SCR-Auth: Secure Call Receiver Authentication on Smartphones Using Outer Ear Echoes

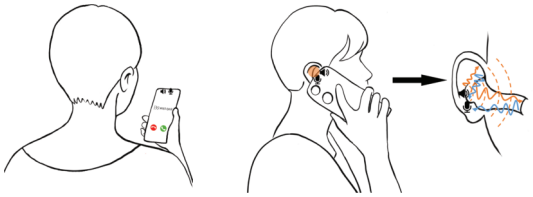
Xiping Sun, Jing Chen, *Senior Member, IEEE*, Kun He, *Member, IEEE*, Zhixiang He, Ruiying Du, Yebo Feng, Qingchuan Zhao, *Associate Member, IEEE*, and Cong Wu, *Member, IEEE*

***Abstract*—Receiving calls is one of the most universal functions of smartphones, involving sensitive information and critical operations. Unfortunately, to prioritize convenience, the current call receiving process bypasses smartphone authentication mechanisms (e.g., passwords, fingerprint recognition, and face recognition), leaving a significant security gap. To address this issue, we propose SCR-Auth, a secure call receiver authentication scheme for smartphones that leverages outer ear echoes. It sends inaudible acoustic signals through the earpiece speaker to actively sense the call receiver's outer ear structure and records the resulting echoes using the top microphone. These echoes are then analyzed to extract unique outer ear biometric information for authentication. It operates implicitly, without requiring extra hardware or imposing additional burden. Comprehensive experiments conducted under diverse conditions demonstrate SCR-Auth's effectiveness and security, showing an average balanced accuracy of 96.95% and resilience against potential attacks.**

***Index Terms*—Call receiver authentication, outer ear echoes, smartphone, user security and privacy.**

I. Introduction

PHONE calls are one of the most widely used and trusted communication forms on smartphones [1], often involving sensitive information and critical operations, such as accessing health records [2] or authorizing financial transactions [3]. While smartphones have adopted various authentication mechanisms to prevent unauthorized access, including passwords [4], fingerprint recognition [5], and face recognition [6], the call receiving process remains an exception. Prioritizing convenience, incoming calls bypass these authentications, allowing anyone with physical access to the smartphone to answer even if it is locked, which fails to meet essential security standards. Therefore, it is crucial to develop an effective call receiver authentication mechanism that ensures only the legitimate smartphone owner can answer incoming calls while maintaining convenience.



(a) An incoming call (b) Call receiver authentication

Fig. 1. Illustration of SCR-Auth. When a call comes in, the earpiece speaker and top microphone serve as an active sonar, authenticating the call receiver's identity by analyzing echoes from the outer ear.

Our work focuses on earpiece-based call reception, in which users hold the smartphone to the ear when answering a call. This mode remains essential in privacy-sensitive, constrained, or urgent situations, where loudspeaker or earphone options are inappropriate or unavailable [7]. Several existing efforts may be adapted for earpiece-based call receiver authentication [11], [12], [13], [14], [15], [16], [17], [18], [19], [22]. Behavioral biometrics methods, such as detecting smartphone pickup gestures [11], [12], [13] or sliding interactions [14], [15], often suffer from low reliability due to behavioral variability [20]. Other approaches focus on ear physiological characteristics, capturing ear images with a camera [16], [17] or pressing the ear against a capacitive touchscreen [18], [19]. However, these approaches require additional user actions, root privileges, or favorable lighting conditions, which limit their practicality.

In this paper, we propose SCR-Auth, a secure call receiver (SCR) authentication method for smartphones based on outer ear echoes. During the natural earpiece-based call receiving process, SCR-Auth emits acoustic sensing signals through the smartphone's earpiece speaker, as illustrated in Fig. 1. These signals interact with the user's outer ear, undergoing absorption and reflection before reaching the top microphone.

Received 10 January 2025; revised 18 May 2025 and 10 June 2025; accepted 26 June 2025. Date of publication 1 July 2025; date of current version 8 July 2025. This work was supported in part by the State Key Laboratory of Intelligent Transportation System under Project 2024-B004, in part by the National Natural Science Foundation of China under Grant 62172303, in part by the Key Research and Development Program of Hubei Province under Grant 2024BAB018, in part by Wuhan Scientific and Technical Achievements Project under Grant 2024030803010172, and in part by the Key Research and Development Program of Shandong Province under Grant 2022CXPT055. The associate editor coordinating the review of this article and approving it for publication was Dr. Naser Damer. *(Corresponding author: Ruiying Du.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB).

Xiping Sun, Jing Chen, Kun He, Zhixiang He, and Ruiying Du are with the Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China (e-mail: xiping@whu.edu.cn; chenjing@whu.edu.cn; hekun@whu.edu.cn; zhixianghe@whu.edu.cn; duraying@whu.edu.cn).

Yebo Feng is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: yebo.feng@ntu.edu.sg).

Qingchuan Zhao is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: cs.qczhao@cityu.edu.hk).

Cong Wu is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: congwu@hku.hk).

Digital Object Identifier 10.1109/TIFS.2025.3584643

1556-6021 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

The resulting echoes carry distinct outer ear biometric information (e.g., auricle shape, ear canal geometry, and tissue properties), which is unique to each individual and can be analyzed for authentication. SCR-Auth achieves seamless and implicit authentication without requiring extra hardware or imposing additional burden, ensuring a smooth call receiving experience.

Realizing SCR-Auth in practice faces several challenges. Firstly, due to the multipath effect, the signals captured by the smartphone's built-in microphone include not only outer ear echoes, but also direct path signals and environmental reflections. These signal components overlap in both frequency and phase, making it difficult to effectively filter the interference caused by the direct path signals and environmental reflections. Secondly, outer ear echoes are sensitive to the relative position between the ear and smartphone due to altered signal propagation properties. This sensitivity leads to unstable echo patterns, making reliable feature extraction a challenge.

To address the first challenge, we propose a two-step denoising method. The process begins with a bandpass filter to remove ambient noises, followed by the Magnitude-Phase Spectrogram Subtraction (MPSS) method to suppress interference. Specifically, for each signal segment derived through synchronization and segmentation, we compute both magnitude and phase spectrograms. A reference segment is then chosen, which primarily contains direct path signals and environmental reflections, free from outer ear echoes. Based on the selected reference segment, we construct differential spectrograms in both the magnitude and phase domains, effectively mitigating unwanted interference. To counteract the position variability between the ear and smartphone, we design a learning-based feature extractor. We first train a Convolutional Neural Network (CNN) model using multi-user data collected under diverse natural smartphone positions at call reception. Through supervised learning, the CNN model is guided to focus on identity-related features while disregarding secondary factors, such as changes in the relative position between the ear and smartphone. Based on the idea of transfer learning, we then transfer the pre-trained model as a generalized feature extractor to obtain reliable features. Finally, SCR-Auth adopts a user-specific one-class classification model to verify the legitimacy of the call receiver.

In summary, the contributions of this paper are as follows:

- We propose SCR-Auth, a novel call receiver authentication scheme for smartphones that leverages outer ear echoes, enabling secure and implicit authentication without the need for extra hardware or user burden.
- To eliminate ambient noise, as well as interference from the direct path signal and environmental reflections, we propose a specially designed two-step denoising method, encompassing bandpass filtering and spectrogram differencing. To further enhance system robustness against smartphone position changes, we introduce a pre-trained neural network model that leverages transfer learning to extract reliable features.
- We conduct comprehensive experiments under various conditions to evaluate the effectiveness of SCR-Auth, e.g., ambient noises, different postures, different periods, smartphone positions and device models. The results show that SCR-Auth can achieve a balanced accuracy of 96.95% and a equal error rate of 1.53%. We demonstrate the security of SCR-Auth by evaluating its resistance to common attacks. Our source code is available at https://github.com/luojiazhishu/SCR-Auth.

II. Related Work

In this section, we review related works on call receiver authentication for smartphones. Additionally, we explore recent advancements in the field of acoustic sensing.

A. Call Receiver Authentication

Authenticating the identity of the call receiver is essential for ensuring both security and privacy on smartphones. Call receiver authentication methods can be broadly categorized into two types: behavioral biometrics-based and physiological biometrics-based methods. Table I summarizes several representative approaches to call receiver authentication on smartphones.

Behavioral biometrics-based methods authenticate the call receiver by analyzing their behavior during phone call interactions [11], [12], [13], [14], [15], [22]. These approaches commonly use motion sensors to capture movement patterns, such as how a user picks up the smartphone and positions it to their ear, to verify their identity. However, these methods often require users to follow specific movement patterns and suffer from low accuracy due to the inherent variability and uncontrollability of user behavior [20].

Physiological biometrics-based methods focus on the unique physiological features of the ear to distinguish users. For example, ear images captured using the smartphone camera during a call are employed for authentication [16], [17], [23], [24], [25]. However, these methods are sensitive to environmental conditions, such as low light intensity. Additionally, active user cooperation is often required to obtain a clear and complete image of the ear. The smartphone touchscreen can also serve as a capacitive sensor to capture a user's earprint [18], [19], [26]. However, they require the user to active position their ear tightly and fully on the smartphone screen to capture capacitive readings, changing user's call receiving habits. Moreover, these methods necessitate rooting the smartphone and modifying the touchscreen module in the kernel source. Additionally, some methods utilize acoustic signals to sense the ear [21], [27]. However, these methods rely on measuring the ear's transfer function for authentication, which is highly sensitive to the smartphone's position. As a result, they encounter substantial challenges in maintaining accuracy when the smartphone's position varies, limiting their practical applicability in real-world scenarios. Recent studies have explored the use of earphones to assist in authentication on smartphones. However, these methods necessitate hardware modifications to existing earphones and the integration of additional sensors, such as cameras [8] or inward-facing microphones [9], [10], [28], [29], [30], which increases costs and leads to incompatibility with commercial earphones. Furthermore, they require users to constantly carry earphones, significantly compromising convenience and practicality.

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISON OF REPRESENTATIVE CALL RECEIVER AUTHENTICATION METHODS ON SMARTPHONES

Device	System	Distinctiveness	No extra hardware[1]	Little usage constraint[2]	No root privileges[3]	Resilient across diverse conditions[4]	Accuracy	Error rate
Smartphones	Conti *et al.* [11]	Hand movements	✓	×	✓	×	N/A	∼ 7%
	Fahmi *et al.* [16]	Entire ear image	✓	×	✓	×	92.5%	N/A
	Bodyprint [18]	Entire ear capacitive image	✓	×	×	×	99.52%	7.8%
	Itani *et al.* [21]	Image & Pinna responses	✓	×	✓	×	N/A	1.6%
	SCR-Auth (ours)	Inaudible outer ear echoes	✓	✓	✓	✓	96.95%	1.53%
With earphones	EarAuthCam [8]	Upper part of ear image	×	✓	✓	✓	84.1%	8.36%
	EarEcho [9]	Audible ear canal echoes	×	✓	✓	✓	94.52%	N/A

[1] : No extra hardware implies that only commodity smartphones are used, without the need for additional devices or sensors.
[2] : Little usage constraint indicates that there are no requirements on movement patterns, additional gestures or usage environments.
[3] : No root privileges means that there is no need to root the smartphone or modify the kernel source.
[4] : Resilient across diverse conditions means that the method is robust in various situations, such as different environments and user postures.

Our method does not require any additional hardware or impose extra burden. It demonstrates resilience to changes in smartphone position and remains effective under various environmental conditions.

B. Acoustic Sensing

Acoustic sensing has garnered significant attention in recent years and finds applications across diverse domains. Leveraging the capabilities of speakers and microphones, it enables environment sensing [31], [32], [33], the monitoring of human activities such as hand tracking [34], [35], [36], lip reading [37], [38], [39], and breathing monitoring [40], [41]. Additionally, acoustic sensing has demonstrated potential in identifying human physiological biometrics, such as hands [42], [43], [44] and faces [45], [46], [47].

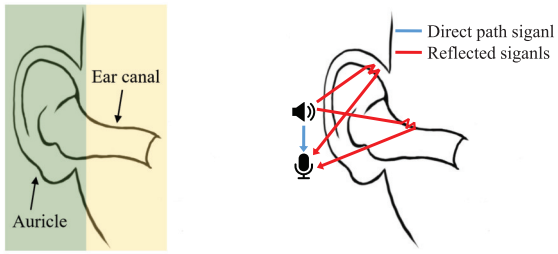
For instance, Cai et al. [32] employ dual microphones to estimate the speed of air-borne sound propagation, allowing for the inference of ambient temperature. Echotrack [34] determines the distance from the hand to the speaker, enabling continuous hand tracking using triangular geometry. Lu et al. [37] extract distinctive behavioral features of users' speaking lips through acoustic signals. EchoHand [43] complements camera-based hand geometry recognition by integrating active acoustic sensing for the other hand. EchoPrint [45] fortifies face authentication against presentation attacks by emitting inaudible acoustic signals to capture 3D facial features.

Our work uses the inaudible acoustic signal to sense the outer ear without interfering with the normal voice conversation. Moreover, it provides implicit protection before the call is answered and supports continuous authentication.

III. PRELIMINARIES

A. Outer Ear Echoes

The outer ear, as the external part of the auditory system, serves as the primary interface between the human body and the acoustic environment. As depicted in Fig. 2(a), it comprises two main components: the auricle and the ear canal. The auricle, composed of cartilage and skin, presents a complex three-dimensional morphology with folds, ridges, and contours that vary significantly across individuals [48]. The ear canal, a narrow tube leading to the eardrum, also exhibits variations in geometry (e.g., length and curvature) and wall



(a) Outer ear structure (b) Echo signal composition

Fig. 2. Illustration of the outer ear structure and the resulting echo signal composition, including both direct and reflected paths shaped by the auricle and ear canal.

composition (e.g., cartilage and bone) across populations [49]. These physiological characteristics affect how acoustic signals are absorbed, reflected, and propagated within the outer ear, resulting in user-unique echo responses.

As shown in Fig. 2(b), when the sensing signal $s(t)$ is emitted from the speaker, it propagates through multiple paths: the direct path to the microphone, and various reflected paths involving the auricle and ear canal. The signal received by the microphone $r(t)$ can thus be modeled as:

$$r(t) = (h_{\text{direct}}(t) + h_{\text{ear}}(t)) * s(t) + n(t) \tag{1}$$

Here, $h_{\text{direct}}(t)$ and $h_{\text{ear}}(t)$ represent the impulse responses of the direct and outer ear-reflected paths, respectively. $n(t)$ denotes ambient sounds, and * denotes the convolution operator.

To understand how outer ear echoes encode user-specific characteristics, we focus on $h_{\text{ear}}(t)$, which can be modeled as a discrete multipath channel:

$$h_{\text{ear}}(t) = \sum_{i=1}^{N} \alpha_i \delta(t - \tau_i) \tag{2}$$

Each path i corresponds to a specific acoustic reflection influenced by the user's auricle or ear canal. α_i is the attenuation factor for path i, which is influenced by the outer ear's geometry and tissue composition. τ_i is the time delay associated with path i, governed by the propagation distance determined by auricle shape and ear canal geometry. The resulting outer

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

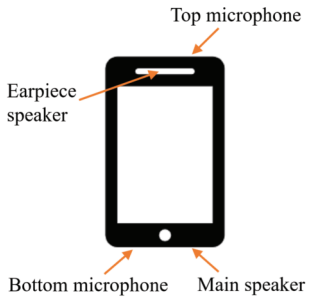


Fig. 3. The typical layout of speakers and microphones on smartphones.

ear echoes $r_{\text{ear}}(t)$ are given by:

$$r_{\text{ear}}(t) = h_{\text{ear}}(t) * s(t) = \sum_{i=1}^{N} \alpha_i s(t - \tau_i) \tag{3}$$

To analyze how the outer ear modifies the frequency-domain characteristics of the echoes, we apply the Fourier transform to Eq. 3, yielding:

$$R_{\text{ear}}(f) = H_{\text{ear}}(f)S(f) = \sum_{i=1}^{N} \alpha_i S(f) e^{-j2\pi f \tau_i} \tag{4}$$

This expression reveals how each reflection path introduces a frequency-dependent transformation to the original signal: the magnitude is scaled by α_i, while the phase is shifted by $2\pi f \tau_i$. As a result, the echo spectrum $R_{\text{ear}}(f)$ is a complex-valued signal exhibiting user-specific variations in both magnitude and phase across frequencies. Accordingly, we represent outer ear echoes using the magnitude spectrum $|R_{\text{ear}}(f)|$ and phase spectrum $\angle R_{\text{ear}}(f)$, which together capture the combined effects of all reflection paths. These frequency-domain representations serve as the foundation for extracting outer ear biometrics in our system.

B. Motivating Examples

We present a toy example to explore the feasibility of distinguishing between different call receivers based on outer ear echoes. Two users are employed to simulate the call answering process. The Google Pixel 3a is selected as the authentication device, and acoustic data is collected at a sampling rate of 48 kHz.

Specifically, we utilize the earpiece speaker to emit inaudible sensing signals and analyze the resulting echoes from the outer ear using the microphone. The sensing signal is a 25-millisecond chirp, ranging from 17 kHz to 23 kHz. After deriving the ear-related signals, we compute their magnitude and phase spectrums. The results are shown in Fig. 4. Fig. 4(a) and Fig. 4(c) present the magnitude and phase spectrums for two instances of the same user, respectively. Fig. 4(b) and Fig. 4(d) show the magnitude and phase spectrums for user 1 and user 2, respectively. We observe that the profiles of two instances for the same user match each other closely. In contrast, the profiles for the two users differ in both magnitude and phase. These results demonstrate the feasibility of using outer ear echoes for authentication, motivating the design of SCR-Auth.

C. Speaker and Microphone Selection

Fig.3 illustrates the typical layout of speakers and microphones on modern commercial smartphones. These devices are generally equipped with two speakers: a main speaker positioned at the bottom and an earpiece speaker located near the ear [50]. They also include two microphones: one at the bottom and another at the top for noise cancellation [51]. For our system, we select the earpiece speaker and the top microphone for sending and receiving signals, as their proximity to the ear supports better sensing.

IV. Overview of SCR-Auth

In this section, we first present the overview of SCR-Auth. Then we introduce the threat model and design goals.

A. System Overview

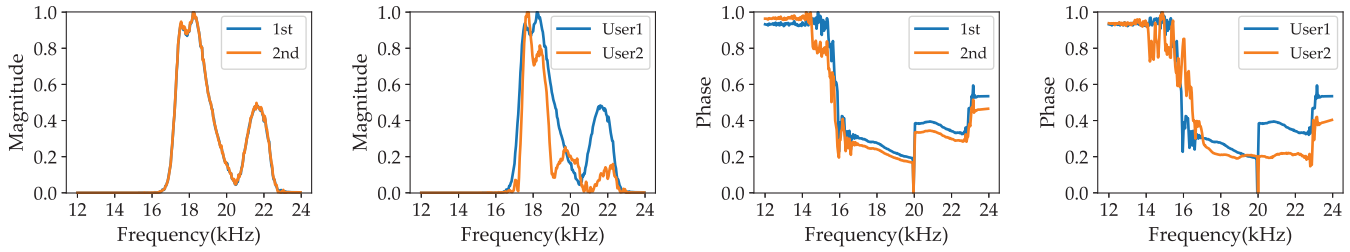
The basic idea of SCR-Auth is to utilize the speaker and microphone on a smartphone for outer ear acoustic sensing, and then analyze outer ear biometric features from the received echo signals to authenticate the call receiver. It consists of two phases: enrollment and authentication. In the enrollment phase, SCR-Auth builds the authentication model of the legitimate user. In the authentication phase, SCR-Auth use the built model to determine whether the call receiver is legitimate.

Fig. 5 illustrates the workflow of SCR-Auth, consisting of four key modules: the data capturer, data preprocessor, feature extractor, and authenticator. The data capturer utilizes the smartphone's earpiece speaker and top microphone as an active sonar system. It sends inaudible chirp signals and captures the resulting echoes. The data preprocessor first synchronizes and segments the echo signals through a correlation-based approach. A two-step denoising process is subsequently applied, which involves the use of a bandpass filter followed by the Magnitude-Phase Spectrogram Subtraction (MPSS) method. This approach eliminates ambient noises and other interferences, thereby enhancing the signal from the outer ear. The feature extractor first performs spectrogram analysis to obtain normalized differential spectrograms. Then it extracts the reliable features using a pre-trained CNN model based on transfer learning. The authenticator trains a one-class classification model during the enrollment phase based on the collected samples from the legitimate user. After enrollment, the model determines whether the user is legitimate.

B. Threat Model

For the sake of privacy and convenience, call receivers typically adopt the earpiece mode on smartphones to answer calls, holding the smartphone against their ear and listening through the earpiece speaker [52]. In this paper, we focus on this natural and realistic call-answering scenario. We assume that the attacker has temporary physical access to the victim's smartphone when an incoming call occurs, such as in cases of theft or when the device is left unattended. The attacker's goal is to bypass the proposed call receiver authentication system in order to answer the call and potentially perform sensitive operations. Based on the attacker's capabilities and goals, we consider the following attacks:

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.



(a) The same user (b) Two users (c) The same user (d) Two users

Fig. 4. The acoustic profiles of outer ear echoes for two users. (a) Magnitude spectrums for the same user at two times. (b) Magnitude spectrums for two users. (c) Phase spectrums for the same user at two times. (d) Phase spectrums for two users.

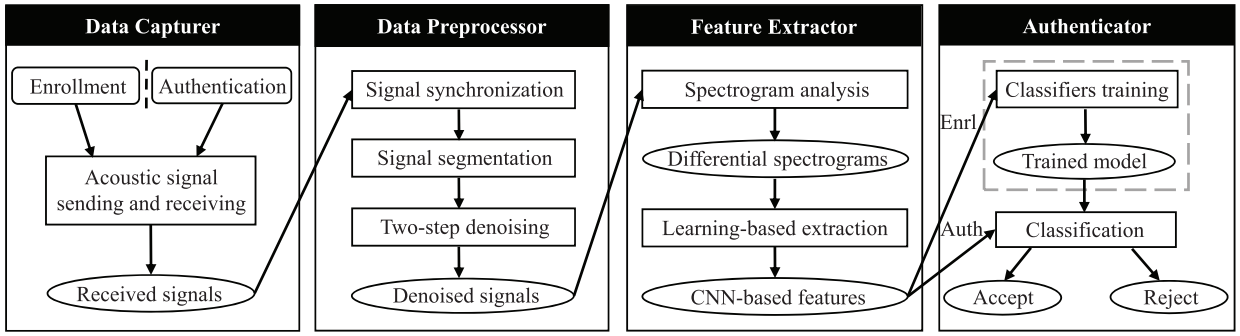


Fig. 5. Workflow of SCR-Auth.

- *Zero-effort attack.* The attacker has no prior knowledge of the legitimate user and simply attempts to hold the smartphone to his/her own ear, hoping to pass authentication by chance.
- *Mimicry attack.* The attacker attempts to impersonate the legitimate user through: (i) Behavioral mimicry, by observing the user during authentication and imitating the smartphone's placement; and (ii) Physical spoofing, by using a fabricated silicone fake ear to deceive the system.
- *Replay attack.* The attacker eavesdrops on acoustic signals during a legitimate authentication process (e.g., using a hidden microphone nearby) and replays the recorded audio to the target smartphone via a speaker.

Other advanced attacks, such as signal injection attacks, are outside the threat model considered in this work, as they require privileged hardware access or specialized external equipment. We provide a further discussion in Section VIII.

C. Design Goals

We think a suitable authentication scheme for a call receiver should satisfy the following goals:

- *Accurate and secure:* The scheme should reliably authenticate the legitimate user with a high success rate while accurately rejecting unauthorized users. It should also defend against common attacks.
- *Implicit:* The authentication process should not impose additional burden and interfere with normal voice conversations.
- *Universal:* It should work on standard commodity smartphones, without requiring additional hardware or root privileges, making it scalable for widespread deployment.
- *Robust:* The scheme should be resilient across varying conditions, such as ambient noises, different postures, different periods, and devices.

V. Design of SCR-Auth

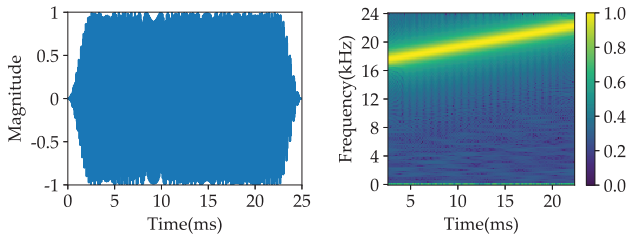
SCR-Auth consists of four modules: data capturer, data preprocessor, feature extractor, authenticator. In this section, we provide a detailed explanation of each module.

A. Data Capturer

SCR-Auth leverages the smartphone's earpiece speaker to emit acoustic sensing signals and the top microphone to receive corresponding echoes. The data capture process integrates seamlessly with the natural call-receiving procedure. Specifically, when a call comes in, the user presses the accept button to answer, which serves as the trigger for the system. Upon this action, the earpiece speaker begins emitting inaudible sensing signals, while the top microphone continuously records the resulting echoes for further processing.

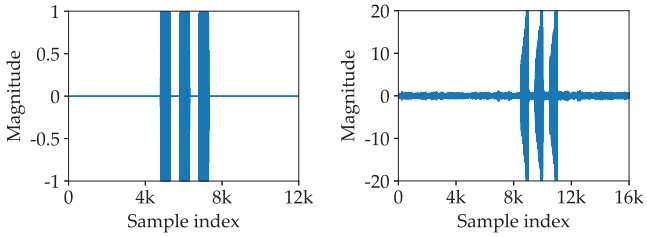
SCR-Auth employs chirp signals as acoustic sensing signals, characterized by a continuously varying frequency over time. Chirp signals are well-suited for acoustic sensing applications due to their excellent auto-correlation properties [53]. Fig.6 illustrates a designed chirp signal used in this study. Research indicates that the upper limit of the human hearing range for adults typically lies between 15-17 kHz [54]. Most smartphones support a maximum sampling rate of 48 kHz [43], which limits the sensing signal's maximum frequency to below 24 kHz in compliance with the Nyquist sampling theorem [55]. To ensure a broad sensing range while remaining imperceptible

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.



(a) A chirp signal (b) The spectrogram

Fig. 6. Illustration of the designed chirp signal in the time and frequency domains.



(a) The transmitted pilot signal (b) The received pilot signal

Fig. 7. The pilot signal for synchronizing the smartphone speaker and microphone.

to users, we adopt the 25-millisecond chirp signal sweeping from 17 kHz to 23 kHz, a range commonly used in acoustic sensing applications [49]. The first and last 120 samples of the chirp are tapered using a Hamming window to reduce potential acoustic annoyance [56]. The interval between two chirps is set to 25 milliseconds, resulting in a sensing signal that alternates between a 1200-sample chirp and a 1200-sample silent period.

B. Data Preprocessor

After capturing the acoustic signals, we proceed with a series of preprocessing steps: synchronization, segmentation, and denoising.

1) Signal Synchronization and Segmentation: To ensure precise segmentation of the acoustic signals, we propose a two-step synchronization approach that aids in the alignment of the signals for further analysis.

Initially, a pilot signal is appended before the sensing sequence to provide coarse synchronization between the smartphone's speaker and microphone [57]. This pilot signal consists of three 500-sample chirps, sweeping from 22 kHz to 18 kHz. An example of the transmitted pilot signal is shown in Fig.7(a), with the corresponding received pilot signal depicted in Fig.7(b). By detecting the presence of this pilot, we can identify the starting point of the sensing process during the call reception. Once coarse synchronization is achieved, the system proceeds to divide the received signals into 50-millisecond segments, each corresponding to a single sensing event.

In the second step, a finer level of synchronization is applied within each segment to counteract any timing drifts or distortions caused by the transmission channel. For each segment, a matched filter is used to precisely determine the arrival time of the transmitted chirp signal [58]. Specifically,

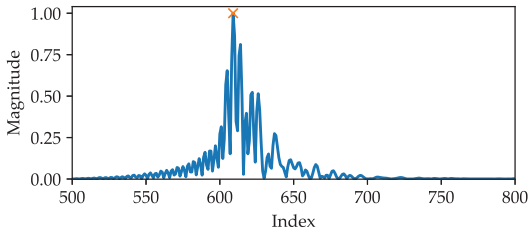


Fig. 8. An example of the cross-correlation result.

the cross-correlation C_{xy} between a received signal segment $r(t)$ and the transmitted chirp signal $s(t)$ is calculated, as expressed in Eq. 5.

$$C_{xy} = r(t) * s^{\star}(-t) \tag{5}$$

Here, * denotes the convolution operator, and $s^{\star}(-t)$ is the complex conjugate of $s(-t)$. Fig. 8 illustrates an example of the cross-correlation result. The index of the highest peak of the cross-correlation result is identified as the start point. Based on the length of the chirp signal, we finally derive 1200-sample segments.

2) Signal Denoising: Due to the multipath effect, the received signals include not only outer ear echoes, but also direct path signals and environmental reflections. Additionally, ambient noises are inevitably introduced during sound propagation. In this study, we propose a two-step denoising approach that combines bandpass filtering with magnitude-phase spectrogram subtraction (MPSS) to effectively suppress unwanted interference.

In the first step, we address ambient noises by applying a Butterworth bandpass filter to remove out-of-band interference [59]. The filter's cutoff frequencies are set at 17 kHz and 23 kHz, corresponding to the expected frequency range of the chirp signal. This selective filtering ensures that only the relevant frequency components are retained, thereby improving the signal-to-noise ratio.

In the second step, we apply the MPSS method to suppress the interference from direct path signals and environmental reflections. The key idea is to carefully choose a reference segment that primarily contains direct path signals and environmental reflections, devoid of outer ear echoes. Since interference components, such as direct path signals and static objects reflections, remain consistent during sensing. By subtracting these interference components, we can highlight echoes from the outer ear.

By analyzing the process of call reception, we select the the first signal segment, captured immediately after the user clicks the "accept" button, as the reference segment. At this point, the smartphone is typically stationary and has not yet been placed on the ear. Once the smartphone is positioned on the ear, changes in the received signal can be attributed to echoes from the ear. The reference segment plays two crucial roles: it acts as a template for the direct path signal, eliminating the need for a quiet environment to detect this signal, and provides a baseline for environmental interference during the call.

To perform MPSS, we use the Short-Time Fourier Transform (STFT) [60] to compute the magnitude and phase

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

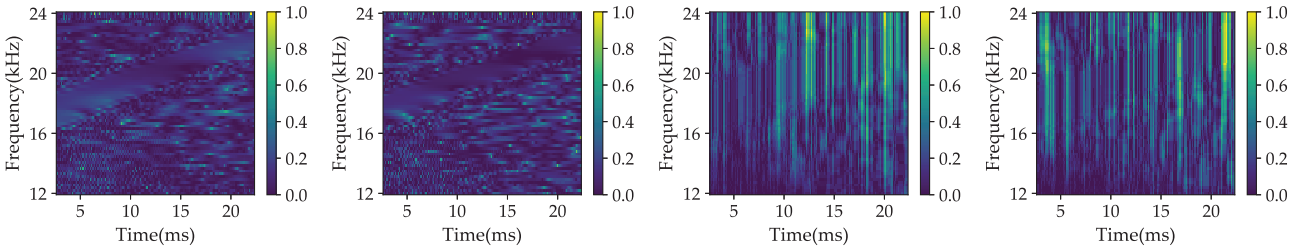


Fig. 9. Normalized magnitude and phase spectrograms for two users. (a) A magnitude spectrogram for user 1. (b) A magnitude spectrogram for user 2. (c) A phase spectrogram for user 1. (d) A phase spectrogram for user 2.

spectrograms for each signal segment, then construct differential spectrograms based on the selected reference segment. Denoting the magnitude spectrogram as $\boldsymbol{S}_m$ and the phase spectrogram as $\boldsymbol{S}_p$, the combined magnitude-phase spectrograms can be expressed as $\boldsymbol{Spec} = [\boldsymbol{S}_m; \boldsymbol{S}_p]$. The differential spectrogram, represented as $\triangle\boldsymbol{Spec} = [\triangle\boldsymbol{S}_m; \triangle\boldsymbol{S}_p]$, is then calculated according to the Eq. 6:

$$\triangle\boldsymbol{Spec} = |\boldsymbol{Spec}_s - \boldsymbol{Spec}_r| \tag{6}$$

where $\boldsymbol{Spec}_r$ represents the spectrograms of the reference segment, and $\boldsymbol{Spec}_s$ corresponds to the spectrograms of one sensing segment. The differential spectrogram serves as the foundation for feature extraction.

C. Feature Extractor

In this section, we perform spectrogram analysis and use a pre-trained convolution neural network model to extract reliable features.

1) Spectrogram Analysis: The acoustic signals captured by the top microphone undergo complex interactions with the user's outer ear, including absorption and reflection. These interactions are primarily governed by the unique physiological characteristics of the auricle and ear canal, such as their shape, geometry, and tissue composition. As shown in the theoretical analysis (Section III-A), the outer ear causes user-specific attenuation and propagation delays across different frequency components of the echoes. To capture these effects, we focus on the magnitude and phase spectrograms: the magnitude spectrogram reflects frequency-selective attenuation shaped by the outer ear's geometry and material properties, while the phase spectrogram encodes the propagation delays introduced by user-specific acoustic paths. Together, these representations preserve user-discriminative outer ear biometric information.

Based on the output of the preprocessor, we construct the normalized differential spectrogram, which includes both magnitude and phase spectrograms. We first reduce computational overhead by focusing on informative spectral regions, followed by min-max normalization [61] to scale all values to the range [0, 1]. Specifically, we retain only the frequency components above a threshold f_{thre}, which is empirically set to 12 kHz. Given a sampling rate of $f_s = 48$ kHz and an FFT size of $N_{fft} = 256$, the corresponding FFT bin index is calculated as $I_{thre} = \frac{f_{thre} \times N_{fft}}{f_s} = 64$. The refined differential spectrogram is denoted as $\triangle\boldsymbol{Spec}_{emp} = [\triangle\boldsymbol{S}_{mr}; \triangle\boldsymbol{S}_{pr}]$, where $\triangle\boldsymbol{S}_{mr} = \triangle\boldsymbol{S}_m(I_{thre}:,:)$ and $\triangle\boldsymbol{S}_{pr} = \triangle\boldsymbol{S}_p(I_{thre}:,:)$. This results in a spectrogram of size $65 \times 158 \times 2$. Finally, the normalized spectrogram $\triangle\boldsymbol{Spec}_{norm}$ is computed as:

$$\triangle\boldsymbol{Spec}_{norm} = \frac{\triangle\boldsymbol{Spec}_{emp} - \min(\triangle\boldsymbol{Spec}_{emp})}{\max(\triangle\boldsymbol{Spec}_{emp}) - \min(\triangle\boldsymbol{Spec}_{emp})} \tag{7}$$

As an example, we present the normalized magnitude and phase spectrograms of two users in Fig. 9. We can observe that spectrograms show differences for different users. These spectrograms are later used as inputs for model training.

2) Learning-Based Feature Extraction: To extract reliable features from magnitude-phase spectrograms, we design a learning-based feature extractor to mitigate the variability caused by smartphone position changes. The foundation of this extractor is a convolutional neural network (CNN) with superior capabilities in feature extraction and representation [62], [63]. Leveraging multi-user data collected under diverse natural smartphone positions during call reception, we train the CNN model using supervised learning to extract identity-related features while disregarding secondary factors, such as changes in the relative position between the ear and smartphone. Based on transfer learning [64], we remove the final layer of the pre-trained CNN and use the output from the 15th layer (as detailed in Table II) as a generalized feature extractor. This approach enables the network to effectively capture effective features of the outer ear.

Table II presents the architecture of our base CNN model, which is designed with multiple convolutional layers to effectively extract features. Each two-dimensional convolutional (Conv2D) layer employs the rectified linear unit (ReLU) as its activation function, mitigating the vanishing gradient problem.

TABLE II
THE STRUCTURE OF OUR BASE CNN MODEL

Layer	Layer type	Output shape	# Param
1	Conv2D + ReLU	(63,156,16)	304
2	Conv2D + ReLU	(61,154,16)	2,320
3	Max Pooling	(30,77,16)	0
4	Dropout	(30,77,16)	0
5	Conv2D + ReLU	(28,75,32)	4,640
6	Conv2D + ReLU	(26,73,32)	9248
7	Max Pooling	(13,36,32)	0
8	Dropout	(13,36,32)	0
9	Conv2D + ReLU	(11,34,16)	4,624
10	Conv2D + ReLU	(9,32,16)	2,320
11	Max Pooling	(4,16,16)	0
12	Dropout	(4,16,16)	0
13	Flatten	(1024)	0
14	Dense + ReLU	(128)	131,200
15	Dropout	(128)	0
16	Dense + Softmax	(30)	3,870

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

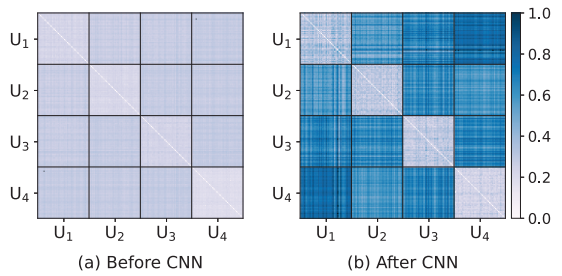


Fig. 10. Pairwise Euclidean distance heatmaps of feature vectors before (left) and after (right) CNN-based feature extraction.

The max-pooling layer is used to down-sample the data from the previous activation layer, which reduces the data dimension and saves computational costs. Dropout layers are added after the max pooling layers to prevent overfitting. The final layer of the model is a dense layer with a softmax activation function, which outputs the probability distribution for each class. The kernel sizes for the Conv2D and max pooling layers are set to 3×3 and 2×2, respectively. The whole model contains 158,526 parameters.

The base CNN model is trained using data from 30 participants, with each contributing 500 acoustic samples. Aligned with natural call reception habits, participants are asked to place the smartphone down and pick it up again, simulating a variety of smartphone positions. We employ the Adam optimizer for parameter optimization and use categorical cross-entropy as the loss function. The training process is performed with a batch size of 50 over 10 epochs. Once trained, the base model serves as the foundation of our feature extractor, eliminating the need for retraining when applied to unseen users. Leveraging the concept of transfer learning, we transform the pre-trained base model into a generalized feature extractor by removing its final layer (i.e., the 16th layer) and retaining the preceding layers. This transformation results in a lightweight 659 kB feature extractor, optimized for deployment on mobile devices. Finally, the feature extractor generates a 128-dimensional feature vector, which is utilized in each authentication process to ensure efficient performance.

To analyze how the CNN contributes to feature extraction and investigate the effectiveness of the extracted features, we conduct a comparative analysis between the input and output representations across different users. Each sample is initially represented as a magnitude-phase spectrogram of shape $65 \times 158 \times 2$, and is mapped to a 128-dimensional embedding through the CNN. We compute pairwise Euclidean distances among 200 samples (4 users $\times$ 50 samples each) under two settings: (i) based on raw spectrograms, and (ii) based on CNN-extracted embeddings. The resulting distance matrices are visualized as heatmaps in Fig. 10, where both axes represent the 200 feature vectors. Brighter (whiter) colors indicate smaller distances, while darker (bluer) colors indicate larger distances. The results demonstrate that, compared to raw spectrograms, CNN-extracted features exhibit improved intra-class compactness and inter-class separability. This confirms that the CNN effectively learns user-discriminative features while suppressing irrelevant variations.

D. Authenticator

In our scenario, the training dataset exclusively consists of samples from the legitimate call receiver. Therefore, the authentication task can be formulated as a one-class classification problem, also known as a novelty detection problem [65]. During the enrollment phase, we use the feature vectors extracted by the pre-trained CNN-based model to train a one-class classifier for the legitimate user. During authentication, the classifier determines whether the incoming sample originates from the legitimate user. We consider two standard novelty detection methods for this task: one-class support vector machine (OCSVM) [66] and local outlier factor (LOF) [67].

VI. Data Collection

To collect the experiment data, we develop an Android data collection app. We use the earpiece speaker to send inaudible sensing signals and the top microphone as the receiver. After receiving approval from our university's institutional review board (IRB), we started our data collection. We recruited 37 participants, aged from 20 to 27 (graduate and undergraduate students), including 19 males and 18 females. We explicitly informed the participants that the purpose of the experiments was to authenticate the receiver of a call. Similar to answering a call, participants were required to click the start button and picked up the smartphone toward their ear. They were allowed to make slight adjustments to the smartphone's position to cover different situations. In our data collection, we compiled the following 8 datasets.

A. Dataset-1

This dataset is used to train our CNN-based feature extraction model. We recruited 30 participants to collect acoustic signals on Google Pixel 3a. For each of them, we collected 500 acoustic signals. In total, we collected $30 \times 500 = 15,000$ acoustic signals for CNN model training.

B. Dataset-2

This dataset is utilized to evaluate the overall performance of our system, which is collected under basic settings. We collected acoustic sensing data from 30 participants on Google Pixel 3a. Participants were seated naturally in a quiet environment. We collected 500 acoustic signals for each participant. Besides, we collected acoustic sensing data from 7 unseen participants to evaluate the performance of the CNN-based feature extraction model for new users. We collected 500 acoustic signals for each new participant.

C. Dataset-3

To evaluate the performance of continuous authentication, we collected acoustic sensing data from two situations: listening and speaking. Therefore, we recruited 5 participants and performed acoustic sensing every 1s. We collected 600 acoustic signals for each participant while they were solely listening and another 600 acoustic signals while they were speaking. In total, we collected $5 \times 600 \times 2 = 6,000$ acoustic signals for dataset-3.

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

D. Dataset-4

To evaluate the influence of ambient noises, we use a laptop as the noise source to simulate the noisy environment. The laptop played the song 'Human Sound/Restaurant2' at 50% volume, which contains common noises in daily life. The sound pressure in this noise environment is about 60-62dB. 30 participants performed this experiment. We collected 500 acoustic signals for each participant in the noisy environment. Dataset-4 involves $30 \times 500 = 15,000$ acoustic signals.

E. Dataset-5

To evaluate the authentication performance over time, we collected data from different time periods. Dataset-2 is collected in the first round of collection. For 30 participants, we collected data one week and two weeks after the first collection round. For each round of collection, the acoustic signals are $30 \times 500 = 15,000$. We finally got 30,000 acoustic signals for dataset-5.

F. Dataset-6

To evaluate the influence of human postures, we consider four common postures: sitting, standing, walking, and running. Dataset-2 was collected under the sitting posture. In this dataset, we recruited 10 participants and collected acoustic data for standing, walking, and running postures. For each participant, we collected 250 acoustic signals for each posture. Finally, we obtained $10 \times 3 \times 250 = 7,500$ acoustic signals.

G. Dataset-7

To evaluate the impact of different smartphone positions, we conducted controlled experiments involving variations in both angle and distance. Specifically, we considered four tilt angles and four distances between the smartphone and the ear. Three participants were recruited, and for each angle and each distance setting, 500 acoustic signals were collected per participant. In total, we obtained $3 \times (4 + 4) \times 500 = 12{,}000$ acoustic signals.

H. Dataset-8

To evaluate the performance of our system on different devices, we collected acoustic data on two extra smartphones: Google Pixel 4 and Vivo S12. 10 participants are recruited to do this experiment. For each participant, we collected 500 acoustic signals on each device. As a result, we got $10 \times 2 \times 500 = 10,000$ acoustic signals.

I. Dataset-9

To evaluate the system's resilience against various attacks, we collected four attack datasets using a Google Pixel 3a: i) Dataset-9A (Zero-effort attack): Seven participants, without prior knowledge, attempted to guess how legitimate users hold the smartphone. We collected $7 \times 500 = 3,500$ acoustic signals. ii) Dataset-9B (Behavioral mimicry attack): The same participants observed and imitated the smartphone placement of legitimate users. We got another $7 \times 500 = 3,500$ signals.

TABLE III
MEAN/STANDARD DEVIATION OF BAC(%), EER(%), AND AUC UNDER TWO DIFFERENT ONE-CLASS CLASSIFIERS

Classifier	Mean/Std BAC	Mean/Std EER	Mean/Std AUC
OCSVM	96.95/1.45	1.53/1.35	0.9982/0.0025
LOF	96.13/3.18	1.90/1.56	0.9972/0.0034

iii) Dataset-9C (Fake ear attack): A fabricated silicone ear was used to spoof the system, resulting in 2,000 acoustic signals. iv) Dataset-9D (Replay attack): We pre-recorded 2,000 legitimate acoustic signals and replayed them to the target smartphone using a speaker to simulate replay attacks.

VII. EVALUATION

In this section, we report the evaluation results of the proposed system. We first present the evaluation metrics, and show the overall performance of SCR-Auth. Additionally, we evaluate its effectiveness under different settings and security against attacks. Finally, we present the authentication latency of our system.

A. Evaluation Metrics

There are four possible results of classification: True acceptance (TA), True rejection (TR), False acceptance (FA), False rejection (FR). We use the following metrics to evaluate the performance of SCR-Auth. True acceptance rate is defined as $TAR = \frac{TA}{TA+FR}$, which measures the proportion of samples classified as positive among legitimate user samples. True rejection rate is defined as $TRR = \frac{TR}{TR+FA}$, which measures the proportion of samples classified as negative among illegal user samples. Balanced accuracy (BAC) is the average of true acceptance rate and true rejection rate, which is defined as $BAC = \frac{1}{2}(TAR + TRR)$. It is used to evaluate the accuracy of imbalanced datasets. A higher BAC means better performance of the system. False acceptance rate ($FAR = \frac{FA}{FA+TR}$) represents the rate at which illegal samples are wrongly accepted. False rejection rate ($FRR = \frac{FR}{FR+TA}$) represents the rate at which legitimate samples are wrongly rejected. Receiver operation characteristic (ROC) shows dynamic changes of TAR against FAR at different classification thresholds. The area under the ROC curve (AUC) is used to measure the probability that prediction scores of legitimate users are higher than illegal users. Equal error rate (EER) is the point on the ROC curve, where FAR is equal to FRR. A larger AUC and lower EER mean better performance of the system.

B. Overall Performance

1) Performance of Different Classifiers: We use 30 users in dataset-2 to evaluate the authentication effectiveness of SCR-Auth. We employ a 5-fold cross-validation for each user to split the data and train a one-class classifier. Then we test the classifier model using the remaining data of the user as well as data from other users.

This study considers two types of one-class classifiers: one-class support vector machine (OCSVM) and local outlier factor (LOF). Parameters such as the *kernel*, γ, and ν significantly

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

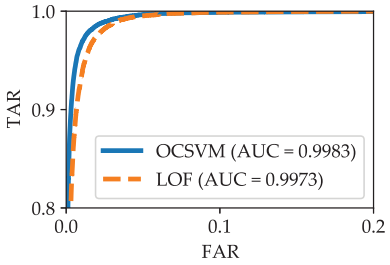


Fig. 11. ROC curves of two classifiers with the best parameters.

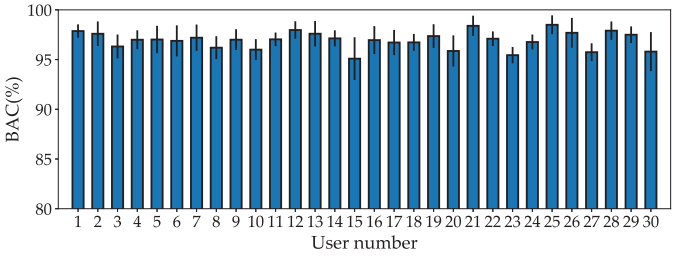


Fig. 12. BAC performance for each user.

TABLE IV
MEAN/STANDARD DEVIATION OF BAC(%), EER(%), AND AUC FOR NEW USERS

Classifier	Mean/Std BAC	Mean/Std EER	Mean/Std AUC
OCSVM	96.48/1.63	2.78/1.95	0.9955/0.0043
LOF	93.49/4.59	2.89/2.21	0.9946/0.0063

impact the results for OCSVM, while for LOF, we consider the *n_neighbors* parameter. We employ grid search to find the best parameter combinations for each classifier. Ultimately, we determine that the radial basis function kernel works best for OCSVM, with γ = 'scale' and $\nu = 0.01$. For LOF, the optimal *n_neighbors* value is 3. Fig. 11 presents the ROC curves of the two classifiers with the best parameters. The AUC for OCSVM is 0.9983, and for LOF, it is 0.9973. A higher AUC value suggests better system performance. The results indicate that the OCSVM classifier outperforms the LOF classifier. Table III shows the mean and standard deviation of BAC, EER, and AUC metrics under two classifiers. OCSVM demonstrates superior BAC and EER metrics compared to LOF, thus we select it as our classifier for subsequent evaluations. This experiment reveals that SCR-Auth achieves an average BAC of 96.95% and an EER of 1.53% using the OCSVM classifier. These results indicate that SCR-Auth is effective in distinguishing users.

2) Per-User Breakdown Analysis: To evaluate the performance of SCR-Auth across 30 different users, we present the BAC of each user under the OCSVM classifier, as shown in Fig. 12. Notably, user #25 achieves the highest BAC of 98.5%, marking the best case among all participants. While the performance of SCR-Auth varies across users, the BAC for every user exceeds 95%, demonstrating the overall effectiveness of SCR-Auth.

3) Performance of Feature Extractor on Unseen Users: To evaluate the performance of the CNN-based feature extractor on new users, we use data from 7 unseen participants, as described in dataset-2, who are not included in the CNN model's pre-training. We use 5-fold cross-validation to split the data. Then we train a one-class SVM (OCSVM) classifier and a local outlier factor (LOF) classifier for each participant. Table IV shows the mean and standard deviation of BAC, EER, and AUC metrics for new users under two classifiers. The BACs for OCSVM and LOF are 96.48% and 93.49%, respectively. Compared to results in Table III, the BAC falls 0.47% for OCSVM and falls 2.64% for LOF. For the OCSVM classifier, the BAC is over 96%, demonstrating the feature extractor's effectiveness for new users. Although the feature extractor is trained on limited data, it is still available to a wide range of users.

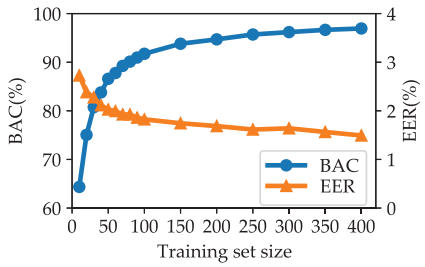


Fig. 13. The BAC and EER for different training set sizes.

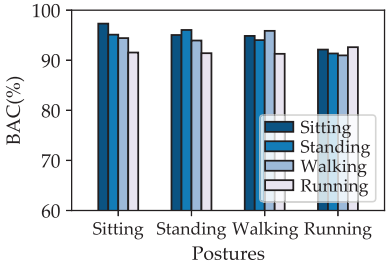


Fig. 14. The BAC of SCR-Auth trained and evaluated under different postures.

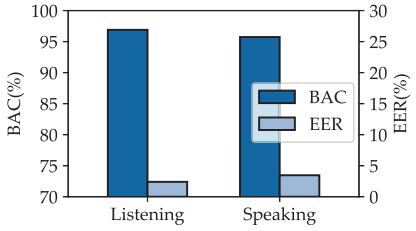


Fig. 15. The BAC and EER performance for continuous authentication.

4) Performance of Continuous Authentication: We analyze two common situations to evaluate the performance of continuous authentication. During the process of answering a call, the receiver will be in one of two states: listening to the caller or speaking to the caller. We train on dataset-2 and test on dataset-3 for evaluation. The results are shown in Fig. 15. For listening and speaking states, the BACs are 96.89% and 95.73%, respectively. The EERs are 2.39% and

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

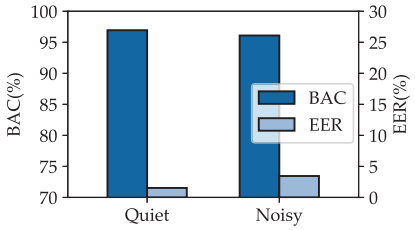


Fig. 16. The BAC and EER performance under different noise conditions.

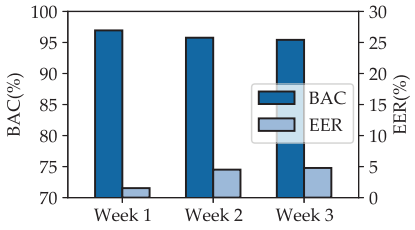


Fig. 17. The BAC and EER performance at different time periods.

3.46%, respectively. The experimental results show that SCR-Auth is available for continuous authentication.

C. Impact Factors Study

1) Impact of Ambient Noises: To assess the impact of ambient noise on system performance, we compare the results under different noise conditions. In this experiment, dataset-2, which is collected in a quiet environment, is used for training. We then evaluate the system's performance on both dataset-2 (for quiet conditions) and dataset-4 (for noisy conditions). Fig. 16 shows the BACs and EERs in both quiet and noisy environments. The BACs are 96.95% and 96.09%, and the EERs are 1.53% and 3.45%, respectively. These results present that SCR-Auth is available for different noise conditions.

2) Impact of Training Dataset Size: To investigate the impact of training set size, we change the amount of training data points for each user on dataset-2. Specifically, for each user, we vary the training data points from 10 to 400 in steps of 10 or 50 to train a one-class SVM classifier. Then we test on the rest of the data. Fig. 13 shows the BAC and EER for different training set sizes. As the size of the training set increases from 10 to 400, the BAC rises from 64.35% to 96.94%. The EER falls to 1.49% from 2.73% when the training set size increases from 10 to 400. That may be because the classifier can learn a better boundary with more legitimate data. The BAC is over 90% with 80 training data points and is over 95% with 200 training data points. With 50 training data points, the EER is less than 2%. These results show that our system is practical on mobile devices.

3) Impact of Different Postures: To evaluate the impact of different postures, we use 10 participants' data in dataset-2 and dataset-6. The data in dataset-2 is collected when the participant is sitting. Dataset-6 contains data on standing, walking, and running. We take turns selecting one posture for training and testing the other postures for each participant. For example, we train on sitting posture data and test on

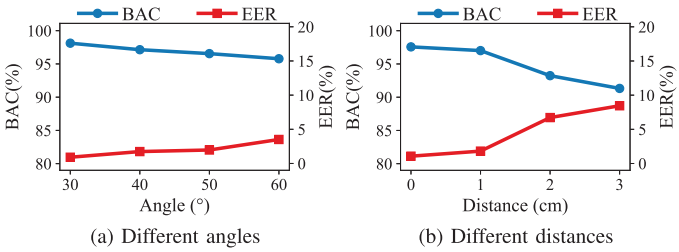


Fig. 18. The BAC and EER performance for different smartphone postions.

sitting, standing, walking, and running posture data. Similarly, we train on the other three postures. Fig. 14 shows the BAC of SCR-Auth under different postures. For example, when we use sitting data for training and testing on the rest of the data, the BACs for the four postures are 97.28%, 95.11%, 94.43%, and 91.55%, respectively. As observed, the highest BAC is achieved when the posture during both training and testing remains the same. The results further reveal that SCR-Auth performs better in sitting, standing, and walking postures than in running. In fact, receiving a call while running is relatively uncommon. Excluding the 'running' condition, SCR-Auth achieves a BAC of over 94% in all other postures, underscoring its applicability across diverse postures.

4) Impact of Different Smartphone Positions: To evaluate the performance of our system under varying smartphone placements, we conducted a dedicated experiment involving three participants. In Dataset-7, we varied two key parameters: the tilt angle of the smartphone and its distance from the ear canal. Specifically, we considered four tilt angles: 30°, 40°, 50°, and 60° counterclockwise from the vertical orientation (0°). In addition, we tested four distances between the smartphone and the ear: 0 cm, 1 cm, 2 cm, and 3 cm. For angle variation evaluation, the authentication model was trained using data collected at 30°, and tested separately on the remaining angles. The resulting BACs are 98.12%, 97.14%, 96.55%, and 95.78%, with corresponding EERs of 0.91%, 1.74%, 1.97%, and 3.51%. For distance variation, the model was trained using data at 0 cm and tested on data from other distances. The resulting BACs were 97.57%, 97.01%, 93.24%, and 91.32%, with EERs of 1.07%, 1.81%, 6.72%, and 8.45%. As shown in Fig. 18, the system maintains strong authentication performance across all tested angles and distances within 2 cm, consistently achieving BAC above 95% and EER below 5%. However, performance begins to degrade when the distance exceeds 2 cm. This degradation may be attributed to the fact that participants naturally tend to place the smartphone within 0-2 cm of the ear during normal usage. To enhance robustness against more extreme placement variations, SCR-Auth can be extended to support model updating based on newly collected samples.

5) Performance Over Time: In this experiment, dataset-2 is used for training, while testing is performed on both dataset-2 and dataset-5. Specifically, data in dataset-2 is collected during the first week, and data from the subsequent two weeks is included in dataset-5. Fig. 17 shows the BACs and EERs across different weeks. For weeks 2 and 3, the BACs are 95.77% and 95.42%, while the EERs are 4.51% and 4.78%,

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

TABLE V

MEAN/STANDARD DEVIATION OF BAC(%), EER(%), AND AUC FOR THREE DIFFERENT DEVICES

Device	Mean/Std BAC	Mean/Std EER	Mean/Std AUC
Pixel 3a	97.32/1.55	0.85/0.80	0.9994/0.0011
Pixel 4	97.62/1.47	0.86/1.29	0.9989/0.0027
Vivo S12	95.20/1.19	4.03/1.02	0.9926/0.0036

respectively. Compared to week 1, the BACs for week 2 and week 3 show slight drops of 1.18% and 1.53%. This decrease may be attributed to changes in users' postures while holding the device. To address this issue, SCR-Auth can be designed to update the authentication model using newly collected data, which is known as the model updating mechanism [68].

6) Impact of Different Devices: We collected data from three smartphones to evaluate the performance on different devices. Specifically, we use Dataset-2, which was collected using the Pixel 3a, and Dataset-8, which contains data from both the Pixel 4 and Vivo S12. For each device, we use the fixed CNN-based feature extractor and train a new one-class SVM for each user using data collected on that device. This mimics the real-world scenario where a user switches to a new device and goes through a lightweight re-enrollment process by providing a small number of samples. As shown in Table V, the mean BACs for the Pixel 3a, Pixel 4, and Vivo S12 are 97.32%, 97.62%, and 95.20%, respectively, and the corresponding average EERs are 0.85%, 0.86%, and 4.03%. The results indicate the effectiveness of our system on different devices.

D. Evaluation of Attack Resistance

To evaluate the system's security against four types of attacks, we test the authentication model trained on Dataset-2 using attack samples from Dataset-9. The false acceptance rate (FAR) is adopted to quantify the percentage of illegitimate samples that were mistakenly accepted. In addition, we analyze the distribution and kernel density estimation (KDE) of the prediction scores for each attack using a Gaussian kernel to visualize the statistical characteristics of the outputs.

1) Zero-Effort Attack: In this scenario, attackers randomly place the smartphone against their own ear. The system yields a FAR of 0.94% with a mean prediction score of -0.399, indicating a low likelihood that random attempts can bypass authentication. This result highlights the discriminative power of the biometric features extracted from outer ear echoes.

2) Mimicry Attack: We evaluate two forms of mimicry: (i) behavioral mimicry attack, and (ii) fake ear attack. In the behavioral mimicry attack, attackers observe the legitimate user and attempt to replicate the smartphone placement. This results in a FAR of 1.40% and a mean prediction score of -0.416. In the fake ear attack, a fabricated silicone ear is used to spoof the system, yielding a FAR of 1.05% and a mean score of -0.429. Despite these efforts, both attacks exhibit low success rates. This is primarily because SCR-Auth captures not only the structural geometry of the outer ear, but also its tissue and material properties, which are difficult to replicate via visual observation or physical fabrication.

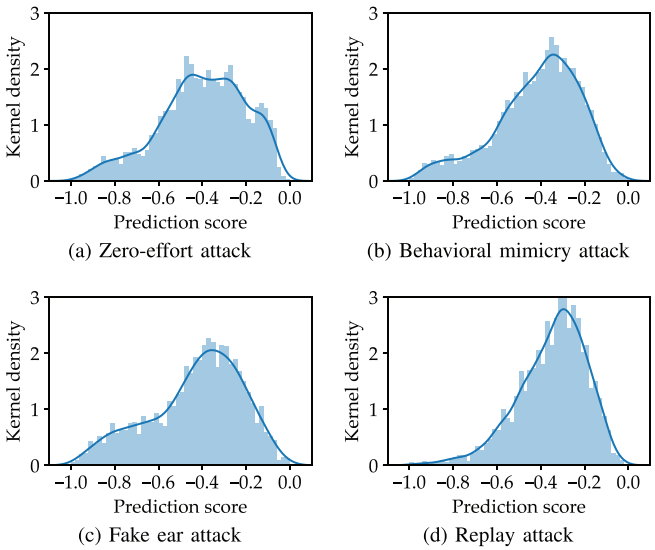


Fig. 19. The kernel density of attack dataset's prediction scores under four different attack types.

TABLE VI

BYPASSED SAMPLES, FAR(%) AND MEAN PREDICTION SCORES UNDER FOUR DIFFERENT ATTACKS

Attack	Bypassed samples	FAR	Prediction scores
Zero-effort attack	33(3500)	0.94	-0.399
Behavioral mimicry	49(3500)	1.40	-0.416
Fake ear attack	21(2000)	1.05	-0.429
Replay attack	31(2000)	1.55	-0.348

3) Replay Attack: In this scenario, attackers replay pre-recorded legitimate signals through a speaker. The system records a FAR of 1.55% and a mean prediction score of -0.348. This is primarily because replayed signals fail to reproduce the real-time physical interactions, such as the phone-to-ear distance and angle. In addition, the recording and playback process introduces distortions and amplifies environmental reflections, which degrade the fine-grained echo features essential for accurate authentication.

The consistently low FARs across all four attack types, as shown in Table VI and Figure 19, demonstrate the effectiveness of SCR-Auth in resisting a wide range of practical spoofing threat.

E. Authentication Latency

We define the authentication latency of our system as the time from recording the received signal to producing the authentication result. Therefore, it consists of time for three modules: data preprocessing, feature extraction, and classification. We developed a prototype system named SCR-Auth on Android to evaluate the authentication latency. We evaluate one sensing process and compute the average latency from 50 tries. On Google Pixel 3a, the average authentication latency for the three modules is 82.8ms, 57.6ms, and 69.6ms, respectively. In total, SCR-Auth requires 0.21s to complete authentication.

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

VIII. Discussion

This section discusses the limitations of our current work and outlines directions for future improvement.

Our study focuses on a realistic call-answering scenario where users hold the smartphone to the ear in earpiece mode. We acknowledge that SCR-Auth does not support all usage modes. However, earpiece mode remains highly relevant in practical situations such as ensuring privacy in public spaces, responding quickly to urgent calls, or when earphones are unavailable. Our goal is to secure this meaningful yet often overlooked call-answering modality. SCR-Auth may reject a legitimate user who registers with one ear but attempts authentication with the other. In future work, we plan to extend the system to support bilateral ear modeling.

When users switch smartphones, only the lightweight one-class SVM requires retraining, while the CNN-based feature extractor remains fixed. This design supports fast and user-friendly re-enrollment. To further enhance cross-device adaptability, we aim to investigate advanced techniques such as device-invariant feature learning and domain adaptation.

We evaluate the system's resilience against four types of practical attacks, assuming adversaries have temporary access to the device during incoming calls. While more advanced threats such as signal injection are theoretically possible, they typically require privileged hardware access and specialized equipment, which are beyond the threat model of this work. Future research could explore integrating liveness detection mechanisms to counter such attacks.

SCR-Auth is designed to operate seamlessly during the natural act of answering a phone call, without requiring any additional user interaction. During our experiments, participants did not report noticeable discomfort or interruption, suggesting good compatibility with natural behaviors. The authentication latency measured in Section VII-E indicates that the system can complete verification promptly, without introducing perceptible delay. Future work will include broader user experience studies to systematically evaluate perceived usability, satisfaction, and trust in real-world scenarios.

While our experiments involved a group of participants and three smartphones, larger-scale studies are essential to confirm SCR-Auth's applicability in diverse real-world scenarios. Future work will expand the user base and device diversity to further assess the system's performance.

IX. Conclusion

In this paper, we propose SCR-Auth, a secure and implicit call receiver authentication scheme for smartphones that leverages outer ear echoes. SCR-Auth utilizes the earpiece speaker to emit inaudible sensing signals and the top microphone to record echoes. In particular, we propose a specially designed two-step denoising method that combines bandpass filtering with magnitude-phase spectrogram subtraction (MPSS) to effectively suppress unwanted interference. Furthermore, we design a learning-based feature extractor to counteract the position variability, while a one-class classifier is used to verify the legitimacy of the call receiver. Comprehensive experiments demonstrate that SCR-Auth achieves an average balanced accuracy of 96.95% and can defend against potential attacks.

References

[1] *Pew Research Center*. Accessed: Jul. 20, 2024. [Online]. Available: https://www.pewresearch.org/internet/2010/09/02/cell-phones-and-american-adults/

[2] *Communications Data*. Accessed: Oct. 2, 2024. [Online]. Available: https://privacyinternational.org/long-read/3176/how-intrusive-communications-data

[3] *Phone Banking Services*. Accessed: Oct. 10, 2024. [Online]. Available: https://www.hsbc.com.hk/ways-to-bank/phone/

[4] P. Moh, A. Yang, N. Malkin, and M. L. Mazurek, "Understanding how people share passwords," in *Proc. 20th Symp. Usable Privacy Secur. (SOUPS)*, 2024, pp. 219–237.

[5] P. Terhörst, A. Boller, N. Damer, F. Kirchbuchner, and A. Kuijper, "MiDeCon: Unsupervised and accurate fingerprint and minutia quality assessment based on minutia detection confidence," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Aug. 2021, pp. 1–8.

[6] B. Meden et al., "Privacy–enhancing face biometrics: A comprehensive survey," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 4147–4183, 2021.

[7] *Phone Call Modes on Smartphones*. Accessed: Apr. 11, 2025. [Online]. Available: https://support.google.com/phoneapp/answer/2811745

[8] M. Conti, I. Zachia-Zlatea, and B. Crispo, "Mind how you answer me! Transparently authenticating the user of a smartphone when answering or placing a call," in *Proc. 6th ACM Symp. Inf., Comput. Commun. Secur. (AsiaCCS)*, 2011, pp. 249–259.

[9] W.-H. Lee, X. Liu, Y. Shen, H. Jin, and R. B. Lee, "Secure pick up: Implicit authentication when you start using the smartphone," in *Proc. 22nd ACM Symp. Access Control Models Technol.*, Jun. 2017, pp. 67–78.

[10] A. Eremin, K. Kogos, and Y. Valatskayte, "Touch and move: Incoming call user authentication," in *Proc. Int. Conf. Inf. Syst. Secur. Privacy (ICISSP)*, 2019, pp. 26–39.

[11] A. Buriro, B. Crispo, and M. Conti, "AnswerAuth: A bimodal behavioral biometric-based user authentication scheme for smartphones," *J. Inf. Secur. Appl.*, vol. 44, pp. 89–103, Feb. 2019.

[12] A. Buriro, B. Crispo, F. D. Frari, J. Klardie, and K. Wrona, "ITSME: Multi-modal and unobtrusive behavioural user authentication for smartphones," in *Proc. Int. Conf. Passwords (PASSWORDS)*, 2015, pp. 45–61.

[13] B. Fan, X. Su, J. Niu, and P. Hui, "EmgAuth: Unlocking smartphones with EMG signals," *IEEE Trans. Mobile Comput.*, vol. 22, no. 9, pp. 5248–5261, Sep. 2023.

[14] P. A. Fahmi, E. Kodirov, D.-J. Choi, G.-S. Lee, A. M. F. Azli, and S. Sayeed, "Implicit authentication based on ear shape biometrics using smartphone camera during a call," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2012, pp. 2272–2276.

[15] N. Damer and B. Führer, "Ear recognition using multi-scale histogram of oriented gradients," in *Proc. 8th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Jul. 2012, pp. 21–24.

[16] C. Holz, S. Buthpitiya, and M. Knaust, "Bodyprint: Biometric user identification on mobile devices using the capacitive touchscreen to scan body parts," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, Apr. 2015, pp. 3011–3014.

[17] J.-L. Cabra, C. Parra, and L. Trujillo, "Earprint touchscreen sensoring comparison between hand-crafted features and transfer learning for smartphone authentication," *J. Internet Services Inf. Secur.*, vol. 12, no. 3, pp. 16–29, 2022.

[18] C. Wu et al., "It's all in the touch: Authenticating users with HOST gestures on multi-touch screen devices," *IEEE Trans. Mobile Comput.*, vol. 23, no. 10, pp. 10016–10030, Oct. 2024.

[19] S. Itani, S. Kita, and Y. Kajikawa, "Multimodal personal ear authentication using acoustic ear feature for smartphone security," *IEEE Trans. Consum. Electron.*, vol. 68, no. 1, pp. 77–84, Feb. 2022.

[20] Y. Mizuho, Y. Kawasaki, T. Amesaka, and Y. Sugiura, "EarAuthCam: Personal identification and authentication method using ear images acquired with a camera-equipped hearable device," in *Proc. Augmented Humans Int. Conf.*, Apr. 2024, pp. 119–130.

[21] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin, "EarEcho: Using ear canal echo for wearable authentication," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–24, 2019.

[22] A. Poosarala, "Uniform classifier for biometric ear and retina authentication using smartphone application," in *Proc. 2nd Int. Conf. Vis., Image Signal Process. (ICVISP)*, 2018, pp. 1–5.

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

[23] A. F. Abate, M. Nappi, and S. Ricciardi, "I-am: Implicitly authenticate me—Person authentication on mobile devices through ear shape and arm gesture," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 3, pp. 469–481, Mar. 2019.

[24] F. Cherifi, K. Amroun, and M. Omar, "Robust multimodal biometric authentication on IoT device through ear shape and arm gesture," *Multimedia Tools Appl.*, vol. 80, no. 10, pp. 14807–14827, Apr. 2021.

[25] M. A. Rilvan, K. I. Lacy, M. S. Hossain, and B. Wang, "User authentication and identification on smartphones by incorporating capacitive touchscreen," in *Proc. IEEE 35th Int. Perform. Comput. Commun. Conf. (IPCCC)*, Dec. 2016, pp. 1–8.

[26] A. H. Akkermans, T. A. Kevenaar, and D. W. Schobben, "Acoustic ear recognition for person identification," in *Proc. 4th IEEE Workshop Autom. Identificat. Adv. Technol. (AutoID)*, Oct. 2005, pp. 219–223.

[27] Y. Wu and J. He, "EarAE: An autoencoder based user authentication using earphones," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2023, pp. 1–6.

[28] T. Arakawa, T. Koshinaka, S. Yano, H. Irisawa, R. Miyahara, and H. Imaoka, "Fast and accurate personal authentication using ear acoustics," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA)*, Dec. 2016, pp. 1–4.

[29] S. Mahto, T. Arakawa, and T. Koshinak, "Ear acoustic biometrics using inaudible signals and its application to continuous user authentication," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 1407–1411.

[30] X. Sun et al., "Earmonitor: In-ear motion-resilient acoustic sensing using commodity earphones," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 4, pp. 1–22, 2023.

[31] Y. Ren et al., "Proximity-echo: Secure two factor authentication using active sound sensing," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2021, pp. 1–10.

[32] C. Cai, H. Pu, L. Ye, H. Jiang, and J. Luo, "Active acoustic sensing for 'hearing' temperature under acoustic interference," *IEEE Trans. Mobile Comput.*, vol. 22, no. 2, pp. 661–673, Feb. 2023.

[33] Y. Su et al., "Embracing distributed acoustic sensing in car cabin for children presence detection," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 8, no. 1, pp. 1–28, Mar. 2024.

[34] H. Chen, F. Li, and Y. Wang, "EchoTrack: Acoustic device-free hand tracking on smart phones," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.

[35] P. Wang, R. Jiang, and C. Liu, "Amaging: Acoustic hand imaging for self-adaptive gesture recognition," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2022, pp. 80–89.

[36] H. Cheng and W. Lou, "PD-FMCW: Push the limit of device-free acoustic sensing using phase difference in FMCW," *IEEE Trans. Mobile Comput.*, vol. 22, no. 8, pp. 4865–4880, Aug. 2023.

[37] L. Lu et al., "LipPass: Lip reading-based user authentication on smartphones leveraging acoustic signals," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2018, pp. 1466–1474.

[38] L. Wu, J. Yang, M. Zhou, Y. Chen, and Q. Wang, "LVID: A multimodal biometrics authentication system on smartphones," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1572–1585, 2020.

[39] R. Zhang et al., "EchoSpeech: Continuous silent speech recognition on minimally-obtrusive eyewear powered by acoustic sensing," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2023, pp. 1–18.

[40] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *Proc. 13th Annu. Int. Conf. Mobile Syst., Appl., Services*, May 2015, pp. 45–57.

[41] X. Song et al., "SpiroSonic: Monitoring human lung function via acoustic sensing on commodity smartphones," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2020, pp. 1–14.

[42] M. Zhou et al., "PressPIN: Enabling secure PIN authentication on mobile devices via structure-borne sounds," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 2, pp. 1228–1242, Mar. 2023.

[43] C. Wu, J. Chen, K. He, Z. Zhao, R. Du, and C. Zhang, "EchoHand: High accuracy and presentation attack resistant hand authentication on commodity mobile devices," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2022, pp. 2931–2945.

[44] Y. Yang, X. Li, Z. Ye, Y. Wang, and Y. Chen, "BioCase: Privacy protection via acoustic sensing of finger touches on smartphone case mini-structures," in *Proc. 21st Annu. Int. Conf. Mobile Syst., Appl. Services*, Jun. 2023, pp. 397–409.

[45] B. Zhou, Z. Xie, Y. Zhang, J. Lohokare, R. Gao, and F. Ye, "Robust human face authentication leveraging acoustic sensing on smartphones," *IEEE Trans. Mobile Comput.*, vol. 21, no. 8, pp. 3009–3023, Aug. 2022.

[46] G. Wang, Q. Yan, S. Patrarungrong, J. Wang, and H. Zeng, "FacER: Contrastive attention based expression recognition via smartphone earpiece speaker," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2023, pp. 1–10.

[47] Z. Xu, T. Liu, R. Jiang, P. Hu, Z. Guo, and C. Liu, "AFace: Range-flexible anti-spoofing face authentication via smartphone acoustic sensing," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 8, no. 1, pp. 1–33, Mar. 2024.

[48] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM Comput. Surv.*, vol. 45, no. 2, pp. 1–35, Feb. 2013.

[49] Z. Wang, S. Tan, L. Zhang, Y. Ren, Z. Wang, and J. Yang, "EarDynamic: An ear canal deformation based continuous user authentication using in-ear wearables," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 1, pp. 1–27, Mar. 2021.

[50] S. Wang, L. Zhong, Y. Fu, L. Chen, J. Ren, and Y. Zhang, "UFace: Your smartphone can 'hear' your facial expression!," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 8, no. 1, pp. 1–27, 2024.

[51] Y. Yang, Y. Wang, Y. Chen, and C. Wang, "EchoLock: Towards low-effort mobile user identification leveraging structure-borne echos," in *Proc. 15th ACM Asia Conf. Comput. Commun. Secur.*, Oct. 2020, pp. 772–783.

[52] C. Wang et al., "MmEve: Eavesdropping on smartphone's earpiece via COTS mmWave device," in *Proc. 28th Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2022, pp. 338–351.

[53] C. Cai, R. Zheng, and J. Luo, "Ubiquitous acoustic sensing on commodity IoT devices: A survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 432–454, 1st Quart., 2022.

[54] D. Purves, G. J. Augustine, D. Fitzpatrick, L. C. Katz, and A. S. Lamantia, *Neuroscience*, 2nd ed. Sunderland, MA, USA: Sinauer Associates, 2001.

[55] Z. Ba et al., "Learning-based practical smartphone eavesdropping with built-in accelerometer," in *Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, 2020, pp. 1–18.

[56] E. C. Ifeachor and B. W. Jervis, *Digital Signal Processing: A Practical Approach*. London, U.K.: Pearson, 2002.

[57] Y.-C. Tung and K. G. Shin, "Expansion of human-phone interface by sensing structure-borne sound propagation," in *Proc. 14th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2016, pp. 277–289.

[58] J. Chen, U. Hengartner, H. Khan, and M. Mannan, "Chaperone: Real-time locking and loss prevention for smartphones," in *Proc. 29th USENIX Secur. Symp. (USENIX Secur.)*, 2020, pp. 325–342.

[59] Y. Xie, F. Li, Y. Wu, H. Chen, Z. Zhao, and Y. Wang, "TeethPass: Dental occlusion-based user authentication via in-ear acoustic sensing," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2022, pp. 1789–1798.

[60] D. Chen, A. B. Wong, and K. Wu, "Fall detection based on fusion of passive and active acoustic sensing," *IEEE Internet Things J.*, vol. 11, no. 7, pp. 11566–11578, Apr. 2024.

[61] J. Liu, W. Song, L. Shen, J. Han, and K. Ren, "Secure user verification and continuous authentication via earphone IMU," *IEEE Trans. Mobile Comput.*, vol. 22, no. 11, pp. 6755–6769, Nov. 2023.

[62] F. Boutros, P. Siebke, M. Klemt, N. Damer, F. Kirchbuchner, and A. Kuijper, "PocketNet: Extreme lightweight face recognition network using neural architecture search and multistep knowledge distillation," *IEEE Access*, vol. 10, pp. 46823–46833, 2022.

[63] F. Boutros, M. Klemt, M. Fang, A. Kuijper, and N. Damer, "ExFaceGAN: Exploring identity directions in GAN's learned latent space for synthetic identity generation," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Sep. 2023, pp. 1–10.

[64] C. Wu, K. He, J. Chen, Z. Zhao, and R. Du, "Liveness is not enough: Enhancing fingerprint authentication with behavioral biometrics to defeat puppet attacks," in *Proc. USENIX Secur. Symp.*, Jan. 2020, pp. 2219–2236.

[65] C. Wu, K. He, J. Chen, R. Du, and Y. Xiang, "CaIAuth: Context-aware implicit authentication when the screen is awake," *IEEE Internet Things J.*, vol. 7, no. 12, pp. 11420–11430, Dec. 2020.

[66] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 1999, pp. 1–7.

[67] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2000, pp. 93–104.

[68] C. Wu, K. He, J. Chen, Z. Zhao, and R. Du, "Toward robust detection of puppet attacks via characterizing fingertip-touch behaviors," *IEEE Trans. Dependable Secure Comput.*, vol. 19, no. 6, pp. 4002–4018, Nov. 2022.

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.

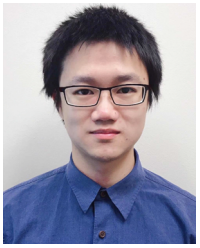
Xiping Sun received the B.E. degree in information security from Wuhan University, China, in 2019, where she is currently pursuing the Ph.D. degree with the School of Cyber Science and Engineering. Her research interests include systems and mobile security.

Jing Chen (Senior Member, IEEE) received the Ph.D. degree in computer science from the Huazhong University of Science and Technology, Wuhan. He was the Vice Chair of the ACM Turing Award Celebration Conference (TURC) 2023. He is currently a Full Professor with the School of Cyber Science and Engineering, Wuhan University. He has published more than 150 research articles in many international journals and conferences, including USENIX Security, ACM CCS, INFOCOM, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON COMPUTERS, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, and IEEE TRANSACTIONS ON SERVICES COMPUTING. His research interests include network security, cloud security, and mobile security. He was twice runner-up for the Best Paper at the INFOCOM 2018 and INFOCOM 2021.

Kun He (Member, IEEE) received the Ph.D. degree from Wuhan University, Wuhan, China. He is currently an Associate Professor at Wuhan University. He has published more than 70 research articles in various journals and conferences, such as S&P, USENIX Security, CCS, INFOCOM, ICSE, UbiComp, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, and IEEE TRANSACTIONS ON MOBILE COMPUTING. His research interests include cryptography and data security.

Zhixiang He is currently pursuing the Ph.D. degree with the School of Cyber Science and Engineering, Wuhan University, Hubei, China. His research interests include human privacy and mobile sensing. His research outcomes have appeared in UbiComp and FCS.

Ruiying Du received the B.S., M.S., and Ph.D. degrees in computer science from Wuhan University, Wuhan, China, in 1987, 1994, and 2008, respectively. She is a Professor at the School of Cyber Science and Engineering, Wuhan University. She has published more than 80 research articles in many international journals and conferences, such as IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, USENIX Security, CCS, INFOCOM, SECON, TrustCom, and NSS. Her research interests include network security, wireless networks, cloud computing, and mobile computing.



Yebo Feng received the Ph.D. degree in computer science from the University of Oregon (UO) in 2023. He is a Research Fellow with the School of Computer Science and Engineering (SCSE), Nanyang Technological University (NTU). His research interests include network security, blockchain security, and anomaly detection. He was a member of the program committees for international conferences, including SDM, CIKM, and CYBER, and has also served on the Artifact Evaluation (AE) committees for USENIX OSDI and USENIX ATC. He was a recipient of the Best Paper Award of 2019 IEEE CNS, Gurdeep Pall Graduate Student Fellowship of UO, and Ripple Research Fellowship. He has served as a reviewer for IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, ACM TKDD, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, and IEEE COMMUNICATIONS SURVEYS AND TUTORIALS.

Qingchuan Zhao (Associate Member, IEEE) received the B.E. degree from South China University of Technology in 2009, the M.S. degree from the University of Florida in 2015, and the Ph.D. degree from Ohio State University in 2021. He is an Assistant Professor with the Department of Computer Science, City University of Hong Kong. He employs both static and dynamic data flow analysis on mobile apps and delves into hardware side channels to uncover a variety of vulnerabilities, including privacy leakage, privilege escalation, and vulnerable access controls. His work has been granted bug bounties from industry-leading companies and has garnered significant media attention. His research focuses on the security and privacy practices in the Android amplified ecosystem.

Cong Wu (Member, IEEE) received the B.E. degree from Xidian University in 2017 and the Ph.D. degree from the School of Cyber Science and Engineering, Wuhan University, in 2022. He is currently a HKU-URC Post-Doctoral at the Department of Electrical and Electronic Engineering, The University of Hong Kong. His leading research outcomes have appeared in USENIX Security, ACM CCS, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, IEEE TRANSACTIONS ON MOBILE COMPUTING, and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. His research interests focus on security and privacy of distributed intelligent systems. He has honored with the Exponential Science Pioneers Award 2025, the ML4CS Best Paper Award, and the TrustCom-SPATI Best Paper Award. He is an Associate Editor of IJCS and SPY and the youth Editor of JII.

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on August 11,2025 at 02:54:24 UTC from IEEE Xplore. Restrictions apply.