2-2023

Lightweight and non-invasive user authentication on earables

Changshuo HU
Singapore Management University, changshuohu@smu.edu.sg

Xiao MA
Singapore Management University, xiaoma.2022@phdcs.smu.edu.sg

Dong MA
Singapore Management University, dongma@smu.edu.sg

Ting DANG
University of Cambridge

Lightweight and Non-invasive User Authentication on Earables

Changshuo Hu*
Singapore Management University
changshuohu@smu.edu.sg

Xiao Ma*
Singapore Management University
xma9612@gmail.com

Dong Ma†
Singapore Management University
dongma@smu.edu.sg

Ting Dang
University of Cambridge
td464@cam.ac.uk

ABSTRACT

The widespread adoption of wireless earbuds has advanced the developments in earable-based sensing in various domains like entertainment, human-computer interaction, and health monitoring. Recently, researchers have shown an increased interest in user authentication using earables. Despite the successes witnessed in acoustic probing and speech based authentication systems, this paper proposed a lightweight and non-invasive *ambient sound* based user authentication scheme. It employs the difference between the in-ear and out-ear sounds to estimate the individual-specific occluded ear canal transfer function (OECTF). Specifically, the {out-ear, in-ear} scaling factors at different frequency bands are captured via linear regression and treated as the OECTF for user authentication. The proposed system is validated using 12 subjects under six different noisy environments and achieves a Balanced Error Rate (BER) of 4.84%. The particularly lightweight system can be easily deployed in earbuds and paves the pathway for more personalized services.

1 INTRODUCTION

Recent years have witnessed a rapid growth of wireless earbuds in the consumer market, thanks to the appealing user experience of active noise cancellation. People use them for various purposes, e.g., entertainment, online education, and meeting, in daily life. By authenticating the wearer, some additional services can be enabled, such as keeping the smartphone unlocked, recommending personalized music genres, or customizing acoustic characteristics to fit each individual's hearing sensitivity/experience at different frequencies [19]. For instance, Huawei FreeBuds 3 incorporated a bone conduction sensor to capture the voiceprint for user authentication, which allows for phone unlocking and online payment. However, the bone conduction sensor is usually more costly and requires additional space and proper mounting on the earbuds to ensure good signal quality.

In academia, different earbuds authentication schemes without requiring the integration of extra sensors have been investigated (basically, only utilizing the speaker and microphone sensor on earbuds), which can be grouped into two categories: acoustic probing (ultrasound [13, 18] or audible sound [7]) and speech [6] based authentication. The underlying principle of probing-based approaches is the unique geometry of each individual's ear canal [17]. By emitting audible/inaudible chirps to measure the unique transfer function between the transmitted and received signals, the ear canal geometry can be captured to identify the user. Speech-based approaches are based on the unique vocal/skull structure of people that leads to different speech characteristics. The unique vocal structure can be captured by the air-conducted speech recorded with out-ear microphones [10, 16], while the skull structure can be sensed by bone-conducted speech recorded with in-ear microphones [6].

However, these approaches have some inherent limitations that hinder their adoption in commercial earbuds for wide applications. Concretely, the probing-based scheme requires the active transmission of audio chirps, which incurs sophisticated processing pipelines, including probing signal generation, adaptive gain control to deal with speaker's frequency selectivity, precise coordination of the transmitted and received pulses, interference elimination from direct-path propagation, etc [7]. Moreover, continuously emitting ultrasound signals in the ear canal might have some health risks [7] and transmission of audible chirps might result in an unpleasant user experience. For speech-based methods, the user has to produce some speech so that the unique bone structure can be measured, which induces extra human effort and may not be applied in scenarios where talking is prohibited (e.g., library). In addition, the external noise from the environment can affect the authentication performance [21].

In this work, we present a lightweight and non-invasive scheme for user authentication using earbuds. Fundamentally, our approach is based on the unique ear canal geometry and eardrum impedance for each individual [8]. However, instead of using acoustic chirps to probe the uniqueness, we leverage the in-ear and out-ear microphones on the earbuds to jointly detect the frequency responses caused by ear canal geometry, referred to as Occluded Ear Canal Transfer Function (OECTF)[1]. Specifically, the out-ear microphone measures any environment sounds, while the in-ear microphone measures its modulated (by the occluded ear canal) version. By mapping the scaling factor between the out-ear and in-ear signals at different frequencies using linear regression, the unique geometry information of the ear canal can be captured and represented by the individual-specific linear regression coefficients (i.e., OECTF).

*First authors with equal contribution in alphabetical order.
†Corresponding author.

HotMobile '23, February 22–23, 2023, Newport Beach, CA, USA

ACM ISBN 979-8-4007-0017-0/23/02…$15.00
https://doi.org/10.1145/3572864.3580332

[1]Note that conventional ear canal transfer function depicts the propagation property of a sound from the entrance of the open ear to the ear drum. In our case, when the ear canal is occluded by the earbuds, the property changes and therefore we call it occluded ear canal transfer function.

During real-world authentication, the OECTF is applied to the instantaneous out-ear signal to obtain its estimated in-ear version, which is then compared against the real signal captured by the in-ear microphone. If the similarity is higher than a predefined threshold, the wearer is accepted as the authentic user, and vice versa.

To demonstrate the feasibility of the proposed system, we developed a prototype to record data from 12 subjects under six noise environments. The experimental results show that our system can achieve 4.84% BER. Moreover, the proposed authentication scheme is lightweight in terms of algorithm complexity and power consumption, and requires no effort (e.g., speaking) from the user.

In summary, we made the following contributions:

- We proposed a novel, lightweight, and non-invasive user authentication scheme for earbuds, by jointly utilizing the out-ear and in-ear microphones to capture the unique ear canal geometry.
- We demonstrated the feasibility of the proposed scheme with real-world data from 12 subjects, with an average authentication BER as low as 4.84%.

2 IN-EAR SOUND VS. OUT-EAR SOUND

An example of the in-ear and out-ear sounds collected under three different scenarios (i.e., woman singing, man singing and traffic sound) is presented in Figure 1. Both time domain (left column) and frequency domain (middle and right columns) are displayed. From the time domain, it can be observed that the in-ear microphone generally receives a lower amplitude signal than the out-ear, due to the earbuds obstructing the external sound. In terms of the frequency domain, we observe a clear energy decrease in high-frequency bands (over 1kHz) of the in-ear sound compared to the out-ear one for all three varieties of sounds. This is due to the occlusion of ear canal that suppresses the frequency component at high frequencies. On the other hand, the closed ear canal chamber also amplifies the low frequencies, so that the energy loss due to earbuds obstruction is compensated to some extent.

We can conclude that there is a mapping (e.g., energy scaling) between the in-ear and out-ear sounds at different frequency bands, which is fundamentally due to the modulation of the occluded ear canal. Thus, we hypothesize that if the scaling factors are (1) distinct for different people and (2) consistent for the same person under different conditions, such that *{out-ear, in-ear} mapping* could be utilized as a unique feature for wearer authentication. Next, we will present our system that realizes and validates the hypothesis.

3 SYSTEM DESIGN

3.1 Overview

Figure 2 shows the overview of the proposed authentication system, consisting of enrollment and authentication. During enrollment, users are required to record some environmental sounds while wearing the earbud to estimate the individual-specific occluded ear canal transfer function (OECTF). In detail, the out-ear and in-ear sound data will first be pre-processed and converted to frequency domain using Fast Fourier Transform (FFT). As some frequencies may not exist for certain sound types, an energy detection module is followed to select the frequency components with acceptable

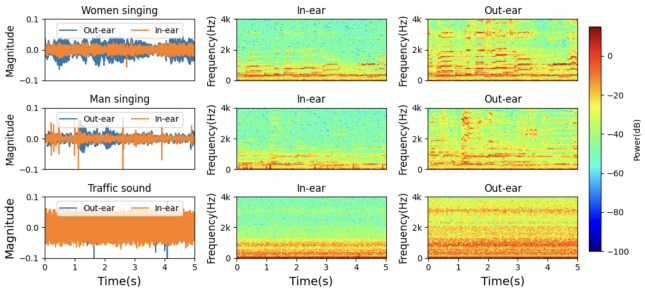


Fig. 1: Time domains and frequency spectrum of the in-ear and out-ear signals of three different environmental sounds.

energy according to the out-ear sound, and the corresponding in-ear frequency components will also be selected. Then, these {out-ear, in-ear} pairs will be used to infer the OECTF through linear regression.

During the authentication phase, our system collects both in-ear and out-ear environmental sounds silently. According to the individual-specific OECTF, the out-ear sound can be converted into its in-ear modulated version, and further compared with the real in-ear data collected by in-ear microphone. A high similarity between them indicates an acceptance of the authentication and vice versa.

3.2 User Enrollment

3.2.1 Pre-processing. The pre-processing steps are applied to both in-ear and out-ear signals.

Data Segmentation and Downsampling: The collected sounds are first segmented into one-second chunks. Since the frequency range of most environmental sounds is below 4 kHz, we further downsample both the out-ear and in-ear sounds to 8 kHz to reduce computational overhead while retaining the useful information.

Fast Fourier Transform (FFT): FFT is employed to convert the sounds in time domain to the frequency domain. To obtain more fine-grained information, a high-resolution FFT with 4096 points is adopted, resulting in a frequency resolution of around 2 Hz. Notably, in order to combat the influence of unexpected noise on a certain frequency bin, we group consecutive frequency bins into a frequency band. Specifically, the energy of every 16 (selected empirically to balance the trade-off between frequency resolution and complexity) continuous frequency bins is averaged and considered as band energy. As a result, 128 frequency bands are calculated to cover the range of 0~4 kHz and each band covers a range of 31.25 Hz.

3.3 Energy detection

As some of the environmental sounds only contain frequency components at certain ranges, part of the bands may not have adequate energy, thus are not reliable to compute the OECTF. Therefore, an energy detection module is developed to filter out the frequency bands with low magnitude and only retain the frequency bands with high energy. Out-ear sound is used for energy detection, as all frequency components are salient compared to in-ear sound that shows a significant attenuation in the high frequency range. In a certain band, if the energy of the out-ear sound is higher than the base value measured under the quiet environment by a threshold th_1 (0.2 is selected empirically in our dataset), we accept the frequency band and find its corresponding band in in-ear sound to

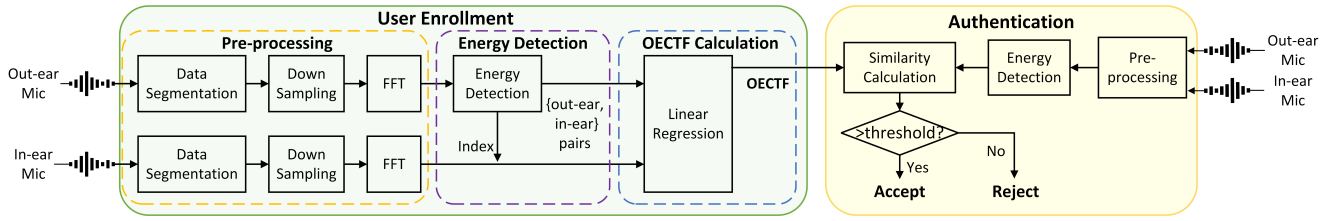
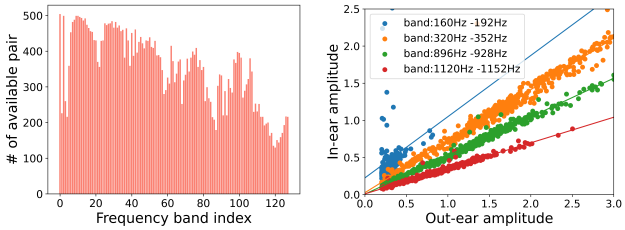
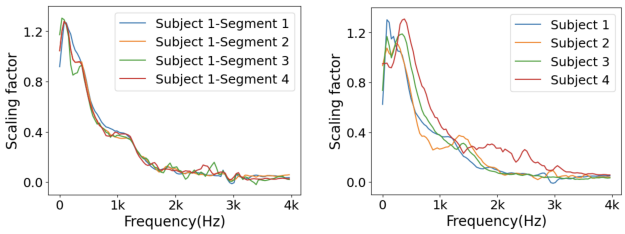


Fig. 2: System overview. The user first enrolls the data to calculate OECTF and then the system can be utilized for authentication.



(a) (b)

Fig. 3: (a) The number of available {out-ear, in-ear} pairs in each frequency band, and (b) scatter plot of the pairs in 4 frequency bands and the fitted lines with linear regression.



(a) (b)

Fig. 4: (a) Scaling factors for four segments from the same subject, and (b) Scaling factors from four different subjects.

form an {out-ear, in-ear} pair. For each one-second segment, we would expect an {out-ear, in-ear} pair at each frequency band ideally. However, after energy detection, some frequency bands with low energy cannot form such a pair, resulting in a different number of {out-ear, in-ear} pairs at different frequency bands. Figure 3(a) plots the distribution of the number of pairs from our dataset for one subject. We can have at least 150 pairs for linear regression, guaranteeing reliable OECTF calculation.

3.4 OECTF Calculation

Figure 3(b) is a scatter plot of the {out-ear, in-ear} pairs for four different frequency bands. As the data includes different sound types and a variety of different sound segments, the magnitude for out-ear and in-ear sounds can cover a relatively wide range.

It is clearly observed that the mapping between out-ear and in-ear sounds is approximately linear for each band. Therefore, we exploit linear regression model to fit {out-ear,in-ear} mappings for each band. In detail, we fit a linear relation between out-ear and in-ear magnitude as:

$$S^i_{in} = a^i * S^i_{out} + b^i, i \in [1, 128], \quad (1)$$

where i represents the frequency band, S_{out} and S_{in} represent the out-ear and in-ear sounds, a^i and b^i are the linear regression coefficients, represented as the straight lines in Figure 3(b). Empirically we found that b^i is approximately zero for all frequency bands. Thus, we ignore b^i and only vectorize the 128 a^i as the OECTF.

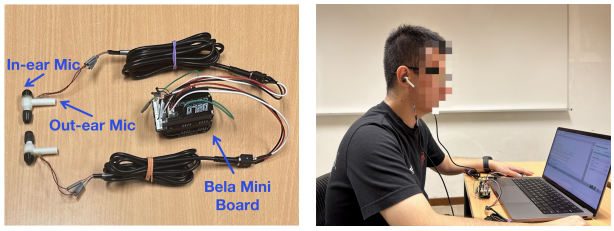
To understand whether derived OECTF is salient enough to capture the individual differences in the ear canal geometry for user authentication, we conduct two additional analyses to verify: (1) the scaling factors between the out-ear and in-ear sounds belonging to the same subject are consistent; and (2) the scaling factors belonging to different subjects show considerably dissimilar patterns. For the first argument, we divide one subject's data into four segments, and estimate its OECTF for each segment, as shown in Figure 4(a). For the second argument, we compare the OECTF for four different subjects as shown in Figure 4(b). We can observe that for the same subject, the scaling factors are more consistent indicating a high intra-subject similarity, while for different subjects, the scaling factors are more distinguishable indicating a high inter-subject dissimilarity. Thus, to this end, we can conclude the feasibility of our system for user authentication.

3.5 Authentication

Once the OECTF is calculated, the system can authenticate users by collecting their in-ear and out-ear data silently. Specifically, the out-ear sound can be converted to the in-ear sound via OECTF, and compared against the real in-ear sound to decide whether accepting or not. However, the frequency components of one sound segment may not cover all the 128 bands. Particularly for certain sound types, the bandwidth may be limited, resulting in many low-amplitude or zero-amplitude bands. As a result, the corresponding {out-ear, in-ear} pairs do not match the scaling factors in the true OECTF, leading to inaccurate estimation. Therefore, only the bands with energy above a certain threshold (th_1) will be adopted for OECTF authentication. Then, if the Euclidean distance between the true and estimated in-ear energy is less than a threshold th_2, the user will be accepted, and vice versa. Additionally, we proposed a voting mechanism to further improve the prediction result. In detail, we separately predict 3 continuous segments and apply majority voting to make the final decision.

4 PROTOTYPING AND DATA COLLECTION

Although the in-ear microphone has been embedded in many off-the-shelf earbuds, its raw data is inaccessible due to the lack of APIs. Thus, we built a prototype, as shown in Figure 5(a), to collect the data for evaluation. Specifically, we printed a 3D earbud shell and integrated an in-ear microphone (in the front of the earbud cavity and facing toward the ear canal) and an out-ear microphone (in the bottom of the earbud stem and facing outward). Then, the two microphones were connected to a Bela Mini Board [1] with audio jacks and sampled at 44.1 kHz.



(a) (b)

Fig. 5: (a) The designed earbuds prototype and data recording board, (b) illustration of a participant wearing the earbuds for data collection.

We recruited 12 subjects (9 males and 3 females) for data collection[2]. In detail, we selected 6 common environmental sounds [3] that cover different frequency ranges to obtain a complete OECTF curve. Then, the subjects wore the developed earbud and sit still while a laptop was playing these environmental sounds. Figure 5(b) illustrates the experiment setup. Each sound was played for 2 minutes for each subject. In total, we collected $12 \times 6 \times 2 = 144$ minutes raw audio data.

As presented in Section 3.2, we downsampled the raw data to 8 kHz, and split them into one-second segments, resulting in 8,640 segments. For each subject, we split the segments of every environmental sound into 70% training, 10% validation, and 20% testing. The training segments are utilized to derive the OECTF curve, and the validation segments are used to determine the threshold th_2 for accepting and rejecting a user. Specifically, the threshold is selected when the system obtains an equal False Acceptance Rate (FAR) and False Rejection Rate (FRR), as discussed next.

5 EVALUATION

5.1 Metrics

We use two typical metrics in an authentication system, False Acceptance Rate (FAR) and False Rejection Rate (FRR), to evaluate our approach. FAR is the probability that an imposter is wrongly accepted as a legitimate user, which reflects the security level of the system. While FRR is the probability that a legitimate user is falsely rejected, which reflects the user experience of the legitimate user. In addition, the average of FAR and FRR is defined as the Balanced Error Rate (BER) and used to indicate the overall authentication performance. FAR, FRR, BER are computed as below,

$$FAR = \frac{FP}{FP+TN}, FRR = \frac{FN}{FN+TP}, BER = \frac{FAR+FRR}{2}, \quad (2)$$

where TP, FP, TN, and FN represent true positive, false negative, true negative, and false positive, respectively.

5.2 Authentication Performance

After obtaining the OECTF of each subject using the 70% training data, we iteratively treat one subject as the legitimate user and the rest are imposters. The resulting FAR and FRR are averaged over all subjects and presented in Table 1. First, we can observe the error

[2]Ethical approval for carrying out the data collection has been granted by the corresponding institution.

[3]Including male song, female song, air conditioning, traffic noise, restaurant background noise, and ocean wave sounds.

Table 1: Overall performance over validation and test sets.

	Validation (10%)			Test (20%)		
	FAR	FRR	BER	FAR	FRR	BER
Without Vote	7.54%	7.54%	7.54%	7.87%	6.77%	7.32%
With Vote	4.77%	4.77%	4.77%	4.22%	5.46%	4.84%

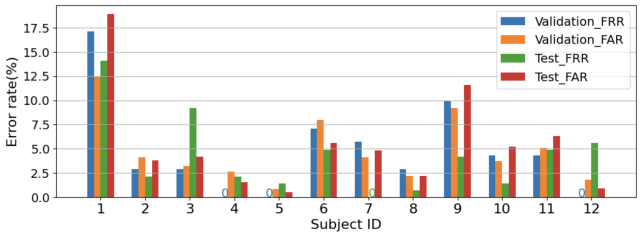


Fig. 6: Individual authentication performance with majority vote (zeros indicates the corresponding error rate is 0).

rates obtained on the validation set and test set are almost equivalent, implying that the threshold th_2 determined by the validation set can be generalized to unseen user data. Second, the proposed vote strategy can further improve the authentication performance, resulting in a BER of 4.84%, which is comparable to other more sophisticated approaches [6, 7, 18]. Note that the threshold is obtained by balancing the FAR and FRR in the validation set, while their values change slightly after applying the same threshold to a different set.

Figure 6 plots the detailed error rates for each individual. We can see that the performance varies among subjects, which is possibly due to that the OECTFs of some subjects show higher similarity with others so it is a harder task for systems to distinguish among them. Specifically, using the Euclidean distance to measure the similarity, Subject 1 shows relatively poorer performance than others. We also explored the use of cosine distance for similarity measurement and found that the error rate for Subject 1 is reduced by 6%. However, cosine distance performs worse on rest of the subjects compared to Euclidean distance, which indicates that the optimal metric to measure similarity needs to be further investigated and such a metric might be user-dependent.

Figure 7 plots the detailed error rates for each noise type. We can observe that different noise types (i.e., real-life scenarios) lead to varying authentication performance, depending on the properties of the noise. In detail, air-conditioning, restaurant, and traffic sounds are relatively stable in the temporal dimension so the calculated scaling factors over one second are more reliable. While the volume of songs and ocean waves might vary significantly over time, resulting in a dramatic change of the strength in certain frequency bands and therefore poor authentication performance.

5.3 System Performance

We measure the run-time overhead (latency and energy consumption) of the proposed system on two platforms: Raspberry Pi (RasPi) and Arduino Nano. Specifically, real-time authentication is decomposed into three stages: pre-processing, energy detection, and similarity calculation. From Table 2, we can observe that (1) the authentication can be completed within 10 ms on RasPi-level device, guaranteeing the real-time operation, (2) most of the overhead comes from pre-precessing as we applied a high-resolution FFT.

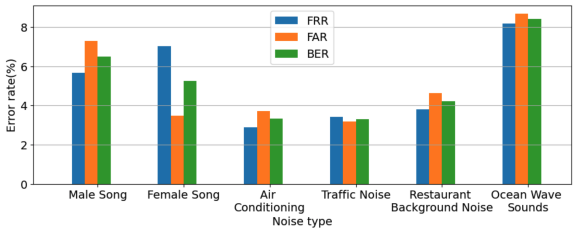


Fig. 7: Authentication performance for each noise type with majority vote.

Table 2: System performance of the proposed method.

Device	Parameter	Pre-processing	Energy detection	Similarity calculation	Total
RasPi	Latency(ms)	8.16	0.12	0.87	9.15
	Power(W)	1.23	0.63	1.06	-
	Energy(mJ)	10.04	0.09	0.92	11.05
Arduino Nano	Latency(ms)	590.39	0.29	1.69	595.97
	Power(W)	0.17	0.13	0.14	-
	Energy(mJ)	100.37	0.04	0.27	100.68

However, Arduino incurs extremely high latency for FFT as the library is not optimized for fast execution. Although the system performance of existing probing and speech based approaches is not compared, we believe our system is much more efficient given the intuitively lightweight operations. Moreover, RasPi and Arduino are general development boards without power optimization, the power consumption would decrease with dedicated microcontrollers and circuit design.

6 RELATED WORK

Recently, earable-based authentication has received increasing attention in academia. Various modalities and bio-metrics have been proposed and demonstrated. The fundamental principles to distinguish users can be classified into three categories: (1) unique ear canal geometry, (2) unique bone structure, and (3) unique brain activity during a certain cognitive task. Ear canal geometry based approaches require the active transmission of audible or inaudible chirps and measure the geometry difference using the echos [7, 13]. In contrast, bone structure based methods usually employ human-generated body sounds, such as speech [6, 9, 20] and walking-induced vibrations [5, 12], to infer the unique frequency modulation after bone conduction. For brain activity-based approaches, the user is asked to perform a specific task during which the electroencephalogram (EEG) response is captured to differentiate people [4, 14, 15]. Our work falls under the category of ear canal geometry based principle. However, instead of using audio chirps to probe the geometry features, we leverage the difference between signals captured by in-ear and out-ear microphones to infer the user-specific ear canal characteristics.

7 DISCUSSION AND FUTURE WORK

In this work, we presented a novel, lightweight, and non-invasive strategy to authenticate the earbuds wearer. By sensing the ambient sounds, it jointly employs the in-ear and out-ear microphones to detect the unique geometry of human ear canal, without the need for active audio chirps emission or human speech. Albeit great performance, this is just a feasibility study and there remain a few limitations and potential solutions for future exploration.

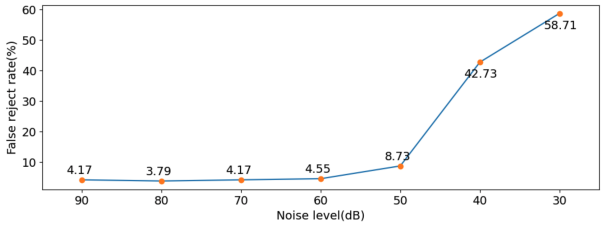


Fig. 8: Authentication performance under different noise levels with majority vote.

Compatibility with music playback. The proposed approach operates under ambient external noise with frequencies between 0-4 kHz, which is overlapped with typical sounds during music playback or voice calls. As a result, how music playback affects our authentication system is unclear. Note that the same issue applies to speech based authentication systems [6, 9]. We plan to explore in two directions. First, since the delivered music is known to the earbud system, it would be possible to design some filtering algorithms to eliminate the interference of the music. Second, as in-ear microphone will also modulate the played music, we plan to discover another transfer function between in-ear and out-ear microphones during music playing. Then, the authentication system can switch between the two modes under different scenarios.

Tightness of earbuds. Our approach relies on the frequency correlation between the in-ear and out-ear sounds, which only exists when the ear canal is sealed property. However, users might wear the earbuds in different ways in daily usage, thereby affecting the sealing quality. To explore the impact of earbuds tightness on the authentication performance, we conducted an experiment with two subjects by manually adjusting the tightness under three levels: tight (insert the earbuds deeply in the ear canal), normal (typical wearing depth with comfort), and loose (slightly fit to the ear canal but might fall off upon movements). The average BERs for the three levels are 5.19%, 5.13%, and 8.84% respectively, which suggests that wearing the earbuds too tight will not affect the performance, while loose fitting will result in higher authentication errors. In addition, the in-ear microphone can also collect some body sounds such as heartbeat and jaw movements [2]. However, as demonstrated in [3, 11], these sounds are below 100 Hz, while our authentication mainly utilizes frequencies from 100-4000 Hz. So the impact of such artifacts is negligible.

Requiring a strong ambient sound. Our system relies on the scaling factors of the external sound to map the wearer's ear canal geometry. To calculate the scaling factor, we have to ensure the strength of the in-ear microphone signal is above noise level, i.e., in-ear microphone can still hear a weak version of the external sound. Consequently, the ambient sound should be relatively strong, which limits the application scenarios. To investigate the minimal sound level required for operation, we test the authentication performance with one subject under various noise levels ranging from 30 dB to 90 dB. As shown in Figure 8, the authentication performance is excellent and stable when the noise level is higher than 60 dB. The error rate starts to increase with 50 dB noise strength, while the system is completely untrustable with 40 dB and lower noise levels. To solve this issue, we can switch to complementary modes such as (1) music-based transfer function as discussed above given that

music is played most of the time, (2) speech-based authentication, or (3) probing-based authentication.

Obtaining a more precise and fine-grained OECTF. Currently, we collected environment sounds under six common scenarios in daily life and utilized them to derive the OECTF for each individual. During data analysis, we spotted two issues: (1) as shown in Figure 3(a), each environmental sound will not contain all frequency components between 0-4 kHz. Thus, depending on the amount of data used for training, some frequency bands might have very few valid {out-ear, in-ear} pairs to learn the OECTF, leading to unreliable regression, and (2) as shown in Figure 3(b), the amplitudes of some {out-ear, in-ear} pairs are very similar (e.g., the blue dots distributed like a cluster instead of a line) because such frequency component has a stable volume in the training data, leading to inaccurate regression. To solve them, we plan to investigate from two aspects. First, from the data perspective, instead of collecting real-world sounds, we will try to synthesize sounds that contain all frequencies so that sufficient pairs at different frequency bands can be used to obtain a more fine-grained OECTF. Second, from the algorithm perspective, we can carry out more appropriate and sophisticated regression techniques (e.g., polynomials) to find a better transfer function to fit the data. In addition, different frequency bands are considered independently in the current system, we plan to explore the correlation between adjacent frequency components and further develop algorithms to jointly optimize the mapping between out-ear and in-ear sounds over a wider frequency range.

Improving authentication performance. In our current test setting, each frequency bands are treated equally. However, the sensitivity at different frequency bands in human ear canal is different. For example, human are more sensitive in perceiving sounds between 2 kHz to 5 kHz. A smaller amplitude variation within this band might be easily perceived. Therefore, the mapping between the out-ear and in-ear sounds at this frequency range might be more saliently different among different individuals. Assigning more importance to these frequency bands could potentially enhance the difference among individuals for better authentication performance. Specifically, we can adopt weighting strategies by assigning different weights to different frequency bands to amplify the inter-user uniqueness. In addition, similar to active-probing based techniques, our design is device-dependent due to the distinct frequency responses of earbud microphones. Consequently, re-enrollment is required when the user switches to a new device.

Performance comparison with probing-based approaches. We demonstrated a completely different mechanism for user authentication on earbuds. Specifically, previous probing-based approaches leverage high spatial resolution achieved through audio chirps (a sweeping signal with a single frequency at one time) to estimate ear canal geometry, while our method relies on the wide bandwidth of environment sounds (multiple frequencies at one time) to capture the in-ear and out-ear correlation. Thus, how the two mechanisms perform compared to each other is unclear. However, our current authentication performance cannot be compared directly with the results reported in probing-based papers [7, 13, 18], as the subjects and data collection procedures are different. To do a fair comparison, we plan to upgrade our prototype (integrate a speaker) and collect data (both passive and probing-based) simultaneously on the same set of subjects for evaluation.

8 ACKNOWLEDGMENTS

This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant (Grant ID: 21-SIS-SMU-036, 001124-00001).

REFERENCES

[1] Online. Bela Mini board. https://learn.bela.io/products/bela-boards/bela-mini/.
[2] Abdelkareem Bedri, David Byrd, Peter Presti, Himanshu Sahni, Zehua Gue, and Thad Starner. 2015. Stick it in your ear: Building an in-ear jaw movement sensor. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. 1333–1338.
[3] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, and Cecilia Mascolo. 2021. Motion-resilient Heart Rate Monitoring with In-ear Microphones. *arXiv preprint arXiv:2108.09393* (2021).
[4] Max T Curran, Jong-kai Yang, Nick Merrill, and John Chuang. 2016. Passthoughts authentication with low cost EarEEG. In *2016 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 1979–1982.
[5] Andrea Ferlini, Dong Ma, Robert Harle, and Cecilia Mascolo. 2021. EarGate: gait-based user identification with in-ear microphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 337–349.
[6] Yang Gao, Yincheng Jin, Jagmohan Chauhan, Seokmin Choi, Jiyang Li, and Zhanpeng Jin. 2021. Voice in ear: Spoofing-resistant and passphrase-independent body sound authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–25.
[7] Yang Gao, Wei Wang, Vir V Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using ear canal echo for wearable authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.
[8] Jinlin Liu, Huiqun Deng, Peifeng Ji, and Jun Yang. 2018. Headphone-to-ear transfer function estimation using measured acoustic parameters. *Applied Sciences* 8, 6 (2018), 918.
[9] Jianwei Liu, Wenfan Song, Leming Shen, Jinsong Han, Xian Xu, and Kui Ren. 2021. MandiPass: Secure and Usable User Authentication via Earphone IMU. In *2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 674–684.
[10] Rui Liu, Reza Rawassizadeh, and David Kotz. 2017. Toward Accurate and Efficient Feature Selection for Speaker Recognition on Wearables. In *Proceedings of the 2017 Workshop on Wearable Systems and Applications*. 41–46.
[11] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. 2021. OESense: employing occlusion effect for in-ear human sensing. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*. 175–187.
[12] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. 2022. Innovative Human Motion Sensing With Earbuds. *GetMobile: Mobile Computing and Communications* 25, 4 (2022), 24–29.
[13] Shivangi Mahto, Takayuki Arakawa, and Takafumi Koshinak. 2018. Ear acoustic biometrics using inaudible signals and its application to continuous user authentication. In *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 1407–1411.
[14] Nick Merrill, Max T Curran, Swapan Gandhi, and John Chuang. 2019. One-step, three-factor passthought authentication with custom-fit, in-ear EEG. *Frontiers in Neuroscience* (2019), 354.
[15] Takashi Nakamura, Valentin Goverdovsky, and Danilo P Mandic. 2017. In-ear EEG biometrics for feasible and readily collectable real-world person authentication. *IEEE Transactions on Information Forensics and Security* 13, 3 (2017), 648–661.
[16] Florentin Thullier, Bruno Bouchard, and Bob-Antoine J Menelas. 2017. A text-independent speaker authentication system for mobile devices. *cryptography* 1, 3 (2017), 16.
[17] Pim T Tuyls, Evgeny Verbitskiy, Tanya Ignatenko, Daniel Schobben, and Ton H Akkermans. 2004. Privacy-protected biometric templates: Acoustic ear identification. In *Biometric Technology for Human Identification*, Vol. 5404. SPIE, 176–182.
[18] Zi Wang, Sheng Tan, Linghan Zhang, Yili Ren, Zhi Wang, and Jie Yang. 2021. EarDynamic: An Ear Canal Deformation Based Continuous User Authentication Using In-Ear Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.
[19] Terry L Wiley, Karen J Cruickshanks, David M Nondahl, Ted S Tweed, Ronald Klein, and Barbara EK Klein. 1998. Aging and high-frequency hearing sensitivity. *Journal of Speech, Language, and Hearing Research* 41, 5 (1998), 1061–1072.
[20] Yadong Xie, Fan Li, Yue Wu, Huijie Chen, Zhiyuan Zhao, and Yu Wang. 2022. TeethPass: Dental Occlusion-based User Authentication via In-ear Acoustic Sensing. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 1789–1798.
[21] Fei Zhao, Hao Li, and Xueliang Zhang. 2019. A robust text-independent speaker verification method based on speech separation and deep speaker. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6101–6105.