2025 IEEE 22nd International Conference on Mobile Ad-Hoc and Smart Systems (MASS) | 979-8-3315-6599-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/MASS66014.2025.00050

Converting Your Bluetooth Headphones into Active Sensing Authenticator: a Bone-Conduction Solution

Jingwei Zhang
Department of Computer Science
Southern Methodist University
Dallas, USA
jingwei@smu.edu

Ruxin Wang
Department of Computer Science
Southern Methodist University
Dallas, USA
ruxinw@smu.edu

Chen Wang
Department of Computer Science
Southern Methodist University
Dallas, USA
cwang6@smu.edu

***Abstract*—Bluetooth headphones are increasingly commonly used in daily life, offering convenience and enhanced audio experience. However, these devices remain underexplored for human-beneficial applications such as serving as authenticator due to the challenges of implementing acoustic sensing on them. Particularly, Bluetooth headphones are limited by low-frequency audio bandwidth and have built-in echo cancellation algorithms, which makes the recorded signal incomplete and cannot be used for acoustic analysis. This work addresses these challenges and achieves acoustic sensing on bone-conduction Bluetooth headphones to extract a unique bone-conducted head biometric for user authentication. Specifically, the proposed system emits a user-friendly signal consisting of a welcome tone followed by a short human voice, and analyzes the received signals using a convolutional neural network developed with residual blocks to derive stable biometrics to verify users. Extensive experiments show that the proposed system can verify users' identities with an average accuracy of 97.51% and can successfully reject 100% of replay attacks, even when an adversary eavesdrops on the authentication sound and the acoustic biometric data.**

***Index Terms*—Acoustic Sensing, User Authentication, Bluetooth Headphones, Bone-conducted Biometric**

I. Introduction

Bluetooth headphones have become widely used due to their wireless convenience, enhanced audio experience, sleek design, and compatibility with smartphones and other devices. Originally used mainly for hands-free calling, they've evolved to support high-quality audio streaming and smart assistant integration. Their popularity surged with the removal of headphone jacks from many smartphones, making them a go-to choice for music lovers, commuters, and fitness enthusiasts. Based on recent trends, the use of Bluetooth headphones continues to increase steadily [1]. This rise is further supported by advancements in Bluetooth technology, improved battery life, and features like voice assistant support.

A few studies explore leveraging passive sensing on head-worn devices for user authentication, such as verifying users with touch behavioral biometrics and voice captured by AR glasses sensors [2], or verifying users with facial dynamics captured by VR headset sensors [3]. However, these methods require active user participants, rely on specific hardware, and are generally more susceptible to environmental noise compared to active sensing approaches. Some studies focus on using active acoustic sensing on head-worn devices for user authentication. For example, one existing study uses active acoustic sensing to authenticate VR headset users by emitting an ultrasonic chirp signal and extracting head biometrics from the recorded sound [4]. Another work used white noise echoes for authentication through the bone-conduction speaker on Google Glass [5]. However, these methods are limited to the restricted frequency range and the limited signal volume due to built-in noise suppression settings. Bone-conduction head-worn devices maintain constant contact with the head, making them better suited for sensing physiological or behavioral signals. Some existing studies apply active acoustic sensing using bone-conduction transducers for biomechanics (joint angle, grip/contact force, pose, material density), but all systems use wired signal transmission and reception [6]–[8]. As we demonstrate in Section III, the built-in echo cancellation mechanism and limited audio bandwidth make it challenging to deploy active acoustic sensing on such devices.

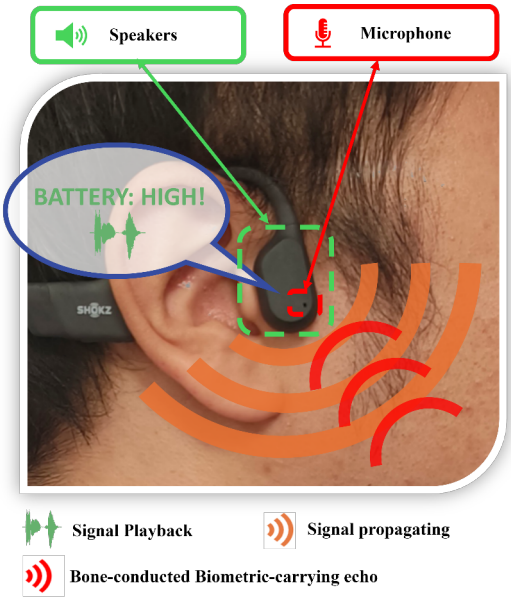


Fig. 1: Illustrating of bone-conducted signal interacting with the user's head.

In this work, we address the challenges and leverage active

2155-6814/25/$31.00 ©2025 IEEE
DOI 10.1109/MASS66014.2025.00050

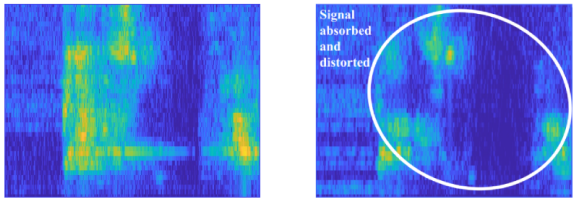
acoustic sensing on bone-conduction Bluetooth headphones to extract unique bone-conducted head biometrics for user authentication. Specifically, we design a sensing signal composed of a welcome tone followed by a very short voice prompt and develop a ResNet-based algorithm to derive stable bone-conducted head biometrics to verify a user's identity. Compared to traditional headphones and head-worn devices, bone-conduction headphones provide robust sensing capabilities in noisy environments, as they maintain continuous contact with the bone. Moreover, the specially designed sensing signal is user-friendly and unobtrusive. To counteract the effects of built-in echo cancellation algorithms, we further develop a ResNet-based model that learns the user's unique bone-conducted head biometrics by utilizing both acoustic and vibration-induced information. For practical usage, a registered user wears a bone-conduction headset with Bluetooth-paired device. The headset can either perform user authentication or serve as a biometric sensing tool for the paired device.

Our contributions are summarized as follows:

- We propose an unobtrusive and replay-resistant user authentication system on bone-conduction Bluetooth headphones. we are the first to deploy active acoustic sensing using bone-conducted media sound, on such non-standalone Bluetooth devices under the limitations imposed by noise cancellation algorithm and volume profile constraints, without any wired hardware overhead.
- We leverage the stable sensing capabilities of bone-conduction devices to address the challenges faced by active acoustic sensing algorithms in noisy environments. Furthermore, to counteract the effects of built-in noise cancellation algorithms, we extract unique bone-conducted head biometrics from received signals that contain both acoustic and vibration information.
- We develop a ResNet-based algorithm to learn unique bone-conducted head biometric for efficient user authentication. Moreover, we simulate acoustic replay attacks and demonstrate that the proposed system can efficiently defend against these replay attacks.
- Extensive experiments with fourteen participants and impact factor studies (noise study, training dataset study) show that our system is robust and efficient in user verification and can be used in practical environments.

II. Related Work

Traditional authentication methods such as PINs [9] and lock patterns [10] are widely used by head-worn devices due to their convenience and low cost deployment. However, these methods are vulnerable to security breaches caused by malware and social engineering attacks [11], [12]. Since biometric authentication does not rely on memorized information, some studies have begun leveraging unique physical or behavioral biometrics for user authentication, such as fingerprints [13], Face ID [14], and hand geometry [15]. However, physical biometrics can be easily forged or reproduced using advanced 3D technologies, making them vulnerable to replay attacks. Moreover, these methods often require active user participation or dedicated hardware. To address reproducibility concerns, some studies focus on acoustic-based authentication methods, such as using voiceprints to recognize a user's identity [16]. Nevertheless, these methods remain vulnerable to replay attacks [17]. Additionally, some studies explore enhanced user authentication using multi-modal sensing, but these approaches are costly, consume significant battery power, and are not suitable for lightweight head-worn devices.



(a) Headphones put on table. (b) User wearing headphones.

Fig. 2: User's influence on sensing signal.

There have been only a few studies on using passive or active acoustic sensing with head-worn devices. For example, passive sensing for user authentication, such as verifying a user's voice [2] or capturing facial dynamics using inertial sensors [3]. However, these methods still require active user participation, dedicated hardware, and are generally more susceptible to environmental noise compared to active sensing approaches. In terms of active acoustic sensing, one study demonstrated that the frequency response of a chirp signal could be used for user authentication while wearing a VR headset [4]. Another study showed that authentication could be achieved using white noise echoes through the built-in bone-conduction speaker on Google Glass [5]. There exists work about using wired earbud for sensing like EarEcho [18], and He et al. leveraged probe signals emitted by the bone-conduction speaker for user authentication, but their approach required wired data collection sensors attached to the device, meaning the authenticator could not operate independently, did not solve sensing problems under Bluetooth challenge, and using long-time chirp signal for sensing is obstructive to the user experience [19], [20]. These methods are limited to using specific frequency signals for sensing. To the best of our knowledge, there are no prior studies deploying acoustic sensing on Bluetooth-based head-worn devices for user authentication independently. A practical solution must accommodate limited user interaction and constrained system resources while also overcoming the limitations imposed by Bluetooth communication protocols.

III. Background and System Models

This work proposes to verify bone-conduction Bluetooth headphones users by capturing their bone-conducted head biometrics in the acoustic domain. Bone-conduction headphones operate by placing their transducer (speaker) in rigid contact with the user's cheekbone. Through this solid-state connection, mechanical vibrations are transmitted directly to the cochlea, effectively bypassing the outer and middle ear [21]. When

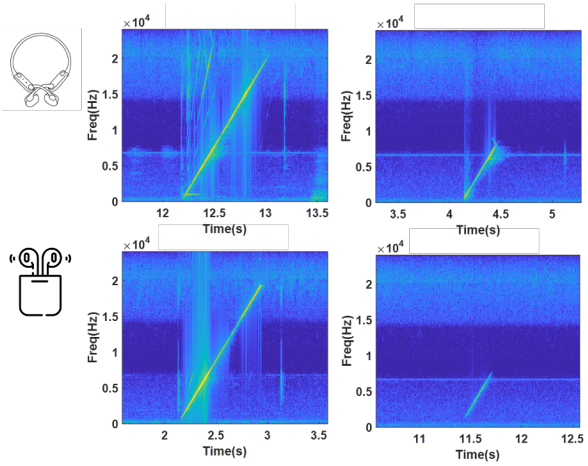


Fig. 3: Speaker capability of two device: down-streaming and dual-streaming.

users wear headphones, the headphones keep tightly contacted to the user's temporal bone, in front of the ear canal. The headphones and the human skull form a rigid body. The headphones' speaker and microphone on this rigid body create an acoustic channel, and a sound traveling on this channel would be absorbed and reflected by it, as shown in Figure 1. Because each human head has a unique size, skull structure, mass, and ear pattern, the corresponding rigid body affects the sound differently before it reaches the microphone. We thus use the headphones' speaker to emit acoustic signals and analyze its speaker-microphone channel responses to extract acoustically presented bone-conducted head biometrics. As shown in Figure 2, we observe the emitted signals are partially absorbed by the human head skull and near-ear skin when the bone-conduction headphones are worn by user. The results demonstrate the impact of a user's biometric contact on the echo signal in the time–frequency domain. The frequency-domain representation of the received signal $Y(f)$ expressed as:

$$Y(f) = X(f) \cdot H(f) + N(f) \tag{1}$$

Where $X(f)$ is the playback signal, $H(f)$ is the acoustic channel response, and $N(f)$ denotes the ambient noise. And the channel response could be divided into two parts: the influence of the environment and device $H_0(f)$, and the influence of the rigid body $H_{\text{bio}}(f)$:

$$H(f) = H_0(f) * H_{\text{bio}}(f) \tag{2}$$

To extract head-related biometric features, we analyze the received signal $Y(f)$ by suppressing the effects of background noise $N(f)$ and compensating for the system-related frequency response $H_0(f)$. This process aims to isolate the individual-specific component $H_{\text{bio}}(f)$, which reflects the user's unique biometric characteristics.

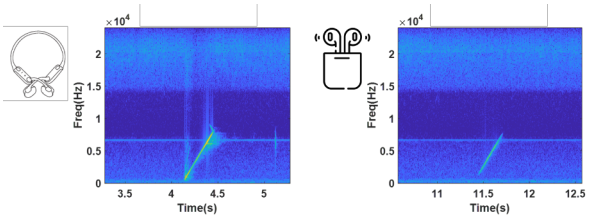


Fig. 4: Microphone capability of two device.

A. Challenges

Prior studies have used bone-conduction speakers for probing signals, but rely on external wired microphones or additional sensors for data collection [19], [20], largely due to the limitations imposed by Bluetooth audio protocols on built-in microphones. According to the official Bluetooth documentation [22], [23], bone-conduction headphones are constrained by limited audio bandwidth. While most Bluetooth headphones manufacturers claim to offer high-quality, high-sampling-rate stereo audio (often up to 48,000 Hz), this applies only in music playback mode. Bluetooth headphones typically operate in two distinct modes: the Advanced Audio Distribution Profile (A2DP), which supports high-quality audio downlink for music [22]; and the Hands-Free Profile (HFP) or Headset Profile (HSP), which enables two-way audio for voice communication but significantly reduces audio quality due to bandwidth constraints [23]. To investigate the usable audio bandwidth on bone-conduction Bluetooth headphones, we conduct two experiments to evaluate the limitations of both the speaker and microphone.

- *Vibration Speakers Test.* We study the headphones' speakers audio bandwidth limitations by playing a 1-second chirp signal ranging 0 24 kHz and recording it using an external smartphone. Since Bluetooth headphones operate under two different audio profiles, we tested two scenarios: (1) playing the chirp only, and (2) playing the chirp while simultaneously recording—i.e., activating the microphone. As shown in Figure 3, the headphones were able to reproduce the full 0–24 kHz chirp signal when operating in the A2DP profile. However, under the HFP/HSP profile—when both playback and recording are active—recorded signal lacked frequency components above 8 kHz, due to the bandwidth limitations of the profile. We also tested on regular Bluetooth earbuds under two Bluetooth profile, as shown in 3. This substantiated the fact that most Bluetooth earbuds/headphones under the two Bluetooth profiles aligns with these findings.
- *Headphone Microphone Test.* We study the headphones' microphones audio bandwidth limitations by using a smartphone to play a 1-second chirp signal ranging from 0 to 24 kHz and analyzing the recorded response. As illustrated in Figure 3, all smartphone speakers are able to reproduce the full frequency range of the signal, while the built-in microphone captured only up to 8 kHz.

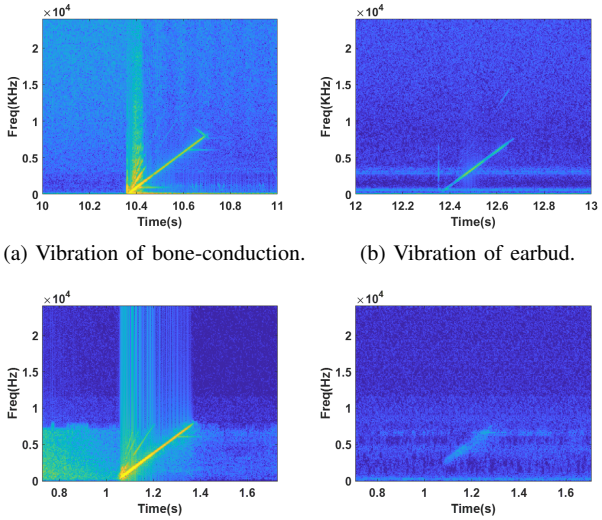
As shown in our experiments and supported by official Bluetooth specifications, Bluetooth headphones operate in

Fig. 5: Vibration measurement experiment on bone-conduction headphones and wireless earbud, with laser vibrometer.

recording mode use the HFP or HSP profile, which limits its audio transmission to 16 kHz mono [23]. According to the Nyquist–Shannon sampling theorem [24], this constraint makes it infeasible to use ultrasonic signals for acoustic sensing in such scenarios.

Audio Volume and Noise Suppression. We measured the actual playback volume of a 1-second chirp signal on two Bluetooth devices under the same volume setting but different working modes. In A2DP mode, the measured sound pressure levels were 57–60 dB and 76–77 dB, respectively. In contrast, under HSP mode, the corresponding levels dropped to 33–34 dB and 43–44 dB. The final recorded signal from a Bluetooth bone-conduction headphone is affected by several factors: The first factor is the volume of the sensing signal. To optimize user experience, host devices (e.g., smartphones or PCs) may automatically adjust playback volume through device-specific implementations in drivers, BIOS, or backend frameworks. For instance, on Android systems, these modifications are implemented at the Audio HAL layer, as well as within Android framework [25]. Similarly, headphones manufacturers may configure different acoustic parameters for different Bluetooth profiles to enhance audio quality [26]. Typically, the media playback volume is reduced, while the voice communication output is boosted to improve the experience during phone or VoIP calls. The second factor is the physical interaction between the headphones and the wearer's head. Soft tissues such as skin and hair absorb part of the mechanical vibration. While this attenuation encodes person-specific biometrics into the signal, it also significantly reduces the amplitude of the recorded signal, making the sensing task more challenging.

In our device testing experiments, we observed that chirp signals are relatively less affected by noise reduction processing. However, due to frequency bandwidth limitations, it is not feasible to utilize high-frequency ultrasonic chirps for low-perceptibility sensing, as would be ideal for minimizing user awareness. Moreover, low-frequency chirps, while feasible, were found to be annoying to users, as confirmed by additional tests. As a result, we opted to use readily available media sounds for sensing purposes. However, media playback signals are more susceptible to acoustic echo cancellation (AEC) effects, as discussed in the next section. Despite these challenges, our system remains capable of performing reliable sensing under AEC conditions.



(a) Vibration of bone-conduction. (b) Vibration of earbud.

(c) Bone-conduction's echo. (d) Earbuds' echo.

Fig. 6: Vibration test result.

B. Bone-conducted Biometric Advantage

Compared to air-conducted acoustic waves from conventional headphones, the solid-state vibrations generated by bone-conduction headphones propagate more effectively through the speaker–head–microphone structure due to superior impedance matching and lower energy loss through solids [27]. We designed an experiment to compare the solid-state vibrations produced by a standard Bluetooth earbud and bone-conduction headphones under the same volume setting. A Polytec laser vibrometer was used to measure the vibrations of the headphone speakers while they played a 1-second chirp signal ranging from 0 to 24 kHz. As shown in Figure 6, the bone-conduction headphones generated significantly stronger vibrations and a broader range of detectable frequency components compared to the conventional earbud.

C. System Overview

Our system architecture is illustrated in Figure 7. After the user wears the bone-conduction headphones, which are paired with a smartphone, the system initiates an authentication session by playing a ringtone followed by a voice prompt such as "connected" or "battery high." As discussed in the previous section, when operating under the HFP/HSP profile, the playback volume is typically reduced compared to the A2DP profile. This adjustment is intended to improve voice call quality, but it results in a recorded signal with insufficient magnitude for extracting reliable acoustic biometric features. To address this issue, we employ the Android API `LoudnessEnhancer` to boost the playback volume to a level comparable to that under A2DP [28]. In practical scenarios, this authentication session is typically triggered when the user powers on the device or checks the battery level by pressing a button. The headphones' microphone

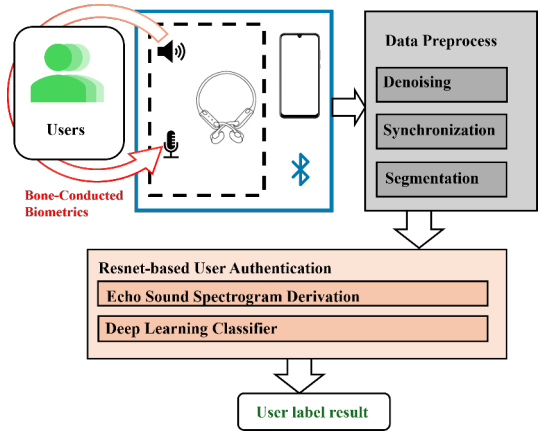


Fig. 7: System architecture.

captures the bone-conducted signal, now enhanced in volume, which contains the user's head-related acoustic biometrics. We then perform data preprocessing, which includes filtering, synchronization, and segmentation.

After preprocessing, we apply a residual block–based classifier as our authentication algorithm [29]. Originally designed for image classification tasks, this ResNet-style architecture has demonstrated strong capability in capturing subtle visual differences between spectrograms. In our study, the algorithm effectively extracts unique biometric features from the spectrograms of preprocessed signals, with an thresholding step applied to further refine binary classification decisions. This helps mitigate the impact of noise, enhances robustness for long-term use, and increases the difficulty of replay attacks. The model can be deployed across various platforms, including smartphones, remote servers, and IoT devices.

D. Attack Models

The problem of system invasion exists in our application scenario, ranging from low-risk cases such as children accessing permission-restricted functions in a smart home environment to high-risk situations where attackers breach personal devices containing sensitive data. In general, the goal of an attack is to gain access to the host device paired with the Bluetooth headphones in order to steal private information or perform unauthorized actions. To achieve this, the adversary must spoof the identity of a legitimate user and bypass the authentication mechanism. Based on the assumption that we trust the security standards of Bluetooth 5.4 and above [30], we consider the following types of attacks:

Zero-effort Attack. Assuming the adversary is unaware of the system's underlying mechanism, a straightforward attempt may involve simply pressing the button on headphones—as what we refer to as a zero-effort attack.

Impersonation Attack. The adversary may then attempt to pass the authentication by wearing the headphones, hoping to be mistakenly recognized as a legitimate user. This type of attack may target specific victims who have similar head biometrics. We evaluate impersonation attacks under two conditions: (1) insider impersonation, where the adversary is a registered user attempting to be authenticated as another registered identity; and (2) outsider impersonation, where the adversary is not registered in the system but attempts to log into a registered user's account.

Acoustic Replay Attack. We must consider the possibility of replay attacks, as this is a common risk faced by all systems utilizing acoustic signals. In this work, we evaluate two types of replay attacks: (1) side-channel eavesdropping replay attack, where the adversary is assumed to have recorded the user's authentication signal from close proximity without being detected, and replays the recording in an attempt to bypass the system, and (2) leaked biometric replay attack, where the adversary gains access to a previously recorded authentication audio file and uses a speaker to play the signal during an authentication session.

IV. Approach Design

A. Sensing Signal Design

Due to the limitations imposed by Bluetooth profiles, signals above 8 kHz cannot be recorded, making ultrasonic sensing infeasible on Bluetooth wearables. Although commonly used sensing signals like chirps or Gaussian noise can still produce wideband responses within the available bandwidth, they tend to be obtrusive and negatively affect the user experience, thus we designed our detection signal based on real-world Bluetooth device notification systems. After investigating multiple products , we observed that notification sounds vary across devices: The typical start-up sounds of Bluetooth headphones include phrases such as "Battery High", "Battery 50%", "Welcome to Brand Name/Device Model", notification ringtones, or a combination of these. By directly reusing these notifications, we can sense the user during the notification process, enabling passive and seamless user authentication. Moreover, a burst of sound may also be emitted by the speaker when the device switches operational modes.

Based on these observations, we designed our signal to begin with the AirPods' notification ringtone, followed by a TTS-generated voice prompt saying "Battery High" [31], with a 1-second interval between the two audio segments, as shown in Figure 8. This design allows us to separate the two signals for independent feature analysis, evaluate a combined-feature model as well as a decision-level fusion model, and utilize the 1-second silent gap to study the effects of ambient noise.

We measured the sound pressure level (SPL) of the original notification sound, which reached approximately 45 dB. Notably, the volume of the headphones' built-in prompt does not vary with the phone's system volume. To ensure consistency during data collection, we used the `LoudnessEnhancer` method provided by the Android platform during the data collection process [28], to adjust the SPL of our playback audio to approximate that of the original notification ringtone under the HSP mode of the bone-conduction headphones.

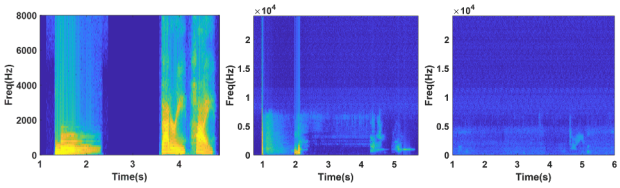


Fig. 8: Original signal and captured biometrics, bone-conduction and earbud.

B. Data Pre-processing

After the microphone captures the bone-conducted sensing signal, the pre-processing stage consists of three steps: signal localization, segmentation, and denoising.

Localization and segmentation: Thanks to the built-in noise suppression of the bone-conduction microphone, the sensing signal can be effectively localized by computing the cross-correlation between the recorded signal and the original reference signal. By shifting the raw data to find the maximum cross-correlation coefficient, The time delay between the received signal $\hat{s}(n)$ and the reference signal $s(n)$ is estimated by maximizing their cross-correlation:

$$\text{delay} = \arg\max_{m} \sum_{n=0}^{N-m-1} \hat{s}(n+m)\, s(n) \tag{3}$$

where m is the candidate delay, N is the total signal length, and $\arg\max_m$ finds the value of m that maximizes the cross-correlation between the two signals. Then we can segment captured signals based on their original length and gap.

Denoising: After cropping the two signal segments from the raw data, we perform denoising. Analysis of the original signals shows that the ringtone primarily occupies the frequency range of 0–2000 Hz, while the voice notification spans 0–4000 Hz. Based on this, we apply a bandpass filter to remove high-frequency noise outside these ranges.

C. Signal Spectrogram Derivation

The primary feature used for our task is the spectrogram, which represents the time–frequency characteristics of the signal. The unique structure of each individual's head affects the signal's frequency components over time, making the resulting spectrogram a form of biometric information. To enable compatibility with image-based classifiers, we convert the audio data into spectrograms.

$$\text{STDTFT}(t, f) = \sum_{\tau=t}^{t+T-1} s(\tau)\, w(\tau - t)\, e^{-j2\pi f \tau} \tag{4}$$

$$\text{spectrogram}(t, f) = |\text{STDTFT}(t, f)|^2 \tag{5}$$

The spectrogram is computed as the squared magnitude of the short-time Fourier transform (STFT) over a sliding window of length T. Here, $s(\tau)$ is the input time-domain signal, $w(\tau - t)$ is a window function (e.g., Hamming window) centered at time t, and f is the frequency bin. The summation iterates over the time window to extract localized frequency information.

D. ResNet-based Authentication Algorithm

We designed a lightweight deep learning algorithm based on the ResNet architecture for the user authentication task. The choice of ResNet is motivated by the effectiveness of its residual block structure in capturing discriminative features.

Residual blocks are the key structural component of ResNet models. In spectrogram-based biometric classification tasks, different classes often differ only in detail, such as slight variations in frequency patterns or localized energy shifts. Residual blocks enable the network to capture these fine-grained differences by allowing deeper architectures without discarding low-level information. The shortcut connections preserve original features while enabling the addition of new transformations, allowing the model to attend to both local details and global structures. This architecture also facilitates stable training and mitigates issues such as vanishing gradients, particularly beneficial when training on limited datasets.

The model consists of four residual stages containing 3, 4, 6, and 3 blocks, respectively. Each residual block comprises two 3×3 convolutional layers followed by Batch Normalization and ReLU activation functions. Downsampling is performed at the beginning of stages 2, 3, and 4 using stride-2 convolutions, as detailed in Table I. An adaptive average pooling layer followed by a fully connected layer produces the final classification logits. The current model has 21 million parameters, slightly fewer than the standard ResNet-50 model [29], which already has open-source deployment solutions available for mobile platforms [32]. Currently, we are using the model in an offline setting on a PC for evaluation. We independently trained the model on the ringtone and voice notification inputs. By combining the classification results of the ringtone s_1 and the notification voice s_2, we obtain a fused score s:

$$s = \alpha \cdot s_1 + (1 - \alpha) \cdot s_2 \tag{6}$$

In this equation, α is the parameter stands for the weight.

V. Performance Evaluation

A. Experimental Setup

Platforms. We evaluated our system using bone-conduction headphones (Shokz OpenRun) paired with a smartphone (Samsung Galaxy S22) as the host device. An Android application was developed to handle sensing signal playback, audio recording for data collection, Bluetooth configuration, and related functionalities. In this application, both audio playback and recording were configured for stereo at a sampling rate of

TABLE I: Structure of ResNet-Based Classification Model.

Layer	Output Shape	Param #
Input	(3, 200, 200)	0
Conv2D (7×7, stride=2) + BN + ReLU	(64, 100, 100)	9,472
Residual Section 1 (3×RBlock)	(64, 100, 100)	112,384
Residual Section 2 (4×RBlock, stride=2)	(128, 50, 50)	475,136
Residual Section 3 (6×RBlock, stride=2)	(256, 25, 25)	1,873,920
Residual Section 4 (3×RBlock, stride=2)	(512, 13, 13)	2,219,008
Adaptive Avg Pooling	(512,)	0
Fully Connected Layer	(14,)	7,182

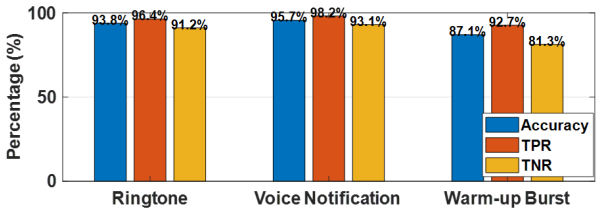


Fig. 9: All participants data classification performance of different sensing signal.

48,000 Hz. However, due to the Bluetooth protocol limitations discussed earlier, the recorded audio was effectively *pseudo stereo*—i.e., the left and right channels contained identical signals. Additionally, we used a Bluetooth earbud (Jabra Elite Active 65t) in the bone-conduction vibration study described in a previous section.

Data Collection. We recruited 14 volunteers (3 females and 11 males, aged 22–36) to participate in experiments using the bone-conduction headphones. Institutional Review Board (IRB) approval was obtained prior to data collection. During the experiment, each participant was asked to wear and remove the headphones six times. For each wear, 20 audio samples were collected, capturing variations caused by slight changes in device position and natural inconsistencies in user behavior. Seventy percent of the collected data was used for training, and the remaining 30 percent was reserved for testing. We also collected a small dataset from three participants to study environmental noise effects and replay attacks.

A sample of the captured bone-conducted biometric signal is shown in Figure 8, from bone-conduction headphones and earbud. As observed, the frequency content is limited to below 8 kHz, and the probing signal exhibits certain modifications within its frequency range after passing through the user's head. In addition to the expected response, we also identified a brief, subtle noise at the beginning of each data collection session of bone-conduction headphones. This signal is barely noticeable to users but consistently appears across sessions and devices. We suspect it is a warm-up artifact generated when the Bluetooth headphones switch operating modes. Given its repeatable nature and potential as an bone-conducted biometric signature, we included this short signal in our evaluation to further validate the effectiveness of our method.

B. Performance

We evaluated the authentication performance in a single-user setting by constructing binary classifiers to distinguish between the registered user and all potential non-user signals. To ensure robust generalization, the negative class consisted of three types of data: (1) data collected when the headphones were not worn and no biometric signal was present; (2) recordings from a silicone dummy head, simulating physical spoofing attempts; and (3) recordings from unregistered volunteers, containing real biometric signals from individuals other than the target user. Figure 9 illustrates the identification results for all 14 users using the bone-conduction headphones. Our system achieved high accuracy across all participants,on both the tone and voice features, with average accuracies of **93.70%** and **96.42%**, and over half of the users achieving accuracy above **95.8%** and **98.5%**. Interestingly, the performance of the warm-up burst feature, while initially expected to be less reliable, still yielded a respectable average accuracy of **86.54%**, outperforming our expectations in several cases.

One biometrics vs. Multiple biometrics. We conducted experiments using the recorded ringtone and voice notification separately for the binary classification task, and further explored two fusion strategies: (1) performing classification on the entire audio recording that contains both signals with a one-second interval in between; and (2) executing two or three separate classification tasks and applying a weighted average to their linearly normalized output scores. We also explored the use of the warm-up burst as an additional biometric signal for classification. As shown in Table II, Our results indicate that different signals can evoke distinct responses from the same head biometric. However, fusing the raw signal features prior to classification may result in the loss of signal-specific information. In fact, although both the **ringtone** and **voice notification** features individually achieved high accuracy (93.70% and 96.42% on average, respectively), their **feature-level fusion** yielded only a marginal improvement (96.75% on average) and even resulted in degraded performance for some users—suggesting the presence of redundant or interfering information when combining raw representations. In contrast, **decision-level fusion** of ringtone and voice notification preserved the discriminative power of each modality, achieving a higher average accuracy of **97.44%**. Extending this to a three-signal fusion—**ringtone**, **voice notification**, and **warm-up burst**—led to a slight increase to **97.51%** on average, with notable improvement observed in only one participant, indicating limited additional benefit from incorporating the warm-up burst in most cases.

TABLE II: Comparison of Single and Multiple Notification Signal Strategies.

Signal Type	Avg. Acc.	Avg. TPR.	Avg. TNR.
ringtone	93.80%	96.4%	91.2%
voice notification	95.72%	99.2%	93.1%
warm-up burst	87.14%	92.7%	81.3%
feature fusion of tone and voice	97.27%	98.48%	96.02%
decision fusion of tone and voice	97.44%	98.62%	94.76%
decision fusion of all	97.51%	98.46%	95.05%

C. Zero-effort and Impersonation Attack

Zero-effort Attack. We evaluate our system against zero-effort attacks, where the adversary attempts to bypass au-

TABLE III: Success Rates of Different Attack Types.

Attack Scenarios	Attack Success Rate
Zero-effort Attack	1.07%
Impersonation Attack *(insider)*	3.37%
Impersonation Attack *(outsider)*	5.20%
Replay Attack - Side-channel	0%
Replay Attack - Leaked Biometric	0%

thentication without presenting any head biometric, in all our attack experiment we use the **ringtone** as an example. To simulate this scenario, we placed the headphones on a table and triggered the data collection process without any user wearing the device. As expected, the captured signals differed significantly from normal usage in both magnitude and frequency characteristics. Table III summarizes the results, showing a 1.07% attack success rate. This strong performance is largely due to the inclusion of such zero-effort scenarios in the training set, where recordings without a user (i.e., headphones placed on a table) were labeled as negative samples. This allows the model to learn and reject inputs with non-biometric signal patterns. These results confirm that our system is highly effective at defending against zero-effort attacks.

Impersonation Attack. We also evaluated the system's resilience against outsider and insider impersonation attacks using the following method: (1) we measured the probability that a volunteer not involved in training was recognized as a target user, simulating an outsider impersonation attack; (2) we measured the probability that a volunteer whose data was used in training was recognized as another target user, simulating an insider impersonation attack. Table III presents the success rates of the two types of attacks: 3.37% insider attacks breach the system, while 5.2% outsider attacks succeeded. The results demonstrate that our system can effectively defend against in-person biometric impersonation attempts.

D. Replay Attack

Side-channel Eavesdropping Replay. We use a smartphone (Samsung Galaxy S22) to record audio during a user's authentication process, and then use a speaker to replay the eavesdropped signal to the bone-conduction headphones in a new authentication session, attempting a side-channel eavesdropping replay attack. The replay success rate of our system in this scenario is 0. The results show that our system performs well in preventing eavesdrop-based replay attacks.

Leaked Biometric Replay. We simulate the leaked biometric replay attack by playing authentication signals previously collected in the test dataset. Table III presents the attack success rates when the system uses the ringtone as authentication biometrics, denoted as 0, respectively. The results demonstrate the replay resistance of our system. A significant difference in success rates is observed across different signal types, which we attribute to the built-in echo cancellation mechanism of the bone-conduction headphones.

The system resists replay attacks because the probing signal undergoes the headphones' noise suppression again during replay, distorting its biometric features.

E. Noise Impact

Effect of Ambient Noise. We evaluated the impact of ambient noise on system performance by testing under a few common indoor sound conditions: normal room ambient noise of 30 dB, background music at 40 dB, room vent noise at 50 dB and media with speech at 60 dB. For a single-user authentication task using only the ringtone signal, the model achieved an accuracy of 98.6% ambient noise of 30 dB. When ambient noise was present, the model maintained 94.44% accuracy under 40 dB music, achieved 99.17% accuracy under 50 dB vent noise and achieved 87.91% accuracy under 60 dB media speech. Ignoring the impact of training randomness, the unexpectedly high performance under vent noise may be attributed to training bias, as well as the fact that vent noise is relatively stationary and continuous. These results suggest that our system remains robust in typical indoor environments.

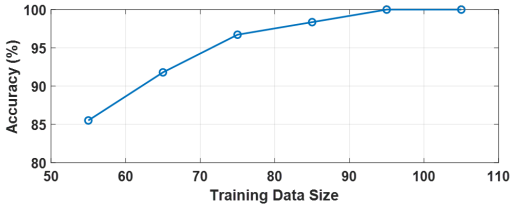


Fig. 10: Impact of training dataset size.

F. Training Dataset Size Impact

In the performance results we present, each label was trained with 80 samples, a choice made to balance performance with the practicality of data collection. To further investigate the impact of the dataset on model performance, we conducted an additional experiment, as shown in Figure 10. The figure illustrates the system's performance when using a single model under varying amounts of training data. The results demonstrate that our model can achieve satisfactory performance even with a relatively small amount of training data.

VI. Discussion and Future Work

Long-Term Stability Study. While our current evaluation focuses on short-term authentication performance, we plan to conduct a long-term study to assess the stability of head-related biometric features over time. Specifically, we aim to examine how typical physical changes — such as variations in body weight, hair length, wearing glasses, or clothing — may influence authentication accuracy. In addition, we will evaluate whether day-to-day factors such as ambient noise, device placement, or headphones positioning lead to more variation in performance than actual biometric changes. This future study will help us understand which features remain consistent over time and guide the development of adaptive models that maintain high accuracy during long-term use.

Signal Edge Analysis and Passive Identification. During data collection process, we conducted preliminary experiments using subtle segments of the audio signal for user classification. Specifically, we evaluated the 1-second silent gap between the ringtone and voice prompt. Using an SVM classifier on data from five users, we achieved a classification accuracy between 40% and 50%. If more advanced feature extraction techniques can be applied, it may serve as a promising foundation for seamless and unobtrusive user authentication.

Practical Deployment. This work shows that bone-conducted acoustic signals, which captured from different users, can be used for identity verification. However, the

current results are based on offline environment. The practical potential for deploying the model on mobile devices — including runtime performance, power consumption analysis, and the exploration of more lightweight model alternatives for this task — will be addressed in our future work.

VII. Conclusion

This work proposes an efficient and robust user authentication system based on acoustic sensing using notification sounds on Bluetooth bone-conduction headphones. The system captures unique head biometrics from the recorded echoes of these signals for authentication. We address the challenges of acoustic sensing over Bluetooth transmission by designing a deep-learning algorithm with a decision-level fusion strategy. The algorithm extracts features from spectrograms to build per-user models and distinguish between users. Experimental results show that the system can effectively authenticate both single and multiple users, and remains reliable under various types of attacks and ambient noise.

Acknowledgment

This work is partially supported by NSF CNS-2450046 and CNS-2440238.

References

[1] Bluetooth SIG, "2024 bluetooth market update," 2024, accessed: Apr. 15, 2025. [Online]. Available: https://www.bluetooth.com/2024-market-update/

[2] G. Peng, G. Zhou, D. T. Nguyen, X. Qi, Q. Yang, and S. Wang, "Continuous authentication with touch behavioral biometrics and voice on wearable glasses," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 3, pp. 404–416, 2017.

[3] C. Shi, X. Xu, T. Zhang, P. Walker, Y. Wu, J. Liu, N. Saxena, Y. Chen, and J. Yu, "Face-mic: inferring live speech and speaker identity via subtle facial dynamics captured by ar/vr motion sensors," in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 478–490. [Online]. Available: https://doi.org/10.1145/3447993.3483272

[4] R. Wang, L. Huang, and C. Wang, "Low-effort vr headset user authentication using head-reverberated sounds with replay resistance," in *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2023, pp. 3450–3465.

[5] S. Schneegass, Y. Oualil, and A. Bulling, "Skullconduct: Biometric user identification on eyewear computers using bone conduction through the skull," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 1379–1384.

[6] K. Takemura, A. Ito, J. Takamatsu, and T. Ogasawara, "Active bone-conducted sound sensing for wearable interfaces," in *Proceedings of the 24th annual ACM symposium adjunct on User interface software and technology*, 2011, pp. 53–54.

[7] N. Funato and K. Takemura, "Estimating contact force of fingertip and providing tactile feedback simultaneously," in *Adjunct Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology*, 2016, pp. 195–196.

[8] G. Saggio, A. S. Santoro, V. Errico, M. Caon, A. Leoni, G. Ferri, and V. Stornelli, "A novel actuating–sensing bone conduction-based system for active hand pose sensing and material densities evaluation through hand touch," *IEEE transactions on instrumentation and measurement*, vol. 70, pp. 1–7, 2021.

[9] P. Markert, D. V. Bailey, M. Golla, M. Dürmuth, and A. J. Aviv, "On the security of smartphone unlock pins," *ACM Transactions on Privacy and Security (TOPS)*, vol. 24, no. 4, pp. 1–36, 2021.

[10] S. Uellenbeck, M. Dürmuth, C. Wolf, and T. Holz, "Quantifying the security of graphical passwords: The case of android unlock patterns," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, 2013, pp. 161–172.

[11] K. Thomas, F. Li, A. Zand, J. Barrett, J. Ranieri, L. Invernizzi, Y. Markov, O. Comanescu, V. Eranti, A. Moscicki *et al.*, "Data breaches, phishing, or malware? understanding the risks of stolen credentials," in *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security*, 2017, pp. 1421–1434.

[12] H. Alhakami *et al.*, "Knowledge based authentication techniques and challenges," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 2, 2020.

[13] K. Inamtisu, "Mobile phone with fingerprint sensor," European Patent Office EP1545102A1, Jun 2005.

[14] A. Wolber, "Apple's face id cheat sheet: What it is and how to use it," Oct 2023. [Online]. Available: https://www.techrepublic.com/article/apples-face-id-everything-iphone-x-users-need-to-know/

[15] M. Choraś and R. Kozik, "Contactless palmprint and knuckle biometrics for mobile devices," *Pattern Analysis and Applications*, vol. 15, no. 1, pp. 73–85, 2012.

[16] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE transactions on speech and audio processing*, vol. 3, no. 1, pp. 72–83, 1995.

[17] J. Lau, B. Zimmerman, and F. Schaub, "Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers," *Proceedings of the ACM on human-computer interaction*, vol. 2, no. CSCW, pp. 1–31, 2018.

[18] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin, "Earecho: Using ear canal echo for wearable authentication," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, Sep. 2019. [Online]. Available: https://doi.org/10.1145/3351239

[19] Z. He, J. Chen, K. He, Y. Gu, Q. Deng, Z. Zhang, R. Du, Q. Zhao, and C. Wu, "Headsonic: Usable bone conduction earphone authentication via head-conducted sounds," *IEEE Transactions on Mobile Computing*, pp. 1–15, 2025.

[20] Z. He, J. Chen, C. Wu, K. He, R. Du, J. Jia, Y. Gu, and X. Sun, "Hcr-auth: Reliable bone conduction earphone authentication with head contact response," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 8, no. 4, Nov. 2024. [Online]. Available: https://doi.org/10.1145/3699780

[21] J. P. Liautaud, "Bone conduction communication system and headset," Patent US20 060 018 488A1, 2006. [Online]. Available: https://patents.google.com/patent/US20060018488A1/en

[22] Bluetooth SIG, "Advanced audio distribution profile 1.4 specification," 2019, accessed: Apr. 6, 2025. [Online]. Available: https://www.bluetooth.com/specifications/specs/advanced-audio-distribution-profile-1-4/

[23] ——, "Headset profile 1.1 specification," 2008, accessed: Apr. 6, 2025. [Online]. Available: https://www.bluetooth.com/specifications/specs/headset-profile-1-1/

[24] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the Institute of Radio Engineers*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[25] Android Open Source Project. (n.d.) Audio architecture. [Online]. Available: https://source.android.com/docs/core/audio

[26] I. Qualcomm Technologies. (2023) Qualcomm® qcc30xx series bluetooth audio socs for stereo headphones or portable speakers. [Online]. Available: https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/Qualcomm-QCC30xx-Series-Bluetooth-Audio-SoCs-for-Stereo-Headphones-or-Portable-Speakers.pdf

[27] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. Wiley, 2000.

[28] Android Developers, "Loudnessenhancer," https://developer.android.com/reference/android/media/audiofx/LoudnessEnhancer, 2025, accessed: Apr. 6, 2025.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[30] Bluetooth SIG, "Bluetooth core specification version 5.4: Part h – br/edr controller: Security specification," 2023, accessed: Apr. 6, 2025. [Online]. Available: https://www.bluetooth.com/wp-content/uploads/Files/Specification/HTML/Core-54/out/en/br-edr-controller/security-specification.html

[31] Narakeet, "Narakeet: Text to voice and video tool," 2024, accessed: Apr. 6, 2025. [Online]. Available: https://www.narakeet.com/

[32] Alibaba Inc., "Mnn benchmark models," https://github.com/alibaba/MNN/tree/master/benchmark/models, 2024, accessed: 2025-07-28.