Eternity in a Second: Quick-pass Continuous Authentication Using Out-ear Microphones

Ming Gao*
Nanjing University of Posts and Telecommunications
gaomingppm@njupt.edu.cn

Xin Tong*
Zhejiang University
xintong22@zju.edu.cn

Jiatong Chen
Zhejiang University
chenjiatong@zju.edu.cn

Yike Chen
Zhejiang University
chenyike@zju.edu.cn

Fu Xiao†
Nanjing University of Posts and Telecommunications
xiaof@njupt.edu.cn

Jinsong Han
Zhejiang University
hanjinsong@zju.edu.cn

Abstract

Continuous authentication is increasingly critical for cyber security. However, existing approaches are time-inefficient due to their simple signal modulation with low-effective feature extraction throughput. In this paper, we propose a continuous authentication technique, OnePiece. OnePiece is free from the requirement of in-ear microphones, which are necessary for existing earphone authentication systems. It exploits out-ear microphones for biometrics extraction, which are ubiquitous on off-the-shelf earphones. We analyze the acoustic response model of ears towards out-ear microphones via the air, which is different from that towards in-ear microphones. A frequency-varying ultrasonic modulation scheme is proposed to characterize in-depth ear biometrics in user-friendly, error-free, and time-efficient ways. Therefore, OnePiece enables quick-pass authentication once users wear the earphones, followed by continuous authentication covering the whole course. Moreover, we propose a wake-up mechanism to reduce the consumed power, which addresses the key power consumption issue in ultrasonic sensing techniques. Particularly, OnePiece can be smoothly deployed on off-the-shelf wired and wireless earphones. It performs good cross-device performance in which users just register only once. Extensive evaluations are conducted to validate its effectiveness under real-world scenarios.

CCS Concepts

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Security and privacy** → **Authentication**.

*Ming Gao and Xin Tong contributed equally to this research.
†Fu Xiao is the corresponding author.

SENSYS '24, November 4–7, 2024, Hangzhou, China

ACM ISBN 979-8-4007-0697-4/24/11
https://doi.org/10.1145/3666025.3699366

Keywords

Continuous User Authentication, Earable Sensing, Acoustic Sensing, Biometrics

ACM Reference Format:
Ming Gao, Xin Tong, Jiatong Chen, Yike Chen, Fu Xiao, and Jinsong Han. 2024. Eternity in a Second: Quick-pass Continuous Authentication Using Out-ear Microphones. In *The 22nd ACM Conference on Embedded Networked Sensor Systems (SENSYS '24), November 4–7, 2024, Hangzhou, China.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3666025.3699366

1 Introduction

As an essential requirement in the security of cyberspace and the Internet of Things (IoT), identification and authentication allow legal users to perform particular functions. Conventionally, authentication is performed in a one-pass way. It authenticates users only once at the login via static credible factors, such as passwords, fingerprints [1], and faces [2]. Apparently, such a one-pass authentication cannot continuously assure that users are who they claim to be throughout the whole course of their working session. On the contrary, continuous authentication provides consistent authorized access, in which the user's identity is guaranteed to be always trusted [3–5].

Continuous authentication solutions are inclined to employ biometrics with dynamic changes. By exploiting dynamic biometric traits (e.g., brainwaves [3], heartbeats [6–8] and ears [9]) or behaviors (e.g., gait [10, 11], handwriting [12] and jaw motion [13, 14]), the continuous authentication provides users with consistent authorized access. However, to measure such biometrics, IoT devices should be equipped with specialized sensors, e.g., electroencephalograph (EEG) electrodes [3], electrocardiography (ECG) electrodes [6], photoplethysmography (PPG) [15–18] or in-ear microphones [19–22]. These sensors are typically expensive and not common on ubiquitous IoT devices. For example, earphones with in-ear microphones cost over 50 USD with a market share of below 7% [23], while the common commodity off-the-shelf (COTS) microphones carrying merely one out-ear microphone are cheaper, around only 2 USD or even less. To make matters worse, the access to in-ear microphones is forbidden by manufacturers [24]. In comparison, exploiting out-ear microphones for authentication would be more user-friendly.

An urgent need for authentication is to achieve user-friendly and secure dynamic credible factors for providing consistent access

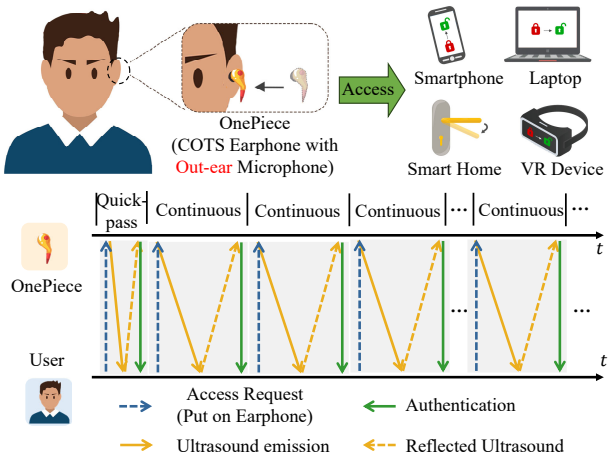


Figure 1: OnePiece and its pipeline. OnePiece can identify the user immediately once the user puts on the earphones, followed by continuous authentication.

throughout the session from login to logout, i.e., in the *whole-course*. This mechanism effectively defends against threats represented by lunch-time attacks, where an adversary gains access to unlocked devices that rely solely on one-pass authentication. Unfortunately, the approach of continuous authentication using out-ear microphones is confronted with three practical challenges as follows.

i) How to capture signals reflected from ears using out-ear microphones? We notice a common phenomenon that COTS earphones are likely to leak acoustics into the air as well as into the out-ear microphones, especially at a high volume setting. However, earable sensing via out-ear microphones suffers from attenuation and multi-path interference over the air, which are typically not encountered with in-ear microphones. The signal-to-noise ratio (SNR) would be too low to meet the requirement of fine-grained sensing.

ii) How to exact biometrics quickly to meet the requirement of whole-course security? Different from the static biometrics collection based on a two-dimensional high-resolution figure capture using advanced specialized sensors (e.g., fingerprint sensors and cameras for faces), the dynamic ones are mainly based on one-dimensional collection with low sampling rates, leading to a low rate of biometrics characterization. Existing dynamic biometrics extraction methods usually cost vast amounts of time (typically over 30 seconds [25]). Therefore, continuous authentication approaches usually start with one-pass solutions based on static biometrics, with the penalty of security degradation, which are suffering from eavesdropping, shoulder-surfing attacks, and spoofing attacks [2]. To improve the whole-course security, authentication systems are required to identify users quickly and continuously.

iii) How to authenticate users at a low power consumption? Keeping collecting biometrics-related signals imposes a huge burden on the capacity-limited battery of IoT devices and mobiles [26]. The authentication system is expected to consume less power and thus work for longer periods.

We propose OnePiece, a novel continuous authentication technique via ultrasound. As illustrated in Figure 1, OnePiece enables whole-course authentication using COTS earphones. By exploiting out-ear microphones for ultrasound transmission, sensing, and identification, OnePiece is user-friendly due to its easy access and inaudibility to humans. With an investigation into the acoustic model of ears towards microphones, we exploit the potential of quick biometrics characterization by designing an ultrasonic modulation scheme. It could expand the sensing dimensions in the frequency domain and benefit dynamic biometrics extraction. Therefore, OnePiece is able to identify the user in short enough time sessions in a non-intrusive way once the user logs in to the device, followed by continuous protection against attacks, during which no user operations are required, including typing passwords, facing cameras, and special actions for enhancing biometrics in prior works, e.g., speaking [14, 27, 28], walking [10], or tooth motion [13, 29, 30]. Furthermore, OnePiece achieves a balance between power consumption and continuity. A wake-up mechanism is proposed with randomly-intermittent ultrasonic emission, which avoids a great waste of power. In addition, considering that users might own multiple earphones, we propose a cross-device authentication scheme. It allows users to register on one pair of earphones and then to be identified on any other earphones.

We implement the prototype of OnePiece on a laptop connected to COTS wired and wireless earphones. Extensive evaluations show that OnePiece performs a high accuracy rate of up to 94.6% in a quick-pass authentication within 1 second and 97.0% for continuous authentication. The average power consumption measures just 75.9 mW (i.e., 25.3 mAh in an hour).

The contributions of OnePiece are summarized as follows:

- We propose OnePiece, a novel ultrasonic authentication method. It enables quick-pass and energy-efficient authentication throughout the whole course. OnePiece supports a cross-device authentication, in which registered users can be directly identified via a new earphone without additional operation.
- We analyze the model of acoustic propagation from ears towards out-ear microphones. Guided by the model, we design an ultrasonic modulation scheme and eliminate interference. It expands the sensing dimensions on dynamic biometrics, which are resilient against sophisticated attackers. With the time-efficient biometrics extraction, OnePiece realizes quick-pass authentication.
- OnePiece can be adopted on COTS wired and wireless earphones using ubiquitous out-ear microphones without hardware modification and specialized sensors (e.g., in-ear microphones). It promotes the real implementation of secure and user-friendly continuous authentication.

2 Background

Before presenting our system design, it is necessary to have a review about earphone-based authentication.

2.1 Earphone-based Authentication

Earphones, especially earbuds, are nearly ubiquitous in our daily lives due to their convenience, small size, and low cost. They have grown dramatically with a global market size of over 10.29 billion dollars in 2022 [23]. Typically, COTS earphones carry built-in out-ear microphones and in-ear speakers. Recently, diverse sensors

have been equipped on advanced wireless earphones, e.g., motion sensors [14, 31–34] and EEG electrodes [3]. These abundant sensors enable earphones the capability of describing the physical world. Generally speaking, earphone-based authentication modes can be divided into two categories: passive and active ones.

Passive authentication solutions depend mainly on the biometric collection without any external stimulus. Motion sensors can measure the vibration of the users‘ mandible during speaking [14, 14] as behavior characteristics for authentication. EEG electrodes can collect the users' brainwaves for identification [3], while EEG sensors are not common among COTS earphones due to their high cost. The in-ear microphones could collect body sounds through ears [4, 19, 27, 35], but this method is vulnerable to ambient noises.

Active authentication methods emit an external stimulus signal and receive the reflected signals for characteristics extraction. In earphones, acoustic signals are usually selected as stimulus signals [36]. Compared to audible stimulus [37, 38] that would seriously disturb users' hearing with an abundance of noise and conflict with other earphone applications, ultrasound-based solutions demonstrate their superiority in inaudibility [39–41]. Existing research has presented the capability of ultrasound sensing for continuous user authentication [25, 39, 42–46]. To highlight dynamic biometrics in the ear, the active authentication system asks users to perform specific actions or behaviors, such as walking [10] or speaking [27, 28]. These actions are always along with the motion of ear muscles. However, this requirement cannot cover all application scenarios and such behaviour-based approaches would become ineffective when the user is stationary or silent, obviously unable to support continuous user authentication. Advanced techniques are concentrated on the inner movement of ear organs for operation-free authentication. However, they face the common issue of high time consumption, e.g., over 30 seconds [37]. The potential of ultrasound in quick-pass and operation-free authentication has yet not been fully probed.

2.2 Application Scenarios

We propose a Hybrid authentication system for whole-course feature monitoring. It consists of two-step authentication, i.e., initial and continuous ones, which are defined as follows:

Initial authentication means the first-time authentication when the user logs in, i.e., puts on the earphone here. Traditional one-pass authentication methods (including PINs, fingerprints, and faces) serve as the initial authentication but can hardly be adopted on earphones. Here, we exploit the potential of ultrasound for quick-pass initial authentication.

Continuous authentication denotes long-time authentication from login to logout. It keeps identifying the legitimate user when the earphone is not removed. Moreover, we propose a wake-up mechanism to reduce the consumed power in ultrasonic continuous authentication.

Combining the above effects together, OnePiece accomplishes a whole-course authentication. It is able to defend against sophisticated attacks, including replay attacks [47] and spoofing attacks [1, 2]. With the aid of OnePiece, the user can put on the earphones and access their device immediately and continuously.

Table 1: Comparison of acoustic Power Spectral Density between in-ear and out-ear microphones (dBm/Hz) and their sensitivity parameters (dB SPL)

Volume Setting		100%	80%	60%	Sensitivity*
In-ear Mic		-55.1	-57.2	-59.3	/
Out-ear Mic	Wired 1	-59.2	-60.5	-61.1	95± 3
	Wired 2	-58.6	-59.8	-60.4	100
	Wireless	-60.8	-60.8	-61.6	99

*: The sensitivity parameters are sourced from the datasheet of these COTS earphones.

3 Feasibility Investigation

To gain a comprehensive understanding of the capability of COTS earphones for sensing, we explore the feasibility of collecting attenuated ultrasonic signals via out-ear microphones. A pilot experiment is conducted to demonstrate that COTS earphones are able to capture ultrasound via out-ear microphones without hardware modification or peripheral.

Limitation of Existing Methods: In-ear Microphones. Typically, an earphone carries an out-ear microphone that is approximately 15 cm away from the user's ear, as illustrated in Figure 2. In prior earable sensing techniques [4, 37], in-ear microphones are fundamental, which allows a high SNR of reflected signals. However, in-ear microphones have not become a necessary component in most COTS earphones yet. To make matters worse, access to in-ear microphones is forbidden by manufacturers [24]. Therefore, the requirement of in-ear microphones limits the practical implementation of acoustic sensing techniques on various COTS earphones.

Our Observation: We observe that acoustic leakage in existing COTS earphones is quite common. Imagine a daily scene where you are likely to hear the sound leaked from earphones on relatively close people, especially when the earphones are set at a high volume. The widely used measure against such a leakage towards microphones in COTS earphones is the acoustic echo cancellation (AEC). It keeps always-on in mobile devices by default to avoid the echo (including the leakage from the earphone to the out-ear microphone) affecting the normal operations of earphones, e.g., calling. If the AEC function is switched off, we observe that the out-ear microphones are capable of picking up acoustic signals from earphones (which is weaker than that captured by in-ear microphones although). A recent work F^2Key [48] leveraging headphones also demonstrates the acoustic leakage. Different from F^2Key [48] using modified headphones, the phenomenon of acoustic leakage in earbuds is more significant without the block from the earmuff of headphones. Therefore, this insight allows us to adapt OnePiece on various COTS earphones and leverage out-ear microphones to perform earable acoustic sensing.

Pilot Experiment: We conduct an experiment in which we compare the intensity of reflected ultrasonic signals collected by in-ear and out-ear microphones. Considering the restricted access to in-ear microphones, we design an earphone with an in-ear microphone using COTS acoustic modules (MAX9814) as the baseline. We use three COTS earphones without in-ear microphones, including two wired earphones (HUAWEI AM115 and SONY MDR-EX15LP) and one wireless earphone (JBL Run BT2). The total four earphones

are connected to a HASEE Z7M-CU5NB laptop, with the default sampling rates, i.e., 44.1 kHz. We ask a volunteer to wear the four earphones respectively. [1] We play an ultrasound tone of 18 kHz via the built-in speakers and collect the acoustic leakage using the out-ear microphones in a quiet room with the ambient acoustic noise of 48.8 dB.

Table 1 lists the received acoustic power of the four microphones under different volume settings. As the baseline, the earphone with an in-ear microphone measures -55.1 dBm/Hz on average at the maximum volume settings. The three COTS earphones with out-ear microphones collect reflected signals with an average power of -59.2 dBm/Hz, -58.6 dBm/Hz, and - 60.8 dBm/Hz respectively. The performance on other ultrasonic bands is similar, with a power of around -60 dBm/Hz. Although out-ear microphones obtain a lower intensity than in-ear ones, 20~30% (around -5 dB) of ultrasonic power residue can still be recorded.

The recorded signals include the direct transmission from the speaker to out-ear microphones, background reflection (removed in Section 4.3.1), and reflected signals by inner ear organs (the target biometrics). With the processed characteristics shown in Figure 3, earphones in a clean space (i.e., no user involved) capture no effective feature. As for the same users, out-ear microphones present a similar (although weak) capability of capturing signals with ear biometrics as the in-ear ones in spite of acoustic attenuation.

4 Time-efficient Biometrics Characterization & Utilization

It is essential to extract dynamic biometrics effectively and efficiently. However, due to the lack of a comprehensive understanding about dynamic biometrics, the characterization based on one-dimensional collection is correspondingly time-consuming. We model the acoustic responses of the human ear, under the guidance of which, we proposed a ramp modulation scheme for the ultrasound. Different from modulation with a fixed frequency, our proposed scheme characterizes ear biometrics from different frequency bands. After the noise suppression (including multipath and motion), the abundant information on more bands but in a short time could distinguish users effectively. Inverse MelSpectrogram is utilized to convert the one-dimensional high-frequency characteristics to two-dimensional matrices, with which advanced deep learning techniques can be directly adopted to deal.

4.1 Ear Response Analysis

We analyze the mathematical physics model of acoustic responses of the human ear, which would guide the following design of ultrasonic modulation.

The acoustic signals enter the auricle to the eardrum via the ear canal. Partial signals are absorbed in the eardrum and then converted into neural signals via ossicles, while the others are reflected and ultimately captured by the earphone microphone. The propagation of acoustic signals within the ear can be equivalently modeled as acoustic propagation in a one-dimensional pipe with a variable cross-sectional area [49]. This one-dimensional pipe model is divided into M sections based on the cross-sectional area, with the cross-sectional area of the m-th section $m(m = 1, 2, \ldots, M)$

[1] All experiments in this paper follow the IRB approval.

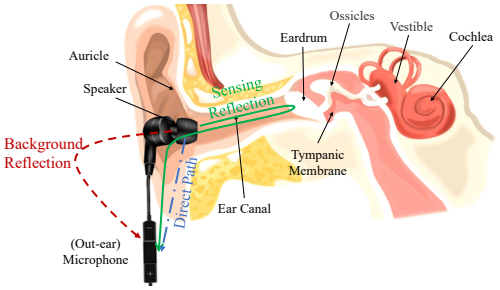


Figure 2: An illustration of a human ear and a COTS earphone's structure, with multipath reflection.

being a constant value of S_m. The lengths of all sections are equal to L. We denote that the acoustic signals' Thévenin equivalent voltage is $u_0(t)$ and the Norton equivalent current is $i_0(t)$, with their Fourier transform as $U_0(f)$ and $I_0(f)$ [50]. The equivalent acoustic impedance at the auricle is z_0, with its Fourier transform as follows,

$$Z_0(f) = \frac{U_0(f)}{I_0(f)} = \frac{\rho c}{S_1} \cdot \frac{1 + r_0(f)}{1 - r_0(f)}, \tag{1}$$

where f represents the acoustic frequency, ρ represents the density of air, c represents the speed of sound, and $r_0(f)$ represents the acoustic reflection coefficient at the auricle. The iterative relationship between the equivalent acoustic impedance of pipes with various cross-sections [51] is

$$Z_m(f) = \frac{\rho c}{S_m} \cdot \frac{Z_{m-1}(f) + j\frac{\rho c}{S_m} tan(k_m L)}{\frac{\rho c}{S_m} + jZ_{m-1}(f) tan(k_m L)}, \tag{2}$$

where j is the unit imaginary number and k_m represents the acoustic damping coefficient of the m-th ear canal as follows,

$$k_m = \frac{2\pi f}{c} - j\frac{c}{10}\sqrt{\frac{f}{S_m}}. \tag{3}$$

We denote that the equivalent acoustic impedance at the eardrum is Z_{TM}. We have the equivalent total impedance Z_T,

$$Z_T(f) = \frac{Z_{TM}(f) Z_{M+1}(f)}{Z_{TM}(f) + Z_{M+1}(f)}. \tag{4}$$

Hence, the overall acoustic response function of the ear is

$$\begin{aligned} H(f) &= \frac{Z_T(f) \cdot e^{-jk_M L} \cdot Z_{top}(f)}{2[1, r_0(f)] Z_b(f) \begin{bmatrix} 1 \\ r_T(f) e^{-j2k_M L} \end{bmatrix}} \\ Z_{top}(f) &= (1 + r_0(f)) \cdot (1 + r_T(f)) \cdot \prod_{m=1}^{M-1} e^{-jk_M L}(1 + r_m(f)) \\ Z_b(f) &= \prod_{m=1}^{M-1} \begin{bmatrix} 1 & r_m(f) \\ r_m(f) e^{-j2k_M L} & e^{-j2k_M L} \end{bmatrix} \end{aligned} \tag{5}$$

where $r_T(f)$ and $r_m(f)$ are the acoustic reflection coefficient of the eardrum and the m-th ear canal respectively. In the response function, the item of $Z_T(f)$ presents the static ear biometrics, while the other items indicate the dynamic ones, in which the inner motion of ear organs acts on acoustic signals via Doppler frequency

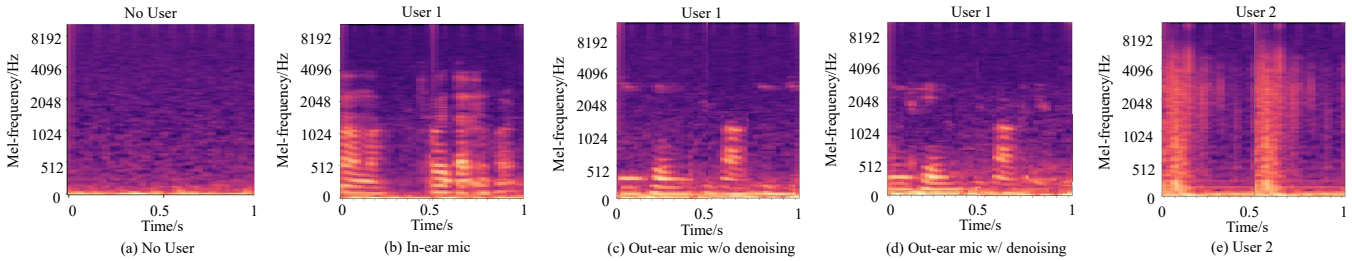


Figure 3: Inverse Mel Frequency Cepstrum Coefficient (IMFCC) profiles of biometrics on ultrasonic signals.

shift. Here, items of S_m, k_m, $r_m(f)$, and $r_T(f)$ are unique to each individual. They could serve as characteristics for authentication.

In particular, the extracted biometrics vary as the frequency of acoustic signals. That is, ear biometrics present diverse characteristics on different frequency bands. In prior earable sensing work, simple acoustic modulation modes, e.g., one tone with a constant frequency [52] or a pulse signal [38] can merely obtain the one-dimensional features in the frequency domain, consequently with a low biometrics extraction rate. In comparison, a wide-band signal with multiple frequencies is a more effective way for efficient extraction. It would characterize the ear in more dimensions without increasing sampling rates.

4.2 Ultrasonic Modulation

Guided by the above model, we modulate the wide-band ultrasonic signals for a fast ear biometrics collection.

A simple and intuitive strategy is to employ multiple ultrasonic tones with different frequencies [53]. However, such a method is inappropriate for earphones. Multiple ultrasounds would result in nonlinear effect [54–57], which generates audible noises. On the other hand, using several ultrasonic signals simultaneously would further increase the power consumption during continuous authentication.

Here, we leverage an ultrasonic signal with a time-varying frequency. Represented by the frequency-modulated continuous wave (FMCW), such a signal can acquire both static (e.g., via the time of flight [46]) and dynamic (e.g., due to the Doppler frequency shift [40]) characteristics. In addition, the ringing effect [54, 55] should be eliminated, in which discrete frequency changes would result in sudden audible impulses. Therefore, instead of a chirp signal, we modulate signals with the frequency changing in a triangular mode, namely the ramp signal. We have

$$x(t) = \begin{cases} sin(2\pi f_0 t + \pi B t^2), & k \le \frac{t}{T} < k + \frac{1}{2} \\ sin(2\pi f_0 t + 2\pi B T t - \pi B t^2), & k + \frac{1}{2} \le \frac{t}{T} < k + 1 \end{cases} \quad (6)$$

where f_0 is the initial frequency, B is the bandwidth, and T is the ramp cycle.

Most COTS earphones support acoustic sampling rates of 44.1 kHz [58]. According to Nyquist sampling theorem, the earphones support signal transmission without distortion at the band of [0, 22.05] kHz. We set the ramp range from 17 kHz to 21 kHz, which is inaudible to humans [40, 46]. The ramp cycle is 0.2 seconds, in which the frequency increases from 17 kHz to 21 kHz in the former 0.1 second and decreases to 17 kHz in the latter. The modulated ramp signals would reflect ear biometrics with features on multiple bands, enabling fast extraction in a short time (e.g., within 1 second).

Our proposed modulation method enables earable sensing to exact biometrics in a time-efficient manner. It lies the foundation for the quick-pass authentication.

4.3 Interference Cancellation

In OnePiece, the leaked signal travels from the speaker, through the ear canal, to the air, and is finally captured by the out-ear microphone. Before reaching the microphone, the ultrasonic signal also suffers from multipath effects over the air along with ambient noise. We also consider the potential noises caused by body movement and hardware.

4.3.1 *Multipath Interference.*

The multipath effects include two categories of interference: i) the direct path from the earphone speaker to the out-ear microphone (not through the ear canal) and ii) the reflected path from the background (e.g., the user's face and surroundings), as illustrated in Figure 2.

Direct Path: To remove the direct transmission, we measure it when the earphone is put in a clean space in advance and subtract the recorded signals during sensing. Considering our adopted ramp modulation, we should synchronize the direct path and current signals. An intuitive method is to align signals according to peaks in the spectrum of acoustic signals. However, it exerts the calculation burden if using the Fourier transform, especially on wearable devices. Instead, we leverage a phase locked loop [59], a tool for seeking signals with a certain frequency, to detect the peaks of 17 kHz, and accordingly synchronize and subtract the direct transmission.

Background Multipath: After leaking into the air, the signals would be affected by the reflected multipath from the background, e.g., the user's face and surrounding objects. The background multipath can be described via a multipath channel model [60] as follows,

$$y[n] = \sum_{i \in S_1} h_i x[n-i] + \sum_{j \in S_2} h_j x[n-j], \quad (7)$$

where $y[\cdot]$ and $x[\cdot]$ are the received and transmitted signals, h is the channel taps, S_1 denotes the set consisting of the indices satisfying $d_d < \frac{i \cdot c}{Fs} < d_r$, and S_2 is composed of the other indices, where Fs=44.1 kHz is the acoustic sampling rates, c is the speed of acoustics, d_d is the distance between the earphone speaker to the microphones, d_r values d_d+7cm, in which 7 cm is longer than twice the length of the ear canal (of below 3.5 cm typically [49]). Here,

the former item represents the target sensing reflection containing biometrics, while the latter denotes the background reflection. Least Square channel estimation (LSCE) [61] is utilized to determine the channel taps and estimate the background reflection. In practice, the out-ear transmission range of ultrasound is below 30 cm empirically, in which few objects are usually adjacent to the user's head. Therefore, the background multipath is mainly caused by the user's face. We use a wiener filter [62] to eliminate the face-caused noise for each individual, which aims at generalized stationary noise of a known distribution. In this case, we are able to estimate the background multipath in advance instead of real-time estimation. Such an approach is beneficial to reducing computation overhead on mobile devices.

4.3.2 *Ambient Noise.*

To focus on the ultrasonic band, we remove the low-frequency environmental noise using a high-pass filter. We adopt a first-order infinite impulse response (IIR) filter, which costs low computing resources [63] and thus suits the implementation on mobile and wearable devices. Its cutoff frequency is set as 12 kHz.

Besides, body motion may affect ultrasonic signals due to the Doppler effect, which distorts biometrics. We empirically observe that the influence of the body motion on different bands over 16 kHz is approximately equal in an extremely short time (e.g., 0.5 s). Therefore, we could use a mean filter with a sliding window of 0.5 s to eliminate the potential interference from users' body movement.

The comparison between the characteristics before and after interference cancellation is presented in Figure 3(c)&(d). The results experimentally demonstrate the effectiveness of our adopted approach.

4.4 Inverse MelSpectrogram

We leverage the inverse MelSpectrogram to characterize the biometrics which are modulated on the one-dimensional ultrasonic signals.

An automatic segmentation technique is fundamental for the practical implementation. Time stamp based methods cannot deal with unintended disruptions of authentication programs. Here, we reutilize the peaks detected by the phase locked loop in Section 4.3.1. Using these peaks as cutting points, we segment each ramp signal and normalize its intensity for the processing.

The Mel Frequency Cepstrum Coefficient (MFCC) has been widely used in speech-related applications [64] and massive deep-learning networks are designed to deal with MFCC features [65]. However, we cannot directly leverage MFCC to characterize the biometrics modulated on ultrasound, which is utilized to emphasize the features in the audible bands. Hence, we adopt the inverse MFCC (IMFCC) [66]. The IMFCC has a high resolution in high-frequency bands. We perform IMFCC using a window length of 1024 and a hop length of 320, and calculate the Melspectrogram with 64 Mel-filter banks. Figure 3(d)&(e) show the significant difference in the extracted features in the inverse Melspectrogram between two users.

The inverse MelSpectrogram characterizes one-dimensional high-frequency biometrics using two-dimensional matrices (images). Therefore, we can further leverage advanced deep-learning network techniques for authentication.

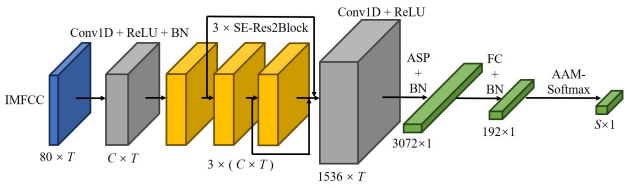


Figure 4: Architecture of ECAPA-TDNN.

4.5 Authentication Model Design

With the aid of the above time-efficient biometrics characterization, we leverage deep learning techniques for quick-pass initial and continuous authentication.

4.5.1 *Initial Authentication.*

During the initial authentication phase, the network should identify users based on short-duration segments lasting a short time (e.g., within 1 second) to ensure a quick-pass authentication process.

We adopt the emphasized channel attention, propagation, and aggregation in the time delay neural network (ECAPA-TDNN) [65] to train the model for the initial authentication. ECAPA-TDNN performs well in user identification according to audio. It consists of multiple modules including conv1D+BN, SE-Res2Block, ASP+BN, FC+BN, and AAM-softmax, with the architecture illustrated in Figure 4. To emphasize the biometrics in acoustic signals, a modified one-dimensional Squeeze-Excitation Res2Blocks is used in the frame layer to model global channel inter-dependencies without increasing the total parameters. The dilated convolutions are utilized to insert some extra spaces between the elements of the convolution kernel. Therefore, ECAPA-TDNN can effectively process the extracted ultrasonic feature signals without significantly increasing the network parameters.

The IMFCC profiles extracted from short-duration segments, empirically set at 1 second, serve as the inputs to the ECAPA-TDNN. The trained network identifies users by leveraging time-efficient characterizations, thereby fulfilling the requirements for a quick pass in real-world scenarios.

4.5.2 *Continuous Authentication.*

After the initial authentication, OnePiece determines authentication based on an N-second signal. It continuously repeats this decision-making process at each N-second interval.

We adjust the ECAPA-TDNN architecture for continuous authentication. Specifically, we split each N-second sample in continuous authentication to N 1-second samples as multi-channel inputs into the network simultaneously. Empirically, we set N=3, and the three 1-second samples act as 3-channel inputs for one authentication decision output. Accordingly, we adjust the input and convolutional layers [67].

We implement the prototype of OnePiece on a laptop connected to COTS wired or wireless earphones. In the prototype system, the COTS earphones transmit and receive ultrasonic probing signals. The signal processing and authentication network run on the laptop. We will adapt OnePiece on mobile phones and earables in the future. In the prototype, the durations of registration, initial,

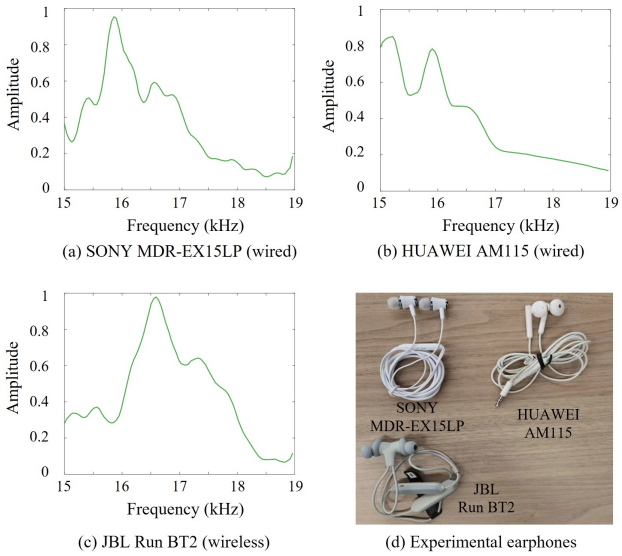


Figure 5: Diversity in the frequency responses among three COTS earphones.

and continuous authentication phases are empirically set as 5 minutes, 1 second, and 3 seconds respectively, with the experiments in Section 6.2.2.

5 Practical Consideration

In this section, we further tackle several practical issues to make OnePiece more practical in real-world settings.

5.1 Scalable Registration

We design a novel cross-device registration method. Once registered via one earphone, the users can smoothly access any other earphones without any additional operation.

In real-world scenarios, the users may possess multiple pairs of earphones, which might be connected to the same device (e.g., one smartphone). These earphones present diverse hardware features. As a result, the biometrics from the same user would distort distinctly when extracted by different earphones. In the traditional registration scheme [2], the user should register each pair before use. However, this time-consuming process would undoubtedly degrade their experience.

The diversity of earphones relies mainly on the differences in their speakers and microphone modules, which directly affect the ultrasonic transmission performance. Ideal earphones present the flat frequency characteristics property in a wide band. In practice, COTS earphones perform various responses across different frequencies due to inherent defects. Therefore, the practical acoustic response should be

$$H_p(f) = H_M(f) \cdot H(f) \cdot H_S(f) \tag{8}$$

where $H_M(f)$ and $H_S(f)$ are the frequency responses of the microphone and the speaker in an earphone respectively. The two responses vary among earphones. Figure 5 illustrates the diversity among three COTS earphones. The hardware diversity prevents the biometrics extracted using ultrasonic signals on one pair of earphones from being directly leveraged by the authentication network model trained using data from another pair.

A straightforward approach is to include biometric signals collected from various commercial headphones for training. However, this approach requires extensive data collection and significant time and computational resources for model training, resulting in relatively high costs.

To address this issue, we propose a frequency response compensation strategy. The basic idea is to measure and normalize the frequency response characteristics of each earphone. In this way, the extracted biometric signals are scalable among devices. Datasheets of some earphones provide detailed information about the frequency response coefficients. Accordingly, we can realize effective compensation. For earphones with unknown coefficients, we suggest users play a 15-second chirp signal when the earphones are idle (i.e, the user does not wear the earphones) to draw the frequency response coefficients $U_m(f) \cdot U_s(f)$. Therefore, we eliminate the impact of earphone diversity. This technique enables cross-device extraction and recognition of biometrics. It realizes the 'one-shot' registration across a wider range of commercial earphones.

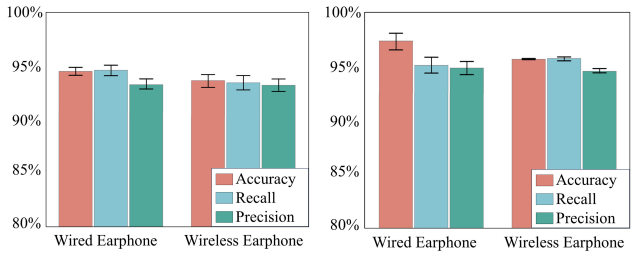
5.2 Wake-up Mechanism

We propose a wake-up mechanism for OnePiece to reduce power consumption. In this mechanism, OnePiece intermittently stops ultrasound emission when the user's conditions remain unchanged. This approach decreases the energy used by ultrasound transmission in wireless earphones, which typically have small batteries, thereby extending their endurance.

Continuous authentication could break if the user's conditions have not changed, in which the earphones are never moved and the users require no access to new data or authority. This is because the earphones must move if an attacker captures the earphones and replaces an authorized user. Advanced earphones equipped with motion sensors or distance sensors [14] can easily detect this change. Also, we design an acoustic-only method on COTS earphones. We utilize the wake-up signal $w(t) = U(t) \cdot cos(2\pi \cdot 18k \cdot t)$ with the gradual intensity $U(t)$. According to the reflected signals, we train a DNN [65] to detect whether the users' earphone-wearing condition alters from one second ago. If detecting the change in users' conditions, the authentication is awakened. Otherwise, there is a random interval in the range of [0,1] seconds before the next wake-up signal. The accuracy reaches up to 99.9% experimentally.

5.3 Volume Setting

Intuitively, a higher volume setting leads to a better sensing performance. However, long-term exposure to high-intensity ultrasound is a potential threat to human health. As a countermeasure, we suggest the volume setting as 50% of the maximum in mobile devices by default, following the common volume suggestion from earphone manufacturers [68]. Under such a setting, the ultrasonic intensity measures 59.3 dB SPL, below the human-exposure limit of 70 dB SPL (sound pressure level) suggested by the International Non-Ionizing Radiation Committee (INIRC) [69]. Note that OnePiece would perform well under most settings of volume. The detailed evaluations refer to in Section 6.4.



(a) Initial authentication (b) Continuous authentication

Figure 6: Overall performance.

6 Evaluation

We evaluate OnePiece on COTS earphones in real-world scenarios. All experiments follow the IRB protocol approved.

6.1 Setup and Metrics

Hardware: We use three COTS earphones, i.e., two wired earphones (HUAWEI AM115 and SONY MDR-EX15LP) and one wireless earphone (JBL Run BT2). They are connected to a HASEE Z7M-CU5NB laptop, whose volume is set as 50% by default. In the laptop, we keep switching off the AEC function and run our developed acoustic collection program, in which echoes in the audible bands are filtered and ultrasonic ones remain. The sampling rates of earphones (both the speakers and out-ear microphones) are 44.1 kHz.

Data Collection: We recruit 36 volunteers (21 males and 15 females). Their ages range from 20 to 50 years old. The participants are asked to put on each earphone for 7 minutes (note that this is not the registration time). The experiments are conducted in a quiet room with environmental noises of 49.7 dB on average. During training, the data from the enrolled user serve as the positive samples, with data of an equal number from the others as the negative. 80% of the data are for training and 20% are for testing.

Network Implementation: The network of OnePiece is implemented in PyTorch and trained in a server with Intel(R) Xeon(R) Silver 4210R CPU@2.40GHz and two Nvidia GeForce RTX 3090.

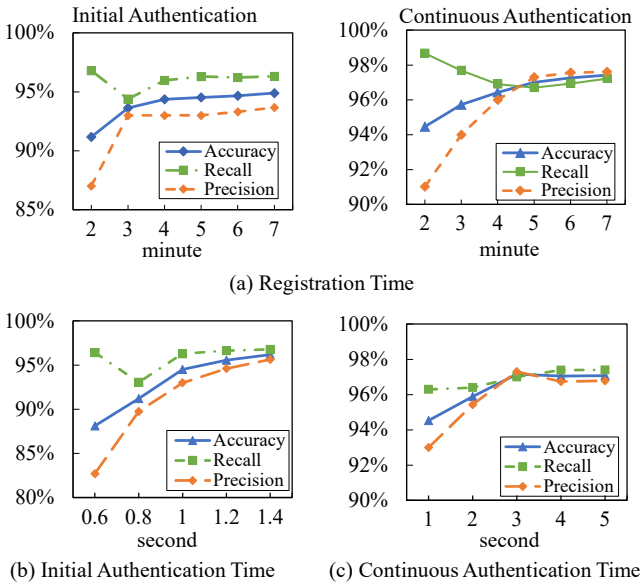
Metrics: We evaluate OnePiece under three frequently used metrics, i.e., accuracy, precision, and recall. The accuracy is defined as the ratio of samples that are predicted correctly to the total samples. The precision and recall are the ratio of correctly-predicted positive samples (i.e., those from legitimate users) to the total number of positive samples and the total number to be predicted as positive samples respectively. The metrics of high values signify a good performance of OnePiece on user authentication.

6.2 Overall Performance

We access OnePiece on COTS earphones with an analysis about the parameter selection.

6.2.1 *Effectiveness.*

We evaluate the two-step authentication, which is combined together for providing whole-course protection. OnePiece presents the capability of easy deployment on COTS earphones. As shown in Figure 6, it achieves an average accuracy of 94.6% in 1-second initial authentication and 97.0% in 3-second continuous authentication.



(a) Registration Time

(b) Initial Authentication Time (c) Continuous Authentication Time

Figure 7: Impact of time parameters.

The equal error rates are 0.046 and 0.032 on the wired earphones and 0.087 and 0.014 on the wireless earphones respectively. The results demonstrate that OnePiece realizes a secure and time-efficient whole-course continuous authentication.

6.2.2 *Parameter Selection.*

In general, the long duration of registration and authentication benefits the performance of OnePiece. To make a trade-off between user-friendliness (requiring a short duration) and performance, we evaluate OnePiece under different parameters as a reference for practical selection.

We first test the impact of registration time at an interval of 1 minute, with results in Figure 7(a). As the increase of the duration, the performance improves consequently. The upward trend of initial authentication slows down with little improvement when the registration time exceeds 2 minutes, while the turning point of slowing in continuous is 5 minutes. Therefore, we set the registration time as 5 minutes. A 5-minute registration phase is reasonable, which only requires one time in practical use.

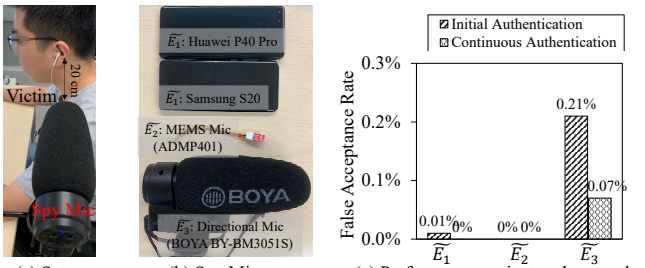
According to Figure 7, we set the authentication duration as 1 s for the initial one, and 3 s for the continuous one, to make a balance between time efficiency and effectiveness.

6.3 Resilience Against Attacks

We consider the following two attacking means to challenge the vulnerability of OnePiece.

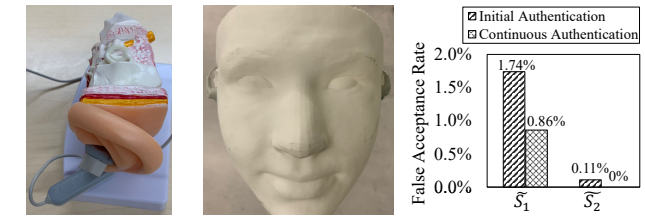
6.3.1 *Replay Attacks.*

Here we consider the most dangerous way, i.e., recording and replaying the malicious recordings to bypass the authentication system maliciously. We set that spy microphones are located towards the victim's ears 20 cm away (the lowest band of the social distance) to record the acoustic leakage during authentication. Otherwise, a closer eavesdropping distance would increase the chance of being detected by victims. Here we consider three kinds of microphones:



(a) Setup (b) Spy Mics (c) Performance against replay attacks

Figure 8: Performance against replay attacks.



$\widetilde{S_1}$: A PVC anatomy model $\widetilde{S_2}$: 3D masks

(a) Fake ear models used in spoofing attacks (b) Performance against spoofing attacks

Figure 9: Performance against spoofing attacks.

$\widetilde{R_1}$: mobile devices represented by smartphones (Samsung S20 and Huawei P40 Pro), $\widetilde{R_2}$: a miniature MEMS microphone (ADMP401), and $\widetilde{R_3}$: a specialized microphone (a directional microphone, BOYA BY-BM3051S). The attacking setup and devices are shown in Figure 8(a)&(b). We consider the most dangerous case that the captured signals are injected directly into the authentication network (as the inputs). This way is more effective than over-the-air replays with the speaker lying next to the out-ear microphone due to the potential issues of signal distortion or out-sync. Thus, if OnePiece would defend this injection, it would be able to resist other replay attacks. We test 300 attempts at initial authentication and 1000 attempts at continuous authentication on five victims. Results show that recordings from the three microphones fail in bypassing authentication. As shown in Figure 8(c), the average false acceptance rates (FARs) in initial authentication are 0.01%, 0, and 0.21%, and those in continuous one are 0, 0, and 0.07%, respectively. The possible reason is that the additional propagation path distorts and attenuates the ultrasonic signals.

6.3.2 Spoofing Attacks.

We adopt two kinds of fake ear models to conduct spoofing attacks: $\widetilde{S_1}$: a well-designed polyvinyl chloride (PVC) anatomy model, and $\widetilde{S_2}$: 1:1 scale 3D masks of two victims made of photosensitive resin. As shown in Figure 9(a) and (b) respectively, the fake anatomy model in $\widetilde{S_1}$ is generic, while in $\widetilde{S_2}$ we use the models of two particular users' ears. We conduct 300 attempts at initial authentication and 100 attempts at continuous authentication from the fake models for each victim. Results in Figure 9 show that OnePiece is free from threats from spoofing attacks with FARs of below 2%. There are two main reasons for the resilience against spoofing attacks. On the one hand, compared with fingerprints and faces, vision-based techniques can hardly imitate and reconstruct real ear biometrics, which requires a higher resolution ratio and lacks enough samples. On the other hand, the inner structure of the ear consists of various biomaterials such as skin and fat, which also affect the acoustic response in the ear. Even though the attacker completely obtains the victims' ear biometrics, the difference in material between fake models (e.g., plastic or photosensitive resin) and human biomaterials makes OnePiece difficult to spoof.

6.4 Volume Setting and Safety

Intuitively, a higher volume would result in a more powerful leakage, which brings about a better performance. For example, OnePiece under the maximum volume could achieve the accuracy of 96.2% and 98.4% in initial and continuous authentication respectively. Here, we mainly evaluate OnePiece under the lower volumes, with the results in Table 2. Its performance degrades sharply at the volume of 10% due to the significant reduction in the acoustic leakage at the extremely low volume. For higher volume settings, OnePiece works robustly. Considering the lowest level (≤10%) of volume settings is relatively seldom, OnePiece could cover most of the requirements for continuous authentication.

The human-exposure limit suggested by the International Non-Ionizing Radiation Committee [69] is 70 dB SPL under 20 kHz and 110 dB SPL over 20 kHz. We measure the sound pressure level using SMART SENSOR Decibelmeter AS804. In OnePiece, the average intensity of the adopted ultrasound at the defaulted volume setting of 50% measures 59.3 dB in the ear, using Brüel & Kjær high-frequency simulator type 5128 and NEXUS CCLD signal conditioner type 2693-A. The results are much lower than the health threshold. That is, OnePiece satisfies the suggested human-exposure limit.

Table 2: Impact of Volume Setting

Volume	10%	20%	30%	40%	50%
Initial	64.5%	91.0%	94.4%	94.3%	94.6%
Continuous	74.7%	95.8%	96.6%	95.2%	97.0%

6.5 Cross-device Performance

We evaluate the scalability of OnePiece among three COTS earphones, including two wired and one wireless. We train on data collected from one earphone respectively, and then test on the others with and without our proposed compensation scheme. The performance degrades significantly on the uncompensated data. The cross-device accuracy could decrease to below 30% without our proposed scheme. In comparison, after the compensation, OnePiece presents good scalability. It achieves an average accuracy of 91.6% in initial authentication and 94.3% in continuous authentication on a new pair of earphones towards users enrolled from the other pairs. Therefore, the users can perform one-shot registration and enjoy a seamless experience among diverse earphones.

6.6 Power Consumption

Power consumption is an essential issue that should not be overlooked, especially for continuous authentication on earables. It consumes additional energy by transmitting and recording ultrasonic signals. Considering the restricted battery capacity in earables,

Table 3: Cross-device Authentication Performance with One-shot Registration

Earphone Category	Phase	Initial Authentication			Continuous Authentication		
		Without Frequency Response Compensation					
	Train / Test	HUAWEI AM115	SONY MDR-EX15LP	JBL Run BT2	HUAWEI AM115	SONY MDR-EX15LP	JBL Run BT2
Wired	HUAWEI AM115	94.7%	71.3%	23.8%	97.7%	70.0%	37.0%
	SONY MDR-EX15LP	66.1%	94.9%	30.1%	82.0%	97.9%	25.0%
Wireless	JBL Run BT2	26.2%	33.3%	94.2%	30.0%	38.0%	95.4%
		With Frequency Response Compensation					
Wired	HUAWEI AM115	94.7%	91.1%	90.7%	97.7%	94.8%	92.1%
	SONY MDR-EX15LP	91.0%	94.2%	88.3%	97.0%	97.9%	91.9%
Wireless	JBL Run BT2	90.8%	89.4%	94.2%	92.3%	90.0%	95.4%

the energy consumed by ultrasound determines the duration of OnePiece. We evaluate the amount of extra power being consumed by the ultrasound transmission and reception in OnePiece on earphones. The test is conducted on the wireless earphone JBL Run BT2, whose battery capability measures 133 mAh. It consumes about 1.5% of the battery (about 2 mAh) in an hour when the idle earphone is connected to a laptop via Bluetooth. The 5-minute registration consumes 6% of the earphone battery (i.e., 7.98 mAh). To test the power consumption during the authentication phase, we ask five volunteers to wear the earphones running OnePiece for an hour, during which they can sit, stand, walk, or run freely. OnePiece consumes 17%, 21%, 22%, 21%, and 24% of the earphone battery. Subtracting the idle consumption of 2 mAh, OnePiece consumes 20.0 mAh, 25.3 mAh, 26.6 mAh, 25.3 mAh, and 29.3 mAh respectively. The average consumption is 25.3 mAh in an hour, i.e., 75.9 mW. Such a low consumption supports long-term use in mobile and wearable devices, e.g., approximately 6 hours on JBL Run BT2. If without the wake-up mechanism, the power consumption during continuous ultrasound transmission can reach up to 52.5 mAh in an hour. The experimental comparison demonstrates the effectiveness of our proposed wake-up mechanism.

As for laptops or mobile phones (connected with the earables), the power consumption would not be a big issue with the aid of a high-capacity battery. The authentication computation consumes 2% of battery capacity in 10 minutes being done on the laptop. In comparison, the idle consumption (by the screen and the system) measures also 2% over the same period. Furthermore, the computation occupies 12% CPU load on average. The latency measures 138 ms on average, with a peak below 226 ms for an authentication decision. In general, the power and computation overhead are acceptable.

6.7 Robustness Analysis

Here, we analyze the impact of various factors on OnePiece, e.g., environmental noise, posture, and permanence. The evaluation involves seven participants unless otherwise stated. These participants are asked to wear the SONY earphones for 2 minutes in each case by default.

6.7.1 Impact of Environmental Noise.

People are likely to wear earphones in various scenarios in which

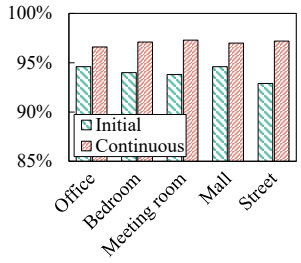


Figure 10: Impact of environmental noise.

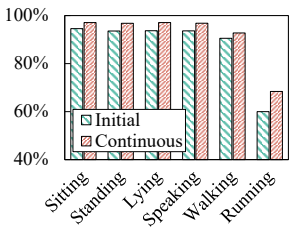


Figure 11: Impact of the user's movement.

OnePiece is confronted with diverse noise. To assess OnePiece in real-world scenes, we conduct experiments covering five common scenarios, including an office, a bedroom, a meeting room, a mall, and a street side. Figure 10 shows that the authentication accuracy remains stable at different scenes. Ambient noise has little influence on authentication.

6.7.2 Impact of Body Posture and Motion.

During the usage, people may perform various postures or body movements. To be practical, we consider six common postures and motions, i.e., sitting, standing, lying, speaking, walking, and running. The seven participants wear the wireless JBL earphones. They can move freely but remain within 4 meters away from the laptop to maintain the Bluetooth connection. Figure 11 demonstrates that OnePiece is robust against various postures and body movements except running. The sensitivity of ultrasound-based techniques to intense and irregular motion [40, 46] is to blame for the bad performance when the users are running. Except for this movement condition, OnePiece is free from other common activities in real-world scenes.

6.7.3 Impact of Earphone Wearing Position and Angle.

People have their individual habits about the wearing manner of earphones. The difference in wearing habits results in a position change in the ear, which may affect the performance of OnePiece. According to Equation 5, the position of the earphones would affect the characterization of the outer ear cancel, e.g., $r_0(f)$ and $Z_0(f)$, but it is independent of the inner biometrics, e.g., $r_m(f)$ and $r_T(f)$.

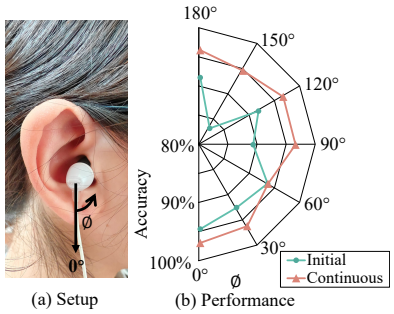


Figure 12: Impact of angle.

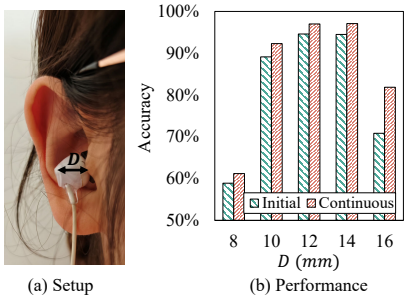


Figure 13: Impact of tightness.

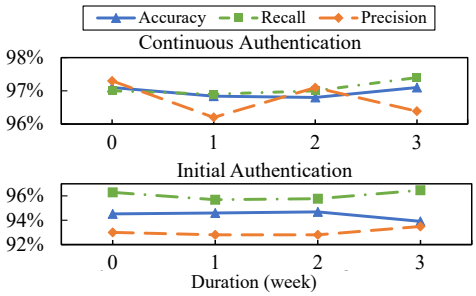


Figure 14: Performance over the time.

Therefore, OnePiece could be robust against the wearing positions to some extent.

For the quantitative analysis about the impact of wearing position, three participants are instructed to follow a certain angle or tightness as follows.

A typical position is with the earphone tail downwards (denoted as Φ=0°) and we evaluate OnePiece under different rotating angles at a step of 30°. The experimental setup and results are shown in Figure 12. The accuracy of initial authentication drops to 83.3% when Φ>90°, while it keeps over 90% under an acute angle. The continuous authentication maintains an accuracy of over 93.5%. Considering that the individual wearing position is relatively fixed, recommending users under their habitual positions would improve the performance.

Furthermore, we evaluate the impact of the tightness of wearing earphones. We denote the distance from the earphone end to the user's ear as D, and test OnePiece under 8 *mm* (Extremely Tight), 10 *mm* (Tight), 12 *mm* (Normal), 14 *mm* (Loose), and 16 *mm* (Extremely Loose). The results are shown in Figure 13. OnePiece performs poorly when the user puts on earphones in an extremely tight or loose manner. In the former case, the acoustic leakage is reduced and weakened. This leads to the poorest performance. In the latter case, fewer ultrasonic signals could go through and be reflected by the ear cancel. Therefore, the biometrics contained in the ultrasonic signals decreases and the performance degrades. Fortunately, the two cases are not common in practice. In the case of external tightness, after being pressed into the ear, the earphone would go loose slowly minutes later (D increases to 10~12 mm experimentally) because of the damping of the ear. If worn extremely loosely (over 16 mm), earphones are likely to fall in the real world. The results imply that OnePiece can cover the most common scenes of earphone usage.

6.7.4 *Permanence.*

We test OnePiece in a longitudinal study lasting for three weeks to verify the stability of our used ear biometrics. Three participants enroll at the start, and test once each week. The result in Figure 14 demonstrates the permanence of OnePiece whose performance just fluctuates in a normal range (at most ± 2%).

6.8 User Experience Study

We assess the user experience of OnePiece via a perceptual quality study to describe users' subjective experiences.

The adopted range of 17~21 kHz is out of the human hearing [40]. There exists variation in the perceived quality among different age groups. Generally speaking, as individuals age, their sensitivity to high-frequency bands gradually decreases [70]. OnePiece is experimentally inaudible towards users of most ages.

We recruit 67 volunteers (39 males and 28 females, aged from 20 to 60) to conduct a subjective evaluation about the user experience. The participants randomly listen to two kinds of audio: application sound (random music and conversation) and application sound with ultrasound under OnePiece. They are not told whether the audios are played along with ultrasound. Then, these participants rate the audios on a 5-point scale ranging from -4 to 0 (of which a high score refers to little influence on the audio quality) for the objective difference grade (ODG) according to the ITU BS.1387 test procedure [71]. The average ODG score of the audio with ultrasound is -0.67, while that without ultrasound is -0.55. An ODG of higher than -1 means that no audio quality degradation can be perceived by human users [71]. In particular, 24 among the 67 volunteers indicate a rating of 0 with OnePiece. The results show that the adopted ultrasound has almost no impact on the audio quality. OnePiece barely affects the earphone applications, and it is imperceptible without disturbing humans in practice.

7 Related Work and Comparison

We compare our proposed system with existing continuous authentication approaches, as listed in Table 4.

In the active authentication, Mahto et. al. [37] demonstrate the feasibility of authentication via acoustic signals. They claim ultrasound could be used for continuous authentication, but it (with a simple ultrasonic modulation mode) performs badly in the phase of initial authentication. EarEcho [22] employs audible sound for authentication on earphones with in-ear earphones. It achieves an accuracy of over 94% using a window time of 3 seconds. However, it would keep disturbing human ears and conflicts with earphone applications, while OnePiece has no such issues instead. EchoPrint [25] combines vision with acoustics together for authentication on smartphones with an accuracy of 93.75%. EarDynamic [28] realizes an ultrasound-only system on earphones but merely for continuous authentication.

As the representatives of passive authentication, EarID [4] passively collects body sound through earphones but requires 60 minutes for enrollment. EarPrint [27] and EarGate [10] improve the

Table 4: Comparison with Existing Acoustic-based Authentication Approaches

System	OnePiece		EarEcho [22]		EchoPrint [25]	EarDynamic [28]
Device	Out-ear Mic		In-ear Mic		Smartphone	In-ear Mic
Mode	Active		Active		Active	Active
Registration Time	300 s		400 s		Not mentioned	Not mentioned
Feature	Ultrasound		Audible sound		Acoustic+Vision	Ultrasound
Procedure	Initial	Continuous	Initial	Continuous	Continuous-alone	Continuous-alone
Accuracy	94.6%	97.0%	94.5%	97.5%	93.75%	93.4%
Authentication Time	1 s	3s	1 s	3 s	120 s	20 words
System	[37]		EarGate [10]		EarPrint [27]	EarID [4]
Device	In-ear Mic		In-ear Mic		In-ear Mic	In-ear Mic
Mode	Active		Passive		Passive	Passive
Registration Time	Not mentioned		Not mentioned		75 s	Not mentioned
Feature	Audible sound	Ultrasound	Body Sound		Body Sound	Body Sound
Procedure	Initial	Continuous	Continuous-alone		Continuous-alone	Continuous-alone
Accuracy	Not mentioned		97.26%		96.36%	95.5%
Authentication Time	60 minutes		1.07 s		75 s	30 s / 2 s

registration time. However, these methods are limited due to the requirement of special users' actions and the vulnerability to body movement, ambient noise, and conflict with earphone applications [27, 35].

In particular, existing earphone-based authentication systems depend fully on in-ear microphones. In comparison, OnePiece can be easily deployed on COTS earphones merely with out-ear microphones. Although a recent work named F^{2}Key [48] claims the usage of out-ear microphones (not specially for authentication), it is implemented on the modified headphones.

OnePiece prevails from the perspectives of both effectiveness and user-friendliness. Users would be never alerted to OnePiece due to its inaudibility and quick-pass authentication (merely consuming 1 second). Its performance is at the same level as the state-of-the-art (SOTA) systems based on audible sound, not to mention that it has other advantages, such as easy implementation on COTS earphones without hardware modification or peripheral, one-shot registration with good cross-device performance, and low power and computation overhead.

8 Discussion

In this section, we discuss the scalability of OnePiece and the potential influence when the users' ears are moist or dirty.

We first discuss the scalability of OnePiece on more categories of earphones. OnePiece performs well on wired earbuds (e.g., SONY MDR-EX15LP) and wireless sports Bluetooth earphones (e.g., JBL Run BT2) in our experiments. Besides, there are several other earphones, including headphones (represented by SONY WH-1000XM5) and true wireless stereo (TWS) earphones (represented by Apple AirPods). A recent work F^{2}Key [48] has demonstrated that headphones would leak acoustics into out-ear microphones only by using an auxiliary spacer to create a gap for leakage intentionally. Under the same conditions, OnePiece can also authenticate users using ear biometrics on headphones. Unfortunately, TWS earphones sample at 16 kHz by default, though they can support a sampling rate of over 44.1 kHz indeed. We will ask manufacturers for a high sampling rate to implement OnePiece on TWS earphones.

The acoustic impedance of the human ears might change significantly when they are moist or dirty, such as after a shower or when exposed to sweat and dirt. Fortunately, these conditions typically last for a short duration. For example, wet ears can typically take about half an hour to an hour to dry completely after a shower or swimming [72]. We will conduct an evaluation to analyze the impact of dampness in future work.

Furthermore, we will compress the network model (using advanced model compression techniques [73]) for implementation on mobile devices with limited computing resources, e.g., smartphones and earphones in future work.

9 Conclusion

We realize OnePiece, a practical whole-course continuous authentication method using COTS earphones. OnePiece presents high effectiveness and resilience against attacks using user-imperceptible ultrasound with low power consumption. OnePiece is capable of extracting ear biometrics accurately and quickly. It also presents excellent cross-device performance, in which the users are merely required to perform a one-shot registration. It addresses the security issue in the initial authentication and the issues of user-unfriendliness and power consumption in the continuous authentication.

Acknowledgment

This paper is partially supported by the National Science Fund for Distinguished Young Scholars of China (Grant No. 62125203), National Natural Science Foundation of China (Grant No. U21A20462, No. 62372400), "Pioneer" and "Leading Goose" R&D Program of Zhejiang (Grant No. 2024C03287), Natural Science Foundation of Jiangsu Province of China (Grant No. BK20240615), and Natural Science Research Start-up Foundation of Recruiting Talents of Nanjing University of Posts and Telecommunications (Grant No. NY224030).

References

[1] A. S. Rathore, Y. Shen, C. Xu, J. Snyderman, J. Han, F. Zhang, Z. Li, F. Lin, W. Xu, and K. Ren, "Fakeguard: Exploring haptic response to mitigate the vulnerability in commercial fingerprint anti-spoofing," in *Proceedings of NDSS*, 2022.

[2] W. Xu, W. Song, J. Liu, Y. Liu, X. Cui, Y. Zheng, J. Han, X. Wang, and K. Ren, "Mask does not matter: anti-spoofing face authentication using mmwave without on-site registration," in *Proceedings of ACM MobiCom*, 2022.

[3] F. Lin, K. W. Cho, C. Song, W. Xu, and Z. Jin, "Brain password: A secure and truly cancelable brain biometrics for smart headwear," in *Proceedings of ACM MobiSys*, 2018.

[4] Y. Zou, H. Lei, and K. Wu, "Beyond legitimacy, also with identity: Your smart earphones know who you are quietly," *IEEE Trans. Mob. Comput.*, vol. 22, no. 6, pp. 3179–3192, 2023.

[5] K. Jiokeng, G. Jakllari, and A. Beylot, "I want to know your hand: Authentication on commodity mobile phones based on your hand's vibrations," *Proceedings of IMWUT/Ubicomp*, vol. 6, no. 2, pp. 58:1–58:27, 2022.

[6] F. Lin, C. Song, Y. Zhuang, W. Xu, C. Li, and K. Ren, "Cardiac scan: A non-contact and continuous heart-based user authentication system," in *Proceedings of ACM MobiCom*, 2017.

[7] L. Wang, K. Huang, K. Sun, W. Wang, C. Tian, L. Xie, and Q. Gu, "Unlock with your heart: Heartbeat-based authentication on commercial mobile phones," *Proceedings of IMWUT/Ubicomp*, vol. 2, no. 3, pp. 140:1–140:22, 2018.

[8] S. M. S. Nirjon, R. F. Dickerson, Q. Li, P. Asare, J. A. Stankovic, D. Hong, B. Zhang, X. Jiang, G. Shen, and F. Zhao, "Musicalheart: a *hearty* way of listening to music," in *Proceedings of ACM SenSys*, 2012.

[9] P. Tuyls, E. Verbitskiy, T. Ignatenko, D. Schobben, and A. Akkermans, "Privacy protected biometric templates: Acoustic ear identification," *Proceedings of SPIE*, 2004.

[10] A. Ferlini, D. Ma, R. Harle, and C. Mascolo, "Eargate: gait-based user identification with in-ear microphones," in *Proceedings of ACM MobiCom*, 2021.

[11] Y. Wang, J. Shen, and Y. Zheng, "Push the limit of acoustic gesture recognition," *IEEE Trans. Mob. Comput.*, vol. 21, no. 5, pp. 1798–1811, 2022.

[12] F. Ding, D. Wang, Q. Zhang, and R. Zhao, "ASSV: handwritten signature verification using acoustic signals," *Proceedings of IMWUT/Ubicomp*, vol. 3, no. 3, pp. 80:1–80:22, 2019.

[13] A. Bedri, D. Byrd, P. Presti, H. Sahni, Z. Gue, and T. Starner, "Stick it in your ear: Building an in-ear jaw movement sensor," in *Proceedings of ACM UbiComp/ISWC*, 2015.

[14] J. Liu, W. Song, L. Shen, J. Han, X. Xu, and K. Ren, "Mandipass: Secure and usable user authentication via earphone IMU," in *Proceedings of IEEE ICDCS*, 2021.

[15] Y. Cao, Q. Zhang, F. Li, S. Yang, and Y. Wang, "Ppgpass: Nonintrusive and secure mobile two-factor authentication via wearables," in *Proceedings of IEEE INFOCOM*, 2020.

[16] Y. Cao, H. Chen, F. Li, and Y. Wang, "Crisp-bp: continuous wrist ppg-based blood pressure measurement," in *Proceedings of ACM MobiCom*, 2021.

[17] T. Zhao, Y. Wang, J. Liu, Y. Chen, J. Cheng, and J. Yu, "Trueheart: Continuous authentication on wrist-worn wearables using ppg-based biometrics," in *Proceedings of IEEE INFOCOM*, 2020.

[18] S. Choi, Y. Gao, Y. Jin, S. j. Kim, J. Li, W. Xu, and Z. Jin, "Ppgface: Like what you are watching? earphones can "feel" your facial expressions," *Proceedings of IMWUT/Ubicomp*, vol. 6, no. 2, pp. 48:1–48:32, 2022.

[19] Y. Cao, C. Cai, F. Li, Z. Chen, and J. Luo, "Heartprint: Passive heart sounds authentication exploiting in-ear microphones," in *Proceedings of IEEE INFOCOM*, 2023.

[20] Y. Cao, C. Cai, A. Yu, F. Li, and J. Luo, "Earace: Empowering versatile acoustic sensing via earable active noise cancellation platform," *Proceedings of IMWUT/Ubicomp*, vol. 7, no. 2, pp. 47:1–47:23, 2023.

[21] D. Ma, A. Ferlini, and C. Mascolo, "Oesense: employing occlusion effect for in-ear human sensing," in *Proceedings of ACM MobiSys* (S. Banerjee, L. Mottola, and X. Zhou, eds.), 2021.

[22] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin, "Earecho: Using ear canal echo for wearable authentication," *Proceedings of IMWUT/Ubicomp*, vol. 3, no. 3, pp. 81:1–81:24, 2019.

[23] Reportlinker, "Earphones headphones market research report." https://www.reportlinker.com/p06393769/Earphones-Headphones-Market-Research-Report-by-Product-Technology-Price-Application-Region-Cumulative-Impact-of-COVID-19-Russia-Ukraine-Conflict-and-High-Inflation-Global-Forecast.html, 2023.

[24] K.-J. Butkow, T. Dang, A. Ferlini, D. Ma, Y. Liu, and C. Mascolo, "An evaluation of heart rate monitoring with in-ear microphones under motion," *Pervasive and Mobile Computing*, vol. 100, p. 101913, 2024.

[25] B. Zhou, J. Lohokare, R. Gao, and F. Ye, "Echoprint: Two-factor authentication using acoustics and vision on smartphones," in *Proceedings of ACM Mobicom*, 2018.

[26] K. Sun and X. Zhang, "Ultrase: single-channel speech enhancement using ultrasound," in *Proceedings of ACM MobiCom*, 2021.

[27] Y. Gao, Y. Jin, J. Chauhan, S. Choi, J. Li, and Z. Jin, "Voice in ear: Spoofing-resistant and passphrase-independent body sound authentication," *Proceedings of IMWUT/Ubicomp*, vol. 5, no. 1, pp. 12:1–12:25, 2021.

[28] 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," *Proceedings of IMWUT/Ubicomp*, vol. 5, no. 1, pp. 1–27, 2021.

[29] 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 *Proceedings of IEEE INFOCOM*, 2022.

[30] Z. Wang, Y. Ren, Y. Chen, and J. Yang, "Toothsonic: Earable authentication via acoustic toothprint," *Proceedings of IMWUT/Ubicomp*, vol. 6, no. 2, pp. 78:1–78:24, 2022.

[31] C. Min, A. Mathur, and F. Kawsar, "Audio-kinetic model for automatic dietary monitoring with earable devices," in *Proceedings of ACM MobiSys*, 2018.

[32] X. Li, S. Liu, Z. Zhou, B. Guo, Y. Xu, and Z. Yu, "Echopfl: Asynchronous personalized federated learning on mobile devices with on-demand staleness control," *Proceedings of IMWUT/Ubicomp*, vol. 8, no. 1, pp. 41:1–41:22.

[33] S. Liu, X. Li, Z. Zhou, B. Guo, M. Zhang, H. Shen, and Z. Yu, "Adaenlight: Energy-aware low-light video stream enhancement on mobile devices," *Proceedings of IMWUT/Ubicomp*, vol. 6, no. 4, pp. 172:1–172:26, 2022.

[34] S. Liu, B. Guo, C. Fang, Z. Wang, S. Luo, Z. Zhou, and Z. Yu, "Enabling resource-efficient aiot system with cross-level optimization: A survey," *IEEE Commun. Surv. Tutorials*, vol. 26, no. 1, pp. 389–427, 2024.

[35] R. Liu, C. Cornelius, R. Rawassizadeh, R. A. Peterson, and D. Kotz, "Vocal resonance: Using internal body voice for wearable authentication," *Proceedings of IMWUT/Ubicomp*, vol. 2, no. 1, pp. 19:1–19:23, 2018.

[36] K. Li, R. Zhang, B. Liang, F. Guimbretière, and C. Zhang, "Eario: A low-power acoustic sensing earable for continuously tracking detailed facial movements," *Proceedings of IMWUT/Ubicomp*, vol. 6, no. 2, pp. 62:1–62:24, 2022.

[37] S. Mahto, T. Arakawa, and T. Koshinaka, "Ear acoustic biometrics using inaudible signals and its application to continuous user authentication," in *Proceedings of IEEE EUSIPCO*, pp. 1407–1411, 2018.

[38] M. Yasuhara, I. Nambu, and S. Yano, "Bilateral ear acoustic authentication: A biometric authentication system using both ears and a special earphone," *Appl. Sci.*, vol. 12, no. 6, 2022.

[39] J. Tan, X. Wang, C. Nguyen, and Y. Shi, "Silentkey: A new authentication framework through ultrasonic-based lip reading," *Proceedings of IMWUT/Ubicomp*, 2018.

[40] L. Lu, J. Yu, Y. Chen, and Y. Wang, "Vocallock: Sensing vocal tract for passphrase-independent user authentication leveraging acoustic signals on smartphones," *Proceedings of IMWUT/Ubicomp*, vol. 4, no. 2, pp. 51:1–51:24, 2020.

[41] Q. Yang and Y. Zheng, "Deepear: Sound localization with binaural microphones," *IEEE Trans. Mob. Comput.*, vol. 23, no. 1, pp. 359–375, 2024.

[42] T. Arakawa, T. Koshinaka, S. Yano, H. Irisawa, R. Miyahara, and H. Imaoka, "Fast and accurate personal authentication using ear acoustics," in *Proceedings of IEEE APSIPA*, 2016.

[43] Y. Gao, Y. Jin, J. Li, S. Choi, and Z. Jin, "Echowhisper: Exploring an acoustic-based silent speech interface for smartphone users," *Proceedings of IMWUT/Ubicomp*, 2020.

[44] J. Tan, C. Nguyen, and X. Wang, "Silenttalk: Lip reading through ultrasonic sensing on mobile phones," in *Proceedings of Annual IEEE INFOCOM*, 2017.

[45] Y. Zhang, W. Huang, C. Yang, W. Wang, Y. Chen, C. You, D. Huang, G. Xue, and J. Yu, "Endophasia: Utilizing acoustic-based imaging for issuing contact-free silent speech commands," *Proceedings of IMWUT/Ubicomp*, 2020.

[46] L. Lu, J. Yu, Y. Chen, H. Liu, Y. Zhu, Y. Liu, and M. Li, "Lippass: Lip reading-based user authentication on smartphones leveraging acoustic signals," in *Proceedings of Annual IEEE INFOCOM*, 2018.

[47] Z. Shi, C. Li, Z. Jin, W. Sun, X. Ji, and W. Xu, "Anti-replay: A fast and lightweight voice replay attack detection system," in *Proceedings of IEEE ICPADS*, 2021.

[48] D. Duan, Z. Sun, T. Ni, S. Li, X. Jia, W. Xu, and T. Li, "F^{2}key: Dynamically converting your face into a private key based on COTS headphones for reliable voice interaction," in *Proceedings of ACM MOBISYS*, 2024.

[49] H. Deng and J. Yang, "Modeling and estimating acoustic transfer functions of external ears with or without headphones," *J. Acoust Soc. Am.*, vol. 138, no. 2, pp. 694–707, 2015.

[50] J. Li, K. Fawaz, and Y. Kim, "Velody: Nonlinear vibration challenge-response for resilient user authentication," in *Proceedings of ACM CCS*, 2019.

[51] M. Stinson, "The spatial distribution of sound pressure within scaled replicas of the human ear canal," *The Journal of the Acoustical Society of America*, vol. 78, no. 5, pp. 596–602, 1985.

[52] J. Gong and G. Laput, "User identification using headphones - united states patent application 20220030345." https://www.freepatentsonline.com/y2022/0030345.html, 2022.

[53] X. Fan, D. Pearl, R. E. Howard, L. Shangguan, and T. Thormundsson, "APG: audioplethysmography for cardiac monitoring in hearables," in *Proceedings of ACM MobiCom*, 2023.

[54] L. Li, M. Liu, Y. Yao, F. Dang, Z. Cao, and Y. Liu, "Patronus: Preventing unauthorized speech recordings with support for selective unscrambling," in *Proceedings of ACM SenSys*, 2020.
[55] N. Roy, H. Hassanieh, and R. Roy Choudhury, "Backdoor: Making microphones hear inaudible sounds," in *Proceedings of ACM MobiSys*, 2017.
[56] K. Sun, C. Chen, and X. Zhang, "'Alexa, stop spying on me!': Speech privacy protection against voice assistants," in *Proceedings of ACM SenSys*, 2020.
[57] Y. He, J. Bian, X. Tong, Z. Qian, W. Zhu, X. Tian, and X. Wang, "Canceling inaudible voice commands against voice control systems," in *Proceedings of ACM MobiCom*, 2019.
[58] X. Fan, L. Shangguan, S. Rupavatharam, Y. Zhang, J. Xiong, Y. Ma, and R. Howard, "Headfi: bringing intelligence to all headphones," in *Proceedings of ACM MobiCom*, 2021.
[59] G.-C. Hsieh and J. Hung, "Phase-locked loop techniques. a survey," *IEEE Trans. Ind. Electron.*, vol. 43, no. 6, pp. 609–615, 1996.
[60] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Nokia Research Center, 2000.
[61] M. Pukkila, *Channel estimation modeling*. Cambridge University Press., 2005.
[62] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Wiley, 1949.
[63] N. Jain, S. Rathore, and S. K. Singh, "Designing and evaluation of the reduced order iir filter design for signal de-noising," in *Proceedings of IEEE CSNT*, 2021.
[64] S. Nakagawa, L. Wang, and S. Ohtsuka, "Speaker identification and verification by combining mfcc and phase information," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1085–1095, 2012.
[65] B. Desplanques, J. Thienpondt, and K. Demuynck, "ECAPA-TDNN: emphasized channel attention, propagation and aggregation in TDNN based speaker verification," in *Proceedings of ISCA INTERSPEECH*, 2020.
[66] D. Sharma and I. Ali, "A modified MFCC feature extraction technique for robust speaker recognition," in *Proceedings of IEEE ICACCI*, 2015.
[67] M. Wortsman, G. Ilharco, J. W. Kim, M. Li, S. Kornblith, R. Roelofs, R. G. Lopes, H. Hajishirzi, A. Farhadi, H. Namkoong, *et al.*, "Robust fine-tuning of zero-shot models," in *Proceedings of IEEE/CVF CVPR*, 2022.
[68] A. Plesa, "Sony wh-1000xm5 controls." https://www.headphonesty.com/2023/02/sony-wh-1000xm5-controls/, 2023.
[69] F. A. Duck, "Medical and non-medical protection standards for ultrasound and infrasound," *Progress in biophysics and molecular biology*, vol. 93, no. 1-3, pp. 176–191, 2007.
[70] K. Slade, C. J. Plack, and H. E. Nuttall, "The effects of age-related hearing loss on the brain and cognitive function," *Trends in Neurosciences*, vol. 43, no. 10, pp. 810–821, 2020.
[71] International Telecommunication Union, "Recommendation bs.1387-0 (12/98) - method for objective measurements of perceived audio quality." https://www.itu.int/rec/R-REC-BS.1387-0-199812-S/en, 1998.
[72] Bouy Health, "Warm or fluid sensation in the ear symptoms, causes statistics." https://www.buoyhealth.com/learn/warm-or-fluid-sensation-ear, 2023.
[73] Z. Li, H. Li, and L. Meng, "Model compression for deep neural networks: A survey," *Comput.*, vol. 12, no. 3, p. 60, 2023.