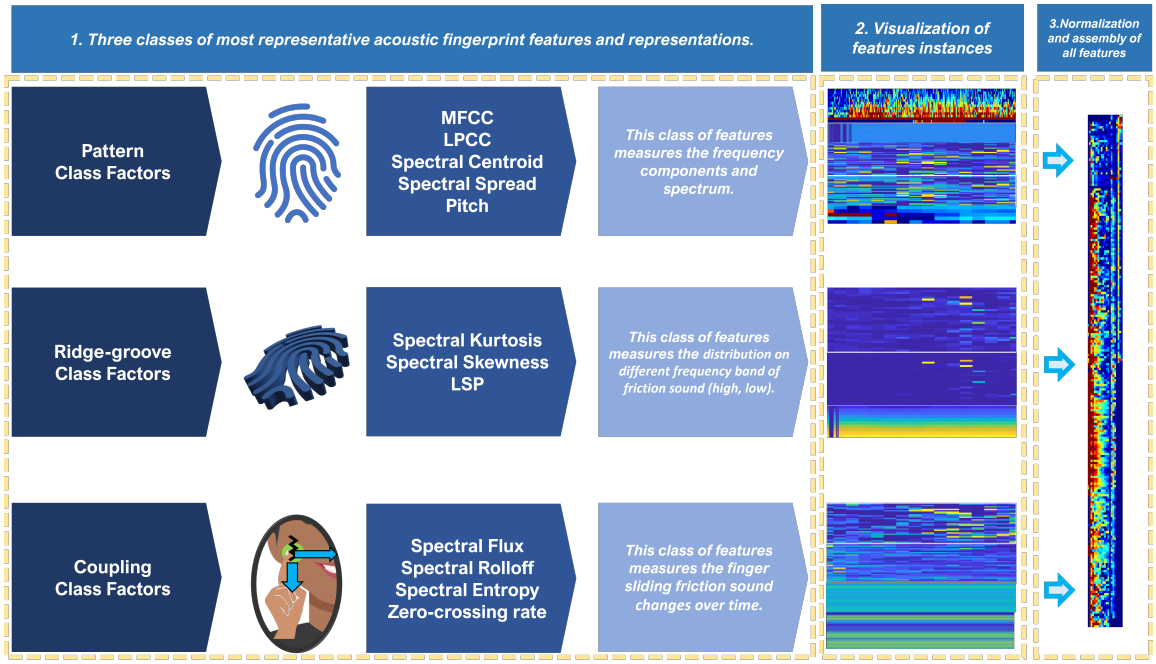
(a) Designed gestures including namely the vertical line ('1'), the circle ('0'), the angular ('L'), the cross ('X'), and the zigzag pattern ('Z').

(b) Face area sensitivity analysis.

Fig. 5. Gesture design and face area sensitivity analysis.

In designing the EarSlide system, careful consideration was given to the selection of finger sliding trajectories or gestures. The vertical line ('1'), the circle ('0'), the angular ('L'), the cross ('X'), and the zigzag pattern ('Z'), were chosen primarily for their ability to capture a broad range of acoustic fingerprint features, as shown in Fig. 5 (a). Each of these simple yet distinct gestures allows for diverse acoustic fingerprint patterns to be captured, thus offering a rich acoustic fingerprint for each user. Importantly, these gestures were not selected solely for their acoustic properties. We also considered the user-friendliness of these gestures. For instance, the circular gesture ('0') facilitates a comprehensive 'scan' of the fingerprint in various directions, thereby enhancing the richness of the captured acoustic fingerprint. This gesture is akin to the multiple scans required during fingerprint registration on a smartphone, thereby providing a robust set of acoustic fingerprint features. The angular ('L') and the cross ('X') are choosen for similar reason caues the provide two orthogonal scanning of the acoustic fingerprint. While the vertical line ('1') is the simplest and most commonly used symbol. And the zigzag pattern ('Z'), is one of the most famous mark with three strokes. Each of these symbols is common and widely recognized, hence users do not need to spend additional time learning new or complex gestures. This feature makes the system more accessible and easy to use, thus enhancing user enrollment.

In addition to gesture design, our system design also included an analysis of the sensitivity of different facial areas to acoustic fingerprint capture. As shown in Fig. 5 (b), the areas closest to the ear tragus demonstrate the highest sensitivity level. This is likely due to their proximity to the inward-facing microphone in the earable



(a) Three categories of most representative features we extract from three class: pattern-class, ridge-groove-class, coupling-class features, and their representative characteristics. (b) Features visualization. (c) Features assembly.

Fig. 6. Acoustic fingerprint hierarchies feature extraction.

device. Consequently, we recommend that users perform sliding gestures in this area for optimal sound capture and user authentication.

4.4 Acoustic Fingerprint Hierarchies Feature Extraction

Our system extracts multi-dimensional acoustic features to represent various levels of information from the acoustic fingerprint. In particular, these multi-dimensional features are distributed into three categories, as illustrated in Fig. 6, based on the impact factors we have discussed: ***Pattern-Class*** factors, ***Ridge-Groove-Class*** factors, and ***Coupling-Class*** factors.

Pattern Class Features. We first extract the Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Cepstral Coefficients (LPCC), Spectral Centroid, Spectral Spread, and Pitch to reveal the pattern-level characteristics of the acoustic fingerprint. MFCC with customed filter bank forcus on frequency range of [20Hz, 4000Hz] better loacte and extract the most representative pattern characteristics of acoustic fingerprint. LPCC are widely used for speech signal representation. They can capture the spectral envelope which is influenced by the underlying pattern structure. Due to the complexity of MFCC and LPCC, MFCC and LPCC are classified as primarily pattern-class features due to their significant responsiveness to the spectral characteristics shaped by fingerprint patterns. However, they also have supplementary impacts on the other two class including the ridge-groove and the coupling dynamics. Both Spectral Centroid and Spectral Spread reflect the major portion of the pattern-induced spectrum components. Pitch is the perceived fundamental frequency of a sound and different fingerprint patterns may lead to different pitch perceptions. These features measure the entire spectrum and frequency components, as the unique pattern of each fingerprint, including the 1st, 2nd, and 3rd level patterns and core types, results in changes in the spectrum and frequency components.

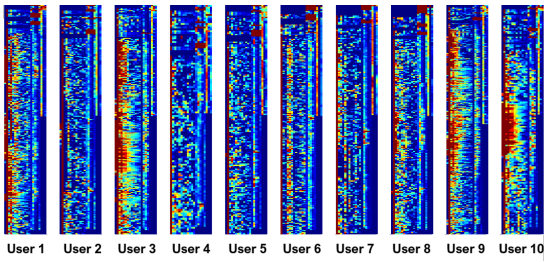
Ridge-Groove Class Features. For the ridge-groove class, we focus on Spectral Kurtosis, Spectral Skewness and Line Spectral Pairs (LSP). The insights of 3D ridge-groove class information is that the ridge-groove-induced sound is a phenomenon due to friction between air and groove, and a large proportion of the ridge-groove-induced friction sound is correlated to air-pumping noise, which is generated by the compression and expansion of air between ridges and grooves. Thus the air pumping produce either high frequency component or low frequency component depand on the compression of air. Therefore, the distribution of ridges and grooves in a fingerprint, including their width, density, thickness, and orientation, impacts the distribution of high and low frequencies. Spectral Kurtosis and Spectral Skewness measure the distribution of different frequency bands components on both high and low frequency band. Also the spectral envelope of these high and low frequency band components are captured by the LSP. This makes these features particularly informative for the ridge-groove class.

Coupling Class Features. The coupling class factors, such as coupling area, loading force, roughness, and sliding speed, influence the coupling status of the two sliding surfaces and consequently affect the frequency response and eigenfrequencies of the acoustic fingerprint. We extract features such as Spectral Flux, Spectral Rolloff, Spectral Entropy, and Zero-crossing Rate to capture the coupling class of the acoustic fingerprint. These features measure acoustic and spectrum changes over time and can be a good representation of coupling status. The Spectral Flux measures the linear changes of acoustic fingerprint in the time domain. And the Spectral Rolloff measures the start and end time of fingerprint sliding on face, which also reflect if the finger is in touch with the face. Zero-crossing Rate is the rate at which a signal changes its sign. This could reflect changes in the coupling state, as different sliding speeds or loading forces could lead to changes in sliding coupling. Spectral Entropy measures the randomness or unpredictability in a signal. As the friction sliding system is more coupling, the system tend to be more predictable, thus impact the Spectral Entropy. By capturing these changes over time, the coupling class features provide valuable insights into the unique acoustic fingerprint of each individual. We also proposed a user behavior mitigation model to mitigate the impact of sliding speed, which is further detailed in section 4.5.

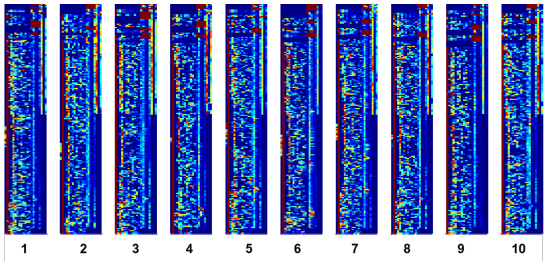
Fig. 6 provides a detailed representation of our acoustic fingerprint hierarchies feature extraction process. We begin by extracting three categories of features: pattern-class, ridge-groove-class, and coupling-class, as shown in Fig. 6 (a). Each class captures different aspects of the acoustic fingerprint, making them essential for characterizing individual fingerprint attributes. Fig. 6 (b) provides sample visualizations for each feature class. These illustrations depict how each feature class represents the characteristics of acoustic fingerprints. For instance, pattern-class features show a distinct frequency spectrum, while ridge-groove-class features illustrate the distribution of high and low frequency components. We further compress and concatenate the captured features into one assembly vector. As shown in Fig.6 (c), it illustrates the assembly of these features as a comprehensive feature vector. This assembled vector, representing the complete acoustic fingerprint of one user, is then input into the Siamese neural network for authentication.

Fig. 7 (a) displays the features assembly of ten different users. Each user's feature vector is visually distinct, showcasing a broad spectrum of patterns. The large variance between these feature vectors demonstrates the ability of our feature extraction process to capture the unique characteristics of each user's acoustic fingerprint. This inter-variance is fundamental to our system's capability to differentiate between different users and provides the basis for effective authentication.

Fig. 7 (b), on the other hand, presents ten instances from the same user. Despite variations inherent to multiple instances of the finger-face sliding action, the feature vectors are remarkably consistent. The small intra-variance reinforces the reliability of our feature extraction process in consistently capturing the unique acoustic fingerprint characteristics of an individual, irrespective of minor variations in the finger sliding action across different instances.



(a) Features assembly of 10 different users.

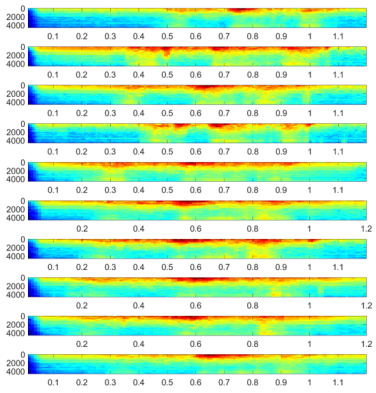


(b) 10 instances from the same user.

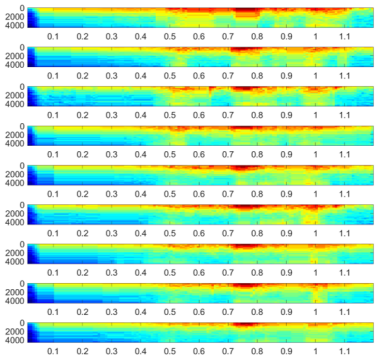
Fig. 7. Features assembly of acoustic fingerprint.

This balance of large inter-variance and small intra-variance underpins the robustness and reliability of our authentication system. It ensures that our system can accurately authenticate users based on their unique acoustic fingerprint, while remaining resilient to minor variations in the acoustic fingerprint across different instances from the same user.

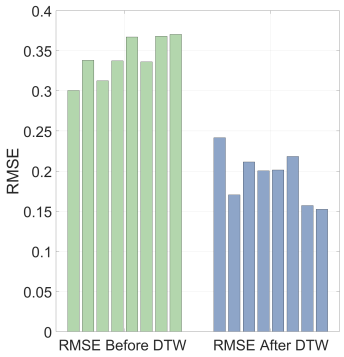
Our feature extraction process is grounded in robust acoustic analysis methods and informed by the physical properties of acoustic fingerprint. By focusing on the key impact factors of acoustic fingerprint and carefully selecting representative features, we ensure that our feature extraction process captures the essential characteristics that make each person's acoustic fingerprint unique. The assembly of these features into a single feature vector allows us to leverage the combined power of these diverse acoustic properties, enhancing the performance of our system in differentiating between individuals.



(a) Mel spectrograms of acoustic fingerprint before mitigation of user behavior.



(b) Mel spectrograms after user behavior mitigation. (The first row acts as a reference and is not mitigated).



(c) RMSE before and after user behavior mitigation.

Fig. 8. Mitigation of user behavior.

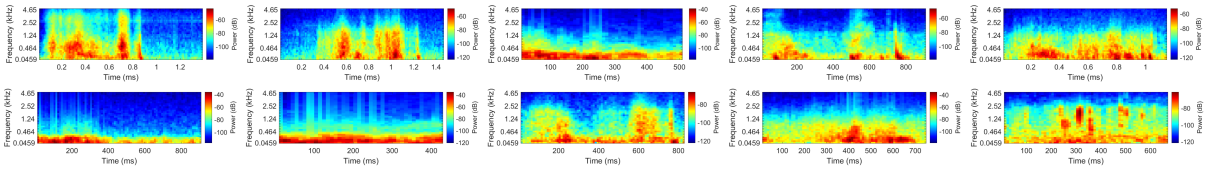
4.5 User Behavior Mitigation

In our acoustic fingerprint based biometric authentication systems, the user's sliding speed forms a important impact on the generated acoustic features. Such variances pose a challenge to the consistency of the extracted

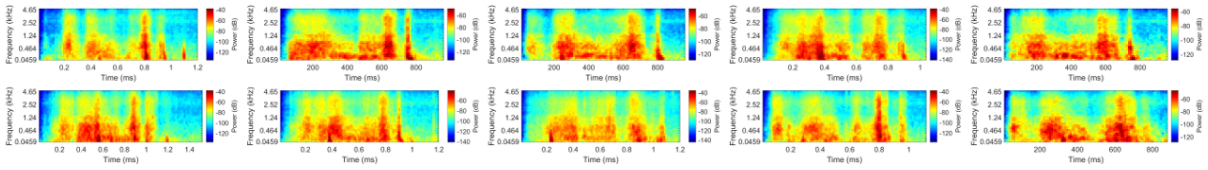
features and subsequently to the reliability of the authentication system. Fig. 8 provides a compelling visualization of how effectively these user behavior variances can be mitigated by our user behavior mitigation to enhance the reliability of the biometric authentication system.

Fig. 8 (a) shows the original Melspectrograms of multiple instances of the same user's acoustic fingerprint. It is evident that despite being from the same user, the time alignments of the spectrograms are not consistent, which is a direct implication of the user behavior variances in terms of sliding speed. Thus, we adopt the Dynamic Time Warping (DTW) to addresses this issue, as depicted in Fig. 8 (b). The alignment of features in the time dimension across the spectrograms suggests that the variances in the sliding speed have been mitigated. To quantify the user behavior mitigation using DTW, we adopt Root Mean Square Error (RMSE). We compared the RMSE values before and after applying DTW. The RMSE values decreased from [0.3004, 0.3381, 0.3125, 0.3374, 0.367, 0.3361, 0.368, 0.3704] to [0.2417, 0.1703, 0.2114, 0.2004, 0.2013, 0.2184, 0.1572, 0.1524] respectively, indicating an improvement in the consistency of the extracted features and thus, the reliability of the authentication system, as shown in Fig. 8 (c).

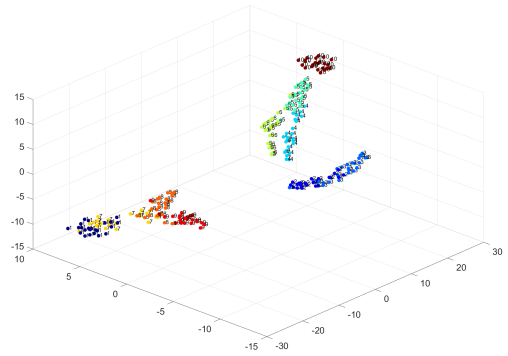
The mitigation of user behavior variations, particularly the sliding speed, is important in enhancing the performance of our system. By mitigating these variances, DTW enhances the reliability and consistency of EarSlide, thereby improving the system accuracy, leading to more consistent and reliable user authentication.



(a) Melspectrograms of 10 different users.



(b) Melspectrograms from the same user.



(c) T-SNE based features dimensionality reduction.

Fig. 9. Feasibility study: acoustic fingerprint features extraction and visualization.

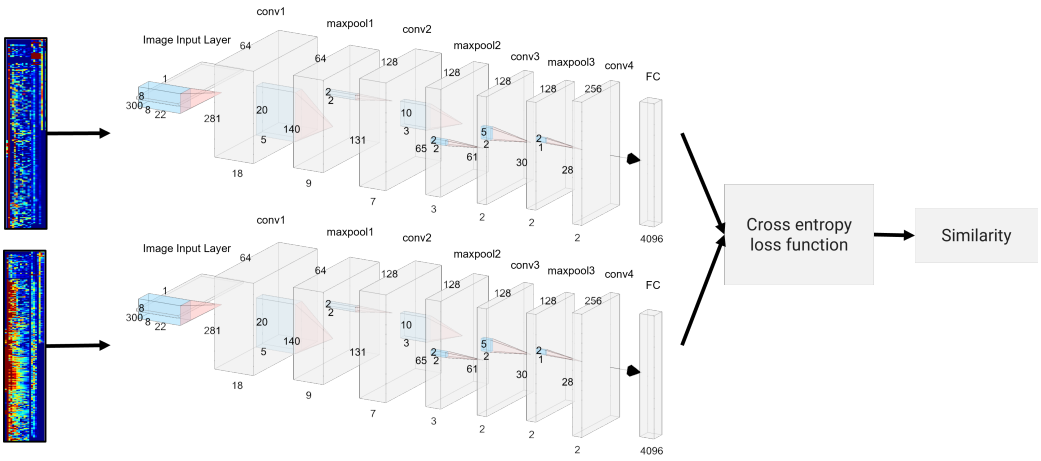


Fig. 10. Architecture of adaptable user authentication model.

4.6 Feasibility Study

Our feasibility study was designed to assess the distinctiveness and consistency of the acoustic signatures generated the different individuals' acoustic fingerprint against a surface. In particular, we focused on the extraction and evaluation of Mel Frequency Cepstral Coefficients (MFCC) and spectrum features, which are commonly used in audio signal processing and particularly suited to the unique characteristics of acoustic fingerprint.

As demonstrated in Fig. 9 (a), the Melspectrograms of 10 different users show a clear distinction in their acoustic fingerprint characteristics. This observation confirms that our feature space is rich enough to effectively differentiate among multiple users. Each individual's unique fingerprint geometry, including ridge spacing and pattern type, contributes to their distinct acoustic fingerprint profile, as reflected in their respective Melspectrograms.

Fig. 9 (b) presents Melspectrograms for multiple instances from the same user. The high degree of similarity among these instances confirms the consistency of the acoustic fingerprint within the same individual. This consistency is crucial to ensure reliable authentication over time, and our results show a promising level of stability in the acoustic fingerprint.

To illustrate the separability of the users in our feature space, we applied a t-SNE dimensionality reduction technique, as shown in Fig. 9 (c). The t-SNE plot demonstrates clear separation among different users, indicating that our feature space is not only high-dimensional but also discriminative. This visualization provides us with the assurance that our selected features can indeed serve as a reliable basis for user identification.

Overall, our feasibility study provides strong evidence supporting the use of acoustic fingerprint for user identification. The distinctiveness and consistency of the acoustic fingerprint, combined with the clear separability in the feature space, demonstrate the potential of our approach for robust user authentication in practical applications.

4.7 Adaptable User Authentication Model

We propose an adaptable authentication model based on a Siamese neural network, utilizing the unique acoustic fingerprint features generated from finger-face interactions. The distinct characteristics of each individual's fingerprint serve as a robust basis for authentication. The Siamese architecture, consisting of twin subnetworks, excels at learning a similarity metric between pairs of input features. This design allows for a scalable authentication system that can accommodate new users and account for variations in user behavior. This adaptability is

crucial for earable authentication systems. The introduction of new users and their unique acoustic fingerprint features can be smoothly integrated into the system without the need for network retraining.

The goal of the Siamese network is to differentiate between the two acoustic fingerprint. The network produces an output probability ranging from 0 to 1, where values closer to 0 suggest a prediction that the acoustic fingerprint features are dissimilar, and values closer to 1 indicate that the acoustic fingerprint features are similar. The loss is calculated using the binary cross-entropy between the predicted score and the actual label value:

$$\text{loss} = -\left[t \log(y) + (1-t) \log(1-y)\right]. \tag{1}$$

where t represents the true label (either 0 or 1) and y is the predicted label.

The architecture of each subnetwork in the Siamese neural network is designed to efficiently extract relevant features from the input data and differentiate between users. The Image Input Layer accepts a 300x22x1 input feature matrix. The subnetwork each contains 4 convolutional layer followed by a Rectified Linear Unit (ReLU) activation function, as well as 3 maxpolling layers. The details of the architecture of the model and parameters are shown in Fig. 10.

4.8 System Implementation

Signal Pre-Processing. Acoustic fingerprint produces a subtle sound that is highly susceptible to interference. We observe that the acoustic fingerprint's primary component is below 4 kHz when analyzing the spectrum of the acquired acoustic fingerprint. Human body vibrations may also interfere with acoustic fingerprint, such as the movements of the limbs. Therefore, we establish a filter band that ranges from 20 Hz to 4 kHz to eliminate interference from out-of-band sources. For a better localization of the acoustic fingerprint, we apply a Butterworth filter with a passband of [20Hz, 4000Hz] to the input signal to remove out-of-band interference.

Finger-Face Gesture Segmentation. To segment each finger-face sliding gesture based on its sound, we employ Hidden Markov Models (HMMs) as our primary method of segmentation. We also use statistical classification based on HMMs and probabilistic rule-based methods to improve our results. To segment the acoustic fingerprint, we first compute a probability distribution for each piece of the sound, and then compare it with its components to find the distribution with the highest probability. We draw inspiration from previous research in automatic segmentation of acoustic signals [23, 45], and combine it with our probabilistic rule-based method to achieve accurate and efficient segmentation of the acoustic fingerprint.

Harmonic Portion of Sliding Friction Sound. To locate the timing of acoustic fingerprint within one segment of acoustic fingerprint and increase the signal-to-noise ratio (SNR), we employ the use of the harmonic ratio. This technique allows us to determine the harmonic portion of finger-face gestures [22]. We begin by computing the normalized autocorrelation of the acoustic fingerprint for each sliding window. Next, we estimate the harmonic ratio by identifying the maximum value of the normalized autocorrelation. To improve the accuracy of our estimation, we use parabolic interpolation. This method provides us with a reliable estimate of the timing of acoustic fingerprint, and allows us to increase the SNR by filtering out unwanted background noise.

5 PERFORMANCE EVALUATION

5.1 Experimental Setup

EarSlide offers opportunities for a range of practical applications in everyday settings such as living room, office, and ourdoor, as shown in Fig. 11 (a). It can serve as a secure method for device authentication, processing payment transactions, or facilitating secure communications. To test the real-world applicability of our system, we conducted evaluations in these varied environments, with participants using earable during their regular activities. This approach provided a realistic testing ground for potential usage scenarios of EarSlide.

(a) Environment evaluation in five different environments: a quiet living room, a somewhat noisy office space with people chatting and moving around, and an outdoor setting with natural background noises mixed with human activities, same living room with background music played and car interior.

(b) Off-the-shelf devices and our prototype.

Fig. 11. Experiment environments and devices.

Commercially available earables with inward-facing microphones, which often come with firmware restrictions hindering access to raw data, our prototype was designed using accessible hardware components, showcasing its practicality and versatility, as demonstrated in Fig. 11 (b). Our design incorporates a single microphone chip with a sensitivity of -28±3 dB into conventional earbuds, complete with a 3.5 mm audio jack and a 12 mm speaker. In addition, we employed an off-the-shelf in-ear microphone to capture the acoustic fingerprint. The overall cost of our prototype hardware is just a few dollors, making it an affordable alternative to more expensive commercial options, and thus, accessible to a broader consumer base.

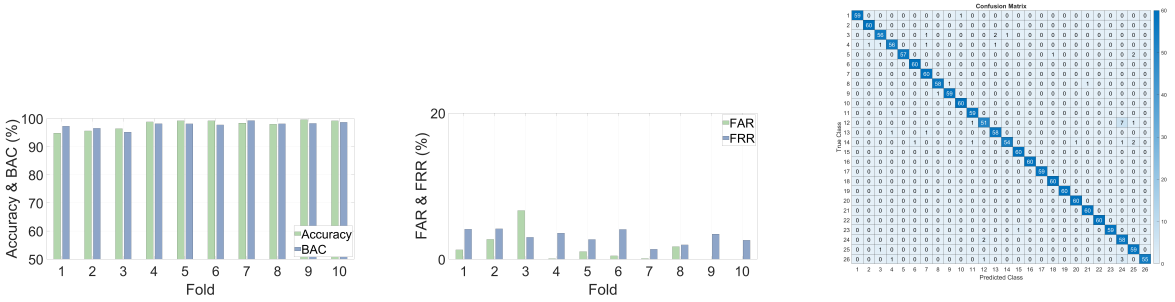
Data Collection and Evaluation Methodology.

For our experiments, we recruited a diverse cohort of 26 participants, comprising 18 males and 8 females, aged between 23 and 36 years (mean age: 28.92, standard deviation: 3.63). All participants were duly informed about the study's objectives and provided their consent prior to participation. Participants were guided to wear the earable device in a comfortable position and practice various finger-sliding gestures until they gained confidence in their execution. The data was collected in different sessions and varied daily environments to encompass a wide range of user behaviors and ambient conditions. They were also instructed to perform sliding gestures with different trajectories, in different environments, and with different motions and makeup. In each session, they were asked to perform each gesture at least 40 times, providing us with a dataset for evaluation. Moreover, 10 participants participated in multiple sessions spread over 6 months, and 6 subjects for 12 months. This enables us to assess the system's performance against temporal variations and potential shifts in user behavior. For system evaluation, we employed a 10-fold cross-validation approach, with a non-overlapping 10% subset reserved for testing. Considering each gesture as an individual authentication attempt, we have over 4,500 attempts for evaluation.

Learning Model and Evaluation Metrics. Our user authentication model is built on a Siamese neural network architecture in MATLAB. This twin network structure learns to distinguish between users by extracting distinct features from the input acoustic fingerprint data. We set the learning rate at 0.01 and used the cross-entropy loss function for model training. The Siamese network yields a similarity score that's thresholded to ascertain user identity. We assessed our model's performance using accuracy, False Rejection Rate (FRR), False Acceptance Rate (FAR), and Balanced Accuracy (BAC) as primary evaluation metrics. FRR quantifies the probability of a legitimate user being wrongly denied access, while FAR measures the likelihood of unauthorized access being incorrectly granted. BAC, on the other hand, provides a measure of performance on imbalanced datasets, computed as the average of the True Positive Rate and the True Negative Rate.

5.2 Overall Performance

In this section, we present the overall performance of our EarSlide system for user authentication based on acoustic fingerprint. As shown in Fig. 12 (a), the average accuracy of the EarSlide system across the 10 folds is



(a) Overall performance of 10 folds. (b) Overall FAR and FRR. (c) Confusion Matrix.

Fig. 12. Evaluation of overall performance of proposed system.

96.86%, with individual fold accuracies ranging from 94.80% to 99.60%. BAC is another important metric, providing a balanced measure of the system's performance across different classes. The average BAC is 98.37%, with values ranging from 95.17% to 99.27% across the 10 folds. Also as shown in Fig. 12 (b), FAR, which measures the percentage of incorrect positive matches, has an average value of 0.13%, while FRR, representing the percentage of incorrect negative matches, has an average value of 3.14%. The confusion matrix as shown as in Fig. 12 (c). From these observations, we can see that the EarSlide system demonstrates a consistently high level of performance in user authentication based on acoustic fingerprint. The system achieves high accuracy and BAC values, indicating its effectiveness in recognizing genuine users and rejecting imposters. The relatively low FAR and FRR values suggest that the system is robust.

We also observed the difference between the FRR and FAR values in our EarSlide system where FRR is slightly higher than FAR. A higher FRR reflects the system's stringent security measures. By maintaining a high threshold for acceptance, the system ensures that only users with a high degree of similarity to the enrolled fingerprint are granted access. This minimizes the risk of unauthorized access and enhances overall security. On the other hand, a lower FAR signifies that the system is less likely to accept imposters, thus providing a strong defense against unauthorized access attempts. While a low FAR is desirable for maintaining security, it is important to strike a balance between security and usability. The reason behind the higher FRR compared to the FAR can be attributed to the system's design, which prioritizes security over user convenience. In scenarios where high security is of utmost importance, such as financial transactions or access to sensitive data, having a higher FRR can be considered a necessary trade-off to ensure robust protection against unauthorized access.

5.3 Evaluation of System Performance under Different Types of Attack

In this section, we evaluate the resilience of our EarSlide system against two potential attack types: mimic attacks and replay attacks.

5.3.1 Mimic Attacks. In the case of mimic attacks, the adversary wears the victim's earable and attempts to perform the same sliding gestures as the victim to bypass the authentication system. Each user's data was utilized as an attacker against all other participants, cycling through this process for every individual user. In total, 20 subjects participated in this evaluation. However, our evaluation shows that this attack is highly ineffective. The success rates for mimic attacks were extremely low, as shown in Fig.13, due to the uniqueness of each individual's fingerprint. The acoustic fingerprint generated by each person's fingerprint during finger-face sliding gestures is unique and cannot be replicated by another person, even when performing the same gestures. Therefore, our EarSlide system effectively counters mimic attacks.

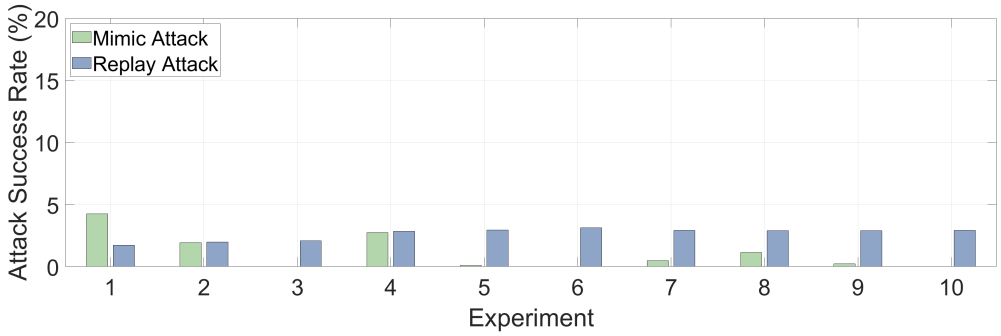


Fig. 13. Comparison of two different attack models: mimic attacks and replay attacks.

5.3.2 Replay Attacks. For replay attacks, an adversary could theoretically use a microphone to secretly record the acoustic fingerprint from a legitimate user and replay this sound to try and trick the system, 6 subjects are involved in this study. However, this attack is also found to be ineffective. The success rates of replay attacks were very low, as shown in Fig.13. This is because the captured signals in the air or near the face are missing the essential information of the face-ear channel. Meanwhile, the microphone can only capture minimal sound without the bone conduction of the face. The sound recorded through an air channel is considerably weaker due to air propagation, resulting in a low attack success rate. Therefore, our system effectively prevents replay attacks.

5.4 Comparative Performance Analysis of Different Finger Sliding Trajectories

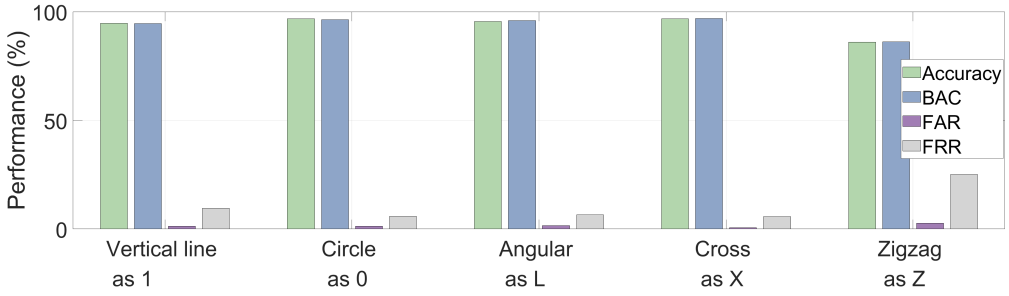


Fig. 14. Average performance metrics for each of the five finger sliding trajectories.

In our assessment of the efficacy of various finger sliding trajectories for user authentication, we have identified several notable trends. We studied five distinct trajectories, namely the vertical line ('1'), the circle ('0'), the angular ('L'), the cross ('X'), and the zigzag pattern ('Z'), each providing a unique exploration of the interaction between acoustic fingerprint and user authentication efficacy.

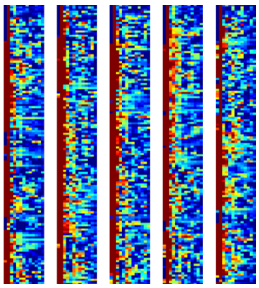
As indicated in Fig. 14, the zigzag (Z) trajectory exhibits the lowest performance metrics, marked by reduced accuracy and BAC, and slightly elevated FAR and FRR values. The complex nature of the zigzag pattern might be a key reason for this diminished performance since it has three strokes while other trajectories only have two. The zigzag pattern introduces greater variability in the recorded sounds due to its intricate structure, thereby challenging the model's ability to generalize and authenticate users accurately.

Conversely, simpler trajectories like '1', '0', 'X', and 'L' display superior performance, characterized by higher accuracy and BAC, and lower FAR and FRR values. The less complex nature of these trajectories may enable

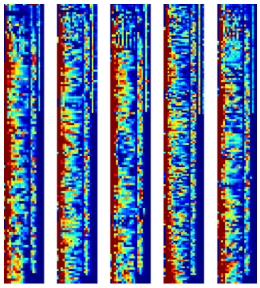
the system to capture more consistent and representative acoustic fingerprint features. Notably, the circular trajectory ('0') is advantageous as it facilitates a more comprehensive scanning of the fingerprint in different directions, similar to multiple scans required during fingerprint registration on a phone. This broader scanning might allow the model to distinguish users more effectively. In sum, our authentication system's performance achieve high performance and can perform robust authentication with different finger sliding trajectories.

5.5 Longivity Study

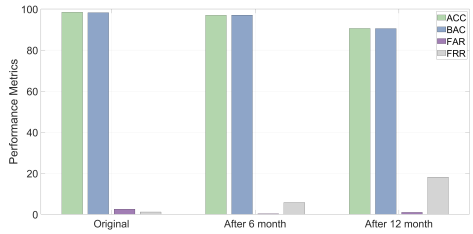
In this section, we evaluate the longevity of the EarSlide authentication system by studying the stability of the system's performance over time. Specifically, we examine how the system's performance changes after 6 months and 12 months compared to the original time of feature recording. 10 subjects participated in the 6 months evaluation and 6 subjects were involved in the 12-month experiments, they were instructed to repeat the same finger sliding in the original experiment at least 40 times, respectively. Over 700 finger sliding are recorded for authentication purposes.



(a) Initial extracted acoustic fingerprint.



(b) The same user after a 12-month period.



(c) System's performance over time, evaluating original, 6-month, and 12-month intervals.

Fig. 15. Longevity study of the system, depicting feature consistency and system performance at different intervals.

Fig. 15 (a) and (b) show the features recorded at the original time and those recorded 12 months later. As shown, the features appear visually identical, suggesting the potential stability of the system over time. Fig. 15 (c) presents the evaluation of the system's performance at the original time, after 6 months, and after 12 months.

Looking at the accuracy and balanced accuracy (BAC), we observe a decline in performance over time. The average accuracy drops from 97.92% at the original time to 96.93% after 6 months and further down to 90.55% after 12 months. Similarly, the average BAC declines from 98.22% to 96.98% and 90.48% at the original time, after 6 months, and after 12 months, respectively. The false acceptance rate (FAR), representing the likelihood of accepting imposters, remains relatively low throughout the period, although there is a slight increase after 12 months. Conversely, the false rejection rate (FRR), which reflects the probability of rejecting genuine users, increases slightly over time, indicating a rising challenge in recognizing genuine users over long period of intervals.

Although there is a decrease in performance over time, it is important to note that even after 12 months, the system's average accuracy and BAC remain above 90%, and the FAR remains low. This suggests that while there may be some degradation in the system's performance over time, the system maintains a relatively high level of accuracy and security even after a year. Another important observation is the increase in FRR over time, which indicates a growing difficulty in recognizing genuine users. This may be due to various factors, such as changes in the user's fingerprint epidermis condition or variations in how the user interacts with the system over time. This observation underscores the need for periodic re-training or updating of the user's template to maintain optimal system performance.

The observation suggests that the EarSlide system demonstrates promising longevity, with a relatively high level of performance even after a year. However, our study also highlights the importance of periodic system updates to mitigate the effects of performance degradation over time.

5.6 Evaluation of Impact of Body Motion

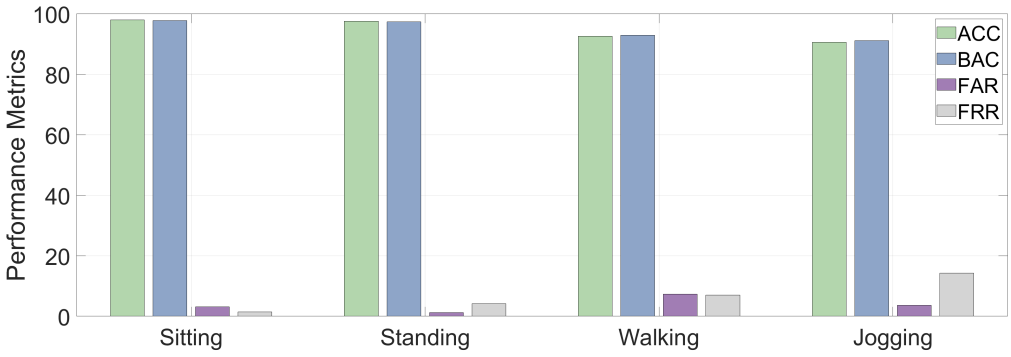


Fig. 16. System performance under the influence of different motions.

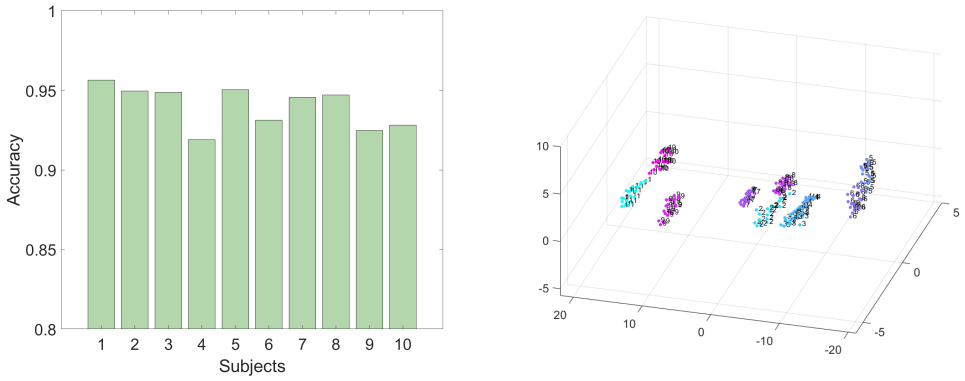
In this evaluation, we evaluate the performance of our system under varying conditions reflective of everyday user behavior. Our evaluation includes scenarios that test the system's resilience when users are engaged in different activities. A total of 12 subjects participated in this study and performed different motions, from the tranquility of sitting and standing to the dynamic actions of walking and jogging.

As shown in Fig. 16, this evaluation reveals that our system demonstrates commendable accuracy in relatively stationary states, exemplified by an accuracy rate of 97.92% for seated users and 97.56% when they are standing. However, this accuracy experiences a slight reduction to 92.56% for walking scenarios and is more significantly impacted during jogging, with a decrease to 90.56%. It appears that the vigorous movements involved in jogging disrupt the steady contact needed between the finger and face for accurate recognition. This disruption is likely intensified by the natural variations in movement and the presence of increased perspiration during intense physical activity.

5.7 Evaluation of Fingerprint Distinction Capability

The focus of this experiment was to explore the variations in the acoustic fingerprint generated by different fingerprints from same user during facial sliding gestures. Two participants, one male and one female, were involved in this preliminary study. Each participant was instructed to perform sliding gestures on their own face using each of their five fingers of right hand. The experiment was designed to isolate the fingerprint as the primary source of variation.

The results, as shown in Fig. 17 (a), suggest that the system has a promising ability to differentiate between fingerprint, with an average accuracy of 94.73%. To further illustrate the distinctiveness of the features extracted, we applied t-SNE for dimensionality reduction, compressing the feature space to three main dimensions. This is visualized in Fig. 17 (b), where each point represents a sliding gesture, color-coded by the finger used. The intra-class instances, gestures performed by the same finger, were closely correlated, forming distinct clusters. On the other hand, inter-class instances, gestures from different fingers, were clearly separated, indicating a significant difference between fingerprint. This three-dimensional representation further emphasizes the system's efficacy in distinguishing between fingerprint.



(a) Accuracy of 10 fold validation.

(b) Classification visulation using T-SNE.

Fig. 17. Evaluation of ability to distinguish fingerprint.

5.8 Evaluation in Different Environments

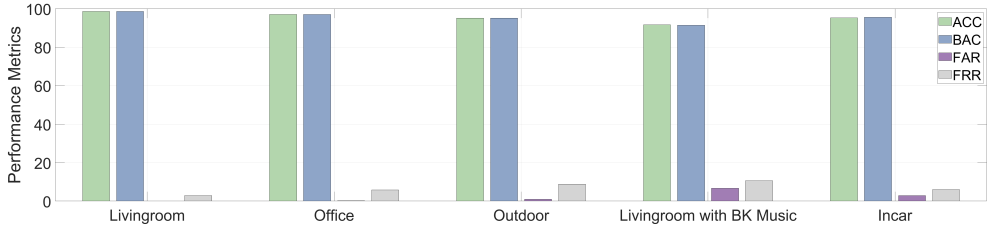


Fig. 18. System performance under varied environments.

In our effort to assess the robustness and reliability of EarSlide under various real-world conditions, we conducted experiments in three different environments: a quiet living room, a somewhat noisy office space with people chatting and moving around, an outdoor setting with natural background noises mixed with human activities, a living room with background music and a car with the engine running, as shown in Fig. 11 (a). The average dB of background music is around 50 dB, and the noise level in office and car are around 40 to 50 dB. For each environment, we ensured that the experiments were conducted at different times of the day to account for potential variations in environmental noises.

Our results showed no substantial differences in the authentication accuracy across first three different environments, as shown in Fig. 18. In a quiet living room environment, the average authentication accuracy was about 98.58%, with a Balanced Accuracy (BAC) of 98.53%. In the office environment, the accuracy was slightly lower at 97.04%, with a BAC of 96.98%. Meanwhile, in the outdoor setting, the accuracy was around 95.11%, with a BAC of 95.07%. In the living room with music, the accuracy of our system was at 91.60%, exhibiting a BAC of 91.36%. Inside a car with the engine on, the system maintained an accuracy of 95.20% and a BAC of 95.58%. Although there was a noticeable decrease in performance, especially in the music-filled environment, the system performance remained within acceptable ranges, confirming the system's resilience in noisy settings.

These statistics indicate that our system performed robustly across all tested environments, with minimal influence from ambient noise. This is primarily due to our earable design, which incorporates an inward-facing microphone that effectively filters out most environmental sounds. Since the microphone is primarily tuned to

capture the human voice, which has huge overlap with the frequency band of acoustic fingerprint generated by the user's sliding gestures, the influence of external noise on the captured sound signals is substantially minimized. Furthermore, the unique face-ear channel used for transmitting the acoustic fingerprint serves as a private and secure communication pathway, further minimizing the effect of ambient noise. This highlights the robustness of EarSlide, demonstrating its ability to deliver consistent performance across diverse environments, thus emphasizing its practicality as a robust solution for earable-based biometric authentication.

5.9 Impact Analysis of Makeup on System Performance

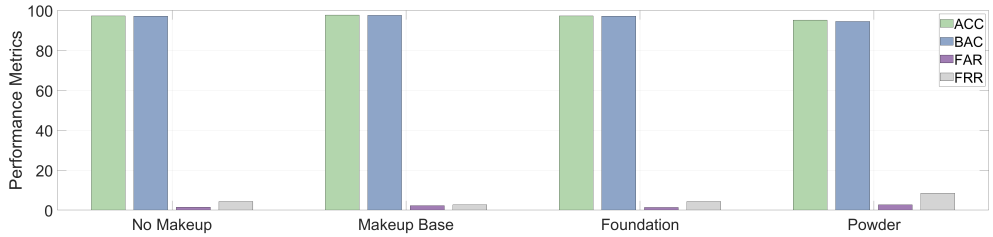


Fig. 19. System performance under the influence of different types of makeup.

Makeup, as an influential factor, has the potential to affect the surface characteristics of the face, thereby impacting the performance of the acoustic fingerprint based user authentication system. Various makeup products such as makeup base, foundation, and powder, each with different textures and functions, were considered in our investigation. Fig. 19 illustrates the system performance under the influence of these makeup products including the makeup based, foundation, and powder.

The makeup base primarily serves as a primer and helps to smooth the skin's surface. Foundation, which is applied over the base, contributes to uniform skin color and hides flaws. The powder, typically applied last, sets and mattifies the makeup. The textures of these products vary: the base and foundation are often creamy or liquid, and powder has a dry, granulated texture.

Our experiment involved participants applying these makeup products in daily usage stages: no makeup, makeup base only, makeup base plus foundation, and makeup base plus foundation and powder. From the results shown in Fig. 19, it can be observed that the application of makeup base and foundation has minimal impact on system performance, suggesting that the primary influence originate from the acoustic fingerprint of the fingerprint rather than the face surface characteristics. A slight decrease in accuracy is noted with the application of powder, which could be attributed to its different texture compared to the other makeup products.

These observations emphasize the robustness of our system in handling variations in makeup application, reinforcing its feasibility for real-world usage. However, the minor impact of powder application indicates the necessity for further improvement in handling diverse skin surface conditions.

6 DISCUSSION

6.1 User Study for Earable Authentication

To gain insights into the user experience of our finger-sliding-based earable authentication system, we conducted a detailed survey among 20 participants. Each participant has prior experience using this finger-sliding method of earable authentication. The participants were asked to rate their experiences on a scale from 1 (most negative experience) to 5 (most positive experience). A summary of the data obtained from the survey is shown in Table. 1.

Firstly, the frequency of earable device usage varied among participants, but a tendency towards more frequent use was observed, with a higher number of respondents indicating usage from 'Often' to 'Always' (13 out of 20),

as opposed to 'Sometimes' to 'Occasionally' (7 out of 20) and no one response as 'Never'. This suggests a growing integration of earable technology into daily life.

In terms of the intuitiveness of the authentication process, the responses were positive, with a total of 19 out of 20 users rating it from 'Intuitive' to 'Very Intuitive'. This demonstrates that these participants find the finger sliding mechanism is perceived as a natural and user-friendly method of authentication.

Regarding skin irritation post-authentication, the survey results indicate a positive outcome, with 80% of respondents reporting 'No irritation at all.' The remaining 20% noted only 'Slight irritation,' suggesting that the physical interaction involved with finger sliding is typically well-tolerated. It is important to consider that these individuals were participating in repeated experiments, subjecting themselves to more frequent sliding than what would be typically needed in regular use for enrollment and authentication. Therefore, it can be inferred that the incidence of irritation could be even less in everyday scenarios.

In assessing the comfort level across different trajectories used in the study, the comfort level of participants remained consistently high regardless of the trajectory followed during finger sliding authentication. Notably, 'Trajectory Z' received one response indicating 'Slightly Uncomfortable,' which could suggest a minor variance in user experience depending on the specific movement pattern. However, this is contrasted by the fact that 'Trajectory Z' also had a high number of participants rating it as 'Very Comfortable,' indicating that the experience may be influenced by individual differences in physiology or preferences rather than the trajectory itself. 'Trajectory X' and 'Trajectory L' showed a slight increase in neutrality, which may reflect a less optimal interaction for some users. Despite this, the majority still felt comfortable to very comfortable, suggesting that while the trajectory might influence the comfort level to a degree, the overall design of the finger sliding authentication is sound and can accommodate various interaction patterns with minimal discomfort.

In evaluating comfort levels, the survey data show that the system provides a high comfort level for the use of finger-sliding authentication in earable devices. For instance, a majority of participants found the device comfortable to wear in the authentication, with 90% rating it as 'Very comfortable' or 'Comfortable.' These high levels of comfort also extend to the act of using the authentication method in public spaces, with 20 out of 20 respondents feeling 'Very comfortable' and 'Comfortable' with public usage. The overall satisfaction with the authentication method is reflected in the willingness of participants to continue using it in the future, highlighting the method's potential for widespread adoption due to its comfortable and user-friendly design.

6.2 System Overhead Discussion

Our system is designed to authenticate users efficiently and train models promptly. However, it's critical to note that the current setup of our prototype relies on a collaborative framework involving mobile phones and PCs to perform sensing and model training tasks. The earables serve as the user interface, while complex computations are delegated to more powerful auxiliary devices.

Memory and Energy Considerations: The Android application at the heart of our system demonstrates remarkable efficiency, consuming less than 15MB of storage and less than 0.5% of 6GB RAM during runtime. This lean operation is due to our system's simplified design, which relies on a single recording function to capture data, which is then sent to the server for processing. In terms of energy, our system's primary consumption occurs during the recording phase on mobile devices. Despite this, the energy usage is relatively low, with a power draw of about 0.31mAh to 0.35mAh, supporting the system's suitability for regular use without significantly impacting battery life.

Server Performance and Scalability: Our evaluation of server performance highlights the system's ability to handle user authentication quickly, taking between 0.359 and 0.781 seconds per attempt. Model training times are equally efficient, requiring between 81.31 and 113.45 seconds for training. These consistent and manageable

Table 1. User experience survey of the finger sliding earable authentication. The number 1 stands for the most negative experience, 5 for the most positive experience. For instance, “never(1), occasionally(2), sometimes(3), often(4), always(5)” for question 1, ”not intuitive at all(1)” to ”very intuitive(5)” for question 2. And ”severe irritation(1)” to ”no irritation(5)” for question 3.

Question	**1**	**2**	**3**	**4**	**5**
1. How often do you use earable devices?	0	3	4	5	8
2. How intuitive did you find the finger sliding authentication process?	0	0	1	8	11
3. After using the finger sliding authentication, did you experience any skin irritation?	0	0	0	4	16
4. How comfortable are you with finger sliding trajectory ”1”?	0	0	1	5	14
5. How comfortable are you with finger sliding trajectory ”0”?	0	0	2	3	15
6. How comfortable are you with finger sliding trajectory ”L”?	0	0	2	7	11
7. How comfortable are you with finger sliding trajectory ”X”?	0	0	3	4	13
8. How comfortable are you with finger sliding trajectory ”Z”?	0	1	3	4	12
9. How comfortable was the device to wear in finger sliding authentication?	0	1	1	6	12
10. How comfortable would you feel using finger sliding authentication in public?	0	0	0	7	13
11. How likely are you to use finger sliding as earable authentication in the furture?	0	0	0	9	11
12. Overall, how satisfied are you with finger sliding authentication?	0	0	1	7	12

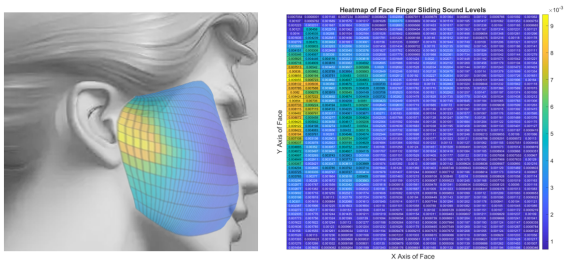
training times are crucial for ensuring the system remains responsive and stable during user enrollment and authentication phases.

In conclusion, by reducing memory and energy requirements on mobile devices, our system remains practical and sustainable. In future developments, we aim to enhance our system by enabling full self-contained system on the earable device, thereby eliminating the need for external device dependencies. Such an upgrade would streamline user experience and bolster privacy with on-device data processing. Achieving this would capitalize on the benefits of edge computing, offering real-time analytics and personalized experiences.

6.3 Internal Sound Interference

One limitation of the earphone-based biometric authentication systems, like our prototype, is that it could be impacted by the presence of internal sound such as music playback, conversations, or chewing. Our system’s on-demand operation limits the exposure to such interference, and internal background sounds differ in frequency from authentication signals that could be separated by our system. However, this does not fully protect against internal interference, leading to a low signal-noise ratio and impacting the system’s performance. To address this, advanced noise filtering techniques can be developed. For instance, if the original audio can be obtained, signal separation and noise-canceling algorithm becomes feasible. These methods could distinguish the authentication attempts from other noises, thus helping enhance the system’s performance. We will refine the system in our future study, which includes exploring acoustic models and machine learning algorithms to adapt to the user’s environment and filter out irrelevant internal sounds. This progress will make the system robust against various internal noise conditions, enhancing its reliability.

6.4 Face Sensitivity Analysis



(a) Face sensitivity model. (b) Face sensitivity analysis.

Fig. 20. Face sensitivity analysis.

The efficacy of acoustic-based authentication systems also depends on the interaction between the user's finger-sliding gestures and the facial area. To study the impact of face sensitivity, we conducted a sensitivity analysis of different facial regions to determine their responsiveness to finger sliding.

Our analysis incorporated the Root Mean Square (RMS) values calculated sliding across various regions of the face to establish a sensitivity profile. The maximum RMS value observed was 0.009425, while the minimum was 0.000633, indicating a discernible variation in the acoustic response elicited by different areas. Notably, the effective area for authentication, typically covering about half of the facial surface, ranged between 25.63% and 35.09%, subject to the selection of appropriate sensitivity thresholds, as shown in Fig. 20 (b).

Overall, as shown in Fig. 20 (a), the face sensitivity model demonstrates this variability, with 25.63% to 35.09% of the face, i.e., the areas near the ear tragus exhibiting the high sensitivity. This finding reveals the areas where sliding gestures are most likely to produce a strong and consistent acoustic signal for reliable authentication.

6.5 Role of Ear Canal Geometry in Earable Authentication

Our current system design primarily focuses on the face-ear channels as the main source of these unique acoustic fingerprints. However, The unique shape and structure of the ear canal can affect the sounds captured by inward-facing microphones, contributing to the acoustic fingerprints used for authentication. We acknowledge that the ear canal, being part of this channel, plays a subsidiary role.

While the ear canal's reflection and the bone-conducted channel are intertwined, the ear canal appears to be a minor contributor to our current setup. Different from some recent studies that actively utilize ear-canal reflections for user authentication, our approach is more passive. We capture the friction sound generated from finger-face sliding, where the ear canal geometry has less influence. Instead, the acoustic fingerprint transmitted by the face-ear channel and bone conductivity act as primary sources of distinctive biometric information. Looking ahead, we plan to investigate advanced techniques for effectively distinguishing between ear canal reflections and other sound sources.

6.6 Sample Size Considerations in Evaluation

The scale of participant involvement in user studies is often a balancing act between research depth and breadth. In our study, the number of participants was constrained by time and resource limitations, hence the number of participant are limited in the current study. The limited sample size, while sufficient for preliminary investigation, may not fully represent the diversity of the wider population. Recognizing this, we aim to extend our research in future studies by inviting more subjects to participate. This planned expansion will not only address the current limitation but will also enhance the validity and applicability of our results to a wider audience.

7 CONCLUSIONS

In this work, we present EarSlide, a secure biometric authentication system leveraging earables implanted with inward-facing microphones to capture acoustic fingerprint for authentication. By investigating the theory of friction and acoustic fingerprint characteristics, we identify three representative classes of acoustic features that serve as the core insights in features extraction and are leveraged for our authentication approach. We capture representive features from pattern-class, ridge-groove-class, and coupling-class to extract 2D patterns, 3D ridge-groove, and coupling information of finger-face sliding interaction. We adopt of a Siamese neural network for adaptable authentication model and a strategy for user behavior mitigation to enhance our system's performance. Evaluations in real-world scenarios show EarSlide's effectiveness, demonstrating our system can achieve average balanced accuracy rate of 98.37%. The results underscore the potential of EarSlide as a robust and secure acoustic fingerprint based authentication system for earables.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful feedback. This work was partially supported by the NSF Grants CNS-2131143 and CNS-1910519.

REFERENCES

[1] Aditya Abhyankar and Stephanie Schuckers. 2009. Integrating a wavelet based perspiration liveness check with fingerprint recognition. *Pattern Recognition* 42, 3 (2009), 452–464.
[2] Adnan Akay. 2002. Acoustics of friction. *The Journal of the Acoustical Society of America* 111, 4 (2002), 1525–1548.
[3] Apple. 2022. *Airpods*.
[4] Takayuki Arakawa, Takafumi Koshinaka, Shohei Yano, Hideki Irisawa, Ryoji Miyahara, and Hitoshi Imaoka. 2016. Fast and accurate personal authentication using ear acoustics. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 1–4.
[5] L Atallah, A Wiik, B Lo, JP Cobb, AA Amis, and GZ Yang. 2014. Gait asymmetry detection in older adults using a light ear-worn sensor. *Physiological measurement* 35, 5 (2014), N29.
[6] Chanavit Athavipach, Setha Pan-Ngum, and Pasin Israsena. 2019. A wearable in-ear EEG device for emotion monitoring. *Sensors* 19, 18 (2019), 4014.
[7] Hind Baqeel and Saqib Saeed. 2019. Face Detection Authentication on Smartphones: End Users Usability Assessment Experiences. In *2019 International Conference on Computer and Information Sciences (ICCIS)*. IEEE, 1–6.
[8] Nam Bui, Nhat Pham, Jessica Jacqueline Barnitz, Zhanan Zou, Phuc Nguyen, Hoang Truong, Taeho Kim, Nicholas Farrow, Anh Nguyen, Jianliang Xiao, et al. 2019. ebp: A wearable system for frequent and comfortable blood pressure monitoring from user's ear. In *The 25th annual international conference on mobile computing and networking*. 1–17.
[9] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, and Cecilia Mascolo. 2023. hEARt: Motion-resilient Heart Rate Monitoring with In-ear Microphones. In *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 200–209.
[10] Yu-Chun Chen, Chia-Ying Liao, Shuo-wen Hsu, Da-Yuan Huang, and Bing-Yu Chen. 2020. Exploring user defined gestures for ear-based interactions. *Proceedings of the ACM on Human-Computer Interaction* 4, ISS (2020), 1–20.
[11] Ivana Chingovska, Andre Rabello Dos Anjos, and Sebastien Marcel. 2014. Biometrics evaluation under spoofing attacks. *IEEE transactions on Information Forensics and Security* 9, 12 (2014), 2264–2276.
[12] Romit Roy Choudhury. 2021. Earable computing: A new area to think about. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*. 147–153.
[13] Phillip L De Leon, Michael Pucher, Junichi Yamagishi, Inma Hernaez, and Ibon Saratxaga. 2012. Evaluation of speaker verification security and detection of HMM-based synthetic speech. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 8 (2012), 2280–2290.
[14] Jianjiang Feng, Anil K Jain, and Arun Ross. 2009. Fingerprint alteration. *submitted to IEEE TIFS* (2009).
[15] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. 2016. On the effects of image alterations on face recognition accuracy. In *Face recognition across the imaging spectrum*. Springer, 195–222.
[16] Yang Gao, Wei Wang, Vir V Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using Ear Canal Echo for Wearable Authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.
[17] Rosa González Hautamäki, Tomi Kinnunen, Ville Hautamäki, Timo Leino, and Anne-Maria Laukkanen. 2013. I-vectors meet imitators: on vulnerability of speaker verification systems against voice mimicry.. In *Interspeech*. 930–934.

[18] Idtechex. 2020. . https://www.idtechex.com/en/research-article/what-does-the-future-hold-for-the-hearables-market/22130.
[19] Artur Janicki, Federico Alegre, and Nicholas Evans. 2016. An assessment of automatic speaker verification vulnerabilities to replay spoofing attacks. *Security and Communication Networks* 9, 15 (2016), 3030–3044.
[20] Yincheng Jin, Yang Gao, Xuhai Xu, Seokmin Choi, Jiyang Li, Feng Liu, Zhengxiong Li, and Zhanpeng Jin. 2022. EarCommand: " Hearing" Your Silent Speech Commands In Ear. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–28.
[21] Fahim Kawsar, Chulhong Min, Akhil Mathur, Alessandro Montanari, Utku Günay Acer, and Marc Van den Broeck. 2018. esense: Open earable platform for human sensing. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*. 371–372.
[22] Hyoung-Gook Kim, Nicolas Moreau, and Thomas Sikora. 2006. *MPEG-7 audio and beyond: Audio content indexing and retrieval*. John Wiley & Sons.
[23] Andreas Kipp, M-B Wesenick, and Florian Schiel. 1996. Automatic detection and segmentation of pronunciation variants in German speech corpora. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, Vol. 1. IEEE, 106–109.
[24] Sang-Kwon Lee, Kanghyun An, Hye-Young Cho, and Sung-Uk Hwang. 2019. Prediction and sound quality analysis of tire pattern noise based on system identification by utilizing an optimal adaptive filter. *Applied Sciences* 9, 19 (2019), 3995.
[25] Xinyu Lei, Guan-Hua Tu, Alex X Liu, Chi-Yu Li, and Tian Xie. 2018. The insecurity of home digital voice assistants-vulnerabilities, attacks and countermeasures. In *2018 IEEE conference on communications and network security (CNS)*. IEEE, 1–9.
[26] Ke Li, Ruidong Zhang, Bo Liang, François Guimbretière, and Cheng Zhang. 2022. Eario: A low-power acoustic sensing earable for continuously tracking detailed facial movements. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–24.
[27] Jonathan Liebers and Stefan Schneegass. 2020. Introducing Functional Biometrics: Using Body-Reflections as a Novel Class of Biometric Authentication Systems. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–7.
[28] Jian Liu, Yingying Chen, Yan Wang, Xu Chen, Jerry Cheng, and Jie Yang. 2018. Monitoring vital signs and postures during sleep using WiFi signals. *IEEE Internet of Things Journal* 5, 3 (2018), 2071–2084.
[29] Qingfan Liu and Ahmed Shalaby. 2017. Relating concrete pavement noise and friction to three-dimensional texture parameters. *International Journal of Pavement Engineering* 18, 5 (2017), 450–458.
[30] Yuxi Liu and Dimitrios Hatzinakos. 2014. Earprint: Transient evoked otoacoustic emission for biometrics. *IEEE Transactions on Information Forensics and Security* 9, 12 (2014), 2291–2301.
[31] Shivangi Mahto, Takayuki Arakawa, and Takafumi Koshinak. 2018. Ear acoustic biometrics using inaudible signals and its application to continuous user authentication. In *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 1407–1411.
[32] Khalid Mahmood Malik, Hafiz Malik, and Roland Baumann. 2019. Towards vulnerability analysis of voice-driven interfaces and countermeasures for replay attacks. In *2019 IEEE conference on multimedia information processing and retrieval (MIPR)*. IEEE, 523–528.
[33] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. 2009. *Handbook of fingerprint recognition*. Springer Science & Business Media.
[34] Swati Mandekar, Lina Jentsch, Dr Kai Lutz, Dr Mehdi Behbahani, and Mark Melnykowycz. 2021. Earable design analysis for sleep EEG measurements. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*. 171–175.
[35] Emanuela Marasco and Arun Ross. 2014. A survey on antispoofing schemes for fingerprint recognition systems. *ACM Computing Surveys (CSUR)* 47, 2 (2014), 1–36.
[36] Gian Luca Marcialis, Fabio Roli, and Alessandra Tidu. 2010. Analysis of fingerprint pores for vitality detection. In *2010 20th international conference on pattern recognition*. IEEE, 1289–1292.
[37] Denys JC Matthies, Bernhard A Strecker, and Bodo Urban. 2017. Earfieldsensing: A novel in-ear electric field sensing to enrich wearable gesture input through facial expressions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1911–1922.
[38] Christian Metzger, Matt Anderson, and Thad Starner. 2004. Freedigiter: A contact-free device for gesture control. In *Eighth International Symposium on Wearable Computers*, Vol. 1. IEEE, 18–21.
[39] Jay Prakash, Zhijian Yang, Yu-Lin Wei, and Romit Roy Choudhury. 2019. Stear: Robust step counting from earables. In *Proceedings of the 1st International Workshop on Earable Computing*. 36–41.
[40] Yanzhi Ren, Yingying Chen, Mooi Choo Chuah, and Jie Yang. 2014. User verification leveraging gait recognition for smartphone enabled mobile healthcare systems. *IEEE Transactions on Mobile Computing* 14, 9 (2014), 1961–1974.
[41] Yanzhi Ren, Chen Wang, Yingying Chen, Mooi Choo Chuah, and Jie Yang. 2019. Signature verification using critical segments for securing mobile transactions. *IEEE Transactions on Mobile Computing* 19, 3 (2019), 724–739.
[42] Tobias Röddiger, Christopher Clarke, Paula Breitling, Tim Schneegans, Haibin Zhao, Hans Gellersen, and Michael Beigl. 2022. Sensing with Earables: A Systematic Literature Review and Taxonomy of Phenomena. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–57.
[43] Tobias Röddiger, Daniel Wolffram, David Laubenstein, Matthias Budde, and Michael Beigl. 2019. Towards respiration rate monitoring using an in-ear headphone inertial measurement unit. In *Proceedings of the 1st International Workshop on Earable Computing*. 48–53.

[44] Ulrich Scherhag, Christian Rathgeb, Johannes Merkle, Ralph Breithaupt, and Christoph Busch. 2019. Face recognition systems under morphing attacks: A survey. *IEEE Access* 7 (2019), 23012–23026.
[45] Florian Schiel. 1999. Automatic phonetic transcription of non-prompted speech. (1999).
[46] Stephanie AC Schuckers. 2002. Spoofing and anti-spoofing measures. *Information Security technical report* 7, 4 (2002), 56–62.
[47] Khairunisa Sharif and Bastian Tenbergen. 2020. Smart home voice assistants: a literature survey of user privacy and security vulnerabilities. *Complex Systems Informatics and Modeling Quarterly* 24 (2020), 15–30.
[48] Sheng Tan, Jie Yang, and Yingying Chen. 2020. Enabling fine-grained finger gesture recognition on commodity wifi devices. *IEEE Transactions on Mobile Computing* (2020).
[49] Hoang Truong, Alessandro Montanari, and Fahim Kawsar. 2022. Non-invasive blood pressure monitoring with multi-modal in-ear sensing. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6–10.
[50] Rudolf M Verdaasdonk and Niels Liberton. 2019. The Iphone X as 3D scanner for quantitative photography of faces for diagnosis and treatment follow-up (Conference Presentation). In *Optics and Biophotonics in Low-Resource Settings V*, Vol. 10869. International Society for Optics and Photonics, 1086902.
[51] Manish Verma and Suneeta Agarwal. 2009. Fingerprint based male-female classification. In *Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS'08*. Springer, 251–257.
[52] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. 2014. E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. 617–628.
[53] Zi Wang, Yili Ren, Yingying Chen, and Jie Yang. 2022. ToothSonic: Earable Authentication via Acoustic Toothprint. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–24.
[54] Zi Wang, Sheng Tan, Linghan Zhang, Yili Ren, Zhi Wang, and Jie Yang. 2021. EarDynamic: An Ear Canal Deformation Based Continuous User Authentication Using In-Ear Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.
[55] Zhizheng Wu, Tomi Kinnunen, Nicholas Evans, Junichi Yamagishi, Cemal Hanilçi, Md Sahidullah, and Aleksandr Sizov. 2015. ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge. In *Sixteenth Annual Conference of the International Speech Communication Association*.
[56] Wentao Xie, Qingyong Hu, Jin Zhang, and Qian Zhang. 2023. EarSpiro: Earphone-based Spirometry for Lung Function Assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–27.
[57] Xuhai Xu, Haitian Shi, Xin Yi, WenJia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K Dey. 2020. Earbuddy: Enabling on-face interaction via wireless earbuds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
[58] ZDNet. 2020. . https://www.zdnet.com/article/apple-airpods-and-other-smart-hearables-are-ruling-the-wearable-tech-market.
[59] Ahed Zeidan, Hany Tallat Abdelgelil, Edward Edwin, and Dhafer Alqarni. 2021. Apple Siri as communication conduit during COVID-19: between inside and outside the OR. *BMJ Simulation & Technology Enhanced Learning* 7, 4 (2021), 274–275.
[60] Linghan Zhang, Sheng Tan, Jie Yang, and Yingying Chen. 2016. Voicelive: A phoneme localization based liveness detection for voice authentication on smartphones. In *Proceedings of the 2016 ACM SIGSAC Conference on CCS*. 1080–1091.