EarLock: Personal Authentication System for Hearables Using Sound Leakage Signals

Takashi Amesaka, Hiroki Watanabe, Yuta Sugiura, Masanori Sugimoto, and Buntarou Shizuki

***Abstract*—Earphone-type wearable devices, also known as "hearables," will have many functions in the future. Some of those functions will require authentication of the wearer for access to the user's privacy information or settlement of payments. In this study, we propose a new personal authentication system for hearables called EarLock. EarLock authenticates the wearer by acquiring and analyzing ear canal and auricle shape information using sound leakage from the device. The system can be implemented using a speaker and external microphone that are highly compatible with hearables. We implemented three prototype devices and investigated EarLock's authentication performance under various practical scenarios, including walking conditions, noisy environments, and situations with object interference. Experimental results showed that the in-ear, open-ear, and bone-conduction devices achieved balanced accuracy (BAC) scores of 87.2–93.7%, 83.4–94.7%, and 85.9–90.0%.**

***Index Terms*—Hearables, authentication, active acoustic sensing, sound leakage, ear biometrics.**

I. Introduction

ALTHOUGH earphone-type wearable devices, also known as "hearables," are gaining attention as a new platform with various functions [1], such as music, communication, vital sensing for health management [2], [3], and voice assistance for daily activities, one key challenge in hearables is enhancing the security. It is anticipated that some hearables applications will require the authentication of the wearer because they may access personal information, such as messages, calendars, and biometric information [4]. Moreover, with advances in hearables authentication technology, offering services that enable users to procure goods and digital content from e-commerce platforms solely through a device's voice assistant may be feasible. Therefore, developing such an authentication method is crucial for advancing the future of hearables.

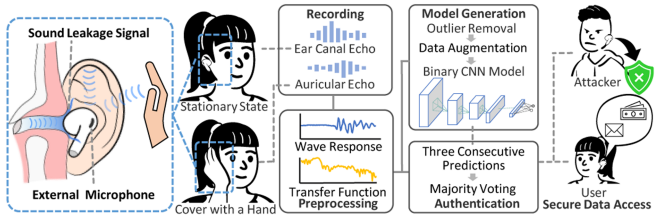


Fig. 1. System overview of EarLock.

However, applying existing authentication methods used in smartphones to hearables poses challenges. Hearables do not have touch screens like smartphones, making entering data like personal identification number (PIN) codes difficult. The compact size of the hearables makes it challenging to incorporate a fingerprint sensor or camera. Consequently, no commercially available devices have these features. Voiceprint authentication, while a potential solution, has drawbacks, especially in noisy environments and public spaces. Previous studies [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20] have mainly proposed authentication methods for hearables using inward-facing microphones (IFMs), photoplethysmography (PPG) sensors, or electroencephalography (EEG) sensors to acquire biometric features of the ear canal, teeth, body, voice, gait, and heartbeat for authentication.

In this study, we propose a personal authentication system, EarLock, that leverages sound leakage signals from hearables captured by an external microphone (Fig. 1). The ear canal and auricle shape varies from person to person [21], leading to individual differences in the acoustic characteristics of the sound leakage signals. EarLock authenticates users by using an external microphone to record leaked sound from hearables in two different user states: when users are wearing the device in a stationary state and when they cover their ears with their hands. External microphones are used for noise cancellation and voice recording, and many commercial products are equipped with them. In a practical scenario, the EarLock system authenticates the user by measuring the signal for approximately 1.7 seconds when the device is worn. After this initial authentication, an automatic ear detection [22] is designed to monitor the user's wearing status and maintain the login state. This approach is similar to the current usage of smartphones. Furthermore, EarLock acquires personal properties of the auricle shape, distinct

Received 30 May 2024; revised 17 January 2025; accepted 3 February 2025. Date of publication 10 February 2025; date of current version 5 June 2025. This work was supported in part by JSPS, KAKENHI, under Grant 23KJ1884/21H03485/24K02968, and in part by JST, PRESTO, under Grant JPMJPR2138/JPMJPR2134. Recommended for acceptance by H.S. Kim. *(Corresponding author: Takashi Amesaka.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Bioethics Committee of Keio University under Application No. 2023-077.

Takashi Amesaka and Yuta Sugiura are with the Department of Information and Computer Science, Keio University, Yokohama 108-0073, Japan (e-mail: amesaka@keio.jp; sugiura@keio.jp).

Hiroki Watanabe is with the Department of Media Architecture, Future University Hakodate, Hokkaido 041-0803, Japan (e-mail: hwata@fun.ac.jp).

Masanori Sugimoto is with the Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0808, Japan (e-mail: sugi@ist.hokudai.ac.jp).

Buntarou Shizuki is with the Department of Computer Science, University of Tsukuba, Tsukuba 305-8577, Japan (e-mail: shizuki@cs.tsukuba.ac.jp).

Digital Object Identifier 10.1109/TMC.2025.3540584

1536-1233 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

from those explored in existing research (e.g., heartbeat, gait, and ear canal shape), thereby enabling its use in conjunction with existing methods to potentially enhance the robustness of hearables authentication. The contributions of this study are as follows.

- We propose "EarLock," a personal authentication system that utilizes sound leakage signals and is highly compatible with hearables equipped with speakers and external microphones.
- By using sound leakage, we expanded the application of personal authentication to open-ear and bone-conduction devices, which have not been the target of previous research.
- Our preliminary study shows that the auricle shape information, a new biometric that has not been captured in previous authentication methods for hearables, is included in the sound leakage signal.
- We conducted an experiment to evaluate the authentication performance of EarLock using three different device types: in-ear, open-ear, and bone-conduction. The results indicated that these devices achieved balanced accuracy (BAC) scores of 87.2–93.7%, 83.4–94.7%, and 85.9–90.0% under various conditions (stable conditions, walking conditions, noisy environments, and situations with object interference).

A. Advancements Beyond Previous Work

In our previous work, we introduced this concept in a short paper format [23]. Our previous work evaluated an authentication system using sound leakage signals limited to stable, noisy, and walking environments with in-ear devices, involving only 16 participants. Our current study expands on this by analyzing a larger dataset with 23 participants across six test environments and improving upon our recognition algorithms. In terms of technical contributions, this study presents a multistream learning model specifically designed to effectively utilize two types of sound leakage signals with distinct characteristics. This enhancement to the recognition system enables the efficient use of diverse features within sound leakage signals. Additionally, we established a mathematical foundation for sound leakage and conducted an in-depth model analysis to assess the impact of shape variations in the ear canal, auricle, and hand on each device type. As for empirical contributions, we assessed the extent to which the shapes of the auricle and hand contribute to the personal identification information contained in sound leakage signals, quantifying their individuality. Furthermore, to expand the applicability of the authentication system, we demonstrated its compatibility with open-ear and bone-conduction devices, which have not been addressed in any previous studies.

To further reinforce the system's performance evaluation in real-world scenarios, we investigated the effects of physical factors, such as gloves and earrings, on authentication accuracy. We also examined the system's resilience to eavesdropping attacks, the influence of required frequency bands for system design, and the impact of training data quantity. These analyses confirm that the EarLock system can maintain robust recognition performance across diverse environments and usage scenarios.

II. Related Work

In this section, we describe previous research on personal authentication methods for hearables, research that uses the individuality of the auricle's shape, and research using active acoustic sensing (AAS) with hearables. Active acoustic sensing is a technology that uses the propagation and reflection of sound signals to sense the state of the object or the environment.

A. Authentication Methods for Hearables

Table I summarizes our EarLock and typical authentication methods for each biometric feature. EarEcho [6], EarDynamic [8], and Earmonitor [16] are personal authentication methods that utilize acoustic characteristics of the ear canal, which are measured by IFMs. EarGate [7] employs IFMs in hearables to capture sounds generated by the wearer's body during walking, using this gait data for personal authentication. TeethPass [11] and ToothSonic [10] use IFMs to record the sounds produced by the meshing of the user's teeth for personal authentication. Vocal Resonance [13] and EarPrint [17] perform personal authentication using voice sounds acquired from the IFM. VAuth [24] is a personal authentication system that employs a contact microphone to assess the vibration characteristics of the body surface when speaking, thereby authenticating the wearer. HeartPrint [9] uses an IFM to record the wearer's heartbeat sounds, also known as phonocardiograms (PCG), for personal authentication. Zou et al. [25] proposed a system that acquires body sounds and performs personal authentication using IFMs. EarPPG [15] employs the device's integrated PPG sensor to authenticate users by analyzing sensor responses to speech-induced physiological changes. Nakamura et al. [26] introduced a personal authentication system employing EEG sensors to utilize individual-specific EEG rhythms as biometric identifiers.

The methods previously discussed utilize IFMs, PPG, and EEG sensors. IFMs are particularly used for high-performance noise cancellation, though they are currently restricted to a limited number of commercially available canal-type devices. EarLock focuses on sound leakage and uses an external microphone to capture echoes from the ear canal and auricle for personal authentication. The external microphone is used for noise cancellation, surrounding sound capturing, and calling, making it highly compatible with hearables. In our research, major manufacturers (20 companies) have included IFMs or similar sensors in only approximately 10–50% of their marketed models, while external microphones are more common and included in over 90% of marketed models. Moreover, EarLock is able to capture individual characteristics of the auricle that conventional authentication methods have not captured. Therefore, it can be used in concert with conventional biometrics such as PPG, EEG rhythms, and body sounds, potentially strengthening the authentication of hearables.

In addition to the proposed method, several biometric authentication methods [27], [28] have been proposed that are

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

TABLE I
LIST OF AUTHENTICATION METHODS FOR HEARABLES BY BIOMETRIC FEATURES USED

Work	EarLock	EarEcho [6] EarDynamic [8] Earmonitor [16]	EarGate [7]	TeethPass [11] ToothSonic [10]	Vocal Resonance [13] EarPrint [17] VAuth [24]	HeartPrint [9]	Zou et al. [25]	EarPPG [15]	Nakamura et al. [26]
Modality	External Microphone	IFMs	IFMs	IFMs	IFMs Contact Microphone	IFMs	IFMs	PPG sensors	EEG
Device Compatibility	○	△	△	△	△	△	△	△	×
Feature	Ear canal & auricular acoustics	Ear canal acoustics	Walking sound	Teeth occlusion sound	Voice sound Body-surface vibrations	PCG	Body sound	PPG	EEG rhythms
Device	In-ear type Open-ear type Bone-conduction type	Canal type	Canal type	Canal type	Canal type	In-ear type	Canal type	In-ear type	Canal type
Participant	23	20/24/120	31	22/25	29/23	45	50	25	15
Result	87.2 - 93.7% 83.4 - 94.7% 85.9 - 90.0%	BAC > 94.5% Acc. = 93.04% BAC = 96.4%	BAC > 80%	Acc. = 98.6% Acc. = 95.0%	Acc. 94.2 - 96.1% EER = 3.64%	FAR = 1.6% FRR = 1.8%	FAR = 3.4% FRR = 4.9%	Acc. = 94.2%	FAR < 2.3% FRR < 32.2%
Action	Covering the ear with a hand	None Mouthing None	Walking	Teeth occlusion	Vocalization	None	None	Mouthing	None
Auth Time	1.7sec.	1 sec. 2∼20 words 1 sec.	Few steps	4 occlusion 1 teeth gesture	Voice command	130ms	30 sec.	One phrase	60 sec.

○: Set up in over 90% of commercial devices, △: Set up in less than 50% of commercial devices, ×: Not available in commercial devices.
Acc. represents accuracy; BAC represents the balanced accuracy.

theoretically applicable to open-ear/bone-conduction devices. However, each of these studies has conducted evaluations using prototype devices, and it remains unclear whether these device types can be practically deployed. In our study, we used a commercially available device as the signal source, adding only an external microphone due to sampling rate limitations. This setup, closer to practical use, allowed us to assess the feasibility of these device types for real-world applications.

B. Individuality of the Auricle

The auricle efficiently collects ambient sound and transmits it to the ear canal [29], [30], which differs from person to person; therefore, some studies use the individuality of the auricle's shape for personal identification/authentication. Holz et al. [31] proposed a personal identification system using a smartphone's touch screen as a sensor to capture spatial feature points of the auricle when the device is placed against the ear. Itani et al. [32] proposed a personal authentication system that acquires the auricular transfer function by positioning a smartphone-type device against the ear. Using headphones, Sulavko et al. [33] acquired and investigated the unique acoustic characteristics of the auricle and ear canal.

Similar to these methods, our system aims to authenticate users using the individuality of the auricle's shape. Specifically, the acoustic characteristics of the auricle are acquired by focusing on the leaked sound and its echo generated by covering the hand. These measures enable the hearables to capture information on the shape of the auricle.

C. Active Acoustic Sensing With Hearables

AAS has been employed in numerous hearables studies beyond personal authentication. For instance, the user's jaw and cheek movements alter the acoustic properties of the ear canal.

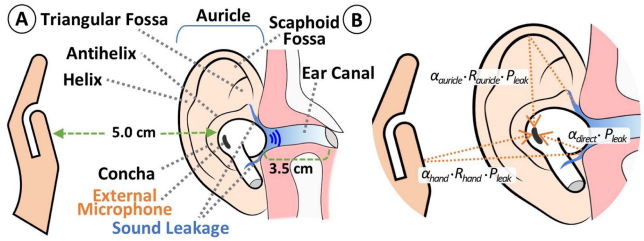


Fig. 2. A: ear structure, B: the sound leakage reflection model.

This characteristic has been used for facial expression recognition [23], [34], [35], [36], [37] and silent speech recognition [38], [39], [40]. Additionally, Yincheng et al. [41] utilized ear canal acoustic properties in ear condition monitoring systems. Although these studies have utilized speakers and microphones on hearables, none has explored sound leakage. By focusing on sound leakage, our method is able to capture the characteristics of both the ear canal and auricle.

III. EARLOCK OVERVIEW

Our system has four processes shown in Fig. 1: 1) Recording, where it captures two types of echoes under different user states; 2) Preprocessing, which involves feature extraction; 3) Model generation, including outlier removal, data augmentation, and learning via a convolutional neural network (CNN); and 4) The authentication process that performs the actual user verification.

A. Recording

EarLock utilizes the sound leakage phenomenon for data acquisition, where sound emitted from hearables leaks outward from the ear canal, as illustrated in Fig. 2. The system acquires

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

two types of sound leakage signals from the user: ear canal echoes (ECE) and auricular echoes (AE). ECE is the sound leakage signal when the user is stationary; the measurement signal reverberates inside the ear canal, leaks outward, and then directly reaches the external microphone. The sound pressure P that is played from the speaker, reflects within the ear canal, and returns to the inside of the device is defined as follows [6]:

$$P = \{p_i \leftarrow d(o_i, C) \mid i = 1, 2, \ldots, m\} \tag{1}$$

Here, each element p_i corresponds to the sound pressure resulting from the sound o_i emitted by the speaker and reflected within the ear canal. The function $d(o_i, C)$ denotes the transfer function from the sound source o_i to the microphone, which depends on parameters C that capture individual characteristics of the ear canal, including its shape and surface properties. The index $i = 1, 2, \ldots, m$ represents the different sound components or propagation paths within the ear canal, each contributing to the unique acoustic profile.

To model the sound pressure of the leakage signal at a specific frequency f, we consider the transmission characteristics of the ear canal and the geometry of the ear canal opening. The sound pressure $P_{\text{leak}}(f)$ that leaks from the ear canal to the external environment can be defined as a function of the base pressure $P(f)$ inside the ear canal, scaled by the area ratio $\frac{S}{A_{\text{ear}}}$ and the frequency-dependent transmission factor $T(f)$, as shown below:

$$P_{\text{leak}}(f) = \frac{S}{A_{\text{ear}}} \cdot T(f) \cdot P(f) \tag{2}$$

$T(f)$ is frequency-dependent, with higher values typically observed for low frequencies, allowing these sounds to leak more easily. Higher frequencies, in contrast, tend to have a lower $T(f)$ value, indicating increased resistance as they propagate from the ear canal to the external environment.

The component of P_{ECE} that reaches the external microphone is a combination of direct and auricle-reflected waves from P_{leak}, and P_{ECE} is defined as follows:

$$P_{\text{ECE}} = \alpha_{\text{direct}}(f) \cdot P_{\text{leak}} + \alpha_{\text{auricle}}(f) \cdot R_{\text{auricle}}(f) \cdot P_{\text{leak}} \tag{3}$$

Here, $\alpha_{\text{direct}}(f)$ and $\alpha_{\text{auricle}}(f)$ are frequency-dependent coefficients representing the proportions of the direct and reflected components, respectively, and $R_{\text{auricle}}(f)$ is the frequency-dependent reflection coefficient of the auricle.

Next, AE is the sound leakage signal when the user covers the ear with their hand; it contains the ECE that further reverberates in the space formed by the hand and ear. P_{AE} is defined as follows:

$$P_{\text{AE}}(f) = \gamma(f) \cdot P_{\text{leak}}(f) \tag{4}$$

Here, the final recorded sound pressure $P_{\text{AE}}(f)$ is represented as the product of the sound leakage $P_{\text{leak}}(f)$ and the composite coefficient $\gamma(f)$.

$$\gamma(f) = \alpha_{\text{direct}}(f) + \alpha_{\text{auricle}}(f) \cdot R_{\text{auricle}}(f) + \alpha_{\text{hand}}(f) \cdot R_{\text{hand}}(f) \tag{5}$$

The composite coefficient $\gamma(f)$ is defined to represent the contributions of the direct component and the auricle- and hand-reflected components for each frequency. This simplified model integrates all reflections and direct components into a single parameter to express the final sound pressure.

Using a measurement signal such as a sweep signal in the playback signal makes it possible to obtain each frequency response of the sound leakage. We defined this frequency response as the sound leakage signal. In this study, we utilize a sweep signal that varies in frequency from f_{start} to f_{end} over a period T to measure the frequency response of the sound leakage signal. The sweep signal $s(t)$ is defined as follows:

$$s(t) = \sin\left(2\pi \left(f_{\text{start}} + (f_{\text{end}} - f_{\text{start}})/T \cdot t\right) t\right) \tag{6}$$

The shape of the ear canal and each part of the auricle differ from person to person [42]. Therefore, these echoes also differ from person to person. The details of the differences in each echo and the changes in the sound leakage signal due to the shape of the ear canal and auricle are described in Section V.

B. Preprocessing

We performed feature extraction to apply each obtained echo to the authentication modeling. The system extracts two types of features: transfer function and impulse response. The transfer function quantifies the response characteristics of each frequency component. The attenuation characteristics of each frequency component vary according to the ear's anatomical structure. The impulse response captures time-series information, including time delay and reverberation characteristics. Similarly to the transfer function, the impulse response also varies according to the ear's anatomical structure.

In our system, we use the time synchronous averaging method for extraction to reduce noise components [43]. Fig. 3 summarizes the time synchronous averaging method and feature extraction flow. Initially, a sound leakage signal with five cycles of up-sweep signal and one cycle of silence period is played, and the response is recorded. The response signal is divided into six segments and combined synchronously to obtain the frequency spectrum of the sound leakage signal using a fast Fourier transform (FFT). This frequency spectrum is convolved with the inverse sweep signal to obtain the impulse response's frequency response, which is then converted to a waveform through an inverse FFT (IFFT) to obtain the impulse response.

In this study, the signal length for one cycle of the up-sweep signal was 2,048 samples, and the sampling rate was 44.1 kHz. Therefore, the measured impulse response length is 1,024 samples ($\approx$ 23.2 ms). The ear canal is approximately 3.5 cm long, and its impulse response length is approximately 0.2 ms. During AE measurements, the space enclosed by the hand is approximately 5 cm high, and the corresponding impulse response length is approximately 0.3 ms. Therefore, in a 1024-sample impulse response, the majority is noise, which should be trimmed. Considering the signal echoes, 128 samples ($\approx$ 2.9 ms) around the peak of the impulse response are extracted. Specifically, 64 samples before and after the peak of the impulse response were extracted. This extracted signal is the impulse response to be used for learning. Next, 128 samples before and after the peak of the impulse response were extracted, Fourier transformed,

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

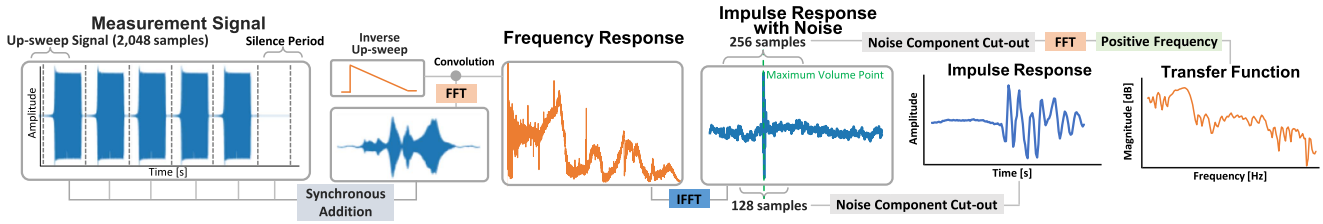


Fig. 3. The time synchronous averaging method and feature extraction flow.

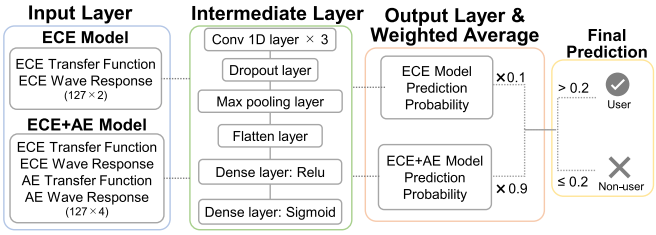


Fig. 4. The architecture of the convolutional neural network model of the proposed method.

and then the positive frequency components were extracted to obtain the transfer function.

C. Model Generation

1) Outlier Removal: Collected data includes outliers data from external noise and data loss during recording. Outlier data are unnecessary for training classification models. We used the isolation forest algorithm [44] for the outlier detection model. The system used impulse response and transfer functions as features and labeled them so that 5% of the data set was considered outliers and removed from the data set.

2) Data Augmentation: We enhanced our dataset by introducing synthetic noise from nine diverse sources, including daily life sounds (chatting, door closing), environmental sounds (cars, wind, cafe), musical sounds (singing, violin, orchestra), and white noise. We randomly varied the signal-to-noise ratio (SNR) between 10 and 20 dB for each data augmentation. This process expanded the original dataset by a factor of 25. We investigated the contribution of accuracy improvement through data augmentation in Section VI-D1.

3) Model Training: The dataset, which has undergone preprocessing, outlier removal, and data augmentation, is used to generate an authentication model that performs a binary classification of users and non-users. Fig. 4 shows the architecture of the authentication model. The authentication model employed a multistream model, using the weighted average of the predictions of two different CNN models as the final prediction. When two different echoes are obtained, several configuration patterns can be considered for model construction. These include combining each echo with its own model or integrating both echoes into a single input for a unified model. We adopt a method that integrates the predictions of the two models: the ECE model with 127×2-dimensional data using the transfer function and impulse response of ECE as the input layer and the ECE+AE model with 127×4-dimensional data using the transfer function and impulse response of ECE and AE as the input layer. The final prediction probability is the sum of the prediction probability of the ECE model multiplied by 0.1 and the prediction probability of the ECE+AE model multiplied by 0.9. If this final prediction probability is above the threshold, the wearer is authenticated as a user; if it is below the threshold, the wearer is rejected as a non-user. In this study, the threshold is set at 0.2. The evaluation experiment for model selection was conducted in Section VI-D2.

D. Authentication Process

To enhance robustness against transient noise, the final prediction is determined by a majority vote of the predictions from multiple measurements. Specifically, the features of ECE are measured three times, followed by three measurements of AE. The signal playback time for a total of six measurements is approximate 1.67 s. The features obtained from these measurements are paired for each of the three instances, and predictions are made for each pair. The final label is determined by the majority vote among the three predicted labels. The number of measurements was decided based on the results presented in Section VI-F.

IV. Implementation

The implementation of our system comprises hardware, which transmits measurement signals and records the sound leakage signals, and software for transmitting and receiving audio data and user authentication.

A. Hardware

While the system assumes the use of a commercially available earphone's built-in external microphone, we attach an additional microphone to evaluate performance over a wider frequency range. This is because built-in microphones in commercial earphones typically apply low-pass filtering (below 8 kHz), limiting the capture of higher-frequency signals.

The device type of hearables is another issue since our user authentication is based on sound leakage from hearables, which differs among types of hearables. Fig. 5 shows the device used in this study. There are four major device types in hearables: canal type, in-ear type, open-ear type, and bone-conduction type. Canal-type devices are mounted with silicone ear tips that cover the ear canal opening and are highly airtight; they are often adopted in devices with noise-canceling functions, such as Sony's wf-1000xm4 and Apple's AirPods Pro. In-ear type devices need to be inserted into the ear canal opening, but they are

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.



Fig. 5. Devices types investigated in this study. A: in-ear type, B: open-ear type, and C: bone-conduction type.

less airtight than canal types because they do not have an ear tip. Examples of such products are Anker's Soundcore Life Note 3S and Huawei's FreeBuds 5. Open-ear type devices are less airtight because they are worn on the auricle and thus do not block the ear canal opening; JVCKENWOOD's Victor HA-NP35 T are examples. Bone-conduction devices transmit sound by vibration and thus do not block the ear canal opening as in the open-ear type, such as Shokz's OpenMove.

Since preliminary investigations have shown that canal-type devices (Apple: Airpods Pro) leak very little sound, we used in-the-ear, open-ear, and bone-conduction devices to study the performance of our system. We used Anker's Soundcore Life Note 3S (in-ear type), JVCKENWOOD's Victor HA-NP35 T (open-ear type), and Shokz's OPENMOVE (bone-conduction type) for signal playback. The microphone was an audio-technica AT9904 and a ZOOM F1 audio interface to convert signals from analog to digital. The microphone is positioned so the line does not interfere when worn, and the sound collection part is placed around the built-in external microphone.

B. Software

The software transmits the measurement signal and records/analyzes the response. The sampling rate for transmitting/recording the signal is 44.1 kHz. The measurement signal used in this study is an up-sweep signal with a frequency transition from 0 to 22 kHz. One cycle of the up-sweep signal used for this measurement signal is approximately 46.4 ms (2,048 samples). Therefore, one cycle of the measurement signal, consisting of five cycles of up-sweep signals and one silence period, is 296.8 ms (12,288 samples). The signal is transmitted using a Realtek(R) Audio driver through earphones connected to a PC via Bluetooth. The signal volume should be set higher than usual to obtain high SNR sound leakage. Using an ear model and a sound level meter (SANWA SUPPLY CHESD1, frequency range: 31.5 Hz–8 kHz), we measured the signal volume with the meter's microphone positioned at the tympanic membrane. The results showed that the signal volume was approximately 75–80 dB. This is about the same noise level as that in a subway or an airplane and is not a problem for hearing health when used for short periods [45]. Signal playback was done from Audacity [46], and recording was done with Python's sounddevice [47]. The authentication model was trained using Python's TensorFlow 2.1.0 [48].

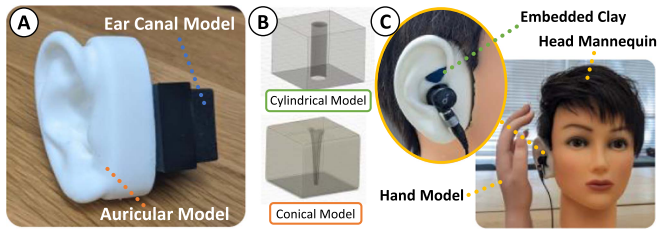


Fig. 6. Physical model for measurement experiments. A: ear and ear canal part, B: head and hand models, and C: ear canal models.

V. Understanding Device Usage and Sound Leakage Signals

In this section, we conducted a crowdsourcing survey to identify the potential user demographics for EarLock. Additionally, we performed model-based analyses and individuality evaluations to investigate the characteristics of sound leakage signals.

A. Crowdsourcing of Device Usage

To understand the prevalence of in-ear, open-ear, and bone-conduction devices among potential users of our method, we conducted a crowdsourcing survey with 332 earphone users who use these devices at least once per week. Of the respondents, 328 were Japanese, and 4 were Chinese, with a gender ratio of 243 men to 89 women. The age distribution was as follows: 92 participants aged 18–39, 180 aged 40–60, and 60 over 60 years old.

The survey results showed that 115 participants (33.9%) used in-ear devices, while 61 participants (18.0%) used open-ear or bone-conduction devices. These findings highlight that a significant proportion of users already utilize devices targeted by EarLock, indicating substantial social demand and the potential for significant impact if implemented.

B. Model Study on Sound Leakage Signals

The characteristics of the sound leakage signal used in this study have been investigated in a simplified manner by existing studies [23]. Still, the hand and head have yet to be reproduced and investigated in detail. In this experiment, we reproduced the actual usage environment of EarLock and confirmed the following three points as proof of concept.

- The ear canal and concha shape changes affect the frequency response of ECE and AE.
- Different frequency responses appear in ECE and AE depending on the presence or absence of a hand.
- The effect of changes in hand shape on frequency response.

Fig. 6(A) shows the physical measurement model used in the investigation of this experiment. The model consists of an auricle model and an ear canal model, and the ear canal model can be replaced. The ear canal models, available in cylindrical

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

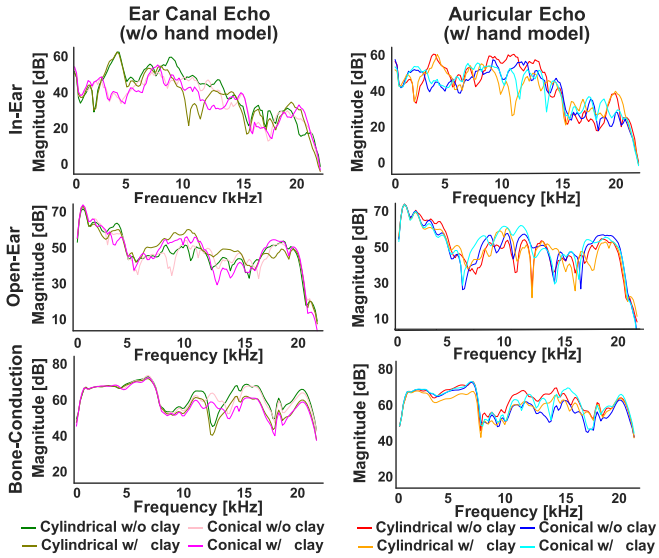


Fig. 7. Measurement results for each device. The left row shows ear canal echoes and the right row shows auricle echoes.

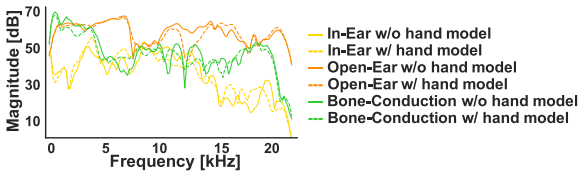


Fig. 8. Change in sound leakage signal with and without hand model. The solid line indicates ECE and the dashed line indicates AE.

and conical forms as shown in Fig. 6(B), were fabricated using a 3D printer from a rubber-like material (Agilus30 shore hardness 50 [49]).

Deformation of the concha is reproduced with and without clay embedding as shown in Fig. 6(C). The measurement model was fitted into a mannequin to reproduce the entire head. The silicon model of the hand was placed as shown in Fig. 6(C). The ears and hands are silicone. The head mannequin is constructed of polyvinyl chloride.

Fig. 7 summarizes the changes in echoes due to geometrical differences in the ear canal and auricle for each device. The left column is measured without the hand model and shows the sound leakage signal characteristics assuming ECE. The hand model measures the right column and shows the sound leakage signal characteristics assuming AE. Each graph plots a combination of ear canal and concha deformation patterns. Within each graph, all four responses were found to be different. Thus, the shape changes in the ear canal and concha affected both ECE and AE. Fig. 8 summarizes the change in echoes with and without the hand model under the same model conditions (cylindrical ear canal and no embedding to the concha). The results with the hand model show a different response than those without the hand model, confirming that different acoustic data can be obtained.

To investigate the effect of hand shape on sound leakage, we reproduced three different hand shapes as shown in the upper left of Fig. 9. Hand Shape 1 (HS1) replicates the hand shape specified

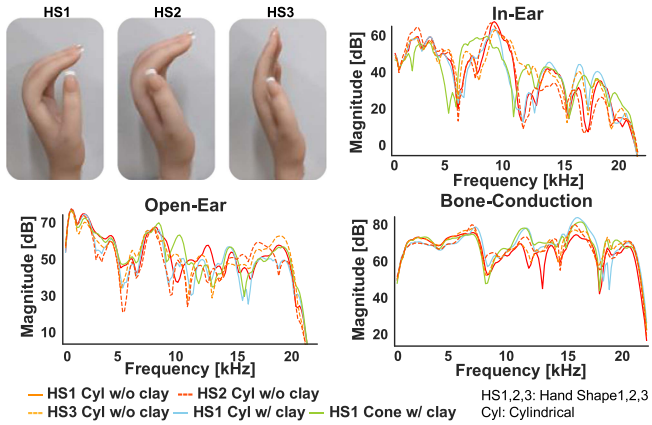


Fig. 9. Hand shape patterns (top left) and each frequency response of sound leakage signal.

for participants in the experiment. Hand Shape 2 (HS2) modifies HS1 by changing the position of the thumb. Hand Shape 3 (HS3) further extends all fingers except the thumb. For each hand shape, sound leakage signals were measured with the ear canal model in a cylindrical shape and without clay embedding. To compare with changes in ear canal and auricle shape, measurements were conducted under HS1 conditions in two states: with a cone ear canal model without clay embedding and with a cone ear canal model with clay embedding. Fig. 9 shows the frequency response variations for each device type under changes in hand shape, ear canal, and auricle shapes. The results indicate that, in certain frequency bands, the response variations due to changes in hand shape were greater than those caused by changes in ear canal and auricle shapes. Therefore, accurately replicating hand shape is essential for stable measurements.

These results indicate that EarLock is capable of acquiring individual ear canal and concha shape differences from the sound leakage signal. Additionally, ECE and AE show different responses and different acoustic information can be obtained from the same ear structure. However, it was also observed that variations due to differences in hand shape have a considerable impact.

C. *Individuality Assessment of Sound Leakage Signals*

The model study above confirmed that the sound leakage signal includes a complex combination of reflection characteristics from the ear canal, auricle, and hand shapes. In this section, we investigate the degree of individuality inherent in each of these anatomical features. Numerous studies have demonstrated the individuality of the ear canal, highlighting its strong discriminative capabilities. Building on this, our study focuses on the auricle and hand shapes to explore biometric potential. Fig. 10 illustrates the setup for experiments designed to capture the unique biometric properties of the auricle and hand independently. To isolate the reflective characteristics of the auricle, participants wore earplugs and covered their ears with a cup equipped with a speaker and microphone. To capture the reflective characteristics of the hand, a speaker and microphone were placed in the center of a box, and participants covered the

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

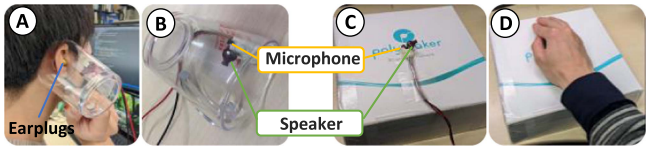


Fig. 10. Individuality assessment of sound leakage signals. A: measurement of auricle individuality test, B: measurement equipment for auricle individuality test, C: measurement equipment for hand individuality test, D: measurement scene.

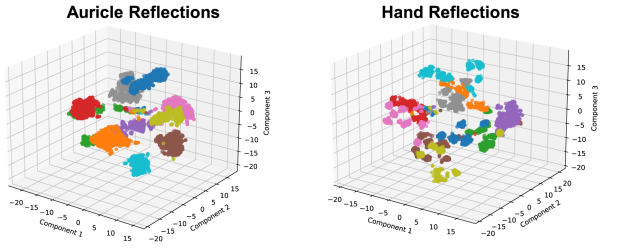


Fig. 11. t-SNE distribution of the echo sound of the auricle reflection and the hand reflection.

box with their hands. These measurements were performed on 10 participants (10 males, mean age 23.2 years). The measurement signal was an upward sweep signal ranging from 0 to 22 kHz over 0.1 s, recorded over eight trials of 5 s each. Using the collected data, we developed and evaluated individual identification models. We utilized a gradient boosting algorithm (LightGBM) for model training and Mel-Frequency Cepstral Coefficients as features. The experimental results showed that the individual identification accuracy for auricle reflections was 96.3%, while that for hand reflections was 93.4%. Fig. 11 presents the t-SNE distribution of samples for each reflection type. The distribution of auricle reflections shows clear separation by a participant, while the distribution of hand reflections exhibits more variance. These findings suggest that the discriminative features of the sound leakage signal derive not only from the ear canal shape but also from the auricle and hand shapes.

VI. Evaluation

We evaluate our system by investigating the authentication performance of practical data obtained in various environments.

A. Outline of Experiment

We collected ECEs and AEs from 23 righthanded participants (13 males and 10 females) between 21 and 67 years. The mean age was 32.2 years ($SD = 15.3$). We experimented in a quiet room. We asked participants to wear the device in their right ear and to sit in a chair, remove earrings, rings, and glasses, and use a hairpin or barrette to keep their ears covered if their hair covered their ears. First, the measurement signal was continuously transmitted, and 17 ear canal echoes (approximately five seconds) were acquired. Then, as shown in Fig. 12, we asked participants to cover the entire right ear with their right hand while the measurement signal was continuously transmitted,



Fig. 12. The user at the time of acquiring auricular echo.

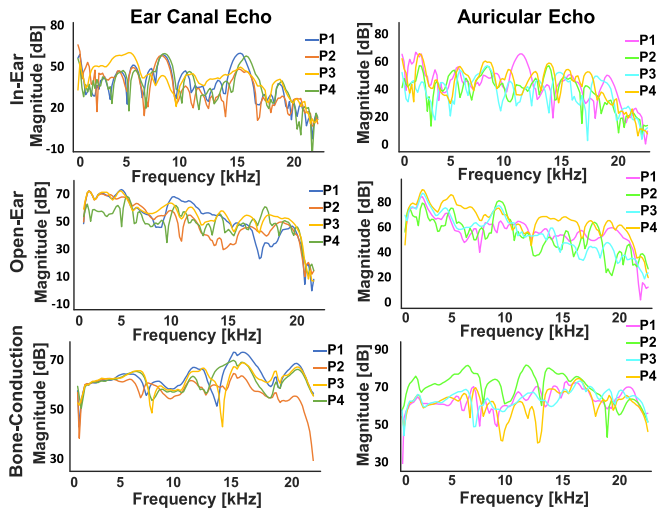


Fig. 13. The transfer functions of ECE and AE for each device.

and 17 auricular echoes were acquired. They kept the shape of the hand constant throughout the experiment. These measurements were considered one round, and between each round, we asked the participants to remove the device and return it to its original position. We collected 15 rounds of data from each participant and used them as standard data. This is data measured under stable conditions, and assumes a general measurement environment during initial setup. We obtained standard data and then recorded practical data that included noise, walking movements, objects, the effects of time passing, and the effects of the angle of the device, assuming a real-world environment. The details of each measurement are explained in each section. The experiment took about two hours for each participant, and we paid approximately 20 US dollars as a reward. This experiment was approved by the ethical board at the author's institution, and informed consent was obtained from the participants. Fig. 13 presents the transfer functions of four experimental participants (P1–P4). The responses are different for each participant.

B. Authentication Performance Under Stable Environment

To investigate the baseline authentication performance of our system, we evaluated the accuracy of leave-one-round-out cross-validation learning using standard data. Specifically, for each classification model evaluation, we designated one participant as the user and two as unknown data (non-user) to create the test

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

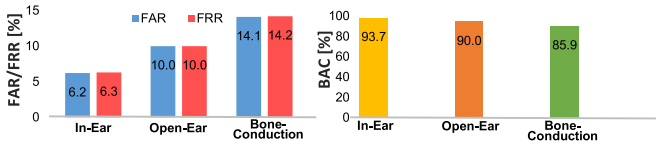


Fig. 14. Baseline authentication performance for each device. A: The false acceptance rate (FAR) and the false rejection rate (FRR), B: The balanced accuracy (BAC).

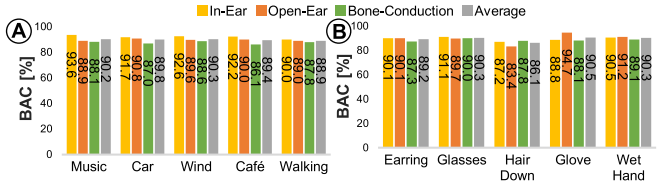


Fig. 15. Results of external factor impact study. A: Noise and walk test results, B: object test results.

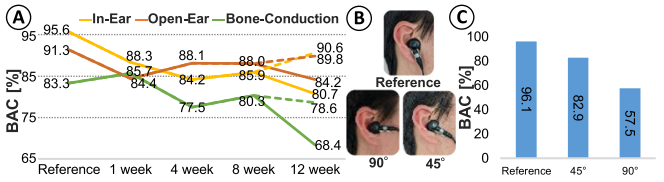


Fig. 16. Results of the time impact study and the angle impact study. A: Time test results, B: device mounting angle, C: angle test results.

dataset. The user data needed to be divided into training and test data, using one set from the 15 standard data sets as test data and the remaining 14 sets for training. For the non-user training data, we used data from the remaining 20 participants out of the 23. This train-test split resulted in an imbalanced dataset between the user and non-user labels. Therefore, we down-sampled the non-user label data to balance the dataset. False acceptance rate (FAR) and false rejection rate (FRR), as well as their balancing accuracy (BAC), were calculated as evaluation measures.

Fig. 14 shows the FAR, FRR, and BAC scores for each device. The experimental results showed that the FAR and FRR for each device were 6.2%/6.3% (in-ear), 10.0%/10.0% (open-ear), and 14.1%/14.2% (bone-conduction), respectively. BAC scores were 93.7% (in-ear), 90.0% (open-ear), and 85.9% (bone-conduction), respectively. Authentication performance of the bone-conduction type was low. EarLock obtains shape information of the ear by reflecting air-conducted sound to the auricle and the ear canal for authentication. Bone-conduction earphones vibrate the bones and the air, so it is possible to measure the air-conducted sound. However, because the earphones and microphone are attached, the vibration propagates directly to the microphone side, which is considered to have blocked accurate measurement of leaked sound.

C. Impact of External Factors

Our system will be used in various environments, such as noisy places, walking, and wearing accessories or glasses. Using practical data, we investigated the impact of these factors on our system.

1) Impact of Noise and Walk: We evaluate the authentication performance of practical data containing four types of noise and walking. Noise-containing data measured the sound leakage signal while we played each noise from a loudspeaker in front of the participant. The noise level near the participant was approximately 50 dB–60 dB as measured by a sound level meter. Walking data measured the sound leakage signal while the participant was walking. We evaluated the authentication model by authenticating the user and three non-users whose data were not used in the training phase.

Fig. 15(A) summarizes the BAC scores for each noise and the authentication performance during the walk. The BAC scores for each device were 90.0–93.6% (in-ear), 88.9–90.8% (open-ear), and 86.1–88.1% (bone-conduction). We did not observe significant authentication performance degradation, indicating that the noise augmentation of the learning modeling worked well.

2) Impact of Objects: We expected several objects to be worn or attached to the ears and hands daily. In this experiment, we measured when the earrings or glasses were worn and when the hair was down to investigate the impact of ear condition changes. To investigate the impact of hand condition changes, we measured when the hands were covered with gloves and when the palms of the hands were wet. The authentication process is the same as in Section VI-C1.

Fig. 15(B) summarizes the BAC scores for each object. The BAC scores for each device were 87.2–91.1% (in-ear), 83.4–94.7% (open-ear), and 87.3–90.0% (bone-conduction). The rate of decline was greater when the hair was let down, possibly due to a significant change in the reflection of the hair in the ECE condition.

3) Impact of Time: After measuring the training data, the sound leakage signals were re-measured at 1, 4, 8, and 12 weeks to investigate the time impact. Re-measurements were taken from 17 of the experiment participants. The authentication process is the same as in Section VI-C1.

Fig. 16(A) summarizes the BAC score after each elapsed time. Authentication performance declined for all devices, with a 15.4-point (in-ear type), 7.0-point (open-ear type), and 16.3-point (bone-conduction type) decline after 12 weeks for each device, respectively. In this experiment, participants were not given specific instructions regarding the position or angle of their hands during the re-measurement process. As a result, many participants could not accurately recall the posture in the previous measurement. Variations in the angle of the thumb alone affected the frequency response, as observed in Section V-B. This suggests that the uncertainty in reproducing the hand shape contributed to the decrease in accuracy. Additionally, human hair grows approximately 0.3 mm to 0.4 mm per day, resulting in about 3 cm of growth over 12 weeks [50]. Some participants also had their hair cut during this period. The hair down experiment also showed that the accuracy was reduced. These factors collectively contributed to a notable reduction in accuracy.

As an additional experiment, we investigated the authentication performance of the data after 12 weeks using an

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

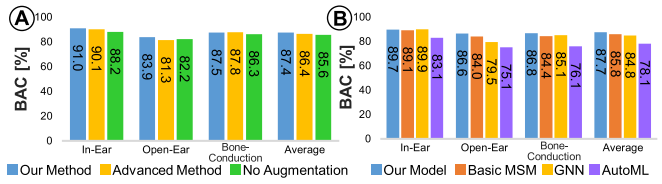


Fig. 17. A: comparison of noise augmentation methods, B: model performance comparison results.

authentication model generated by incorporating the data measured after 1, 4, and 8 weeks into the training data. The results of the experiment showed that the BAC scores for each device increased to 90.6% (in-ear), 89.8% (open-ear), and 78.6% (bone-conduction). As can be seen from this result, continuous improvement of the model is expected to prevent a decline in accuracy over time. Some studies have proposed improvements to the model over time for behavioral authentication [51], [52]. In the future, we will consider improvements using incremental learning systems to reduce the effects of variation in hand shape reproducibility and the effects of hair growth.

4) Impact of Mounting Angle: We investigated the impact of the device angle. We measured the sound leakage signal for in-ear devices with the angle shifted 45° and 90° degrees, as shown in Fig. 16(B). We did not investigate open-ear and bone-conduction devices because changing the device angle was impossible. The authentication process is the same as in Section VI-C1.

Fig. 16(C) summarizes the BAC scores for each mounting angle. When the device angle was 45 degrees, the BAC score was 82.9%; at 90 degrees, 57.5%. The insertion part of the in-ear device is not circular. As a result, the fitting condition of the ear canal changes with the angle of the device, which changes the characteristics of the leakage signal. Future actions include incorporating data from different angles into the training data set to increase the model's versatility or integrating systems that guide users to the correct fitting using sound leakage signals or IMU sensors.

D. Analysis of Noise Augmentation and Model Comparison

1) Impact of Noise Augmentation: We investigated the extent to which the proposed system's noise augmentation method contributes to authentication performance. In this section, we compare three approaches: no noise augmentation, the proposed noise augmentation method, and an advanced noise augmentation method. The advanced noise augmentation method divides the frequency range into high (13.2–22 kHz), mid (4.4–13.2 kHz), and low (0–4.4 kHz) frequency bands, and applies random noise addition rates to each band, along with reverberation-based noise augmentation. Fig. 17(A) shows the average performance for each device and the overall average for each method. Experimental results confirmed that the proposed noise augmentation method achieved a 1.87-point improvement in accuracy compared to no noise augmentation. Additionally, the proposed method, which employs simple SNR adjustments for noise augmentation, was found to be more effective than the advanced noise augmentation method. This result suggests that the current simple noise augmentation method effectively enhances model robustness by introducing general variability without overfitting to complex noise patterns, which may have occurred with the advanced noise augmentation.

2) Model Comparison: In this study, a comparison of learning models was conducted, focusing on four types of models: automated machine learning (AutoML), graph neural networks (GNN), the proposed method (multistream model), and a basic multistream model (Basic MSM). The characteristics and expected benefits of each model are as follows.

First, AutoML was adopted as an approach to explore the optimal combination of learning models and parameters using the Flaml framework [53]. We used a candidate set of seven algorithms, including random forest, gradient boosting, decision trees, support vector machines, and neural networks, to select the best model. Next, the GNN model was designed to capture the continuity and similarity between adjacent frequency components by representing ECE and AE frequency responses as nodes, with edges connecting adjacent nodes in a graph structure. The GNN model uses GCNConv layers to extract features from adjacent nodes, with a fully connected layer allowing binary classification. The Basic MSM trains CNN models on ECE and AE frequency responses and integrates the outputs of each model. In contrast, our multistream model combines the output of a CNN model trained only on ECE with a CNN model and on both ECE and AE, allowing it to efficiently exploit the complementary features of both frequency responses to improve authentication performance.

The results of the model comparison are shown in Fig. 17(B). On average, our multistream model achieved the highest accuracy, showing a 1.9-point improvement over the basic MSM. On the other hand, the GNN model showed lower accuracy for open-ear devices and exhibited variability in response to the training data compared to other models. These results indicate that the proposed multistream model could achieve stable authentication performance by effectively exploiting the complex information within the ECE and AE frequency responses.

E. Resistance to Replay Attacks

We investigated the system's resistance to replay attacks by measuring the sound leakage signal during authentication. A microphone identical to the one used in the prototype device was placed approximately 20 cm behind each participant to record the ECE and AE signals. Among the data collected from 10 participants, 6.9% of the samples were incorrectly accepted as belonging to the registered user. Considering that the current proposed method has a false rejection rate of approximately 10% for attacker data, this error rate is comparable, suggesting that the impact of eavesdropping attacks is minimal.

F. Number of Measurements and Authentication Performance

Our system measures ECE and AE three times each and performs user authentication. The fewer measurement times, the higher the usability since the time required for authentication decreases. By contrast, the greater the number of measurements,

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

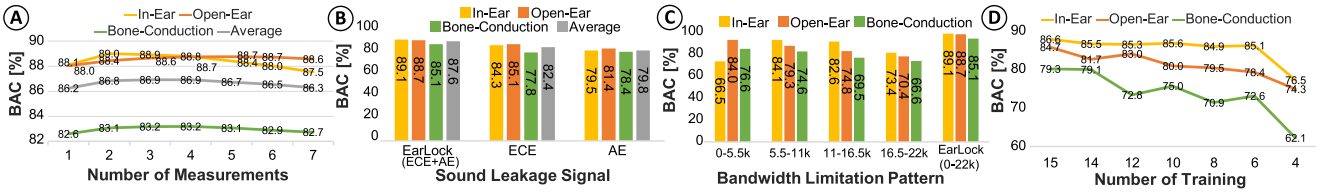


Fig. 18. A: Change in BAC score for each device when the number of measurements is varied from 1 to 7. The total signal playback time per measurement is approximately 0.56 s. B: BAC scores from the ECE-only and the AE-only data, C: BAC scores with each bandwidth limitation, D: Number of training data sets, and BAC score trends.

the less susceptible to noise and the more stable the authentication. In this experiment, we investigated the relationship between the number of measurements and authentication performance.

The number of measurements and BAC are summarized in Fig. 18(A). From the results, we determined that the number of measurements in this study would be three.

G. Authentication Performance of Each Sound Leakage Signal

We investigated the authentication performance of each sound leakage signal individually. The authentication performance of each signal is summarized in Fig. 18(B). The BAC scores were higher for ECE, averaging 82.4%, compared to AE, which averaged 79.8%, across all device types. ECE is highly individualized because it primarily includes the acoustic properties of the ear canal, which are also used in existing studies. Moreover, AE measurements' susceptibility to slight deviations from hand shape reproduction impacts signal stability. Combining both sound leakage signals effectively increased the BAC score to 87.6%. These findings illustrate that each sound leakage signal uniquely contributes to identifying the participant.

H. Authentication Performance With Bandwidth Limitation

We investigated the authentication performance of the sound leakage signal in each frequency band, examining the contribution of each frequency band to authentication performance as well as the potential use of high-frequency bands to reduce auditory strain on users. Specifically, we applied band-pass filtering to the sound leakage signal in the 0–5.5, 5.5–11, 11–16.5, and 16.5–22 kHz ranges, and used these filtered signals to train and evaluate our model. Fig. 18(C) shows the authentication performance for each frequency band. The authentication performance in the low-frequency band (0–5.5 kHz) was highest for the open-ear device. As shown in Fig. 13, the sound pressure in the low-frequency range was also the highest for the open-ear device, suggesting that stable measurements were achievable. For bone-conduction devices, authentication performance across all bands was around 75% or lower, likely due to the impact of device vibrations across the entire frequency range, as discussed in Section VI-B. For the in-ear device, performance in the low-frequency band was quite low. This may be because in-ear devices have a sealed structure within the ear canal, and when the signal leaks through small gaps, the longer-wavelength low-frequency components are less likely to pass through. As a result, these low-frequency components tend to repeatedly reflect and resonate within the ear canal, potentially leading to a loss of individual characteristics. The 16.5–22 kHz frequency range is largely inaudible to users, minimizing interference with music/voice listening. However, the authentication performance using only this range was currently around 70%, which was insufficient.

I. Impact of Training Data Size on Authentication Performance

We investigate the relationship between the amount of training data and authentication performance. The accuracy of the learning model depends on the amount of training data, though if authentication can be achieved with less training data, the burden on users for data collection can be reduced. We evaluated the BAC as the number of training data sets was reduced from 15 to 4 sets.

As shown in Fig. 18(D), a decrease in accuracy of over 5% from the baseline (15 sets) was observed in each device type when the training data was reduced to 4 sets for in-ear devices, 10 sets for open-ear devices, and 12 sets for bone-conduction devices. While the rate of accuracy reduction varied by device type, we observed a significant drop in accuracy across all devices when the number of training data sets reached 4 rounds.

J. Authentication Performance of a Different Model Device of the Same Type

In this section, we investigated the authentication performance of sound leakage signals measured from different models within the same device type. If authentication across models of the same device type is feasible, users would not need additional data measurements or model training even when switching between devices or using multiple models. Experiments were conducted using Apple's Earpods (in-ear type), JVC Kenwood's Victor HA-NP50 T (open-ear type), and audio-technica's ATH-CC500BT (bone-conduction type). To this end, we collected and evaluated ECE and AE from 10 participants (six males and four females, with an average age of 24.1 years), three times each.

Consequently, the false acceptance rates for each device type were 90.2% (in-ear type), 88.3% (open-ear type), and 95.7% (bone-conduction type). These findings indicate that the characteristics of sound leakage signals vary significantly even among models of the same type, with considerable differences in frequency responses observed. It is likely that these variations are due to differences in housing shape, fit, and speaker characteristics. This confirms that users would need to acquire data for model training specific to each device.

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

VII. Discussion

We discuss system improvements, resistance to attacks, and future potential for multimodal usage.

A. System Improvement

1) Authentication Performance: The authentication performance was several points lower than that of the method using IFM to acquire the acoustic characteristics of the ear canal [6], [8], [16]. This is caused by the external microphone's susceptibility to ambient noise and the lack of information due to signal attenuation. For example, a method to obtain frequency response from music signals has been proposed [54]. By incorporating this method into our system, we can measure ECE over a long period of time while the user is listening to music, leading to performance improvement. Simultaneously measuring ECE from the left ear would also be a simple but effective way to improve performance. We plan to improve the authentication system in the future.

2) Usability: Our system uses sweep signals in the audible range. This is a limitation due to the frequency properties of the speaker and microphone. In principle, it is also possible to use ultrasonic band signals. If the use of ultrasonic band signals becomes possible with future developments in a device, it is expected that the usability of our method will improve. Although the volume of the signals was kept uniform in this study, it is also possible to reduce the signal strength in quiet environments, for example. In the future, we plan to implement an algorithm for adjusting the volume of the signal according to the surrounding environment and evaluate its effectiveness.

3) Ear Structure Deformation and Surrounding Objects: The structure of the ear remains relatively stable from ages 8 to 70 [55]. Furthermore, implementing system-side measures such as incremental model learning can mitigate the effects of gradual structural changes in the ear due to aging. However, there is a need to investigate whether the system can accommodate significant changes in ear structure caused by conditions such as cauliflower ear, which is common in fighters, or swelling due to otitis externa. Future research is needed to address these concerns.

In this study, we evaluated performance by focusing on changes in the wearer's own state, such as glasses, gloves, and hair. On the other hand, the effects of reflections from surrounding objects, which occur in places such as crowds and near walls where the device is expected to be used, have not been investigated. These effects also need to be investigated in the future.

B. Resistance to Attacks

We confirmed the effectiveness of EarLock against standard replay attacks through eavesdropping experiments, as reported in Section VI-E. Nonetheless, more advanced replay attacks could be explored. For instance, attackers might attempt to predict the transfer function from a photograph of the ear and apply this to modify the eavesdropped sound leakage signal, crafting an attack more closely resembling the target's acoustic profile. Huang et al. [56] suggested a method to resist replay attacks through random signal decoding during authentication, and we plan to investigate the applicability of this approach within EarLock to improve robustness against more sophisticated replay attacks.

Additionally, we recognize the potential for spoofing attacks, which could involve scanning the geometry of the ear and ear canal to create a 3D replica. Previous research has explored these methods by acquiring the acoustic characteristics of the ear canal for authentication [6]. They reported that even with a physical replica, the variation in reflection characteristics due to material differences creates significant challenges for fraudulent authentication. We expect similar challenges to apply to our system.

In addition, allowing for flexible hand shapes during authentication could serve as an added layer of security. While participants in this study were asked to maintain a consistent facial expression and hand shape during the ECE and AE acquisition, introducing the option to freely configure hand shapes may function similarly to a password, adding further customization to the system's authentication process. Such an approach can potentially strengthen the system's robustness against advanced replay and spoofing attacks. Future work will investigate the feasibility of integrating flexible hand shape configurations into the authentication system, assessing both authentication performance and resilience against attacks.

As an advanced countermeasure, multi-factor authentication using biometric data could enhance the system's security against sophisticated attacks. For example, incorporating PPG or temperature sensors allows the system to verify the user's live biological signals—such as blood flow or temperature variations—thus making it challenging for a silicon replica or other physical model to pass authentication. This layered approach would increase the system's robustness by combining sound leakage signals with unique, real-time biometric data, effectively strengthening defenses against high-level spoofing attempts.

C. Multimodal Usage

EarLock uses an external microphone as a sensor to acquire biometric characteristics contained in the leaked sound. Although our system requires an authentication action, but does not require long-term measurements. Related studies include systems requiring continuous heart rate [9], [14] and gait data [7] acquisition. These systems require consideration of the time it takes to complete authentication and the effects of attaching and detaching the device. We believe there are new prospects for our system to be used cooperatively with such continuous authentication systems. First, since each system acquires different physical and behavioral characteristics, authentication performance is expected to be improved by using both biometric characteristics. It is also considered possible to prevent the degradation of authentication performance by using the sound leakage signal even when continuous measurement is insufficient. In addition, usability could be improved by recording only ECE for the sound

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

leakage signal used in this system when continuous measurement provides sufficient authentication assurance. In the future, we will consider the development of a multimodal authentication system and its effective use method.

VIII. Conclusion

In this study, we proposed "EarLock," a system to authenticate the wearer of hearables by acquiring ear canal and auricular echoes from the sound leakage signal of the hearables. A model study confirmed that changes in the auricle and ear canal shape affect the sound leakage signal. Experiments with in-ear, open-ear, and bone-conduction devices resulted in BAC scores of 87.2–93.7%, 83.4–94.7%, and 85.9–90.0% under multiple conditions (stable environment and motion/sound/reflected-object noise). These results suggest that the EarLock system can provide a reasonably effective method for user authentication, potentially adapting well to various types of devices and environmental conditions.

References

[1] J. Plazak and M. Kersten-Oertel, "A survey on the affordances of "hearables"," *Inventions*, vol. 3, no. 3, pp. 1–16, 2018. [Online]. Available: https://www.mdpi.com/2411-5134/3/3/48

[2] A. Ferlini, A. Montanari, C. Min, H. Li, U. Sassi, and F. Kawsar, "In-ear PPG for vital signs," *IEEE Pervasive Comput.*, vol. 21, no. 1, pp. 65–74, First Quarter, 2022.

[3] D. D. He, E. S. Winokur, and C. G. Sodini, "An ear-worn vital signs monitor," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2547–2552, Nov. 2015.

[4] R. R. Choudhury, "Earable computing: A new area to think about," in *Proc. 22nd Int. Workshop Mobile Comput. Syst. Appl.*, 2021, pp. 147–153, doi: 10.1145/3446382.3450216.

[5] Y. Cao, C. Chao, A. Yu, F. Li, and J. Luo, "EarAcE: Empowering versatile acoustic sensing via earable active noise cancellation platform," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 7, no. 2, 2023, Art. no. 47.

[6] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin, "EarEcho: Using ear canal echo for wearable authentication," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 3, no. 3, 2019, Art. no. 81, doi: 10.1145/3351239.

[7] A. Ferlini, D. Ma, R. Harle, and C. Mascolo, "EarGate: Gait-based user identification with in-ear microphones," in *Proc. 27th Annu. Int. Conf. Mobile Comput. Netw.*, 2021, pp. 337–349, doi: 10.1145/3447993.3483240.

[8] 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," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 5, no. 1, 2021, Art. no. 39, doi: 10.1145/3448098.

[9] Y. Cao, C. Cai, F. Li, Z. Chen, and J. Luo, "HeartPrint: Passive heart sounds authentication exploiting in-ear microphones," in *Proc. 2023 IEEE Conf. Comput. Commun.*, 2023, pp. 1–10.

[10] Z. Wang, Y. Ren, Y. Chen, and J. Yang, "ToothSonic: Earable authentication via acoustic toothprint," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 6, no. 2, 2022, Art. no. 78, doi: 10.1145/3534606.

[11] Y. Xie, F. Li, Y. Wu, H. Chen, Z. Zhao, and Y. Wang, "TeethPass: Dental occlusion-based user authentication via in-ear acoustic sensing," in *Proc. 2022 IEEE Conf. Comput. Commun.*, 2022, pp. 1789–1798, doi: 10.1109/INFOCOM48880.2022.9796951.

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

[13] R. Liu, C. Cornelius, R. Rawassizadeh, R. Peterson, and D. Kotz, "Vocal resonance: Using internal body voice for wearable authentication," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 1, 2018, Art. no. 19, doi: 10.1145/3191751.

[14] S. Choi et al., "PPGface: Like what you are watching? Earphones can "feel" your facial expressions," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 6, no. 2, 2022, Art. no. 48, doi: 10.1145/3534597.

[15] S. Choi, J. Yim, Y. Jin, Y. Gao, J. Li, and Z. Jin, "EarPPG: Securing your identity with your ears," in *Proc. 28th Int. Conf. Intell. User Interfaces*, New York, NY, USA, 2023, pp. 835–849, doi: 10.1145/3581641.3584070.

[16] X. Sun et al., "Earmonitor: In-ear motion-resilient acoustic sensing using commodity earphones," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 6, no. 4, Jan. 2023, Art. no. 182, doi: 10.1145/3569472.

[17] Y. Gao, Y. Jin, J. Chauhan, S. Choi, J. Li, and Z. Jin, "Voice in ear: Spoofing-resistant and passphrase-independent body sound authentication," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 5, no. 1, Mar. 2021, Art. no. 12, doi: 10.1145/3448113.

[18] H. Bi, Y. Sun, J. Liu, and L. Cao, "Smartear: Rhythm-based tap authentication using earphone in information-centric wireless sensor network," *IEEE Internet of Things J.*, vol. 9, no. 2, pp. 885–896, Jan. 2022.

[19] N. J. Grabham et al., "An evaluation of otoacoustic emissions as a biometric," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 1, pp. 174–183, Jan. 2013.

[20] Y. Mizuho, Y. Kawasaki, T. Amesaka, and Y. Sugiura, "EarAuthCam: Personal identification and authentication method using ear images acquired with a camera-equipped hearable device," in *Proc. Augmented Humans Int. Conf.*, New York, NY, USA, 2024, pp. 119–130, doi: 10.1145/3652920.3653059.

[21] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM Comput. Surv.*, vol. 45, no. 2, Mar. 2013, Art. no. 22, doi: 10.1145/2431211.2431221.

[22] Apple, "*Automatic Ear Detection*," 2024. Accessed: Nov. 04, 2024. [Online]. Available: https://support.apple.com/en-us/108764#::text=By&20default&2C&20AirPods&20sense&20when,resume&E2&80&94when&20you&20remove&20both

[23] T. Amesaka, H. Watanabe, M. Sugimoto, Y. Sugiura, and B. Shizuki, "User authentication method for hearables using sound leakage signals," in *Proc. ACM Int. Symp. Wearable Comput.*, New York, NY, USA, 2023, pp. 119–123, doi: 10.1145/3594738.3611376.

[24] H. Feng, K. Fawaz, and K. G. Shin, "Continuous authentication for voice assistants," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, 2017, pp. 343–355, doi: 10.1145/3117811.3117823.

[25] Y. Zou, H. Lei, and K. Wu, "Beyond legitimacy, also with identity: Your smart earphones know who you are quietly," *IEEE Trans. Mobile Comput.*, vol. 22, no. 6, pp. 3179–3192, Jun. 2023, doi: 10.1109/TMC.2021.3134654.

[26] T. Nakamura, V. Goverdovsky, and D. P. Mandic, "In-ear EEG biometrics for feasible and readily collectable real-world person authentication," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 3, pp. 648–661, Mar. 2018.

[27] H. Shin et al., "SkullID: Through-skull sound conduction based authentication for smartglasses," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2024, Art. no. 834, doi: 10.1145/3613904.3642506.

[28] T. Srivastava, S. Pan, P. Nguyen, and S. Jain, "Jawthenticate: Microphone-free speech-based authentication using jaw motion and facial vibrations," in *Proc. 21st ACM Conf. Embedded Netw. Sensor Syst.*, 2024, pp. 209–222, doi: 10.1145/3625687.3625813.

[29] J. Middlebrooks and D. Green, "Sound localization by human listeners," *Annu. Rev. Psychol.*, vol. 42, pp. 135–159, 1991.

[30] D. J. Hurley, B. Arbab-Zavar, and M. Nixon, "The ear as a biometric," in *Handbook of Biometrics*, Berlin, Germany: Springer, 2008, pp. 131–150.

[31] C. Holz, S. Buthpitiya, and M. Knaust, "Bodyprint: Biometric user identification on mobile devices using the capacitive touchscreen to scan body parts," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, 2015, pp. 3011–3014, doi: 10.1145/2702123.2702518.

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

[33] A. Sulavko, "Biometric-based key generation and user authentication using acoustic characteristics of the outer ear and a network of correlation neurons," *Sensors*, vol. 22, no. 23, pp. 1–27, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/23/9551

[34] X. Song, K. Huang, and W. Gao, "FaceListener: Recognizing human facial expressions via acoustic sensing on commodity headphones," in *Proc. 21st ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, 2022, pp. 145–157.

[35] Y. Cao, F. Li, H. Chen, X. Liu, and Y. Wang, "Tongue-jaw movement recognition through acoustic sensing on smartphones," *IEEE Trans. Mobile Comput.*, vol. 23, no. 1, pp. 879–894, Jan. 2024.

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.

[36] S. Zhang, T. Lu, H. Zhou, Y. Liu, R. Liu, and M. Gowda, "I am an earphone and i can hear my users face: Facial landmark tracking using smart earphones," *ACM Trans. Internet of Things*, vol. 5, Aug. 2023, Art. no. 1, doi: 10.1145/3614438.
[37] K. Nagayama and R. Takada, "Unvoiced vowel recognition using active bio-acoustic sensing for silent speech interaction," in *Artificial Intelligence in HCI*, Cham, Switzerland: Springer, 2023, pp. 150–161.
[38] Y. Jin et al., "EarCommand: "hearing" your silent speech commands in ear," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 6, no. 2, Jul. 2022, Art. no. 57, doi: 10.1145/3534613.
[39] T. Srivastava, P. Khanna, S. Pan, P. Nguyen, and S. Jain, "Muteit: Jaw motion based unvoiced command recognition using earable," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 6, no. 3, Sep. 2022, Art. no. 140, doi: 10.1145/3550281.
[40] R. Zhang et al., "HPSpeech: Silent speech interface for commodity headphones," in *Proc. ACM Int. Symp. Wearable Comput.*, New York, NY, USA, 2023, pp. 60–65, doi: 10.1145/3594738.3611365.
[41] Y. Jin, Y. Gao, X. Guo, J. Wen, Z. Li, and Z. Jin, "EarHealth: An earphone-based acoustic otoscope for detection of multiple ear diseases in daily life," in *Proc. 20th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, New York, NY, USA, 2022, pp. 397–408, doi: 10.1145/3498361.3538935.
[42] A. Iannarelli, *Ear Identification*. Fremont, CA, USA: Paramount Publishing Company, 1989. [Online]. Available: https://books.google.co.jp/books?id=jgPkAAAACAAJ
[43] K. Iida, *Fundamentals of Head-Related Transfer Function and its Application to 3-D Sound System*. Tokyo, Japan: CORONA PUBLISHING CO.,LTD., 2017. [Online]. Available: https://www.coronasha.co.jp/np/isbn/9784339011333/
[44] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Mining*, 2008, pp. 413–422.
[45] World Health Organization, "Safe listening devices and systems," 2019. Accessed: Jan. 15, 2025. [Online]. Available: https://www.who.int/publications/i/item/9789241515276
[46] Audacity, "*Audacity 3.3*," 2023. Accessed: Nov. 04, 2024. [Online]. Available: https://www.audacityteam.org/
[47] sounddevice, "*sounddevice 0.4.6*," 2023. Accessed: Jan. 15, 2025. [Online]. Available: https://python-sounddevice.readthedocs.io/en/0.4.6/
[48] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: https://www.tensorflow.org/
[49] DMM.make, "*3D PRINT*," 2013. Accessed: Jan. 15, 2025. [Online]. Available: https://make.dmm.com/
[50] G. Loussouarn, I. Lozano, S. Panhard, C. Collaudin, C. El Rawadi, and G. Genain, "Diversity in human hair growth, diameter, colour and shape. an in Vivo study on young adults from 24 different ethnic groups observed in the five continents," *Eur. J. Dermatol.*, vol. 26, pp. 144–154, 2016.
[51] Z. Shen, S. Li, X. Zhao, and J. Zou, "IncreAuth: Incremental-learning-based behavioral biometric authentication on smartphones," *IEEE Internet of Things J.*, vol. 11, no. 1, pp. 1589–1603, Jan. 2024.
[52] J. Chauhan, Y. D. Kwon, P. Hui, and C. Mascolo, "ContAuth: Continual learning framework for behavioral-based user authentication," in *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 4, no. 4, Dec. 2020, Art. no. 122, doi: 10.1145/3432203.
[53] C. Wang, L. Wang, Q. Zhang, J. Tang, and D. Luo, "FLAML: A fast and lightweight AutoML library," 2021, *arXiv: 1911.04706*.
[54] H. Kawahara and K. Yatabe, "Safeguarding test signals for acoustic measurement using arbitrary sounds: Measuring impulse response by playing music," *Acoustical Sci. Technol.*, vol. 43, no. 3, pp. 209–212, 2022.
[55] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to Biometrics*. New York, NY, USA: Springer, 2011.
[56] L. Huang and C. Wang, "PCR-auth: Solving authentication puzzle challenge with encoded palm contact response," in *Proc. IEEE Symp. Secur. Privacy*, 2022, pp. 1034–1048.



Takashi Amesaka received the BE and ME degrees in engineering from Hokkaido University, in 2018 and 2020, respectively, and the PhD degree in systems and information engineering from the University of Tsukuba, in 2023. From 2021 to 2023, he was a research fellow of the Japan Society for the Promotion of Science (DC2), University of Tsukuba. Currently, he is a post doctoral researcher with Keio University and the Japan Society for the Promotion of Science (JSPS). His research interests include wearable computing and human-computer interaction.



Hiroki Watanabe received the BE, ME, and PhD degrees in engineering from Kobe University, Hyogo, Japan, in 2012, 2014, and 2017. From 2015 to 2017, he was a research fellow of the Japan Society for the Promotion of Science (DC2), Kobe University. From 2017 to 2024, he was an assistant Professor with the Graduate School of Information Science and Technology, Hokkaido University, Hokkaido, Japan. Since 2024, he has been an associate Professor with Future University Hakodate, Hokkaido, Japan. His research interests include wearable computing and ubiquitous computing.



Yuta Sugiura received the PhD degree from the Graduate School of Media Design, Keio University, in 2013. He is an associate Professor of the Department of Information and Computer Science, Keio University. Before joining the Department, he worked as a post doctoral researcher with the National Institute of Advanced Industrial Science and Technology (2015–2016). His research interest is in user interfaces and ubiquitous computing. He has received several awards including the IPSJ Prize, UIST Best Talk Award, and the Good Design Award. He served as a program committee member for various international conferences including ACM UIST, TEI, and SIGGRAPH ASIA E-tech.



Masanori Sugimoto received the BE, ME, and DE degrees in aeronautics and astronautics from the University of Tokyo, in 1990, 1992 and 1995, respectively. Currently, he is a professor with the Graduate School of Information Science and Technology, Hokkaido University, Japan. His research interests focus on acoustic engineering, signal processing, artificial intelligence and human computer interaction technologies for designing smart systems and environments.



Buntarou Shizuki received the BSc, MSc, and DSc degrees from the Tokyo Institute of Technology, in 1994, 1996, and 2000, respectively. Currently, he is a professor with the Department of Computer Science, University of Tsukuba. His research interests include human-computer interaction.

Authorized licensed use limited to: Tsukuba Univ Lib. Downloaded on October 15,2026 at 05:52:51 UTC from IEEE Xplore. Restrictions apply.