EarEcho: Using Ear Canal Echo for Wearable Authentication

YANG GAO, University at Buffalo, State University of New York, USA
WEI WANG, University at Buffalo, State University of New York, USA
VIR V. PHOHA, Syracuse University, USA
WEI SUN, University at Buffalo, State University of New York, USA
ZHANPENG JIN, University at Buffalo, State University of New York, USA

Smart wearable devices have recently become one of the major technological trends and been widely adopted by the general public. Wireless earphones, in particular, have seen a skyrocketing growth due to its great usability and convenience. With the goal of seeking a more unobtrusive wearable authentication method that the users can easily use and conveniently access, in this study we present EarEcho as a novel, affordable, user-friendly biometric authentication solution. EarEcho takes advantages of the unique physical and geometrical characteristics of human ear canal and assesses the content-free acoustic features of in-ear sound waves for user authentication in a wearable and mobile manner. We implemented the proposed EarEcho on a proof-of-concept prototype and tested it among 20 subjects under diverse application scenarios. We can achieve a recall of 94.19% and precision of 95.16% for one-time authentication, while a recall of 97.55% and precision of 97.57% for continuous authentication. EarEcho has demonstrated its stability over time and robustness to cope with the uncertainties on the varying background noises, body motions, and sound pressure levels.

CCS Concepts: • **Security and privacy → Authentication**; **Biometrics**; • **Human-centered computing → Ubiquitous and mobile devices**.

Additional Key Words and Phrases: Acoustic, echo, ear canal, biometric, authentication, wearable devices

ACM Reference Format:
Yang Gao, Wei Wang, Vir V. Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using Ear Canal Echo for Wearable Authentication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 81 (September 2019), 24 pages. https://doi.org/10.1145/3351239

1 INTRODUCTION

Biometrics using traditional human physiological and behavioral characteristics, like fingerprints, faces, voices, and touch, have been widely applied on state-of-the-art mobile devices for user identification and authentication. Recently, as the advances of wearable technologies, a wide variety of smart wearable devices have been proposed and adopted by the general public, such as smart watches, smart wristbands, and wireless earphones. Looking

Authors' addresses: Yang Gao, University at Buffalo, State University of New York, Department of Computer Science and Engineering, Buffalo, NY, 14260, USA, ygao36@buffalo.edu; Wei Wang, University at Buffalo, State University of New York, Department of Computer Science and Engineering, Buffalo, NY, 14260, USA; Vir V. Phoha, Syracuse University, Department of Electrical Engineering and Computer Science, Syracuse, NY, 13244, USA; Wei Sun, University at Buffalo, State University of New York, Department of Communicative Disorders and Sciences, Buffalo, NY, 14214, USA; Zhanpeng Jin, University at Buffalo, State University of New York, Department of Computer Science and Engineering, Buffalo, NY, 14260, USA.

2474-9567/2019/9-ART81 $15.00
https://doi.org/10.1145/3351239

forward, this new trend is expected to continue and may also bring new opportunities of more convenient and secure authentication solutions.

Fingerprint is probably the most widely used biometric modality given its great usability and accuracy. However, it has also been criticized for its vulnerability to spoofing attacks [55]. Voice is another biometric modality which has low hardware requirement and good usability, but meanwhile it cannot achieve a high level of accuracy and robustness as a sole authentication method. Voice is also highly susceptible to noise. Recently, face recognition has emerged as a more popular way to verify the user's identity on smartphones, such as the "FaceID" by Apple [4]. Although multi-level sensors (including a dot projector, a flood illuminator, and an infrared depth camera) have been integrated to enhance the security of FaceID, it has been proven that FaceID is still not safe enough and the depth and IR camera could be possibly deceived [10].

In addition to those popular biometric authentication approaches, various bio-electrical signals of human body (e.g., electrocardiograph (ECG) and electroencephalograph (EEG)) have also been explored as biometric identifiers for authentication purpose. A smart wristband named "Nymi Band" that used ECGs to authenticate user identity and any conceivable device has been used for wearable credit card payment [1]. With the increasing popularity of Virtual Reality (VR), researchers have integrated EEG electrodes into the VR headset to collect the user's brainwave signals as the continuous authentication credentials [37]. However, those solutions either require additional gestures to implement authentication or are not suitable for long-time wearing. Thus, we ask a question: *is there a more unobtrusive wearable authentication method or biometric modality that the users can easily use and conveniently access?* In the communities of audiology and clinical otolaryngology, the unique physiology and dimension of ear canal has been a widely accepted fact [66], as claimed in [30] that "*A real ear will also have a unique eardrum impedance, as well as unique ear canal dimension.*" This finding is even considered as the main obstruction and challenge to the diagnosis and treatment of ear diseases such as otitis externa [41]. In addition, taking advantage of the recent growth of voice assistants and extended battery capacity, among those smart wearable devices, wireless earphones (especially those true wireless earbuds with smaller form factors) have seen a skyrocketing growth in the consumer electronics market recently, which is probably faster than anything else [57]. It was reported that [54] the global earphones and headphones market is expected to grow at a CAGR of 7.31% during 2017-2023 with 20 billion dollars in revenue by 2023. Thus, the increasing popularity and pervasiveness of wireless earphones brings a potential entrance of human-computer interface, and moreover, a new modality of user authentication.

In this paper, we focus on the wireless earphones which have grown dramatically in recent years, and propose a novel user authentication system — *EarEcho* — that packs a small microphone into the earphone (containing the original earpiece speaker). The design has low requirements of form factor and cost, and thus can be easily deployed on most existing earphone products. In general, *EarEcho* extracts the unique features by comparing the original sound emitted from the earpiece speaker and the sound propagating, reflected, and absorbed through the ear canal which can be recorded by the built-in microphone. The comparison and authentication will take place on the mobile device with which the earphones are connected through Bluetooth. With the popularization of wireless earphones, more and more users are getting used to wearing earphones while working, studying or strolling. As shown in Fig. 1, existing popular mobile and wearable authentication solutions (e.g., Face IDs, voiceprints, or fingerprints) usually have two major limitations in terms of demanding active interactions with devices and being easy to be stolen or spoofed. Those biometric traits (e.g., faces, voices, fingers) are widely exposed to the public and the Internet, which increases their vulnerability to spoofing threats. To provide a more secure and usable authentication solution, we propose the EarEcho that enables a passive authentication channel while a user is wearing the earphones. Given an acoustic stimulus, our EarEcho captures the uniqueness of the user's ear canal morphology, and the entire authentication process occurs in the user's auditory canal which is relatively isolated and concealed. Compared with face IDs, fingerprints and voiceprints, the EarEcho presents a more unobtrusive authentication approach with great usability potentials. For example, *EarEcho* could be used as

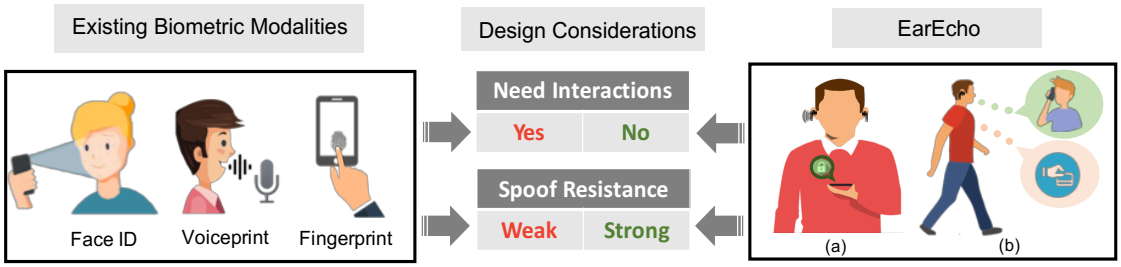


Fig. 1. Application scenarios of *EarEcho*. For example, (a) Once the user is wearing the earphones, *EarEcho* can perform authentication without requiring the user to making additional tasks (e.g., pressing the thumb, or facing to the camera). (b) Only by listening, a user can verify his identity during the entire phone call (i.e., bank call preventing from the fraud), or verify the mobile payment through voice assistants (e.g., Siri, Google, Alex) without worrying privacy information leakage.

a potential ubiquitous authentication method including unlocking personal mobile devices (i.e., smartphones), authorizing over-the-phone payments, and verifying identities during sensitive remote conversations.

Compared with existing mobile and wearable authentication solutions (i.e., fingerprint, faceID, voiceprint), *EarEcho* possesses the following advantages. (1) *Ubiquitous*: Microphones and earpiece speakers are two basic components for wireless earphone designs with small form factor and low cost. (2) *User-friendly*: *EarEcho* doesn't require active authentication operations (i.e., facing the front camera or pressing fingertips on the fingerprint sensor). It is capable of automatically processing the authentication requests while the user is wearing the earphones. (3) *Unobtrusive*: Spoofing is one of the major concerns for fingerprint and face based mobile authentications due to the low-effort illegitimate acquisition of users' information. However, when it comes to *EarEcho*, because the data collection and authentication process is conducted in a more unobtrusive and unnoticeable manner, it is very hard for malicious attackers to steal the enrolled user's ear (physical and physiological) information through side-channel attacks. (4) *Reliable*: *EarEcho* has demonstrated its stability over time and robustness to cope with the uncertainties on the varying background noises, body motions, and sound pressure levels.

Specifically, we make the following contributions in this work:

- We develop the acoustic signal processing techniques for universal noise interference cancellation of echoes through the ear canal.
- We propose an end-to-end system framework, including a context-free acoustic feature generative model by using transfer functions between the emitted and reflected signals, and the final authentication model.
- We design the prototype that packs a commercial earphone with a microphone, and perform extensive experimental evaluations about the system robustness under diverse application scenarios.

To the best of our knowledge, *EarEcho* is the first to leverage the unique ear canal geometry and acoustic features propagating in the ear with context-free audible signals for mobile and wearable authentication. It has been demonstrated that *EarEcho* has robust performance without any specific usage requirement and additional sensor, and possesses superior advantages of effectiveness, unobtrusiveness, ease-of-use, and cost-effectiveness.

The rest of this paper is organized as follows. In Section 2, we review the state of the arts of existing wearable and acoustic-based authentication solutions. We present our application scenarios and design considerations in Section 3. In Sections 4 and 5, we elaborate the design overview and rationale of *EarEcho*. Then we describe our implementation of an earphone prototype in Section 6, and present the performance evaluation of our solution with human subjects in Section 7. We discuss the limitations and conclusions in Sections 8 and 9.

2 RELATED WORK

2.1 Mobile and Wearable authentication

Existing authentication solutions on mobile and wearable devices can be generally divided into three major categories: knowledge, physiological characteristics, and behavioral characteristics.

Knowledge. Personal Identification Numbers (PINs) and graphical passwords/patterns are the most traditional and still the primary solutions for mobile authentication. Despite the simplicity and popularization, the vulnerability to eavesdropping makes PINs and patterns the most unsafe identifiers. Some researchers have also explored the vulnerability of credential information leakage from side-channel attacks including vibration, acoustics, thermal information, and even Wi-Fi signals. TapPrints [47] sensed the letters typed on the smartphone with QWETRY keyboard using the built-in accelerometer and gyroscope. PatternListener [72] cracked the user's smartphone locking pattern by analyzing signals recorded from the built-in speaker and microphone. Abdelrahman *et al.* [2] leveraged the thermal camera to capture thermal residues on the touchscreen to infer the most possible PINs and patterns pressed on mobile devices. Li *et al.* [36] proposed a new potential threat by eavesdropping user's passwords using Channel State Information (CSI) of Wi-Fi signals transmitted from the user's smartphone.

Physiological Characteristics. Compared with the traditional knowledge-based solution, physiological biometrics are more secure and user-friendly and thus become the new trend for mobile and wearable authentication such as fingerprints [53, 58], iris [65], face [17, 23], ear shape [13, 52], Photoplethysmography (PPG) [24, 51], ECG [19, 28], and EEG [22, 37]. Fingerprint sensors can achieve very high precision and are prevailing on many commercial smartphones, but fingerprint's static characteristics make sensors easy to be spoofed. Given an optical fingerprint dataset, a malicious user is able to generate MasterPrint [55] which is a synthetic or real partial fingerprint that serendipitously matches one or more of the enrolled templates for a significant number of users. Arteaga-Falconi *et al.* [5] proposed an ECG-based authentication method with a two-electrodes sensor attached at the back of smartphones. Wang *et al.* [68] proposed to use the smartphone to capture the user's chest vibration corresponding to each heartbeat as a biometric characteristic for the mobile authentication. Zou *et al.* [73] designed the BiLock which extracted the dental occlusion biometric while user performs an occlusion gesture using the smartphone.

Behavioral Characteristics. Instead of extracting subject's physiological uniqueness, behavioral biometrics can represent the identity of an individual through a sequence of the subject's activities which may include keystrokes, finger gestures on the touchscreen [14, 59], voice [18], breath [27], and gestures [35]. Liu *et al.* [38] proposed a robust and multi-expert based gesture recognition for mobile authentication. VibWrite [39] provided a low-cost and tangible finger-input authentication solution which leveraged the unique physical vibration on any solid surface for smart access systems. Hoang *et al.* [26] presented a gait-based authentication system on mobile devices by analyzing the smartphone's accelerometer data and employing a fuzzy commitment scheme. BreathPrint [12] caught the public attention by capturing the acoustic features hidden in three daily breath behaviors including sniff, normal breath, and deep breath to verify the legitimate user.

2.2 Sensing on Mobile and Wearable Devices

2.2.1 Acoustic Sensing on Smartphones. Acoustic sensing has been mostly used on mobile and wearable devices as a solution for distance approximation and floor plans estimation [70]. Mao *et al.* [45] developed an acoustic imaging system using the speaker and microphone in the smartphone. To mimic the Synthetic Aperture Radar (SAR), they moved a smartphone with a pre-defined trajectory around the object to obtain the acoustic image. Even though many studies have been conducted to leverage the built-in microphone and speaker in the mobile device as a sensing technique, only a few works investigated the potential of authentication. EchoPrint [71] utilized the acoustic and vision sensors on a commodity smartphone to verify the user's face with a well-designed acoustic emitted signal from the earpiece speaker. Compared with commercial products (e.g., faceID), this technique had

the strict requirement for face alignment, and as the authentication was exposed by the public environment, it may get affected by the strong background noise. Lu *et al.* [42] proposed a lip-reading based mobile authentication solution, they utilized pre-defined ultrasonic acoustic signal to capture the unique Doppler profiles caused by user's mouth movement when the user is speaking the same passphrase. Machto et al. [43] leveraged a pre-tuned probe signal within the audible frequency range and ultrasound to authenticate the user's identity using microphones and ultrasound speakers built in earphones. However, since ultrasound is inaudible to human, users might have potential safety risk if being exposed to the airborne ultrasound with 120 dB or higher in a unconscious manner [33].

2.2.2 In-Ear Sensing. Recent advances in wearable devices and IoTs have brought great potentials [31] to monitor and sense human physiological functions (e.g., PPG, EEG) [20, 67] and behaviors (e.g., jaw movements, eating habits, eye gestures, facial expressions) [8, 44, 46, 48] out of the clinic and in people's daily lives. Park *et al.* [50] designed a Piezoelectric sensor measuring the pressure variances of the ear canal surface, and those variances can be used to estimate the heart rates. Bi *et al.* [9] developed "Auracle", a wearable earpiece that can recognize a user's eating behaviors by capturing and analyzing the sound patterns of chewing through the bone and tissue of the head. Laput *et al.* [32] proposed the "SweepSense" that used the reflected swept-frequency ultrasonic to detect whether the earphone was in ear or not. Goverdovsky *et al.* [21] proposed the "Hearables," an earpiece with multimodal sensors including EEG electrodes and a microphone, which can measure the user's brain, cardiac and respiratory activities.

3 THREAT MODEL

3.1 Attack Scenarios

In this section, we briefly introduce two major attack scenarios for off-the-shelf mobile and wearable authentication schemes.

Side-channel Attacks. PINs or lock patterns could be easily stolen through eavesdropping [64]. As the existing knowledge-based authentication process is mostly conducted in an open space (i.e., typing passwords on the touchscreen), the malicious attackers have a higher chance to eavesdrop those credential information by various side-channel attacks (e.g., vibration, acoustic, thermal information, and wireless signals).

Replay Attacks. Some biometric authentication identifiers (e.g., fingerprint, face, voice, ECG) might be too complex to be directly inferred by side-channel attacks, there still are possibilities of information leakage under other occasions that hackers can utilize to implement replay attacks. For example, researchers claimed that they can spoof the wearable ECG authentication system by linking the ECG captured by the wearable sensor and the ECG templates stored in the hospital even if they were recorded from different body locations [15]. Fingerprints were also proved to be not safe. German defence minister's fingerprints were faked by hackers with only a few high-definition photographs of her hand [25].

3.2 Design Considerations

To design *EarEcho* as a qualified authentication identifier, we make the following considerations.

Ubiquitous. Considering the wireless earphones are more prevailing on the market due to the small form factor and low cost, we propose to pack a small and cheap microphone into an off-the-shelf earphone, so that it can be easily deployed by large manufacturers with minimum hardware costs.

Unique. Ear could be used as a passive biometric modality according to some recent studies [6]. It is also claimed that ear is more stable than face because it is less affected by users' emotions and ages [16]. Some existing image-based authentication systems [7, 56] captured the shape of the outer ear and took advantage of the uniqueness of the geometry for authentication purpose. However, this technique might not be effective for

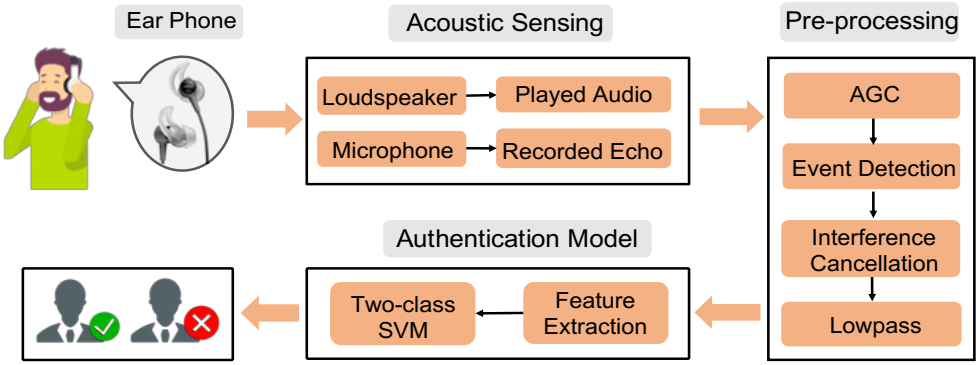


Fig. 2. The methodological flow of *EarEcho* authentication system. *EarEcho* uses the transfer function between the played audio and the recorded echo to extract acoustic features. An SVM classifier is trained to verify a registered user.

mobile authentication scenarios because the performance could be largely affected by the alignment and quality of the captured ear pictures.

Unobtrusive. To relieve the risk of information leakage and guarantee the authentication in a user-friendly manner, a qualified biometric trait should be processed during the daily use (e.g., user is making the phone call, listening to the radio or music) without specific interruptions. Thus, we choose to leverage the echos that travel through the ear canal with the context-free input audio signals to represent the in-ear morphological uniqueness.

Robust. Robustness is an indispensable design factor for any reliable authentication systems. The existing acoustic-based mobile authentication approaches, such as EchoPrint and BreathPrint, were exposed in an open environment and thus easier to be affected by surrounding noises. Our *EarEcho* utilizes the acoustic propagation in a near hermetic space formed by the eardrum, ear canal, and the accessorial silicone tips of the earphones which can largely isolate the environmental noises.

4 OVERVIEW

EarEcho aims at verifying the identity of the user while wearing the earphones. It utilizes the earpiece speaker and microphone in the earphone for acoustic sensing. It extracts acoustic features from the played audio file and the recorded echos using transfer function estimation. Fig. 2 shows the overview of the system design, which consists of three major components: acoustic sensing, signal pre-processing, and user authentication.

In the acoustic sensing phase, the built-in microphone captures the sound emitted by the earpiece speaker that propagates through the user's ear canal.

In the pre-processing phase, the actual output sound from the earpiece speaker is estimated by the Adaptive Gain Control (AGC) module based on the frequency selectivity of target earpiece speakers. *EarEcho* detects the high power-density acoustic activity and filters out undesired noisy segments. The detected audio segments are fed into the interference cancellation process to reduce the influence of direct-path propagation from the earpiece speaker to the microphone within the earphone cavity. In addition, a low-pass filter is designed to remove the high frequency (>6 kHz) noises.

In the authentication model phase, the transfer function based features are extracted based on the estimated output sound and the noise-removed echo. Then a two-class SVM is trained to distinguish between legitimate users and unauthorized users. During the authentication, the user doesn't need to do any additional action and just keeps listening to the earphones as usual. The acoustic features are extracted from the played audios and the recorded echos, and fed into the trained SVM classifier for final authentication.

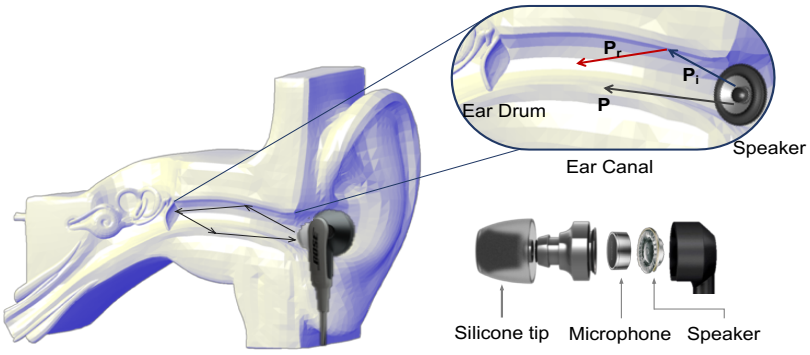


Fig. 3. An illustration of the simplified sound wave modelling for the acoustic propagation and reflection in the ear canal with the proposed earphone design. P is the direct propagation of the sound wave, P_i denotes the incident sound wave, P_r represents the reflected sound wave.

5 ECHO SENSING

Echoes propagating through the ear canal are highly unique: i) the echoes are sensitive to the relative emitting direction of the sound source. The varieties of people's external acoustic meatuses or concha sizes would result in a user-preferred earphone wearing behavior for each individual, which might affect the relative position of the earphone in the ear canal. ii) Each ear canal is a unique closed space consisting of many different reflection and absorption surfaces that make echoes in the ear more distinctive [62].

5.1 Acoustic Sensing Modeling

Sound emitted by the earpiece speaker propagates within the ear which is reflected and attenuated by the canal wall. Essentially, given the wave equation describes the propagation of acoustic waves through material medium, which is defined as:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 p \tag{1}$$

where ∇^2 is the spatial Laplacian, p is the acoustic pressure, and c is the wave speed.

We model sound waves travelling within the ear canal as two major behaviors: propagation and reflection/absorption, as shown in Fig. 3. For the direct propagation, let p be an acoustic scalar function $p = p(x, y, z; t)$ in three space dimensions, so we get:

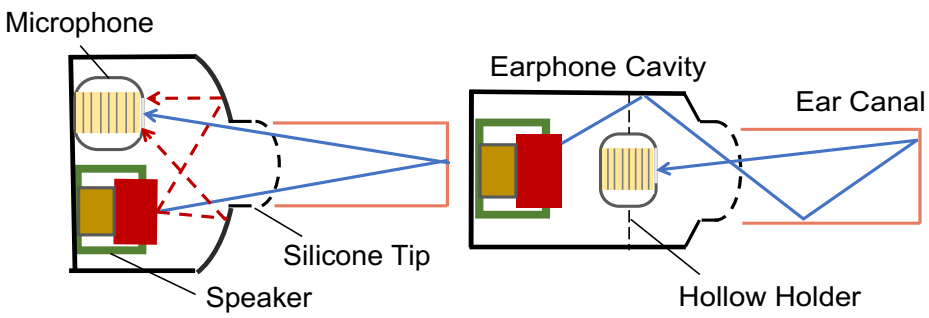
$$\frac{\partial^2 p}{\partial t^2} = c^2 \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \tag{2}$$

In a small space such as the ear canal, we can consider the air medium is isotropic, and c is a constant which equals to 343 m/s in the 20°C environment. For the reflection/absorption, it may significantly vary given the specific texture of the surface (i.e., ear canal wall). Some energy will be absorbed, and some energy will be scattered or reflected. Thus, we have:

$$\frac{p_r}{p_i} = |r_p| e^{j\sigma\pi} \tag{3}$$

where $|r_p|$ is the sound reflection coefficient with respect to the surface material and roughness, σ represents the phase difference between P_i and P_r.

In general, we use $C = \{c_1, c_2, ..., c_k\}$ to denote the user's ear canal characteristics, and c_i represents each independent attribute such as the auditory canal geometry size, surface property, earphone relative placement.



(a) (b)

Fig. 4. Two layout designs of the earphone prototype; (a) paralleled placement of microphone and earpiece speaker, and (b) sequential placement of microphone and earpiece speaker. The blue line indicates echoes containing ear canal information, and the red lines represent interference acoustic from the earpiece speaker without propagation through ear canal.

Let $O = \{o_1, o_2, ..., o_m\}$ represent the set containing emitted sound rays from the earpiece speaker, and $g(.)$ is the general transfer function from the earpiece speaker to the microphone, including wave propagation and reflection. Thus, we have the recorded acoustic set P defined as:

$$P = \{p_1 \leftarrow g_1(o_1, C), p_2 \leftarrow g_2(o_2, C), ..., p_m \leftarrow g_m(o_m, C)\} \tag{4}$$

5.2 Earpiece Speaker and Microphone Design

Typically, the average diameter of adult human ear canal is about 0.75 *cm* [60], an off-the-shelf microphone diameter is around 0.60 *cm* and the diameter of earpiece speaker in the earphone is around 0.3 to 0.5 *cm*. It is hard to directly fit both microphone and earpiece speaker in the cross-section area of ear canal without customization. Fig. 4 shows two different layout designs for the commercial microphone and earpiece speaker built in the earphone. Fig. 4(a) describes a paralleled placement for microphone and earpiece speaker in the earphone cavity. As the regular earphone speaker is not directional, the emitted sound waves will not only travel trough the silicone tip and the ear (as shown by those blue lines), but also be directly reflected by the cavity wall and recorded by the microphone (as indicated by the red dash lines). Those reflected sound waves will be mixed with the echoes containing the user's ear canal information and captured by the microphone, which would largely affect the echo quality.

In this paper, we adopt the second layout design as shown in Fig. 4(b), the microphone and earpiece speaker are sequentially located, facing the ear canal. The emitted sound waves from the loudspeaker will pass through the empty space (hollow holder) around the microphone, propagate in the ear canal, and be reflected back. This will significantly reduce the interference noises caused by massive cavity reflections. Even though this layout (i.e., the microphone is placed in front of the loudspeaker.) will slightly block and attenuate the sound from the speaker, it can provide the user with a better wearing experience and the echo's quality.

5.3 Feasibility Analysis

Given the theoretical modeling of the acoustic propagation in the ear canal, in this section, we will verify the uniqueness of users' ear canals by analyzing recorded echoes. We test two types of sound stimuli: single tone and conversation.

Single Tone. Fig. 5(a) shows the result when we use a 200 Hz sinusoidal tone as the input stimulus and test among 4 different human subjects, which validates the uniqueness of echoes corresponding to users' ear canal information. We asked four subjects to wear the same earphone with comfort, and play the same probe signal.

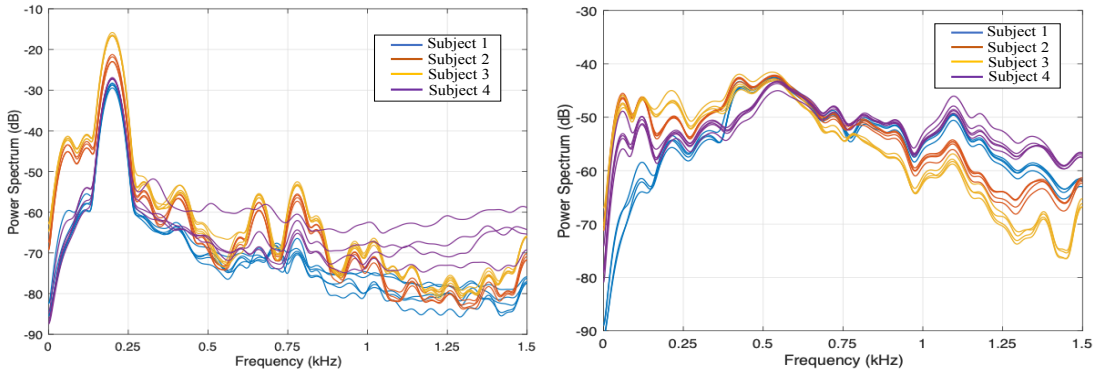


Fig. 5. The power spectrum (0-1500Hz) of five trails from four users under two sound stimuli. (a) Input stimulus: a 5-second single sinusoidal tone at 200 Hz; (b) Input stimulus: a 10-second conversation.

Then we extracted the power spectrum over the frequency domain from the recorded echoes. We repeated the process for 5 times and calculated the corresponding power spectrum. It is observed that different subjects' spectrums show some certain levels of difference around 200 Hz. However, a stimulus of one single frequency can't provide sufficiently distinguishable uniqueness among all subjects (e.g., subject 1 and subject 4). Hence, a stimulus containing more frequency information (e.g., conversation) is needed to elicit more clear distinctions of the ear canal echo. Note that the patterns above 400 Hz are caused by several high-frequency harmonics and noises from the microphone.

Conversation. To further validate our hypothesis about the uniqueness of in-ear acoustic characteristics and improve the stimulus's frequency information, we also evaluated using a short period of conversation that was a mix of numerous tones including voices, laughs, and background noises. As shown in Fig. 5(b), over the dominant frequency range of the conversation stimulus (50 Hz - 550 Hz) and the harmonics high-frequency range (1000 Hz - 1500 Hz), different subjects' echo spectrums clearly demonstrate observable distinctions.

6 METHODS

6.1 Acoustic Pre-processing

6.1.1 Adaptive Gain Control. It is known that, given a certain volume level, the frequency response of output sound is controlled and amplified by the Automatic Gain Control (AGC) module in smart devices(e.g., smartphones, smart speakers). However, those smart devices are mostly frequency selective [34], which means that different loudspeakers might have different frequency responses given the same audio stimulus and volume setting. To eliminate the front-end gain interference caused by hardware, we measure the frequency selectivity of our prototype earpiece speaker using a chirp signal. Based on the measured frequency response and the audio's volume, for each echo recording, we estimate the output sound from the earpiece speaker by compensating the digital audio file.

6.1.2 Event Detection. During the earphone's daily usage, there are always short intermissions during audio's playing. For example, music have short intervals of silence, or there are scattered pauses during a speech. To eliminate the effects of the low SNR segments and to increase the energy efficiency especially for the continuous authentication scenarios, we adopt a Likelihood Radio Test (LRT) and Hidden Markov Model (HMM)-based event

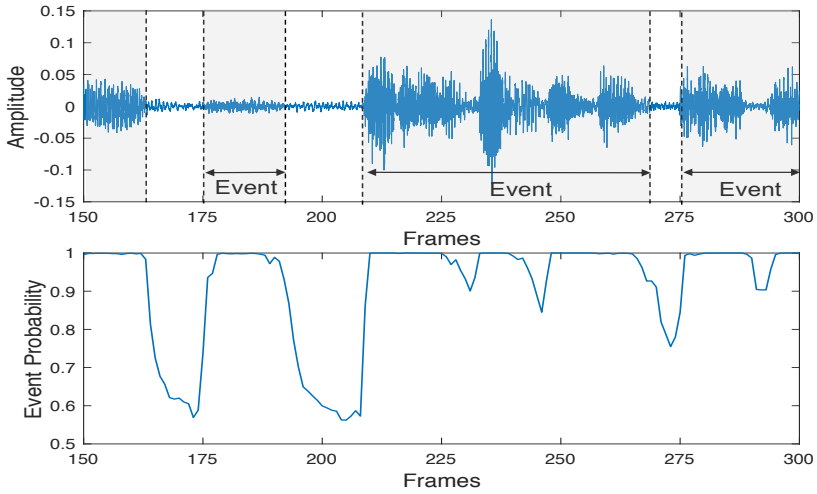


Fig. 6. Example of event detection in a recorded echo from a playback conversation and the corresponding event occurrence probability.

detection module to filter out undesired low-power density echo segments [61]. Given that:

$$\begin{aligned} H_0 &: \text{event absence} : X = N \\ H_1 &: \text{event presence} : X = N + S \end{aligned} \tag{5}$$

where S, N, X are Discrete Fourier Transform (DFT) coefficient vectors of audio, noise, and noisy audio, with their kth elements S_k, N_k, and X_k, respectively.

The probability density functions conditioned on H_0 and H_1 are given by:

$$\begin{aligned} p(X|H_0) &= \prod_{k=0}^{L-1} \frac{1}{\pi\lambda_N(k)} \exp\{-\frac{|X_k|^2}{\lambda_N(k)}\} \\ p(X|H_1) &= \prod_{k=1}^{L-1} \frac{1}{\pi[\lambda_N(k)+\lambda_S(k)]} \cdot \exp\{-\frac{|X_k|^2}{\lambda_N(k)+\lambda_S(k)}\} \end{aligned} \tag{6}$$

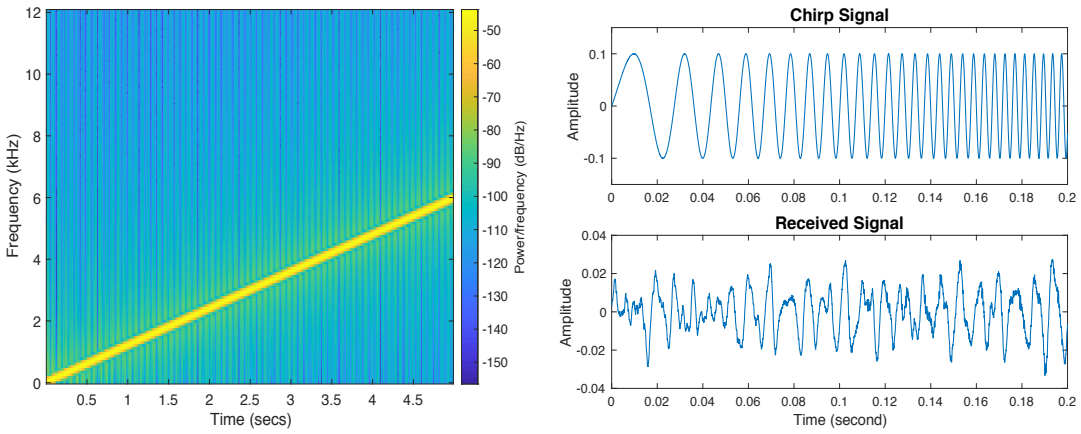
where $\lambda_N(k)$ and $\lambda_S(k)$ represent the variances of N_k and S_k. The likelihood ratio for the kth frequency band is

$$\Lambda_k \triangleq \frac{p(X_k|H_1)}{p(X_k|H_0)} = \frac{1}{1+\xi_k} \exp\{\frac{\gamma_k \xi_k}{1+\xi_k}\} \tag{7}$$

where ξ_k and γ_k are called prior and posterior Signal-to-Noise Ratios (SNR's). The decision rule is obtained from the average likelihood ratio for each band, which is given by

$$\log\Lambda = \frac{1}{L}\sum_{k=0}^{L-1} \log\Lambda_k \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{8}$$

As shown in Fig. 6, given a raw echo signal collected by the microphone, we will first calculate the event probability for every frame, and then filter out the noise segments with lower probability.



(a) Spectrogram of the chirp signal

(b) Chirp signal and received audio over time

Fig. 7. The chirp signal used for interference cancellation that sweeps over the mostly used audible frequency bands from 20 Hz to 6 kHz.

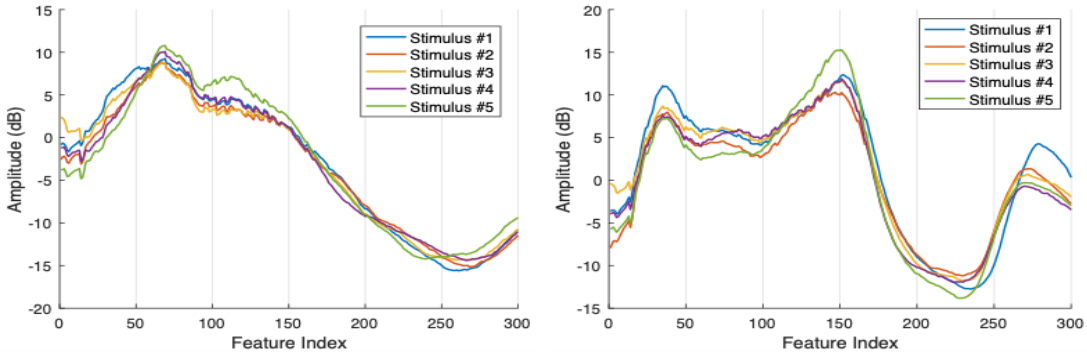
6.1.3 Interference Cancellation. This step is designed to remove interferences of the direct sound transmission between the earpiece speaker and the microphone. In our prototype design, the earpiece speaker is placed behind the microphone. Once the earpiece speaker plays the sound, some of the sound waves will directly propagate to the microphone and reach the membrane. There are also a small amount of echoes that are reflected only in the earphone cavity without traveling into the ear canal. To achieve a better interference cancellation, we record the direct transmission by putting the earphone in a clean space, where no major reflector is within one meter distance in front of the microphone. We design a chirp probe signal that ranges between 20 Hz to 6 kHz which covers most of the frequency range of human voice and music, as shown in Fig. 7. By analyzing the transfer function between the played probe audio and the received audio, we can obtain the features from the direct path interference. When we authenticate the user's identity, we estimate the interference noise of the played audio from the direct path and subtract it from the entire received echo.

6.1.4 Low-pass Filter. As the classical theorem of music and speech, the frequency covers from 20 Hz to the highest tone C8 at 4,186 Hz [3]. We thus design a Butterworth low-pass filter with stop frequency at 6 kHz and feed the interference cancelled signal into the low-pass filter to remove undesired high-frequency noises.

6.2 Authentication Model

6.2.1 Feature Extraction. As discussed in Section 5.1, the relationship between the emitted acoustic signal and the recorded echo signal is dominated by the geometry information of the user's ear canal, and can be modeled as a linear, time-invariant system. Given the input x as the estimated earpiece speaker sound and the output y as the noise-removed recorded echo, in the frequency domain, we have $Y(f) = H(f)X(f)$. For a single-input (i.e., the earpiece speaker) and single-output (i.e., the microphone) system, the H_1 estimate of the transfer function is given by

$$H_1(f) = \frac{P_{yx}(f)}{P_{xx}(f)} \tag{9}$$



(a) The averaged feature vector of user A (b) The averaged feature vector of user B

Fig. 8. The averaged feature vectors of two users given five different stimuli.

where P_{yx} and P_{xx} are the cross power spectral density of x and y respectively, and can be estimated using Welch's averaged, modified periodogram method [69] as follows:

$$P_{xy}(\omega) = \sum_{m=-\infty}^{\infty} R_{xy}(m)e^{-j\omega m} \tag{10}$$

And the cross-correlation sequence $R_{xy}(m)$ is defined as

$$R_{xy}(m) = E\{x_{n+m}y_n^*\} = E\{x_n y_{n-m}^*\} \tag{11}$$

where x_n and y_n are jointly stationary random processes, and $E\{.\}$ is the expected value operator. This estimate assumes that the noises including system noises and background noises are not correlated with the system input.

Based on the played acoustic signal (input) and the recorded echo (output), we extract the feature with 2,048 Discrete Fourier Transform (DFT) points and a Hann sliding window (window length 150 ms; overlap length 100 ms). And we obtain the feature set $G = \{H_1(f_1), H_1(f_2), ..., H_1(f_k)\}$ where $H_1(f_k)$ represents the transfer function estimate of the kth frequency band ($f_k \leq 6$ kHz). As shown in Fig. 8, given different input audio stimuli played by the speaker in earphones (i.e., five 2-minute long conversation episodes), it is observable that the extracted features are highly correlated and context-free for the same subject, and distinctive for different subjects.

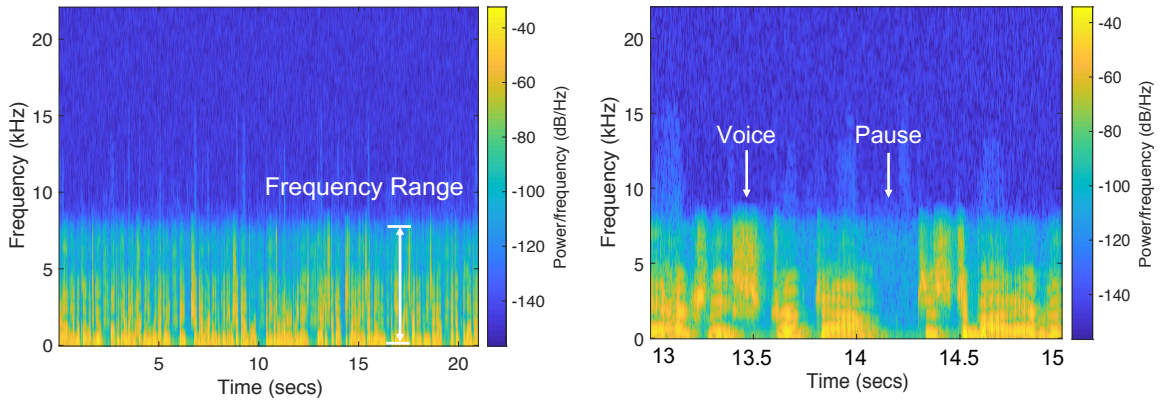
6.2.2 Classifier. The nature of mobile and wearable authentication is typically considered as a classification problem to distinguish legitimate and illegal users. We adopt a two-class SVM classifier with Radial Basis Function (RBF) kernel in our *EarEcho* design, due to its relatively high computational efficiency. The prohibitive computational complexity and costs make many advanced classifiers (e.g., deep neural networks) less beneficial for resource-constrained mobile devices. During the real-world usage, *EarEcho* needs to collect samples from the legitimate users combined with stored benchmark impostors' samples to train the SVM classifier.

7 EVALUATIONS

7.1 Experimental Setup

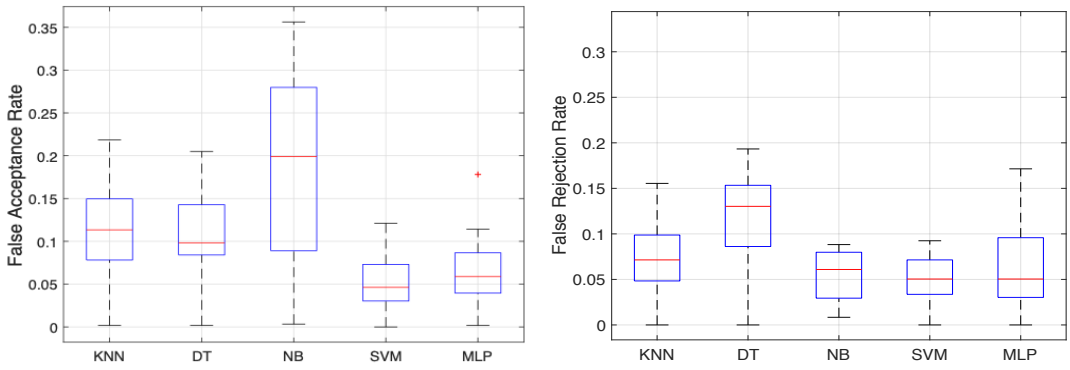
We implemented *EarEcho* on a earphone prototype to evaluate the uniqueness of acoustic behaviors in the ear canal.

Hardware. We embedded the MS-TFB microphone into the Bose SoundSport in-ear headphone, which has a high sensitivity of -32 dB and can capture a maximum sound pressure level of 115 dB. To ensure a good noise isolation and comfortable fit, we used the Bose silicone earbud tips with user-specific sizes. Also, we used a 3.5



(a) Spectrogram of the stimulus audio segment (b) A detailed segment of stimulus consisting of voices and pauses

Fig. 9. An example of the spectrogram of the audio signal used for evaluation stimulus with the major frequency range from 20 Hz to 6 kHz.



(a) False Acceptance Rate (b) False Rejection Rate

Fig. 10. Performance of different classifiers. The box plot shows the median, quartiles, and error ranges.

mm Jack cable adapter to merge the record channel and the play-back channel and connect it to the MacBook Pro laptop with a 2.5 GHz Intel i7 CPU and 16GB memory.

Stimuli. To ensure the diversity and representativeness of stimulus signals, we collected 5 different conversation audio records from a popular American podcast *"West Wing Weekly"*, with respect to content, speakers' genders, pitch, and speech speed, as shown in Fig. 9. It can emulate the real cases when the user is picking up a phone-call or is listening to the radio or music. Each audio trail lasted for 2 minutes and was sampled at a rate of 44,100 Hz for a full coverage of human audible frequency range (20 Hz to 20 kHz).

Participants. 20 subjects (age ranges 24-30 years old, 6 females and 14 males) were recruited in the experiments. During the data collection, each subject was asked to wear the earphone prototype and listen to those 5 audio records (each audio lasts for 2 minutes). Participants were asked to take out and put on the earphone between each audio record. In addition, to mimic the daily usage scenario in real environments, we also asked participants to perform different postures (e.g., sitting and standing) and body motions (e.g., mouse and head movements) during the experiments. Also, in order to ensure the environment diversity, the data was collected under various background noises (e.g., room, shopping mall, cafe, street). In total, we collected 11,900 samples (each sample was

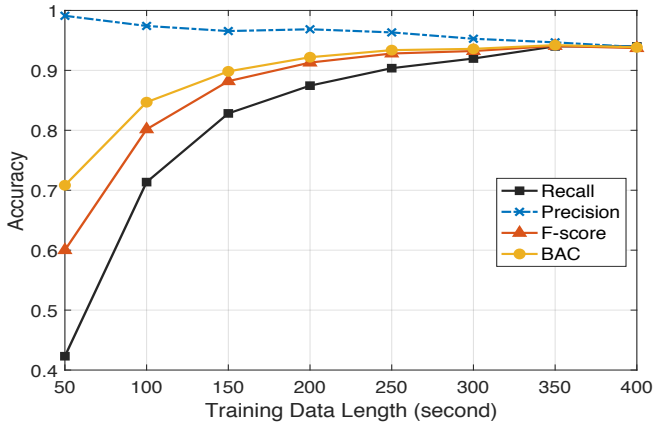


Fig. 11. Authentication performance for different time periods of collecting training data at 44,100 sampling rate.

Table 1. **Authentication accuracy of the legitimate user and intruders with 1-second authentication window size**

	Mean	Median	Standard Deviation
Precision (%)	95.16	97.56	5.78
Recall (%)	94.19	96.30	6.63
F-score (%)	94.46	95.35	4.53
BAC (%)	94.52	95.56	4.49

a one-second segment for 5 recordings with no overlap) from 20 subjects, and divided it into two parts, 80% for model training and 20% for evaluation.

7.2 Classifier Performance

To evaluate the performance of the SVM classifier with the transfer function based features, We compare different classifiers using the test dataset including K-Nearest Neighbor (KNN), Decision Tree (DT), Naive Bayesian (NB), Support Vector Machine (SVM) and Multi-layer Perceptron (MLP). Fig. 10 shows the detailed error rates for different classifiers. It is seen that SVM outperforms all other classifiers in both FAR and FRR, and also takes shorter time for training (1.3 seconds compared with 12.8 seconds for MLP which has similar accuracy).

7.3 Authentication Performance

7.3.1 Precision, Recall, F-score, and BAC. In an authentication problem, we introduce the precision, recall, F-score, and Balanced Accuracy (BAC) as the evaluation metrics. Given the true positive (TP), false negative (FN), and false positive (FP), We define the precision and recall as below:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

where a high precision means that only the authorized users can successfully pass the verification, and a high recall indicates that most legitimate users will not be rejected. Sometimes in an imbalanced testing set, precision and recall may not perform well. Thus, we also introduce F-score and BAC which overcome this problem, and

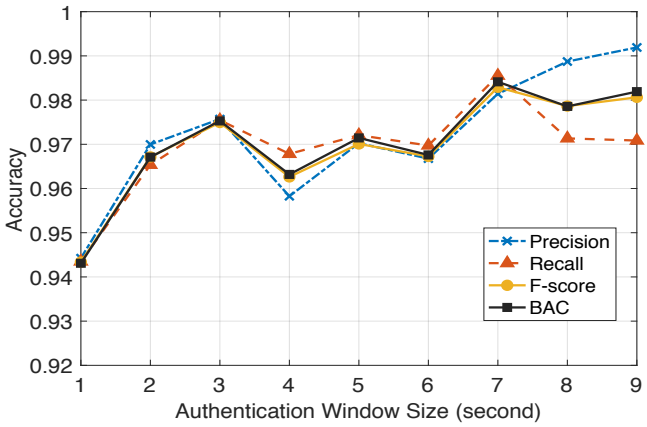


Fig. 12. Continuous authentication performance with different window sizes at 44,100 sampling rate.

Table 2. **Continuous authentication accuracy of the legitimate user and intruders with 3-second window size**

	Mean	Median	Standard Deviation
Precision (%)	97.57	100	3.96
Recall (%)	97.55	100	3.40
F-score (%)	97.49	97.50	2.75
BAC (%)	97.53	97.47	2.69

defined as:

$$\text{F-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{14}$$

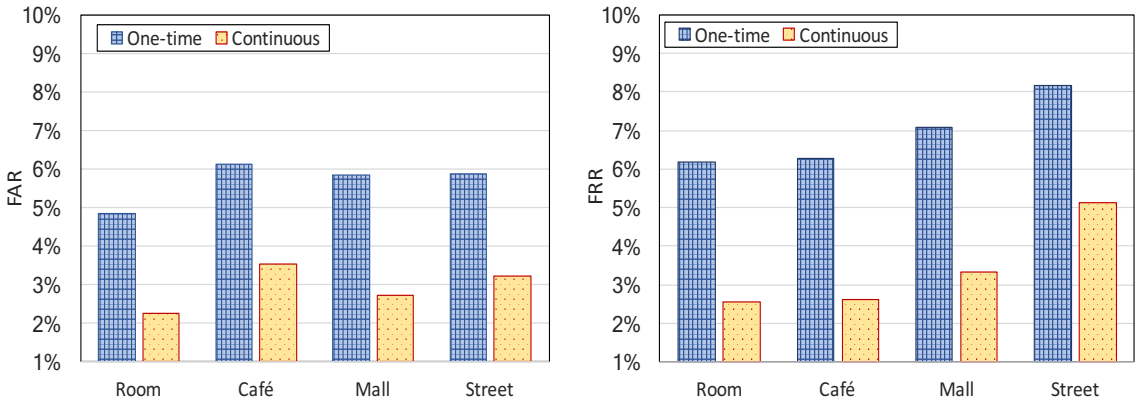
$$\text{BAC} = \frac{TPR + TNR}{2} \tag{15}$$

where TPR means the true positive rate, and TNR means the true negative rate.

In this evaluation, we train a two-class SVM model for each subject as the legitimate user and the rest 19 subjects are divided into two groups, 9 out of 19 subjects are selected as the imposters that are used for training, and the rest 10 subjects are selected as the intruders that are used for unknown users testing. For each SVM model, to achieve a better practicality with respect to context-free and a balanced training set, We choose four conversation episodes' data collected from the authorized user as the enrolled training data and randomly pick one episode from other 9 imposter users as the benchmark training data. For the evaluation, the remaining untrained episodes for the testing group are chosen as the testing data.

Fig. 11 shows the performance metrics along with increasing training data with fixed one-second authentication window length. It is observable that, when we only collect a small amount of the user's echo data, the captured features corresponding to the user are underrepresented, and hold very low variance and large bias compared with the true user's feature distribution. Thus, it results in less false acceptance rate but generates more false rejection cases. When we collect over 400 second echo data for training, the performance becomes stable at the level of around 95%, which can provide a better user experience. The detailed performances are listed in Table 1.

7.3.2 Performance of Authentication Modes. We evaluate two authentication modes under different application scenarios including the one-time authentication and the continuous authentication which covers the major two authentication modes in the current IoT environment.



(a) False Acceptance Rate (b) False Rejection Rate

Fig. 13. The impact of different background noises on the performance including FAR and FRR.

One-time Authentication. One-time authentication means that the users are asked for a single verification request of their identities (e.g., typing password, scanning face, or swiping ID card). As a passive acoustic ear authentication solution, it is not convenient for users to interrupt their on-going audios and stimulate a probe signal for verification purpose. Thus, we randomly select one-second audio segments from untrained conversation audios to mimic the scenario that users are requesting for verification while using earphones without interruption. As discussed in Section 7.3.1, we can achieve accuracy of 95.16% and 94.19% for precision and recall respectively.

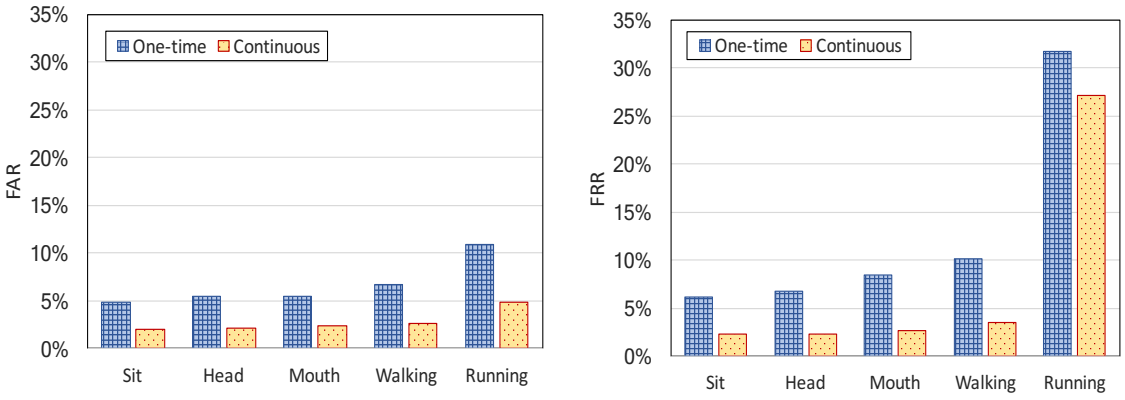
Continuous Authentication. For one-time authentication, the required time for authentication is always a major concern and constraint from the perspective of user experience. However, for the long-term continuous authentication solution that is conducted in a passive manner, we explore the performance with different authentication times. As shown in Fig. 12, the window size starts from 1 second which is the default setting for one-time authentication, then the accuracy increases with longer window sizes. This is because the longer window size allows more data to be used for extracting features with a wider coverage of frequency range. When the window size reaches 3 seconds, the accuracy gradually converges to the level of about 98%. Thus, we select the window size of 3 seconds as an authentication cycle for the continuous authentication scenario and we can achieve 97.55% recall and 97.57% precision (see Table 2).

7.4 Impact Quantification

In this section, we evaluate the robustness of *EarEcho* in terms of background noises, body motions, sound pressure levels of audio, permanence, and wearing positions.

7.4.1 Background Noise. With the popularity and convenience of wireless earphones, people are wearing earphones in more and more scenarios, However, acoustic sensing is also known to be sensitive to background noises. Thus, we examine our *EarEcho* in four different environments: a quite room (40 dB), a normal cafe (55 dB), a crowded shopping mall (65 dB), and a noisy street (75 dB). To ensure the replicability and controllability of the experiments, during the enrollment phase, we use data collected in the quite room for training and then simulate the testing scenarios by playing background noises at the corresponding effective sound pressure levels [11, 29, 49]. We use a smartphone to play the background sound as the noise source with 1 meter distance to participants.

From Fig. 13, we can observe that the FARs of noisy environments would slightly increase compared with a quite environment but the background noises' sound pressure level (SPL) doesn't have significant impacts on



(a) False Acceptance Rate (b) False Rejection Rate

Fig. 14. The impact of different body motions (e.g., sit, head motions, mouth motions, walking, and running) on the performance including FAR and FRR for both one-time and continuous authentication modes.

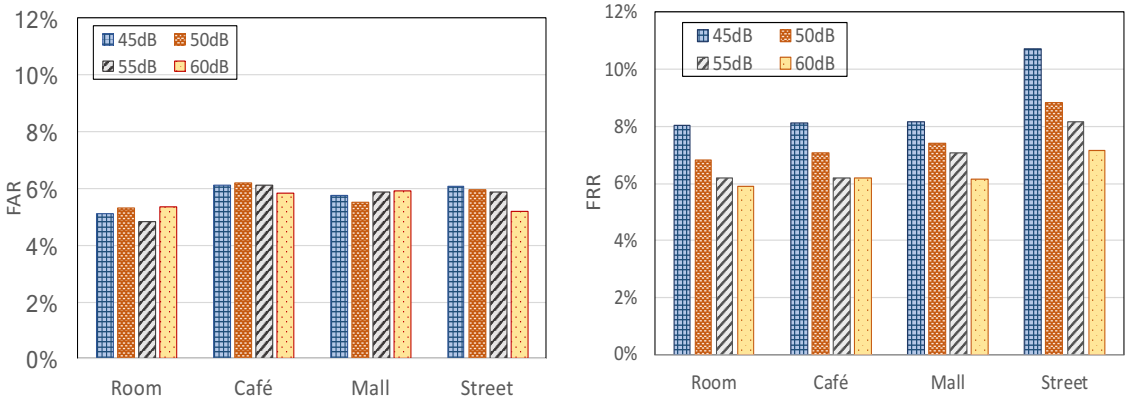
FARs. However, the FRRs indeed increase with higher SPLs. This is because the background noises can propagate through the isolation layer (the earphone's silicone tips) with some loss, then travel in the user's ear canal and mix with echos collected by the microphone. If the noise SLP is high enough, it can pass through the event detection and affect the extracted transfer function features.

7.4.2 Body Motions. Besides the background noises, noises may also come from users' activities during the usage (e.g., listening to music while walking, yawning, tuning the head around while making a phone call). To evaluate the robustness of our system, we test our *EarEcho* for multiple user body motions: head movement, mouse motion, walking, and running. To best emulate the real-life scenarios, we ask the participants to perform selected body motions with a high degree of freedom (e.g., randomly tuning head in vertical and horizontal direction, walking at a common pace) while wearing the earphones.

Fig. 14 shows the performance of system robustness against various body motions. Except for running, other activities only cause slight increases on FARs and FRRs (2% $\sim$ 4%) for one-time authentication mode. In the continuous authentication mode, the performance remains quite stable. This demonstrates that *EarEcho* is very robust to moderate motion noises. The reason for more significant performance influence resulted from dramatic motions like running is that, during the strenuous exercise, the soft interface will cause the earphone to have slight displacement relative to the ear canal which would change the acoustic propagation characteristics. However, it is worth noting that it represents an extreme scenario to have authentication requests during running.

7.4.3 Sound Pressure Levels of Stimuli. Volume is one of the major user preference settings for earphone usage. Different users might have different comfort Sound Pressure Levels (SPLs), and SPLs are also content-environment-dependent (e.g., people may prefer to raise the SPLs when listening to pop and rock music; SPLs are often set up higher when walking on the street compared to staying in a quite room). A higher SPL means a better audio quality which brings a higher Signal-to-Noise Ratio (SNR). To explore the impacts of SPLs on authentication accuracy, we evaluate the testing dataset from the low volume (45 dB SPL) to the high volume (60 dB SPL) and with different background noises. The default SPL for training and testing is 55 dB.

Fig. 15 shows the impacts of SPLs on both FAR and FRR. In Fig. 15(a), similar as background noises, given different SPLs, the FARs don't have significant variations. However, unsurprisingly, the higher SPLs result in the lower FRRs especially in a noisy environment (e.g., mall and street).



(a) False Acceptance Rate (b) False Rejection Rate

Fig. 15. Impacts of different stimulus sound pressure levels (dB) on the one-time authentication performance under different environments.

Table 3. Recall performance of longitudinal study over one month

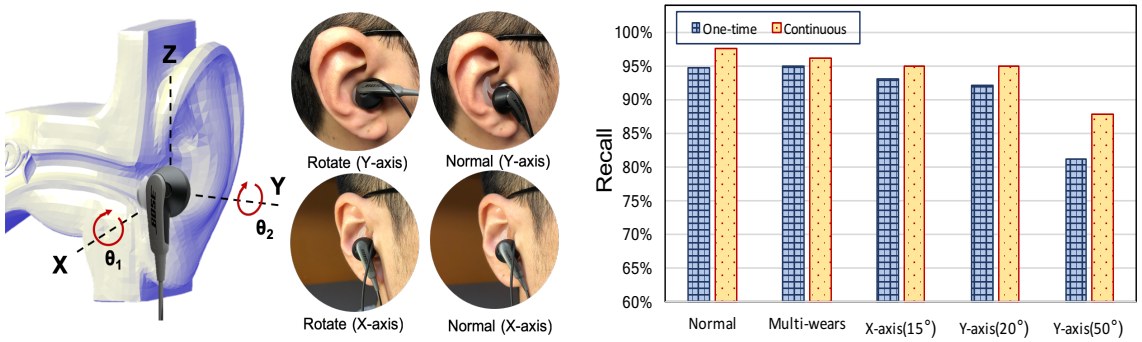
Recall (%)	Reference	1 day	1 week	1 month
One-time Authentication	95.0	94.3	93.6	91.4
Continuous Authentication[1]	97.1	97.1	96.3	95.4

[1] The authentication window size for continuous authentication is 3 seconds.

7.4.4 Permanence. It is important to provide a stable performance for any biometric design. To evaluate the permanence of our system, we randomly selected a sub-group consists of four subjects to participate in the longitudinal study lasting for one month. This study had two phases: enrollment and testing. All participants finished data collection for 10-minute audio on the first reference day. For the testing part, all subjects were asked to collect data again for another 10-minute audio after a certain time duration (i.e., one day, one week, and one month). The results of recall are shown in Table 3. After a period of one month, the recall has a slight drop of 2% to 3% (which might be caused by the different levels of in-ear cleanliness). It is demonstrated that our *EarEcho* can be considered as stable over a period of one month. In the real-world scenario, we can always augment the new collected echos into the training dataset to update the slowly changing features for the enrolled user.

7.4.5 Earphone Wearing Positions. Putting on and taking out your earphone are two most common actions in the daily usage of earphones. Each wear would cause a slight relative position change of the earphone in the ear canal. Besides the permanence over time, we also evaluated the system performance in terms of the effects of various earphone wearing behaviors including slightly varied positions of regular multiple wears and purposely rotated wearing positions along the X-axis and Y-axis, as shown in Fig. 16(a).

Among the 20 participants, we chose a subgroup of 10 participants (3 females and 7 males) as the enrolled, genuine users to investigate the effects of various earphone wearing behaviors. We designed two sessions, named the regular multi-wear session and the irregular rotation session. For the multi-wear session, in the lab environment, all ten participants were instructed to hear 5 audio records (2 minutes per record), and for every 20 seconds, participants needed to take out and put on their earphones repetitively. Thus, in total we collected 30 different wears for each participant. We used 20 of them to train the SVM classifier, and the rest 10 wears' data was used for testing. In addition to the regular multi-wear behaviors, we further evaluated several irregular earphone wearing behaviors including rotating the earphone by 15° along the X-axis and by 20°and 50° along the Y-axis. Each participant was asked to listen to one test audio record (2 minutes) with all three different earphone



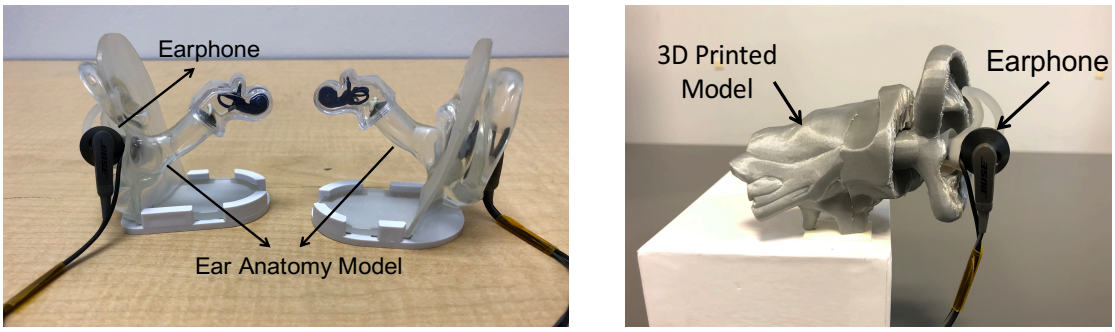
(a) (b)

Fig. 16. (a) Simulations of earphone wearing rotations. θ_1 ranges from 0 to 15°, θ_2 ranges from 0 to 50°. (b) Impacts of different earphone positions on both one-time and continuous authentication recall performance.

rotation angles respectively (15° on the X-axis, 20° and 50° on the Y-axis). Based on the classifier trained through the data collected from the regular multi-wear session, we evaluated the recall for the same participant with those irregular earphone wearing positions. As shown in Fig. 16(b), we can observe that regular multi-wear behaviors would not significantly affect the system performance if we involve the multi-wear data into model training. Irregular earphone rotation behaviors within a small range (e.g., 15° and 20°) would cause slight recall drops. However, once the rotation angle along the Y-axis increases up to 50 °, it would cause a certain degree of unfit interface between the earphone and the ear canal with noise leakage and volume change of inside cavity (ear canal). The recall drops to 81.2% and 88.0% for one-time and continuous authentication scenarios. Therefore, earphone displacement resulting from regular multiple wearing attempts would not affect the recall performance, and some irregular wearing behaviors would cause different levels of recall drops depending on the degree of rotation.

7.5 Vulnerability Study

Similar as other mobile and wearable authentication solutions, it is critical to investigate the security vulnerability of *EarEcho*. Even though the ear canal information is more hidden and hard to be stolen compared with fingerprints and faces information, there still could be potential threats from spoofing attacks by presenting a fake ear-canal anatomy model which has the similar geometry as the human ear canal. To imitate the spoofing attacks, we used two fake models: a well-designed anatomy plastic model and a 3D-printed PLA-based model as shown in Fig. 17. Both models were fabricated with the normal size and an accurate canal design. To better investigate the potential risks from spoofing attacks, we also slightly adjusted the placement of the earphone in the ear canal for multiple times. We tested 720 attempts from the fake models for each enrolled subject, and we only obtained, on average, 0.22% and 0.18% FARs for one-time and continuous authentication respectively. This result indicates that our *EarEcho* authentication solution provides a high level of resistance to spoofing attacks. The reason is that the material of ear canal inner surface also affects the acoustic propagation behaviors like absorption and reflection. Even though the attacker obtains the enrolled user's ear canal geometry information and designs a fake model with high resolution. The difference of material between fake models (e.g., plastic, PLA, or silicone) and human tissue still makes *EarEcho* hard to be spoofed.



(a) A well-designed anatomy plastic model (b) A customized low-cost 3D printed PLA model

Fig. 17. Spoofing attack study using two fake ear models.

Table 4. Comparison with Existing Acoustic-based Mobile and Wearable Authentication Biometrics

Biometrics	EarEcho	SilentKey[63]	Vocal Resonance [40]	EchoPrint [71]	BiLock [73]
Features	acoustic	acoustic	acoustic	acoustic+vision	acoustic
No.subjects	20	50	29	45	50
Modes[1]	Passive	Active	Passive	Active	Active
BAC[2]	94.5%-97.5%	78%-87%	94.2%-96.1%	93.75%	97%
Devices	earphones	smartphones	throat-mounted microphones	smartphones	smartphones

[1] The modes are categorized into two ways. Active means users need to perform certain actions to verify identities. Passive means the authentication process is implemented in an unobtrusive way.
[2] Balanced Accuracy.

8 DISCUSSIONS

8.1 Comparison with Other Mobile and Wearable Biometric Authentication Approaches

We compare the performance of *EarEcho* with other emerging acoustic-based biometrics deployed on mobile and wearable devices including SilentKey [63] (mouth motions), Vocal Resonance [40] (body sounds), EchoPrint [40] (face), and BiLock [73] (dental occlusion). As shown in Table 4, those active biometric solutions involve larger group of participants with lower BACs. The performance of BiLock was evaluated based on only legitimate users and imposters, without involving intruders (unknown new users). Compared with Vocal Resonance, *EarEcho* aims to provide an accurate and also a ubiquitous solution that can be deployed in existing wearable devices of large popularity.

8.2 Comparison with Ultrasound-based Ear Acoustic Continuous Authentication

Different from our proposed ear acoustic biometric using only audible frequency range, a hybrid system combining the fixed audible sound (for initial authentication) and ultrasound (for continuous authentication) has been explored by Machto et al. [43]. Particularly, under the long-term authentication modes, due to the high SNR and resolution of ultrasonic sensing, it can achieve a higher accuracy than audible-based solution. However, as claimed by the authors, ultrasonic sensing is too sensitive to the earphone displacements caused by body motions. Compared with our system using only built-in regular earpiece speaker, the ultrasound-based solution also requires additional ultrasound speaker that keeps emitting high frequency audios that would significantly increase the power consumption of earphones. In addition, according to our survey, most users are concerned about the potential health risk of long-term exposure to ultrasound. Lastly, it may also raise some security threats. For example, users are unperceptive if earphones' ultrasonic speakers are hacked by malicious attackers.

8.3 Limitations

EarEcho is only a proof-of-concept prototype at the current stage and still far from a well-engineered product. We list several main limitations as follows:

Earphone Wearing Position. In our current design, we adopt the commercial silicone tips as the interface between the earphone and ear to provide a secure fixed position. As discussed as Section 7.4.5, regular multi-wear behaviors would not affect the accuracy. However, excessive irregular rotations of the earphone and significant body motions like running would still cause non-negligible changes of the earphone's displacement. Therefore, a mechanism for the alignment of varying earphone wearing positions is needed to address such a challenge.

Varying User Ear Canal Conditions. The current extracted features were trained based on a limited set of data, far from sufficient to be robust against the user's varying ear canal conditions (e.g., ear infection, earwax, and other cleanliness states). Regularly retraining the SVM model with newly collected data could be an effective way to adapt to those possible changes.

8.4 Future Work

Our *EarEcho* proposes a new potential way to verify the user's identity in the mobile and wearable application scenarios. To provide *EarEcho* with a better accuracy, attack resistance and usability, we are planning to improve our design from the following aspects:

Earphone Position Alignment. As discussed in limitations section above, to achieve a higher true positive rate and lower false negative rate, we will integrate the earphone with additional sensors (e.g., accelerometers and gyroscopes) to estimate the irregular wearing position of the earphone in the ear canal. It may require the user to adjust the earphone position for the initial authentication or compensate the acoustic feature variations caused by body motions during continuous authentication. This will further improve the authentication performance.

Two-factor Authentication. Due to the rapid improvement on bio-signal based authentication (e.g., PPG, ECG, EEG) which can provide the liveness detection and reliable performance in a noisy environment, our *EarEcho* can be integrated with some other biometric sensors to enhance the robustness against acoustic noises and replay attacks.

Large-scale and Mobile-platform Evaluation. We tested only 20 subjects in the current experiments. A larger user group is needed to further validate the performance towards a finely engineered product. Moreover, in our design, to reduce the computation complexity, we use the SVM-based authentication, we still need to transplant our system to a smartphone platform to have a more precise evaluation on mobile power consumption and computation latency.

9 CONCLUSION

In this paper, we propose a novel authentication scheme — *EarEcho* — which leverages the acoustic characteristics in the user's ear canal through the integrated microphone and loudspeaker on commodity earphones. We validate that the extracted acoustic features between the emitted audios from the earpiece speaker and the received echoes from the microphone can be used as a unique and reliable identifier for user authentication. After acoustic signal pre-processing, the transfer function based features are feed into an SVM classifier for identity verification. Our results show that *EarEcho* can achieve 97.55% recall and 97.57% precision.

REFERENCES

[1] 2019. Nymi. https://nymi.com/. [Online; accessed 6-Jan-2019].

[2] Yomna Abdelrahman, Mohamed Khamis, Stefan Schneegass, and Florian Alt. 2017. Stay cool! understanding thermal attacks on mobile-based user authentication. In *Proceedings of the 2017 ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, 3751–3763.

[3] Abdullah I. Al-Shoshan. 2006. Speech and music classification and separation: a review. *Journal of King Saud University-Engineering Sciences* 19, 1 (2006), 95–132.

[4] Apple. 2018. About Face ID advanced technology. https://support.apple.com/en-us/HT208108. [Online; accessed 6-Jan-2019].

[5] Juan Arteaga-Falconi, Hussein A. Osman, and Abdulmotaleb E. Saddik. 2016. ECG authentication for mobile devices. *IEEE Transactions on Instrumentation and Measurement* 65, 3 (2016), 591–600.

[6] Uttara Athawale and Manoj Gupta. 2018. Survey on recent ear biometric recognition techniques. *International Journal of Computer Sciences and Engineering* 6, 6 (2018), 1208–1211.

[7] Sarah A. Bargal, Alexander Welles, Cliff R. Chan, Samuel Howes, Stan Sclaroff, Elizabeth Ragan, Courtney Johnson, and Christopher Gill. 2015. Image-based ear biometric smartphone app for patient identification in field settings.. In *VISAPP (3)*. 171–179.

[8] Abdelkareem Bedri, David Byrd, Peter Presti, Himanshu Sahni, Zehua Gue, and Thad Starner. 2015. Stick it in your ear: Building an in-ear jaw movement sensor. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 1333–1338.

[9] Shengjie Bi, Tao Wang, Nicole Tobias, Josephine Nordrum, Shang Wang, George Halvorsen, Sougata Sen, Ronald Peterson, Kofi Odame, Kelly Caine, et al. 2018. Auracle: Detecting eating episodes with an ear-mounted sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 92.

[10] Jennings Brown. 2019. Researcher who said he hacked iPhone X Face ID with a printed image cancels talk after employer shaming. https://gizmodo.com/researcher-who-said-he-hacked-iphone-x-face-id-with-a-p-1831490792. [Online; accessed 6-Jan-2019].

[11] Alfredo Calixto, Fabiano B. Diniz, and Paulo H. Zannin. 2003. The statistical modeling of road traffic noise in an urban setting. *Cities* 20, 1 (2003), 23–29.

[12] Jagmohan Chauhan, Yining Hu, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. 2017. BreathPrint: Breathing acoustics-based user authentication. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 278–291.

[13] Hui Chen and Bir Bhanu. 2007. Human ear recognition in 3D. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 4 (2007), 718–737.

[14] Alexander D. Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann. 2012. Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 987–996.

[15] Simon Eberz, Nicola Paoletti, Marc Roeschlin, Marta Kwiatkowska, I Martinovic, and A Patané. 2017. Broken hearted: How to attack ECG biometrics. In *Processings of the 2017 Network Distributed System Security Symposium*.

[16] PN A. Fahmi, Elyor Kodirov, Deok J. Choi, Guee S. Lee, A M. Fikri Azli, and Shohel Sayeed. 2012. Implicit authentication based on ear shape biometrics using smartphone camera during a call. In *Proceedings of the 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2272–2276.

[17] Mohammed E. Fathy, Vishal M. Patel, and Rama Chellappa. 2015. Face-based active authentication on mobile devices. In *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1687–1691.

[18] Huan Feng, Kassem Fawaz, and Kang G Shin. 2017. Continuous authentication for voice assistants. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. ACM, 343–355.

[19] Yang Gao, Wei Wang, Borui Li, Omkar R. Patil, and Zhanpeng Jin. 2018. Replicating your heart: Exploring presentation attacks on ECG biometrics. In *Proceedings of the IEEE Conference on Communications and Network Security (CNS)*. 1–9.

[20] Valentin Goverdovsky, David Looney, Preben Kidmose, and Danilo P Mandic. 2016. In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring. *IEEE Sensors Journal* 16, 1 (2016), 271–277.

[21] Valentin Goverdovsky, Wilhelm V. Rosenberg, Takashi Nakamura, David Looney, David J. Sharp, Christos Papavassiliou, Mary J. Morrell, and Danilo P. Mandic. 2017. Hearables: Multimodal physiological in-ear sensing. *Scientific reports* 7, 1 (2017), 6948.

[22] Qiong Gui, Maria V. Ruiz-Blondet, Sarah Laszlo, and Zhanpeng Jin. 2019. A Survey on Brain Biometrics. *Comput. Surveys* 51, 6 (2019), 112:1–112:38.

[23] Guodong Guo, Lingyun Wen, and Shuicheng Yan. 2014. Face authentication with makeup changes. *IEEE Transactions on Circuits and Systems for Video Technology* 24, 5 (2014), 814–825.

[24] Takahiro Hashizume, Takuya Arizono, and Koji Yatani. 2018. Auth'n' scan: Opportunistic photoplethysmography in mobile fingerprint authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 137.

[25] Alex Hern. 2014. Hacker fakes German minister's fingerprints using photos of her hands. https://www.theguardian.com/technology/2014/dec/30/hacker-fakes-german-ministers-fingerprints-using-photos-of-her-hands. [Online; accessed 6-Jan-2019].

[26] Thang Hoang, Deokjai Choi, and Thuc Nguyen. 2015. Gait authentication on mobile phone using biometric cryptosystem and fuzzy commitment scheme. *International Journal of Information Security* 14, 6 (2015), 549–560.

[27] Chenyu Huang, Huangxun Chen, Lin Yang, and Qian Zhang. 2018. BreathLive: Liveness Detection for Heart Sound Authentication with Deep Breathing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 12.

[28] Pei Huang, Borui Li, Linke Guo, Zhanpeng Jin, and Yu Chen. 2016. A robust and reusable ECG-based authentication and data encryption scheme for eHealth systems. In *Proceedings of the IEEE Global Communications Conference*. 1–6.
[29] Jian Kang. 2006. *Urban sound environment*. CRC Press.
[30] James M. Kates. 1988. A computer simulation of hearing aid response and the effects of ear canal size. *The Journal of the Acoustical Society of America* 83, 5 (1988), 1952–1963.
[31] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Allesandro Montanari. 2018. Earables for personal-scale behavior analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89.
[32] Gierad Laput, Xiang'Anthony' Chen, and Chris Harrison. 2016. Sweepsense: Ad hoc configuration sensing using reflected swept-frequency ultrasonics. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 332–335.
[33] B. W. Lawton. 2013. *Exposure limits for airborne sound of very high frequency and ultrasonic frequency*. Technical Report 334. University of Southampton, Institute of Sound and Vibration Research.
[34] Hyewon Lee, Tae H. Kim, Jun W. Choi, and Sunghyun Choi. 2015. Chirp signal-based aerial acoustic communication for smart devices. In *Proceedings of the 2015 IEEE Conference on Computer Communications*. IEEE, 2407–2415.
[35] Ho-Man C. Leung, Chi-Wing Fu, and Pheng-Ann Heng. 2018. TwistIn: Tangible authentication of smart devices via motion co-analysis with a smartwatch. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 72.
[36] Mengyuan Li, Yan Meng, Junyi Liu, Haojin Zhu, Xiaohui Liang, Yao Liu, and Na Ruan. 2016. When CSI meets public WiFi: Inferring your mobile phone password via WiFi signals. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1068–1079.
[37] Feng Lin, Kun W. Cho, Chen Song, Wenyao Xu, and Zhanpeng Jin. 2018. Brain password: A secure and truly cancelable brain biometrics for smart headwear. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 296–309.
[38] Can Liu, Gradeigh D. Clark, and Janne Lindqvist. 2017. Where usability and security go hand-in-hand: Robust gesture-based authentication for mobile systems. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 374–386.
[39] Jian Liu, Chen Wang, Yingying Chen, and Nitesh Saxena. 2017. VibWrite: Towards finger-input authentication on ubiquitous surfaces via physical vibration. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 73–87.
[40] Rui Liu, Cory Cornelius, Reza Rawassizadeh, Ronald Peterson, and David Kotz. 2018. Vocal resonance: Using internal body voice for wearable authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 19.
[41] Dawn B. Logas. 1994. Diseases of the ear canal. *Veterinary Clinics of North America: Small Animal Practice* 24, 5 (1994), 905–919.
[42] Li Lu, Jiadi Yu, Yingying Chen, Hongbo Liu, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. Lip reading-based user authentication through acoustic sensing on smartphones. *IEEE/ACM Transactions on Networking* (2019).
[43] Shivangi Mahto, Takayuki Arakawa, and Takafumi Koshinak. 2018. Ear acoustic biometrics using inaudible signals and its application to continuous user authentication. In *Proceedings of the 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 1407–1411.
[44] Hiroyuki Manabe, Masaaki Fukumoto, and Tohru Yagi. 2015. Conductive rubber electrodes for earphone-based eye gesture input interface. *Personal and Ubiquitous Computing* 19, 1 (2015), 143–154.
[45] Wenguang Mao, Mei Wang, and Lili Qiu. 2018. AIM: Acoustic imaging on a mobile. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*. ACM, 468–481.
[46] Denys J. C. Matthies, Bernhard A. Strecker, and Bodo Urban. 2017. EarFieldSensing: A novel in-ear electric field sensing to enrich wearable gesture input through facial expressions. In *Proceedings of the 2017 ACM Conference on Human Factors in Computing Systems (CHI)*. 1911–1922.
[47] Emiliano Miluzzo, Alexander Varshavsky, Suhrid Balakrishnan, and Romit R. Choudhury. 2012. Tapprints: your finger taps have fingerprints. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*. ACM, 323–336.
[48] Mark Mirtchouk, Christopher Merck, and Samantha Kleinberg. 2016. Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 451–462.
[49] Christopher C. Novak, Joseph L. Lopa, and Robert E. Novak. 2010. Effects of sound pressure levels and sensitivity to noise on mood and behavioral intent in a controlled fine dining restaurant environment. *Journal of Culinary Science & Technology* 8, 4 (2010), 191–218.
[50] Jang-Ho Park, Dae-Geun Jang, Jung Park, and Se-Kyoung Youm. 2015. Wearable sensing of in-ear pressure for heart rate monitoring with a piezoelectric sensor. *Sensors* 15, 9 (2015), 23402–23417.
[51] Omkar R. Patil, Wei Wang, Yang Gao, Wenyao Xu, and Zhanpeng Jin. 2018. A non-contact PPG biometric system based on deep neural network. In *Proceedings of the IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. 1–7.
[52] Anika Pflug and Christoph Busch. 2012. Ear biometrics: A survey of detection, feature extraction and recognition methods. *IET Biometrics* 1, 2 (2012), 114–129.
[53] Nalini K. Ratha, Ruud M. Bolle, Vinayaka D. Pandit, and Vaibhav Vaish. 2000. Robust fingerprint authentication using local structural similarity. In *Proceedings of the 5th IEEE Workshop on Applications of Computer Vision*. IEEE, 29–34.

[54] ReportBuyer. 2018. Earphones & headphones market - Global Outlook and Forecast 2018-2023. https://www.reportbuyer.com/product/5289128/earphones-and-headphones-market-global-outlook-and-forecast-2018-2023.html. [Online; accessed 4-Feb-2019].
[55] Aditi Roy, Nasir Memon, and Arun Ross. 2017. Masterprint: Exploring the vulnerability of partial fingerprint-based authentication systems. *IEEE Transactions on Information Forensics and Security* 12, 9 (2017), 2013–2025.
[56] Daniela Sánchez and Patricia Melin. 2014. Optimization of modular granular neural networks using hierarchical genetic algorithms for human recognition using the ear biometric measure. *Engineering Applications of Artificial Intelligence* 27 (2014), 41–56.
[57] Vlad Savov. 2018. Wireless headphones are improving faster than anything else in tech. https://www.theverge.com/2018/8/31/17803728/wireless-headphones-usb-c-google-assistant-siri-ifa-2018. [Online; accessed 4-Feb-2019].
[58] Nabilah Shabrina, Tsuyoshi Isshiki, and Hiroaki Kunieda. 2016. Fingerprint authentication on touch sensor using phase-only correlation method. In *Proceedings of the 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES)*. IEEE, 85–89.
[59] Muhammad Shahzad, Alex X. Liu, and Arjmand Samuel. 2013. Secure unlocking of mobile touch screen devices by simple gestures: you can see it but you can not do it. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*. ACM, 39–50.
[60] Edgar A. G. Shaw. 1974. The external ear. In *Auditory System. Handbook of Sensory Physiology, vol 5/1*, Wolf D. Keidel and William D. Neff (Eds.). Springer-Verlag, Chapter 14, 455–490.
[61] Jongseo Sohn, Nam S. Kim, and Wonyong Sung. 1999. A statistical model-based voice activity detection. *IEEE Signal Processing Letters* 6, 1 (1999), 1–3.
[62] Michael R. Stinson and BW Lawton. 1989. Specification of the geometry of the human ear canal for the prediction of sound-pressure level distribution. *The Journal of the Acoustical Society of America* 85, 6 (1989), 2492–2503.
[63] Jiayao Tan, Xiaoliang Wang, Cam-Tu Nguyen, and Yu Shi. 2018. SilentKey: A new authentication framework through ultrasonic-based lip reading. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 36.
[64] Furkan Tari, Ant Ozok, and Stephen H. Holden. 2006. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *Proceedings of the second Symposium on Usable Privacy and Security*. ACM, 56–66.
[65] Shejin Thavalengal, Petronel Bigioi, and Peter Corcoran. 2015. Iris authentication in handheld devices-considerations for constraint-free acquisition. *IEEE Transactions on Consumer Electronics* 61, 2 (2015), 245–253.
[66] Pim T. Tuyls, Evgeny Verbitskiy, Tanya Ignatenko, Daniel Schobben, and Ton H. Akkermans. 2004. Privacy-protected biometric templates: Acoustic ear identification. In *Biometric Technology for Human Identification*, Vol. 5404. International Society for Optics and Photonics, 176–183.
[67] Boudewijn Venema, Johannes Schiefer, Vladimir Blazek, Nikolai Blanik, and Steffen Leonhardt. 2013. Evaluating innovative in-ear pulse oximetry for unobtrusive cardiovascular and pulmonary monitoring during sleep. *IEEE Journal of Translational Engineering in Health and Medicine* 1 (2013), 2700208–2700208.
[68] Lei Wang, Kang Huang, Ke Sun, Wei Wang, Chen Tian, Lei Xie, and Qing Gu. 2018. Unlock with your heart: Heartbeat-based authentication on commercial mobile phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 140.
[69] Peter Welch. 1967. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* 15, 2 (1967), 70–73.
[70] Bing Zhou, Mohammed Elbadry, Ruipeng Gao, and Fan Ye. 2017. BatMapper: acoustic sensing based indoor floor plan construction using smartphones. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 42–55.
[71] Bing Zhou, Jay Lohokare, Ruipeng Gao, and Fan Ye. 2018. EchoPrint: Two-factor authentication using acoustics and vision on smartphones. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 321–336.
[72] Man Zhou, Qian Wang, Jingxiao Yang, Qi Li, Feng Xiao, Zhibo Wang, and Xiaofen Chen. 2018. PatternListener: Cracking android pattern lock using acoustic signals. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1775–1787.
[73] Yongpan Zou, Meng Zhao, Zimu Zhou, Jiawei Lin, Mo Li, and Kaishun Wu. 2018. BiLock: User authentication via dental occlusion biometrics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 152.