Who's Wearing? Ear Canal Biometric Key Extraction for User Authentication on Wireless Earbuds

Chenpei Huang, *Student Member, IEEE,* Lingfeng Yao, *Student Member, IEEE,* Hui Zhong, *Student Member, IEEE,* Kyu In Lee, *Member, IEEE,* Lan Zhang, *Member, IEEE,* Xiaoyong Yuan, *Senior Member, IEEE,* Tomoaki Ohtsuki, *Senior Member, IEEE,* and Miao Pan, *Senior Member, IEEE*

***Abstract*—Ear canal scanning/sensing (ECS) has emerged as a novel biometric authentication method for mobile devices paired with wireless earbuds. Existing studies have demonstrated the uniqueness of ear canals by training and testing machine learning classifiers on ECS data. However, implementing practical ECS-based authentication requires preventing raw biometric data leakage and designing computationally efficient protocols suitable for resource-constrained earbuds. To address these challenges, we propose an ear canal key extraction protocol, EarID. Without relying on classifiers, EarID extracts unique binary keys directly on the earbuds during authentication. These keys further allow the use of privacy-preserving fuzzy commitment scheme that verifies the wearer's key on mobile devices. Our evaluation results demonstrate that EarID achieves a 98.7% authentication accuracy, comparable to machine learning classifiers. The mobile enrollment time (160 ms) and earbuds processing time (226 ms) are negligible in terms of wearer's experience. Moreover, our approach is robust and attack-resistant, maintaining a false acceptance rate below 1% across all adversarial scenarios. We believe the proposed EarID offers a practical and secure solution for next-generation wireless earbuds.**

***Keywords*—User Authentication, Acoustic Sensing, Usable Authentication**

I. Introduction

Biometric authentication utilizes human physiological or behavioral characteristics to recognize users, offering a secure, reliable, and user-friendly authentication system. Due to their uniqueness across individuals, biometric methods only require a user to enroll their biometric data once for later authentication [1–3]. Examples include fingerprint authentication on smartphones (TouchID [4]), voice recognition for smart home control (Alexa Voice Control [5]), and face-based online payment systems (Pay By Face [6]). Recently, innovative biometric approaches leveraging advanced wearable technologies, such as wireless earbuds, smart glasses, and virtual reality headsets, have gained significant attention [7–11]. These wearables exhibit remarkable biometric sensing capabilities and introduce new requirements concerning usability for authentication [12, 13].

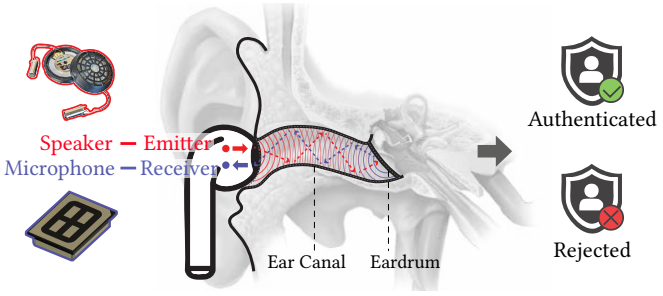

Fig. 1. Illustration of ear canal scanning/sensing. The user inserts the earbuds to run ear canal scanning, acoustic sensing the unique ear canal structure, for biometric user authentication.

Biometrics using ear canal scanning.
Offering an appealing "hands-free" feature, wireless earbuds are particularly suitable for new biometric authentication designs [14]. Positioned inside the ear canal, ear-based authentication can: (1) passively capture signals such as tooth vibrations [8], footsteps [7], and bone-conducted audio [9]; and (2) actively sense otoacoustic emissions [10], facial patterns [15] along with their dynamics [11], and structural characteristics of the ear canal or middle ear [16–21]. Consequently, wireless earbuds can function effectively as biometric scanners, authenticating users for paired mobile devices, such as unlocking smartphones. In this paper, we focus on a specific earable biometric: *ear canal scanning/sensing (ECS)* for wireless earbuds. Typically, wireless earbuds are equipped with a speaker and inward-facing microphones. When placed inside the occluded ear canal, as illustrated in Fig. 1, they can emit acoustic energy into the ear canal and capture the resulting reflections, similar to the radar system. These reflection signals are shaped by the size and structure of the user's ear canal and middle ear, producing unique patterns that can be used to distinguish different individuals.

Recently, many user authentication schemes have been proposed using ECS biometrics. For instance, Mahto *et al* [16] study the advantages of using human inaudible frequency ranges during the ear canal scanning. EarDynamic [18] operates at the inaudible frequencies, and exploits the ear canal deformation as extra discriminative features. HeadFi [19] ex-

C. Huang, L. Yao, H. Zhong and M. Pan are with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204, USA. (e-mail: chuang30@uh.edu; lyao12@uh.edu; hzhong5@uh.edu; mpan2@uh.edu).

K. Lee is with the Department of Information Science Technology, University of Houston, Houston, TX 77204, USA. (e-mail: klee48@central.uh.edu).

L. Zhang and X. Yuan are with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634, USA (e-mail: lan7@clemson.edu; xiaoyon@clemson.edu).

T. Ohtsuki is with Department of Information and Computer Science, Keio University, Kanagawa 223-8522, Japan (e-mail: ohtsuki@keio.jp).

tends the in-ear sensing capacity to low-end devices without an in-ear microphone. While these studies focus on different aspects of ear-scanning technologies, they all employ machine learning classifiers trained on ECS data to make authentication decisions.

Wireless earbuds authentication challenges.
Thanks to Bluetooth technology, wireless earbuds can be easily paired with mobile devices, enabling hands-free audio and voice interaction. With those salient features, wireless earbuds are gaining immense popularity among consumers, expecting a USD 563.2 billion market by 2030 [22]. Since wireless earbuds are becoming more powerful and versatile, one may ask: can we also develop wireless earbuds as biometric scanner for user authentication? Can we implement ECS on wireless earbuds? However, most existing studies rely on training machine learning classifiers, which presents several challenges for practical protocol design.

① Limited computation capacity. Wireless earbuds have limited computational capacity due to their compact size, power efficiency requirements, and battery constraints. For instance, the microcontroller unit (MCU) in a typical wireless earbud operates at a frequency roughly ten times slower than that of a standard laptop processor[1]. However, existing ECS classifiers typically require training after user enrollment, which involves computationally intensive iterative optimization. Although some high-end earbuds support on-device machine learning, such capabilities are mainly reserved for tasks like active noise cancellation and voice command detection using lightweight inference models, not for biometric authentication. Conversely, if a binary key effectively captures the user's ear scan, verification can be performed using efficient binary operations, such as XOR.

② Raw biometric data transmission during authentication. Raw biometric data is immutable and sensitive; if compromised, it can be reused indefinitely for unauthorized access [24]. To mitigate this risk, systems must avoid exposing such data beyond the user's control. For instance, Apple's Touch ID processes fingerprints within a hardware-isolated Secure Enclave [4, 25]. In contrast, wireless earbuds rely on Bluetooth, making them more vulnerable during authentication. While enrollment may occur in trusted settings, authentication often takes place in less secure environments, where adversaries may exploit Bluetooth attacking tools to capture and replay audio signals, such as BlueSpy [26]. This motivates our approach: to extract the ECS key solely on the earbuds and perform verification on the mobile device.

Motivation: extract on earbuds, verify on mobile.
Towards our motivation, we propose **EarID**, a key extraction protocol that encodes distinctive biometric features into a fixed-length binary key. Instead of training classifiers during enrollment, EarID identifies user-specific uniqueness by comparing the user's feature distribution against population-level statistics. The most distinctive features are binarized into keys compatible with privacy-preserving schemes such as fuzzy commitment [27] or FIDO2 [28]. During authentication, only the binary keys are transmitted over Bluetooth. Because raw biometric data is never shared, intercepted communication cannot reveal any ear canal information. Moreover, EarID avoids the overhead of model training, enabling efficient on-earbud computation and responsive authentication.

Contributions.
A. Learning-free ear canal key extraction. To address challenge ①, we propose a learning-free ECS biometric key extraction algorithm to generate unique keys for individual users. By applying denoising and physically-meaningful transformation, we leverage biometric information [29] for feature selection and utilize the random matrix method [30] for binarization. Such a process speeds up about 3× to 90× enrollment time, comparing to different ML classifiers training time on mobile. Meanwhile, it enables 160 ms processing time on earbuds MCU[2], which is negligible to user authentication experience.
B. Privacy-preserving earbuds-to-mobile authentication. To address challenge ②, we eliminate raw data and explicit key transmission during the authentication phase, when the wireless environment may be untrusted (e.g., subject to interception or eavesdropping). Specifically, the binary key enables the use of Fuzzy Commitment [27], a privacy-preserving biometric authentication scheme well-suited for secure verification between wireless earbuds and mobile devices—without exposing raw biometric data.
C. EarID testbed and evaluation. We develop a complete hardware-software testbed for ear acoustic scanning, data collection, and processing. Using real-world collected data, we evaluate EarID in terms of accuracy, efficiency, robustness, and security under various simulated and emulated conditions.

II. Background

A. Biometric Authentication System

Framework. A standard biometric authentication system consists of four main modules: a biometric data acquisition device, a biometric feature extractor, a biometric database, and an authentication decision-making module, such as a classifier or matcher. Specifically, the data acquisition step involves using a biometric sensor (e.g., a camera for face recognition, a microphone for voice authentication, or an earbud's speaker and in-ear microphone for ECS) to collect physiological or behavioral signals from the legitimate user, known as the *biometric sample*. The *biometric sample* is then extracted into the *biometric feature*, which contains discriminative information specific to the user and is resistant to unrelated factors. A *biometric template* stores the legitimate user's information in a secure database for further authentication. This template can consist of (un)transformed features, feature distributions, or machine learning models derived from the features. Finally, the decision-making is done by a classifier or a matcher that verifies whether the current attempt belongs to the claimed user.

[1]Cortex-M4 @ 160 MHz for earbuds [23] vs. 1.6 GHz laptop CPU. Other architectural factors also influence processing speed.

[2]Unless otherwise specified, the EarID performance is reported based on BCH(255,123,19).

Enrollment and Authentication. In the enrollment phase, the system captures the raw *biometric samples* from the user, extracts *biometric features*, and stores the *biometric template* in the database. In the authentication phase, the system will capture the new *biometric sample*, follow the same data processing, and make decisions based on the stored *biometric template*.

B. Ear Canal Scanning

Ear Canal Scanning Front-End. Many ECS authentication papers [17–19, 21] adopt a radar/sonar-style in-ear acoustic sensing approach. A speaker emits a wide-band waveform into the ear canal, and the in-ear microphone records the reflected signal, as shown in Fig. 1. The acoustic channel is estimated via cross-correlation:

$$h(n) = \sum_{t=0}^{\infty} s(n-t)r(t), \tag{1}$$

where $h(n)$ is the target in-ear impulse response at sample index n, $s(t)$ is the transmitting waveform at time t, and $r(t)$ is the reflected signal captured by the in-ear microphone. Nevertheless, the magnitude transfer function, which is the magnitude of the Fourier Transform of $h(n)$, is employed by most papers due to the smaller intra-user variation by removing the noisy phase information. Prior work [16–18, 21] has also explored waveform design, re-insertion effects, device response, and canal deformation.

Decision Making Back-End. Most ECS authentication studies rely on machine learning classifiers such as Support Vector Machine (SVM) [17, 19, 20], Linear Discriminant Analysis (LDA) [16], and ensemble learning methods. In contrast, template matching, a commonly used back-end in fingerprint and iris recognition, has not yet been adopted in ECS systems. Template matching operates on fixed-length feature vectors without model training, making it lightweight and compatible with privacy-preserving authentication frameworks. However, it requires compact, stable, and user-specific representation from the raw biometric samples, which has not been explored yet for the ear scan data. In this work, we, for the first time, propose the ECS key extraction protocol that can practically run on wireless earbuds.

III. EarID System Design

A. System and Threat model

System Model. The goal of the proposed *EarID* system is to enable ECS-based biometric authentication that supports a hands-free, listen-to-authenticate experience. The system is designed for execution on commercial wireless earbuds, allowing ECS biometric key extraction directly on the device. We consider a typical use case in which a user wears wireless earbuds paired with a personal mobile device, such as a smartphone. The ear canal features serve as the user's biometric trait for authentication. To enable this, we assume that the wireless earbuds are equipped with in-ear microphones and operated by an embedded microcontroller (MCU) and a digital signal processing (DSP) unit capable of audio computation [23]. A Bluetooth module allows communication between the earbuds and the paired mobile device. The mobile device, owned by the user, is equipped with a general-purpose processor and sufficient storage. It is responsible for storing relevant authentication data, including population-level feature statistics and user-specific keys. We assume this sensitive data is stored in a protected location such as the Secure Enclave [25], which cannot be accessed or modified by unauthorized parties.

To use EarID, the user must wear the earbuds firmly inserted into the ear canal and ensure the mobile phone is within Bluetooth communication range. We assume the enrollment process takes place in a secure environment where the user has physical control of both devices and a trusted wireless channel. During enrollment, the mobile device receives raw ECS data from the earbuds to generate and store the user's reference key. The mobile also provides the corresponding helper data to the earbuds. For subsequent authentication, the same earbuds and mobile device are used. The earbuds collect a fresh ECS sample, regenerate the biometric key locally, and transmit commitment data to the mobile device. The mobile verifies the user's identity based on the received data and the stored reference.

Threat Model. We consider an adversary whose goal is to impersonate a legitimate user and gain unauthorized access to the user's mobile device, knowing that the user relies on wireless earbuds for ECS-based authentication. Although wireless earbuds and mobile devices are typically personal belongings used exclusively by the owner, we also account for cases where these devices may be temporarily accessed by close individuals, such as co-workers or curious family members. We further consider more sophisticated scenarios in which the devices are stolen by the adversary.

We categorize the threat model into two main scenarios, namely with physical access to devices or without physical access. In the first case, the attacker has direct access to both the victim's mobile device and wireless earbuds. The attacker may attempt the following:

- *Passive attack.* The attacker inserts the victim's earbuds into their own ear and attempts authentication using their own ECS response.
- *Synthetic ECS attack.* The attacker inserts the victim's earbuds into a silicon ear model in order to spoof the ECS response.
- *Universal ECS attack.* A more sophisticated synthetic attack, where the attacker constructs a silicon ear model designed to emulate common ECS patterns extracted from public datasets.

In the second case, the attacker does not possess the devices but remains within Bluetooth range to launch an attack, for example using tools like BlueSpy [26]. A common strategy is:

- *Key guessing attack.* The attacker attempts to authenticate by transmitting a guessed or self-generated ECS key via a compromised or previously paired Bluetooth connection.

It is important to note that under proximity-based conditions, the adversary could also attempt biometric stealing if raw ECS data were transmitted over the Bluetooth channel. However,

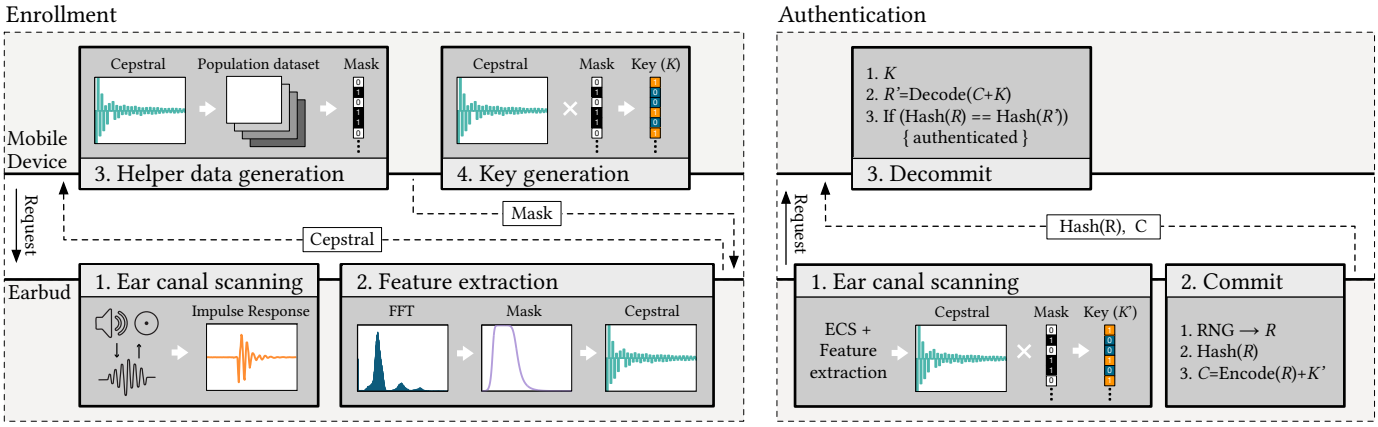

Fig. 2. EarID authentication protocol for enrollment and authentication stage. Enrollment: earbuds locally run (1) ear canal scanning (ECS), (2) cepstral feature extraction and data transmission; and then mobile devices runs (3) helper data generation (send to earbuds) for (4) key extraction (stored at mobile devices). Authentication: wireless earbuds run (1) on-device key extraction based on new ECS data, (2) commitment generation and transmission; and then mobile devices runs (3) decommitment and verification.

our design eliminates this threat by avoiding the transmission of raw biometric information altogether, thereby rendering such attacks infeasible.

B. Proposed Protocol

Overview. EarID is a biometric authentication protocol based on ear canal scanning key extraction, designed for seamless integration with commercial wireless earbuds and mobile devices. Unlike conventional approaches that rely on classifiers or transmit raw biometric data during authentication, EarID generates a user-specific binary key directly on the earbuds, enabling secure and private authentication without exposing sensitive information. The system consists of two core components: the wireless earbuds, which capture and process acoustic signals from the user's ear canal, and the mobile device, which verifies user identity based on a previously stored key. By avoiding machine learning models and minimizing computational burden, EarID delivers a lightweight, privacy-preserving, and user-friendly authentication experience.

Wireless Earbuds. The wireless earbuds serve as the sensing and processing front end of the EarID system. When worn by the user, the earbuds utilize their built-in audio unit to perform ECS by emitting acoustic energy and capturing the impulse response, as defined in Eq. (1). The captured ECS signal is then processed locally to extract relevant features. With the assistance of helper data (i.e., user-specific mask) received during enrollment, the earbuds can reliably regenerate the user-specific key during authentications.

Mobile Device. The mobile device plays two primary roles within the EarID framework. First, it has access to population-level ECS statistics and assists in user enrollment by generating and transmitting helper data specific to the user's wireless earbuds. Second, it functions as the identity verifier: upon receiving authentication data from the earbuds, the mobile device performs verification using the fuzzy commitment protocol, comparing the received commitment against the stored credential to determine authenticity.

C. Authentication Process

Enrollment Phase. The enrollment process begins when the user wears the wireless earbuds and pairs them with the mobile device. As shown in Fig. 2, the process is initiated by the mobile device. ①: The earbuds perform ear canal scanning to capture the ear canal impulse response. ②: The earbuds extract features from the scanned signal using cepstral analysis and transmit the resulting cepstral features to the mobile device. ③: The mobile device analyzes the user's ECS feature uniqueness based on population-level statistics and generates corresponding helper data, including a feature mask and binarization parameters. ④: The mobile generates and securely stores the user's key, then sends the helper data back to the earbuds for future authentications.

Authentication Phase. When the user wears earbuds to request authentication, such as unlocking the device or authorizing a transaction, the following steps occur: ①: The earbuds perform ECS and extract features, then use the stored helper data to generate a new biometric key. ②: Following the fuzzy commitment protocol, the earbuds encrypt a random secret (e.g., binary string) using the newly generated key and transmit the resulting commitment and a hash of the secret to the mobile device. ③: The mobile device receives and decodes the random secret with stored key. If the recovered hash matches the transmitted hash, the user is successfully authenticated. All steps above are illustrated in Fig. 2.

IV. Learning-Free ECS Key Extraction

A. Feature Extraction

ECS Denoising. ECS noise can originate from external environments or in-body sounds. External noise, such as car horns, birds singing, and background music, is less significant for the in-ear microphones. Since the earbuds are inserted firmly into the ear canal for better experiences, the ear tips provide significant attenuation, up to 40 dB as shown in Fig. 3. On the other hand, when the ear canal is occluded, it also amplifies

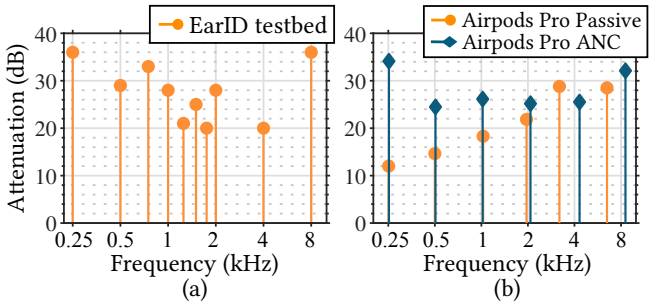

Fig. 3. Impact of passive and active noise cancellation (ANC) for (a) EarID testbed and (b) Apple Airpods Pro [31].

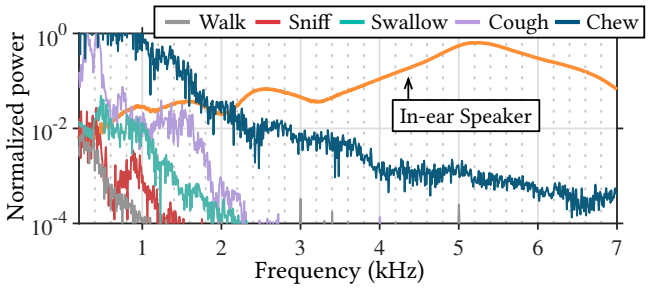

Fig. 4. Impact of occlusion effect amplified body sound, interfering acoustic sensing at low frequencies (<2 kHz)

low-frequency in-body sounds by up to 40-dB [32], including sounds from breathing, walking, and eating activities. These sounds can have power levels comparable to or even exceeding the ear canal response below 2 kHz.

In Fig. 4, we observe that noises mostly interfere with the low-frequency band (below 2 kHz) due to their impact on the occlusion effect. To mitigate this, we apply a bandpass filter to exclude the vulnerable low-frequency band below 2 kHz for user authentication applications. As a result, we set the low cut-off frequency at 2 kHz and recommend setting the high cut-off frequency within the device frequency range, e.g., 8 kHz, which offers flat frequency response. Next, we perform power normalization to ensure consistent signal strength and improve the robustness of the biometric features against variations. In addition, data aggregation by averaging multiple scanning results can mitigate the random variation caused by the dynamic deformation, especially for short waveform duration. Therefore, the normalized band-pass FFT feature is given by

$$|H^{\text{norm}}(f)| = \frac{\sum_{k=1}^{K} \left|H_k^{\text{est}}(f)F^{\text{bp}}(f)\right|}{\sqrt{\sum_{f_l}^{f_h}\sum_{k=1}^{K}\left|H_k^{\text{est}}(f)F^{\text{bp}}(f)\right|^2}}, \quad (2)$$

where $|H^{\text{norm}}(f)|$ is the normalized magnitude of the frequency response, $H_k^{\text{est}}(f)$ is the ear canal frequency response measured in k-th trial, $F^{\text{bp}}(f)$ represents the bandpass filter, K is the total number of trials, and f_l and f_h denote the lower and upper bounds of the frequency range.

Cepstral Analysis. The FFT-normalized feature in Eq. (2) can be highly correlated and redundant, i.e., inconvenient to add or remove one dimension based on its importance. We aim to compress this ECS representation and preserve the most useful information. Therefore, we apply cepstral analysis to the FFT feature vector, which reveals periodic structures that indicate echoes and reverberations in the quefrency domain [33]. This process generates a new cepstrum feature vector whose entries are uncorrelated and can be interpreted as unique multipath reflections within the ear canal. The cepstrum feature is defined as:

$$c(n) = \mathcal{F}^{-1}\left\{\log|H^{\text{norm}}(f)|\right\}, \quad (3)$$

where $c(n)$ denotes the cepstrum and $\mathcal{F}^{-1}$ is the inverse Fourier Transform with ceptrum index n. Because the FFT feature is a real-value vector, we use the inverse Discrete Cosine Transform (IDCT) instead. Each entry of the transformed feature $c(n)$ represents a unique multipath strength from the ear canal. Given the ear canal length and the speed of sound, we can safely remove cepstral components outside the reasonable time-of-flight range (e.g., 10 ms). This step is called time-of-flight liftering [33], which reduces feature dimension and removes irrelevant noises. As a result, the cross-correlation among features in a user's feature vector can be reduced, as shown in Fig. 5.

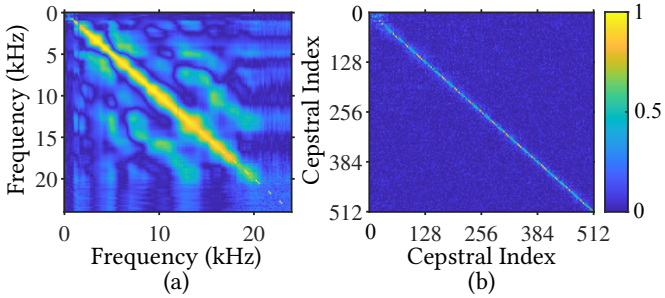

Fig. 5. Cross-correlation of (a) FFT features (strong correlation) and (b) cepstral features (independent).

B. Key Generation

Biometric Information. EarID must identify the informative multipath from a user's enrollment cepstrum features. This "information" is considered high if a feature exhibits low intra-subject variation and high inter-subject variation. Using handwriting recognition as an analogy, Alice's distinct handwriting style makes her writing easily recognizable, and her handwriting remains consistent across different instances. To this end, we utilize the *biometric information* (BI) defined by Adler et al. [29] as a metric, which is the distribution divergence between the feature distribution from the user enrollment data and that from the population. We modify the original definition using the Rényi divergence as:

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1}\log_2\left(\sum_{x\in\mathcal{X}} P(x)^\alpha Q(x)^{1-\alpha}\right), \quad (4)$$

where $P(x)$ represents the probability distribution of the user's feature, $Q(x)$ represents the probability distribution of

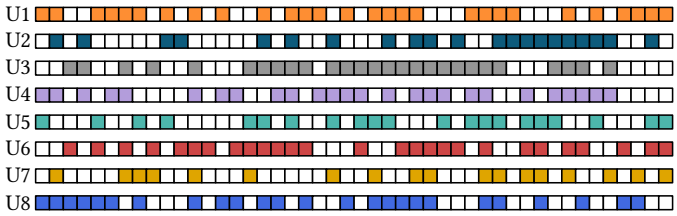

Fig. 6. The binary ECS codes transformed from frequency responses from 8 users, preserving the uniqueness among users.

the population's feature, and α is the order of the Rényi divergence. This modification allows us to focus on different aspects of distribution divergence by tuning α. The empirical study shows that when $\alpha \to 0$, the EarID system achieves the best overall performance.

Feature Masking. It is straightforward to generate a feature mask, select cepstrums with relatively high BI, and discard those with low BI that may lead to errors. However, setting a hard BI threshold for all users is not desirable. For instance, if a user has uniformly high BI cepstrums, only a subset of the highest BI features is sufficient. Conversely, for a user with generally low BI, the system should still attempt to recognize them using their most informative cepstrums. Therefore, we propose the masking approach based on each user's BI across cepstrum features by automatic thresholding using the Otsu's method [34]. The resultant user-specific binary feature mask is then applied to the cepstrum feature, where '1' means the high-BI cepstrum to retain and '0' means the low-BI cepstrum to discard.

Feature Binarization. Given the selected features with high biometric information, the final step is to generate a fixed-length binary key by transforming the feature vector from continuous space into a binary string. Since features are standardized using population-level mean and variance, i.e. $\tilde{c} = (c - \bar{c})/\text{var}(c)$, each dimension has approximately a 50% chance of being positive or negative, making sign-based binarization a natural choice. However, using only the sign discards magnitude information and limits the key length to the number of selected features.

To address this, we adopt a random projection approach, also known as BioHashing [30]. A random matrix $\mathbf{R} \in \mathbb{R}^{L\times d}$ with entries drawn from a standard normal distribution is multiplied with the standardized feature vector $\mathbf{x} \in \mathbb{R}^{d\times 1}$, producing a projected vector $\mathbf{y} = \mathbf{R} \cdot \mathbf{x}$. The final binary key $\mathbf{K} \in \mathbb{B}^{L\times 1}$ is then obtained by applying a sign function to each element:

$$\mathbf{K}[i] = \begin{cases} 1, & \text{if } \mathbf{y}[i] \geq 0, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

and the visualization of the derived key is shown in Fig. 6. In the figure, eight different users all have their unique key extracted from the ECS data, where similar number of colored boxes (bit '1') and white boxes (bit '0') can be observed.

V. Earbuds-to-Mobile Authentication

Since the biometric key representing the legitimate user's identity has already been generated, we now describe the authentication protocol based on these keys. As introduced in previous sections, the mobile device securely stores the enrolled key $\mathbf{K}^{\text{enroll}}$, while the wireless earbuds are enabled to generate a fresh key $\mathbf{K}^{\text{auth}}$ from ECS during each authentication attempt. While this design avoids transmitting raw biometric data over potentially insecure channels, the challenge is shifted to: how can the mobile device verify that the key generated on the earbuds matches the enrolled one on the mobile side?

To address this, we adopt the fuzzy commitment scheme, which allows verifying key similarity from two devices (earbuds and mobile). During each authentication, the earbuds use a random number generator (RNG) to produce a random (binary) secret $\mathbf{R} \in \mathbb{B}^{T\times 1}$, where $T < L$. This random sequence provides $\log_2(T)$ bits of security. The hash of random sequence $\text{Hash}(\mathbf{R})$ is transmitted to the mobile device for verification. Besides, wireless earbuds use an error-correcting code (ECC), such as BCH or Reed-Solomon, to encode $\mathbf{R}$ into a codeword of length L, denoted $\text{Enc}(\mathbf{R})$. The encoded sequence is then combined with newly generated key:

$$\mathbf{C} = \text{Enc}(\mathbf{R}) \oplus \mathbf{K}^{\text{auth}},$$

where $\mathbf{C}$ is the commitment transmitted to the mobile device. The ECC enables correction of up to t bit errors, compensating for noise and variability in ECS-based key extraction.

Upon receiving the commitment $\mathbf{C}$ and hash $\text{Hash}(\mathbf{R})$, the mobile device uses the stored key $\mathbf{K}^{\text{enroll}}$ to attempt recovery of the original message:

$$\mathbf{R}' = \text{Dec}(\mathbf{C} \oplus \mathbf{K}^{\text{enroll}}),$$

where $\text{Dec}(\cdot)$ denotes ECC decoding. The mobile then computes $\text{Hash}(\mathbf{R}')$ and compares it to the received $\text{Hash}(\mathbf{R})$. If the two match, it implies that the key generated during authentication is close to the enrolled key within the ECC correction capability, i.e.,

$$\text{dist}(\mathbf{K}^{\text{auth}}, \mathbf{K}^{\text{enroll}}) \leq t.$$

The primary benefit of adopting the fuzzy commitment scheme is that it eliminates the need to transmit raw biometric data, thereby mitigating the risk of biometric theft during communication. Instead of directly revealing the biometric key or ear canal impulse response, the scheme verifies the user's identity through indirect comparison. The choice of ECC in the commitment process reflects a trade-off between security and tolerance to intra-user variability; specifically, it determines the number of protected bits (entropy) and the system's ability to correct bit-level errors. In this work, we adopt BCH codes for the EarID fuzzy commitment system. For example, a BCH(255,123,19) code implies a key and codeword length of $L = 255$, provides $T = 123$ bits of protection, and corrects up to $t = 19$ bits of error.

VI. Evaluation

In the following section, we will evaluate the proposed system and use the acronyms explained in Table I.

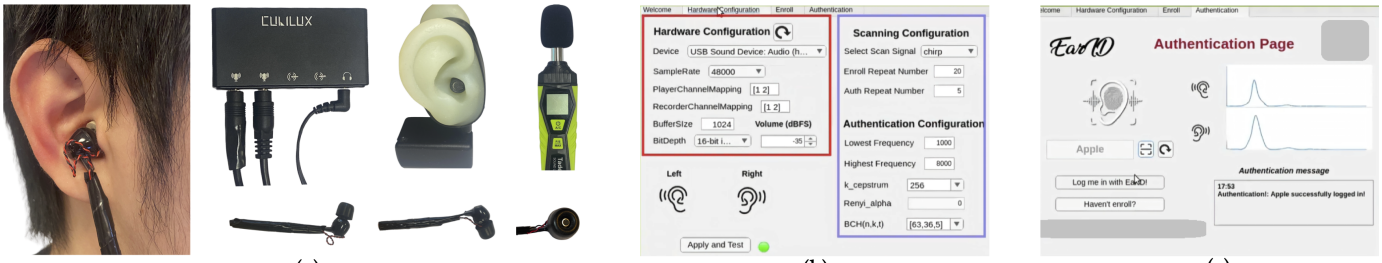

Fig. 7. (a) EarID hardware platform (b) EarID software configuration page and (c) EarID authentication page (dummy user)

TABLE I. LIST OF ACRONYMS

Acronym	Description	Acronym	Description
ECS	Ear Canal Scanning	SVM	Support Vector Machine
BER	Bit Error Rate	MLP	Multi-Layer Perceptron
FRR	False Rejection Rate	RF	Random Forest
FAR	False Acceptance Rate	ASR	Attack Success Rate
EER	Equal Error Rate	RNG	Random Number Generator
BAC	Balanced Accuracy	ECC	Error-Correcting Code
BCH	Bose–Chaudhuri–Hocquenghem Code		

A. Data Collection

EarID Hardware Platform. While commercial wireless earbuds do not currently provide access to in-ear microphone APIs, we developed a custom hardware setup to emulate their functionality. Specifically, we embedded mini microphones (CMC-3015-44L100 [35]) into the foam tips of off-the-shelf (COTS) earbuds (Sony MDREX15AP [36]). These microphones, positioned toward the ear canal, worked in tandem with the earbuds' built-in speakers to perform ear canal scanning, functionally replicating the ECS capability of wireless earbuds. Both the speaker and microphone were connected to a full-duplex USB sound card, which served as a proxy for the DSP typically found in wireless earbuds. The sound card was interfaced with a laptop for data acquisition and analysis. The complete hardware platform is illustrated in Fig. 7(a). In addition, we utilized a silicon ear model to simulate the *Synthetic ECS attack* scenario, and a Sound Pressure Level (SPL) meter was employed to monitor ambient noise levels during experiments.

TABLE II. ACOUSTIC AND EARID SYSTEM PARAMETERS

Acoustic Parameter	Value	Acoustic Parameter	Value
Volume	User-adjustable	Dataset Samples	44×40=1760
Sampling Frequency	48 kHz	Scans/Enroll in Wild	8
MLS Duration	1 second	Scans/Auth in Wild	2
Chirp Frequency	20 Hz - 20 kHz	Quiet Room SPL	30-40 dBA
Chirp Duration	1 second	Indoor/Outdoor Noise	∼60/70 dBA

EarID Software Platform. To facilitate the testing process, we developed a software platform for data collection and evaluation. The configuration page, shown in Fig. 7(b), allowed the developer to adjust system parameters, including hardware configuration (in red box) and authentication configuration (in purple box). During enrollment, users simply registered their identifier (non-sensitive last 4 digit of their phone number) and captured their ECS signal for key extraction. The extracted key was then stored. For authentication, as depicted in Fig. 7(c), users typed in their identifier and performed quick ECS scans. The system provided real-time visual feedback to help users monitor signal quality and displayed final authentication results on the interface. With the user consent, we collected the enrollment and authentication data for algorithm development.

IRB-Approved Subject Recruitment. This experiment, approved by IRB protocols, involved recruiting 44 participants from a university campus, ensuring diversity in gender, race, age, and roles (students, faculty, and staff). Our user pool surpasses the median size (26) of similar studies [16–21] and is second only to the largest pool (52) in [20]. For cross-day evaluation, due to scheduling challenges, data was collected from 8 participants.

Ear Scanning and Recording. Human subject data collection for ear canal scanning was conducted in a quiet indoor environment with ambient noise levels ranging from 30 to 40 dBA. Participants were instructed to insert the earbuds until their ears were fully occluded, remain silent, and behave naturally throughout the recording process using our platform in Fig. 7. Each data collection session consisted of 20 repetitions of ECS, and each user completed two sessions using both Maximum-Length Sequence (MLS) and chirp signals. The specific parameters for these excitation signals are summarized in Table II.

Robustness Experiments. For robustness evaluations, only chirp signals were used in the false triggering, cross-day and cross-session tests. To evaluate false-triggering, which are scenarios referred to as the "not-in-use" condition, data were collected while the earbuds were held in hand, placed on various surfaces, or covered by cloth. Cross-day testing involved five authentication attempts per day over five consecutive days, while cross-session testing consisted of one enrollment session followed by nine authentication sessions conducted within a short time window. A silicon ear model was also employed as a baseline for comparison with human users and to simulate passive attacks in the cross-session evaluation.

Evaluation Parameters. To evaluate the performance of

TABLE III. PERFORMANCES OF EARID AUTHENTICATION

EarID w. ECC	Accuracy				Robustness	Security			
Metrics (%)	EER ↓	FAR ↓	FRR ↓	BAC ↑	FT-Acc. Rate ↓	P-ASR ↓	S-ASR ↓	U-ASR ↓	K-ASR ↓
BCH(127,64,10)	1.2	1.0	2.2	98.4	0.1	1.0	0.8	0.2	0.0
BCH(255,123,19)	0.7	0.6	2.0	98.7	0.0	0.6	0.8	0.1	0.0
BCH(511,241,30)	1.0	1.0	2.3	98.4	0.0	1.0	0.7	0.0	0.0

EarID, we focus on authentication accuracy, efficiency, robustness against unintended false triggering, and security against attacks. The default setup of the experiment is as follows. The number of FFT points is set to 1024, and the number of cepstrum coefficients after liftering is 256. The bandpass denoising filter is designed to allow frequencies between 2 kHz and 8 kHz. The order of the Rényi divergence α is 0.

B. EarID Accuracy Evaluation

Bit Error Interpretation. To evaluate the performance of key extraction, we first directly compare the enrolled key stored on the mobile device with fresh keys generated during authentication attempts to gain insight. A low bit error rate (BER) is expected for legitimate authentication, indicating strong key consistency for the same user. In contrast, a high BER is desirable when comparing keys from different users (i.e., passive attackers), reflecting good user discrimination. We evaluate three key lengths L: 127, 255, and 511. Results are shown in "Passive attack" in Fig. 8, where the gray and red lines represent BERs for legitimate users and passive attackers, respectively. In the figure, above 97% percentage legitimate user only have approximately 5% BER. When BER increases to more than 10%, the percentages of the legitimate user gradually drops to 0. In contrast, red line shows that the percentage of attackers is extremely low at low BER region and gradually increases above 15% BER for passive attack (human intruders). Since line plots in Fig. 8 reflect the changes of false rejection rate (FRR) and false acceptance rate (FAR), their intersection (red circle) refers to the optimal threshold, achieving equal error rate (EER).

TABLE IV. ACCURACY OF ML CLASSIFIER BASELINES

ML Classifiers	Accuracy			
Metrics (%)	EER ↓	FAR ↓	FRR ↓	BAC ↑
SVM	0.9	1.3	0.7	99.0
Random Forest	0.7	2.9	0.5	98.3
MLP	1.0	1.2	0.9	99.0
Adaboost	2.0	1.7	2.3	98.0

Authentication Accuracy and ECC. In practice, the EarID system does not have the bit error analysis data to make decision. Instead, the decision is made by a selected ECC. If the bit error goes beyond certain error correction capacity, the decommitment results will never match (see Fig. 2), leading to reject the attempt. Therefore, the ECC should be selected to provide sufficient key length as well as a suitable error correction ability near the exact EER point. For example, with a key length $L = 127$, the corresponding EER corresponds to tolerating 8.9% BER (about 11 bits). Using BCH(127,64,10), which can correct 10 bit errors, it will achieve a FRR of 2.2% and FAR of 1.0%, close to the ideal EER. Similarly, performance with BCH(255,123,19) and BCH(511,241,30) is reported in Table. III. We also report the Balanced Accuracy (BAC) defined as $\text{BAC} = (2 - \text{FAR} - \text{FRR})/2$ in the table.

Classifier Baseline Configurations. We evaluate EarID against four baseline machine learning classifiers: Random Forest (RF), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and AdaBoost. All models are implemented using scikit-learn [37] and tuned to achieve optimal performance. The RF model uses 50 decision tree estimators with a maximum depth of 10. The SVM adopts a linear kernel with a regularization parameter $C = 50$. The MLP consists of two hidden layers with 16 ReLU units each and is trained using the Adam optimizer for up to 1000 iterations. AdaBoost integrates 10 weak decision trees, each with a maximum depth of 10.

Comparison with Classifiers. To train the classifiers, we follow the protocol described in [17]. The target user's enrollment data is combined with data from 30 randomly selected users (also referred to as built-in users) to train the model. An additional 10 users are designated as passive attackers. For evaluation, we mix the target user's authentication data with that of the passive attackers to assess the system's ability to correctly accept the legitimate user while rejecting unauthorized attempts. As shown in Table IV, all baselines achieve above 98.0% BAC on our data. For comparison, EarID achieves a comparable BAC greater than 98.4% across three key lengths, which is smaller only than that derived from MLP (99%).

C. Efficiency

Evaluation Platform. We evaluate the efficiency of EarID by emulating the authentication algorithm on two platforms: a mobile device and the Arduino Due, which serves as a proxy for the computational capacity of wireless earbuds. For the mobile device, we use a Google Pixel 5a smartphone equipped with a Snapdragon 765G processor, featuring 8 CPU cores running at 1.8 GHz and operating on Android 14. All programs, including EarID and baseline classifiers, are executed via Termux Android terminal emulator [38]. For the earbud-equivalent platform, we use an Arduino Due featuring an 80 MHz AT91SAM3X8E chip, due to the limited availability of open earable computing platforms.

Enrollment Time on Mobile. To emulate EarID's enrollment process, one user's enrollment data and population-level feature distribution are loaded onto the smartphone. When using all available cores, the computation time for EarID with key length 127/255/511 is approximately 155/160/178 ms. When

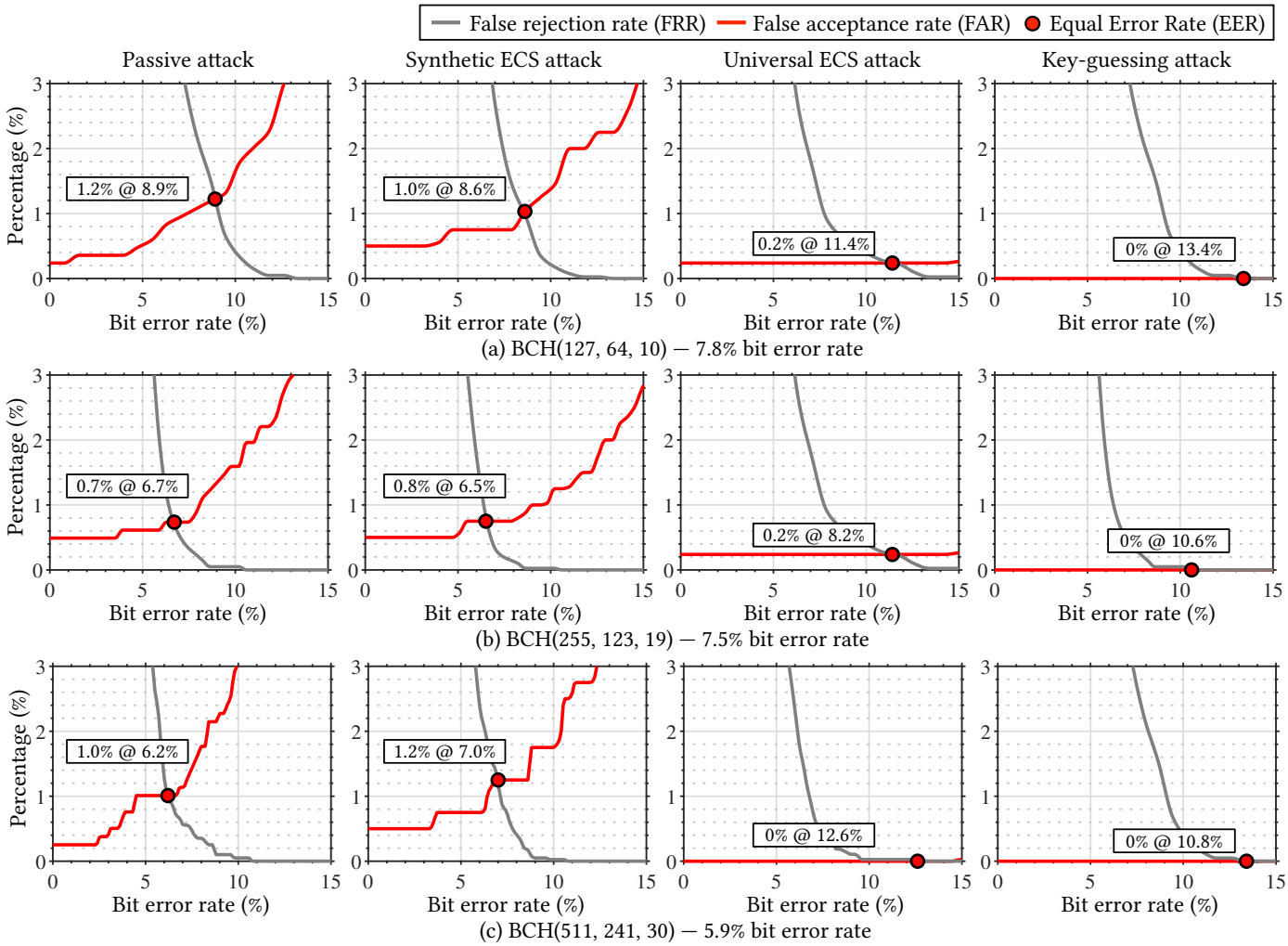

Fig. 8. The false rejection rate (FRR, in gray) and false acceptance rate (FAR, in red) plots against the bit error rate (BER) between enrolled and authentication keys. The intersection of FRR and FAR is the point achieving equal error rate (EER), where the corresponding BER is the ideal error correction threshold. The actual error correction codes (ECC) are (a) BCH(127, 64, 10), (b) BCH(255, 123, 19), and (c) BCH(511, 241, 30).

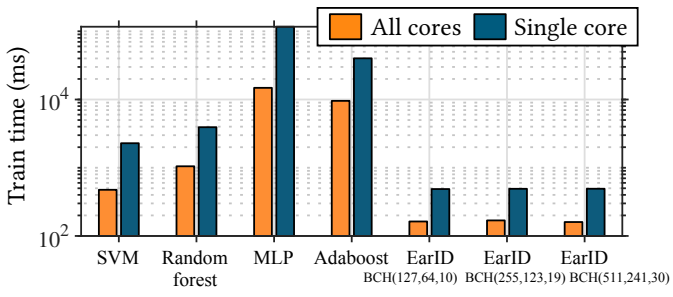

Fig. 9. Baseline classifier training vs. EarID mobile enrollment time.

restricted to a single core, this time increases to 491 ms. With the same device and training strategy stated above, machine learning classifiers generally require longer training time than EarID, as shown in Fig. 9. With all cores, they range from 475 ms (SVM) to 14,824 ms (MLP), about 3x to 90x slower than EarID. The performance gap widens under single-core conditions, where classifiers need between 2,286 ms and 116,579 ms, i.e. up to 4x and 230x slower. This strong advantage in computational efficiency makes EarID suitable for personal on-device applications.

Key Extraction Time on Earbuds. We then measure our key extraction time on the Arduino Due, representing the comparable (or even lower) computing capability of commercial earbuds. Since only feature extraction and binarization are required at the earbuds during authentication, our emulations report the required time for these steps. For BCH(127,64,10), 145 ms feature extraction time and 40 ms binarization result in a total 185 ms execution time. By extending key length to 255/511, it follows the same feature extraction but exhibit more bit extraction. Therefore, with the same feature extraction time (145 ms), BCH(255,123,19) and BCH(511,241,30) have longer key extraction time of 226 ms and 322 ms, respectively.

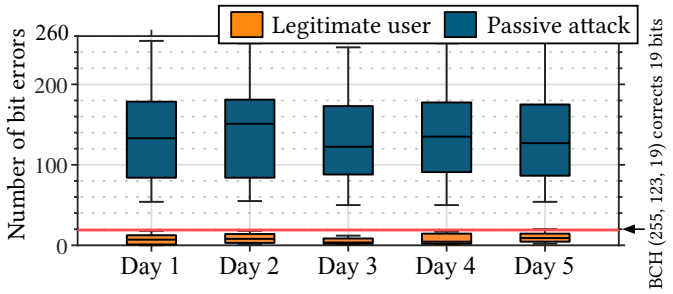

Fig. 10. Cross-day evaluation: bit error of legitimate user vs. passive attacker vs. error correction from BCH(255, 131, 19).

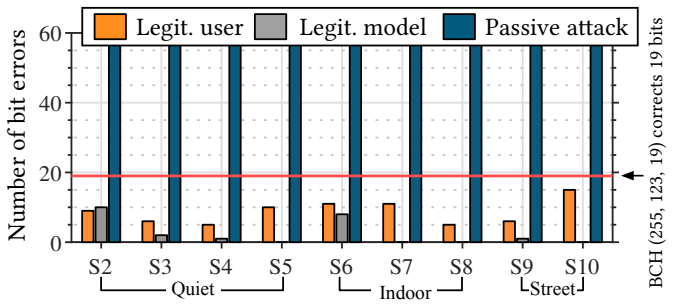

Fig. 11. Cross-session evaluation: bit error of legitimate user/ear model vs. passive attacker vs. error correction from BCH(255, 131, 19).

Since the wireless module on earbuds have the coding and encryption capability for Bluetooth communication, no extra time is introduced to leverage this module for commitment.

D. Robustness

Robustness of our proposed EarID is evaluated through: resistance to unintended falsely triggered attempts (FT), cross-day and cross-session performances. The corresponding metrics include the falsely triggered acceptance rate, bit errors under cross-day and cross-session evaluations. We do not conduct extra experiments on in-body noise because the in-body noise can be picked up during the ECS process and our system is designed to be immune to it.

Resistance to Unintended FT Attempts. False triggering occurs when the earbuds are not intentionally worn by the legitimate user. Common conditions include being placed on a table, held in hand, or covered by cloth. Ideally, the authentication system should reject these attempts. To evaluate this, we deliberately collected the FT samples (40) and mixed them with the legitimate user attempts. The results, shown in Table III, indicate that all EarID configurations can effectively reject FT attempts.

Cross-Day Evaluation. To evaluate long-term stability, we asked participants to wear the test earbuds, performing five authentication attempts per day for five successive days. Here we analyze the range of bit errors to evaluate system margin. As shown in Fig. 10, using a moderate BCH(255,123,19), legitimate users typically exhibit 0–20 bit errors, with only 2 of 40 trials exceeding the correction threshold by 1 bit on Day 5. In contrast, passive attack samples show significantly higher errors, ranging from 50 to 255 bits. While the slight increase in error over time may suggest subtle aging effects, the large margin between legitimate users and attackers enables dynamic thresholding to maintain 100% accuracy, e.g. with a 20-bit correction. Moreover, the lightweight design allows quick re-enrollment if the ECS key expires.

Cross-Session Evaluation. For cross-session stability within short time frames, participants first enrolled themselves in the initial session (S1). Subsequently, participants repeatedly re-wore the earbuds for 9 times (S2-S9) and tested EarID each time while simulating noise conditions using a speaker and SPL meter in 20 minutes. Additionally, silicon ear model data were collected for reference in the same environment. An illustration of bit error performance in the cross-session evaluation for a human user and an ear model is shown in Fig 11. For cross-session stability within short time frames, participants first enrolled themselves in the initial session (S1). They then repeatedly re-wore the earbuds in 9 follow-up sessions (S2-S9), and tested EarID under different noise conditions displayed by a smartphone: Quiet, Indoor, and Street, refer to Table II. For reference, silicon ear model data were also collected under the same conditions. As shown in Fig 11, the cross-session evaluation demonstrates EarID's robust performance across multiple insertions. Both legitimate users and silicon ear models consistently show bit errors below the 19-bit correction threshold (red line) of BCH(255,123,19). In contrast, passive attacks exhibit significant higher bit errors across all sessions, making them easily to detect by EarID.

E. Security

We assess EarID's security by evaluating its resistance to various attackers outlined in the threat model. The metric used for this evaluation is the attack success rate (ASR) for each attack, which indicates the proportion of successful bypass attempts by a specific type of attacker, granting them access as the legitimate user.

①: Passive Attack. Passive attack samples are selected from other user's data other than the legitimate users. Specifically, in each independent test, we randomly designate 30 participants in the dataset as the gallery dataset, providing population-level statistics and feature probability distribution. Meanwhile. the remaining 14 users enroll with their own ECS data. These enrolled users attempt (a) legitimate authentication on their own account and (b) passive attacks by attempting to access another user's account. The result is aggregated in 20 independent trails. Table III shows that the EarID across all three BCH settings successfully defends against such intrusions, with passive attacker ASR (P-ASR) consistently below 1%.

②: Synthetic ECS Attack. To simulate synthetic ECS attacks, we wear the wireless earbuds on a silicon ear model (Fig. 7(a)) to collect 40 synthetic samples (S-attacks). In each independent test, we randomly re-group the user similar to passive attack, and use S-attack samples to to attempt authentication into enrolled accounts. The resulting S-ASR remains low and

TABLE V. COMPARISON OF RELATED EARABLE/HEAD-WEARABLE AUTHENTICATION DESIGNS.

Study	Purpose	Type	Auth Model	Subjects	Accuracy (Conditions)
EarGate [7]	Acoustic footstep	Phy	SVM	31	97.3%
ToothSonic [8]	Tooth gesture	Beh	Neural Net	25	95%
Voice-in-Ear [9]	In-ear voice	Phy	GMM-UBM + AdaBoost	23	96.3%
Earprint [10]	TEOAE	Phy	LDA + Pearson Dist.	54	99.4%
F2Key [11]	Acoustic mouth	CR	cGAN + Siamese NN	26	95.3%∼99.9%
SonicID [15]	Ultrasound face	Phy	ResNet	40	97.44% (x-ses / days)
Inaudible ECS [16]	Hybrid ECS	Phy	LDA	25	98.6% (A-x-ses), 99.9% (I-cont.)
EarDynamic [18]	Dynamic ECS	Phy	Classifier Boosting	24	93%
HeadFi [19]	ECS device	Phy	SVM	27	95%
EarEcho [17]	ECS classification	Phy	DT/SVM/MLP/etc	20	95.1% (x-ses), 97.6% (cont.)
Bilateral ECS [20]	Bilateral ECS	Phy	SVM	52	99.6%
EarID (ours)	ECS over wireless earbuds	Phy	Key extraction	44	98.7% (A-cont / x-ses / x-day)

Phy: physiological; Beh: behavioral; CR: challenge response;
TEOAE: Transient-Evoked Otoacoustic Emissions;
x-ses/days: cross session and days; cont.: continuous authentication.
A: audible band; I: inaudible band.

comparable to P-ASR, indicating that the system detects the ear model as an unauthorized source.

③: Universal ECS Attack. Universal ECS attacks target users with ECS features that are common or non-distinct, when attacker can exploit a public gallery dataset to construct generic ECS patterns. The attacker selects the most common ECS feature from the public dataset and uses it to attempt access. While this approach mimics human data more closely than synthetic ear models, it still fails against EarID. This is because EarID explicitly selects high-biometric-information features (Eq. 4) during key extraction, while discarding common features. As a result, features used in universal attacks cannot match the legitimate user's feature mask, and U-ASR remains near 0% shown in Fig. 8.

④: Key Guessing Attack. Instead of random guess the extracted binary key, attackers may user their own EarID for key guessing, expecting higher success rate if EarID cannot emphasize each user's uniqueness and generate low-entropy keys. However, this strategy proves ineffective, which leads to a 0% K-ASR and a FAR curve that hugs the x-axis in Fig. 8, confirming that such key guessing attempts fail completely.

VII. USER HEARING EXPERIENCE

To incorporate human auditory perception into our evaluation, we employed Zwicker's psychoacoustic annoyance (PA) model [39] to quantify the subjective listening experience of the transmitted ECS signal. Using the SQAT toolbox [40][3], we obtained PA values of 14.8 for the chirp signal and 10.7 for the MLS signal (parameters detailed in Table II). For context, 20 commonly used "notification" and "alarm" sound effects sourced online [41] exhibited an average PA value of 15.4. Additionally, a Mean Opinion Score (MOS) assessment was conducted to capture participants' subjective experiences of the ECS sounds. Participants rated the sounds on a 5-point scale (1: Very Annoying, 5: Very Pleasant). The average MOS scores for the chirp and MLS signals were 3.2 and 2.2, respectively. Although the MOS results contrast with the psychoacoustic annoyance (PA) values, they indicate that the ECS signals effectively notify the user while maintaining an auditory experience within an acceptable range. It is worth noting that ECS can operate in the inaudible frequency band [16]. However, this approach has been reported to be suitable primarily for continuous authentication, which falls outside the scope of this study, as our focus is on improving efficiency.

[3]The minimum audio duration required by the toolbox is 2 seconds. This criterion is met as both the 1s-MLS and 1s-chirp signals are played twice (totaling 2 seconds) for data aggregation.

VIII. DISCUSSION

Impact of Bandpass Filtering. As discussed in Section V, the bandpass filter plays a crucial role in mitigating additive and convolutional noise. We evaluated the impact of different passband parameters on EarID-BCH(255,123,19) by testing (0 Hz - 8 kHz), (2 kHz - 4 kHz), (2 kHz - 8 kHz), (2 kHz - 16 kHz), and (8 kHz - 16 kHz). The corresponding BAC results were 68.5%, 97.3%, 98.7%, 97.9%, and 96.3%. These results indicate that low frequencies below 2 kHz adversely affect system performance. Also, large bandwidth does not always enhance accuracy in EarID system.

Impact of Biometric-Informative Masking. The purpose of biometric-informative masking is to filter out cepstral features that lack uniqueness or exhibit high intra-subject variability. To study its impact, we excluded this module and generate ECS code for matching. Without biometric-informative masking, the FAR dropped to 0%, but the FRR soared to 86.9%. This demonstrates that the proposed feature selection module effectively identifies useful cepstral features, significantly improving EarID system performance.

Selection of Error Correction Code. We tested only BCH as the error correction code (ECC) for EarID, while other options including Reed-Solomon (RS) [42] and Low-Density Parity-Check (LDPC) [43] codes are more complex to perform. This selection can be optimized in future work, as EarID relies on ECC to tolerate random bit errors caused by variations

in the user's ear canal and the way earbuds are worn. It is noteworthy that the *flexible* (yet sufficient) error-correction capacity, rather than a high error-correction capacity, benefits EarID. Additionally, complexity is an important consideration when implementing ECC on earable devices. Nevertheless, ECC is fundamental for wireless systems and might be utilized without incurring extra costs.

System Update. During the warm-up stage of ECS-based biometric authentication, it is essential to collect a sufficiently annotated gallery dataset of biometric features to develop an accurate and robust system. Continuous data collection from an increasing number of consenting users facilitates system updates. When an update is necessary, all users opt to re-enroll based on the new gallery dataset to enhance security. In this context, EarID's learning-free and low-complexity design is advantageous for system updates and can cease once the population distribution converges.

Potential Extensions. The proposed EarID system can be considered a promising biometric authentication solution for personal mobile device paired with earables. However, the fundamental challenge of the new biometric modality studies is still the lack of a public ECS dataset. To address this issue, we will release details of our hardware and software platforms to support ECS data collection for new research. Additionally, our proposed EarID can also be improved in the future. For example, using index-based IoM-hashing [44] can potentially be more efficient compared with the current random projection-based BioHashing. Lastly, the new ECS representations proposed in this work can effectively capture in-ear sensing information, potentially extending their application to other human in-ear sensing research, for example, detecting the ear canal deformation similar to EarDynamic [18] with EarID's binary representation.

IX. Related Work

Wearable Authentication Methods. Listed in Table. V, several related works achieve high authentication accuracy using non-ECS (Ear Canal Scanning) acoustic sensing modalities. For example, EarGate [7], ToothSonic [8], and Voice-in-Ear [9] authenticate users by passively recording in-body sounds, while Earprint [32], F2Key [11], and SonicID [15] actively emit acoustic signals to drive authentication. Notably, SonicID stands out for its evaluation across sessions and days, demonstrating long-term consistency.

ECS-Based Authentication Methods. Among ECS-based authentication designs, different approaches emphasize specific aspects, such as inaudible ECS [16], dynamic deformation [18], device modification [19], classifier ensemble [17], and bilateral ear [20]. These works commonly use classifiers to perform the authentication task and are reproduced using our data, as shown in Table. V.

Our Contribution. Comparing with existing authentication method, our unique contributions include:

- Practical biometric authentication by fully extracting key on earbuds and verifying at mobile.
- No raw ECS data transmission to mobile device to mitigate biometric stealing over insecure channels.
- Efficiency in enrollment and ability to authenticate on resource-limited wireless earbuds.

X. Conclusion

In conclusion, we present EarID, a learning-free biometric authentication system that leverages ear canal scanning (ECS) key extraction. Our target is to implement the ECS authentication directly on wireless earbuds, in order to avoid raw biometric data transmission over insecure wireless environment during authentication. Unlike existing methods that rely on ML classifiers, we design the lightweight learning-free key extraction approach that can execute on wireless earbuds, condensing the user-specific unique features into binary keys. Furthermore, the derived key is integrated into the fuzzy commitment scheme for secure earbuds-to-mobile authentication. Our extensive evaluations demonstrate that EarID achieves high authentication accuracy up to 98.7%, which is comparable with classifiers but is faster in mobile enrollment (160 ms) and enables earbuds key extraction (226 ms). EarID also shows strong resilience against false triggering and multiple threat models, with all false acceptance rate lower than 1%. These results suggest that EarID is a practical and secure solution for seamless biometric authentication on wireless earbuds.

REFERENCES

[1] Davide Maltoni, Dario Maio, Anil K Jain, Salil Prabhakar, et al. *Handbook of fingerprint recognition*, volume 2. Springer, 2009.

[2] Frédéric Bimbot, Jean-François Bonastre, Corinne Fredouille, Guillaume Gravier, Ivan Magrin-Chagnolleau, Sylvain Meignier, Teva Merlin, Javier Ortega-García, Dijana Petrovska-Delacrétaz, and Douglas A Reynolds. A tutorial on text-independent speaker verification. *EURASIP Journal on Advances in Signal Processing*, 2004:1–22, 2004.

[3] Wenyi Zhao, Rama Chellappa, P Jonathon Phillips, and Azriel Rosenfeld. Face recognition: A literature survey. *ACM computing surveys (CSUR)*, 35(4):399–458, 2003.

[4] Apple Inc. About touch id advanced security technology. https://support.apple.com/en-us/105095, 2023. Accessed: 2024.

[5] Amazon. Alexa smart home. https://www.amazon.com/alexa-smart-home/, 2024. Accessed: 2024.

[6] PayByFace. Paybyface – face recognition payment system. https://paybyface.io, 2024. Accessed: 2024.

[7] Andrea Ferlini, Dong Ma, Robert Harle, and Cecilia Mascolo. Eargate: gait-based user identification with in-ear microphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, pages 337–349, 2021.

[8] Zi Wang, Yili Ren, Yingying Chen, and Jie Yang. Toothsonic: Earable authentication via acoustic toothprint. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(2), July 2022.

[9] Yang Gao, Yincheng Jin, Jagmohan Chauhan, Seokmin Choi, Jiyang Li, and Zhanpeng Jin. Voice in ear: Spoofing-resistant and passphrase-independent body sound authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1):1–25, 2021.

[10] Yuxi Liu and Dimitrios Hatzinakos. Earprint: Transient evoked otoacoustic emission for biometrics. *IEEE Transactions on Information Forensics and Security*, 9(12):2291–2301, 2014.

[11] Di Duan, Zehua Sun, Tao Ni, Shuaicheng Li, Xiaohua Jia, Weitao Xu, and Tianxing Li. F2key: Dynamically converting your face into a private key based on cots headphones for reliable voice interaction. In *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services*, MOBISYS '24, page 127–140, New York, NY, USA, 2024. Association for Computing Machinery.

[12] Pınar Kürtünlüoğlu, Beste Akdik, and Enis Karaarslan. Security of virtual reality authentication methods in metaverse: An overview. *arXiv preprint arXiv:2209.06447*, 2022.

[13] Hui Zhong, Chenpei Huang, Xinyue Zhang, and Miao Pan. Metaverse can: Embracing continuous, active, and non-intrusive biometric authentication. *IEEE Network*, 2023.

[14] Sophie Stephenson, Bijeeta Pal, Stephen Fan, Earlence Fernandes, Yuhang Zhao, and Rahul Chatterjee. Sok: Authentication in augmented and virtual reality. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 267–284. IEEE, 2022.

[15] Ke Li, Devansh Agarwal, Ruidong Zhang, Vipin Gunda, Tianjun Mo, Saif Mahmud, Boao Chen, François Guimbretière, and Cheng Zhang. Sonicid: User identification on smart glasses with acoustic sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(4):1–27, 2024.

[16] Shivangi Mahto, Takayuki Arakawa, and Takafumi Koshinak. Ear acoustic biometrics using inaudible signals and its application to continuous user authentication. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 1407–1411. IEEE, 2018.

[17] Yang Gao, Wei Wang, Vir V Phoha, Wei Sun, and Zhanpeng Jin. Earecho: Using ear canal echo for wearable authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–24, 2019.

[18] Zi Wang, Sheng Tan, Linghan Zhang, Yili Ren, Zhi Wang, and Jie Yang. 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):1–27, 2021.

[19] Xiaoran Fan, Longfei Shangguan, Siddharth Rupavatharam, Yanyong Zhang, Jie Xiong, Yunfei Ma, and Richard Howard. Headfi: bringing intelligence to all headphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, pages 147–159, 2021.

[20] Masaki Yasuhara, Isao Nambu, and Shohei Yano. Bilateral ear acoustic authentication: A biometric authentication system using both ears and a special earphone. *Applied Sciences*, 12(6):3167, 2022.

[21] Yetong Cao, Chao Cai, Anbo Yu, Fan Li, and Jun Luo. Earace: Empowering versatile acoustic sensing via earable active noise cancellation platform. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(2):1–23, 2023.

[22] Yahoo Finance. Global true wireless stereo earbuds market, 2022-2030: Health meets entertainment - the fusion of hearing aid functionalities in tws earbuds, 2023. Accessed: 2025-05-26.

[23] NXP Semiconductors. True wireless earbuds - development boards and designs, 2023. Accessed: 2025-05-26.

[24] Apple Inc. Biometric security, 2024. Accessed: 2025-05-26.

[25] Apple Inc. Secure enclave, 2024. Accessed: 2025-05-26.

[26] Tarlogic. Bluespy, 2023. Accessed: 2025-05-26.

[27] Ari Juels and Martin Wattenberg. A fuzzy commitment scheme. In *Proceedings of the 6th ACM conference on Computer and communications security*, pages 28–36, 1999.

[28] FIDO Alliance. FIDO2. https://fidoalliance.org/fido2/, 2025. Accessed: May 19, 2025.

[29] Andy Adler, Richard Youmaran, and Sergey Loyka. To-

wards a measure of biometric information. In *2006 Canadian conference on electrical and computer engineering*, pages 210–213. IEEE, 2006.

[30] Andrew Teoh Beng Jin, David Ngo Chek Ling, and Alwyn Goh. Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern recognition*, 37(11):2245–2255, 2004.

[31] Nicky Chong-White, Jorge Mejia, Brent Edwards, et al. Evaluating apple airpods pro 2 hearing protection and listening. *Canadian Audiologist*, 10(6), 2024.

[32] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. Oesense: employing occlusion effect for in-ear human sensing. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, pages 175–187, 2021.

[33] A.V. Oppenheim, A.S. Willsky, and S.H. Nawab. *Signals & Systems*. Prentice-Hall signal processing series. Prentice Hall, 1997.

[34] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[35] CUI Devices. Digikey page: Cmc-3015-44l100 product, 2024. Accessed: 2024-05-25.

[36] Sony. Amazon page: Sony mdrex15ap/b black earbuds, 2024. Accessed: 2024-05-25.

[37] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[38] Termux Community. Termux. https://termux.dev/en/, 2024. Accessed: 2025-05-29.

[39] Eberhard Zwicker and Hugo Fastl. *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media, 2013.

[40] Gil Felix Greco, Roberto Merino-Martínez, Alejandro Osses, and Sabine C Langer. Sqat: a matlab-based toolbox for quantitative sound quality analysis. In *Inter-Noise and Noise-Con Congress and Conference Proceedings*, volume 268, pages 7172–7183. Institute of Noise Control Engineering, 2023.

[41] Mixkit. [mixkit]: free sound effect. https://mixkit.co/free-sound-effects/notification/. Accessed: 2024-12-02.

[42] Stephen B Wicker and Vijay K Bhargava. *Reed-Solomon codes and their applications*. John Wiley & Sons, 1999.

[43] Robert Gallager. Low-density parity-check codes. *IRE Transactions on information theory*, 8(1):21–28, 1962.

[44] Zhe Jin, Jung Yeon Hwang, Yen-Lung Lai, Soohyung Kim, and Andrew Beng Jin Teoh. Ranking-based locality sensitive hashing-enabled cancelable biometrics: Index-of-max hashing. *IEEE Transactions on Information Forensics and Security*, 13(2):393–407, 2017.