An Ear Canal Deformation Based Continuous User Authentication Using Earables

Zi Wang
Florida State University
Tallahassee, FL, USA
ziwang@cs.fsu.edu

Sheng Tan
Trinity University
San Antonio, TX, USA
stan@trinity.edu

Linghan Zhang
Florida State University
Tallahassee, FL, USA
lzhang@cs.fsu.edu

Yili Ren
Florida State University
Tallahassee, FL, USA
ren@cs.fsu.edu

Zhi Wang
Florida State University
Tallahassee, FL, USA
zwang@cs.fsu.edu

Jie Yang
Florida State University
Tallahassee, FL, USA
jie.yang@cs.fsu.edu

ABSTRACT

Biometric-based authentication is gaining increasing attention for wearables and mobile applications. Meanwhile, the growing adoption of sensors in wearables also provides opportunities to capture novel wearable biometrics. In this work, we propose EarDynamic, an ear canal deformation based user authentication using ear wearables (earables). EarDynamic provides continuous and passive user authentication and is transparent to users. It leverages ear canal deformation that combines the unique static geometry and dynamic motions of the ear canal when the user is speaking for authentication. It utilizes an acoustic sensing approach to capture the ear canal deformation with the built-in microphone and speaker of the earables. Specifically, it first emits well-designed inaudible beep signals and records the reflected signals from the ear canal. It then analyzes the reflected signals and extracts fine-grained acoustic features that correspond to the ear canal deformation for user authentication. Our experimental evaluation shows that EarDynamic can achieve a recall of 97.38% and an F1 score of 96.84%.

CCS CONCEPTS

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

KEYWORDS

Mobile Authentication, Wearable, Biometrics, Ear Canal, Acoustic Sensing

ACM Reference Format:
Zi Wang, Sheng Tan, Linghan Zhang, Yili Ren, Zhi Wang, and Jie Yang. 2022. An Ear Canal Deformation Based Continuous User Authentication Using Earables. In *The 27th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '21), January 31–February 4, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3447993.3482858

1 INTRODUCTION

Recently, the new-generation mobile devices and ear wearables (earables) have integrated various sensors (e.g., microphones, vibration and motion sensors) to provide better user experience and to support a wide range of emerging applications [1, 7]. For example, the Apple AirPods Pro equips multiple outward-facing and inward-facing microphones to improve the ability of the active noise cancellatio [2]. These embedded sensors in the earables can both support various mobile applications and provide opportunities to sense and capture new types of biometrics. For example, it is possible to utilize the embedded sensors in the ear wearable to capture a user's ear canal structure, which is unique to each individual for user authentication. Moreover, many emerging applications that enabled by the earables, such as the voice assistant for smart home and IoT devices, require secure user authentication to protect sensitive and private information. Traditional voice-based user authentication is convenient but has been proven vulnerable to the voice spoofing attack [8, 9]. Reusing ear wearable to capture the ear canal structure for user authentication thus provides a novel and promising approach to enhance system security.

Comparing to traditional biometrics, the ear canal based authentication has several advantages. First, it relies on the unique geometry of the ear canal, which is hidden within the human skull. It is thus more resilient to the spoofing attack than the traditional biometrics, such as fingerprint, voice, or face. In addition, the ear canal based authentication is transparent to the users as it doesn't require any user cooperation. Traditional biometrics, however, require explicit user operation, such as pressing the fingertip on the fingerprint reader or posing the face to the camera [5]. The unique advantages of the ear canal based authentication could also potentially benefit many emerging applications, such as Virtual Reality (VR) and Audio Augmented Reality (AAR).

In this paper, we introduce EarDynamic, a continuous user authentication system that leverages the ear canal deformation sensed by the earables. Specifically, the ear canal deformation reflects the ear canal dynamic motion caused by jaw joint or articulation activities, for example, when the user is speaking. Thus, the ear canal deformation not only contains the static geometry of ear canal that represents the physiological characteristic of the user but also

---

ACM MobiCom '21, January 31–February 4, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-8342-4/22/01…$15.00
https://doi.org/10.1145/3447993.3482858

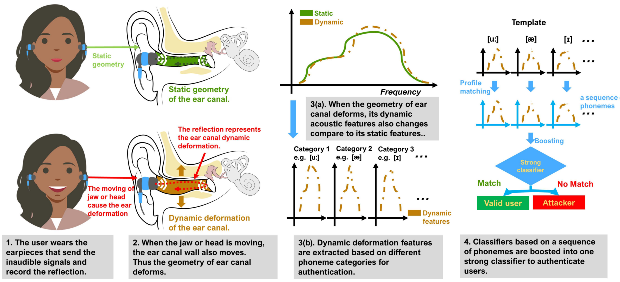

Figure 1: The core idea of the system

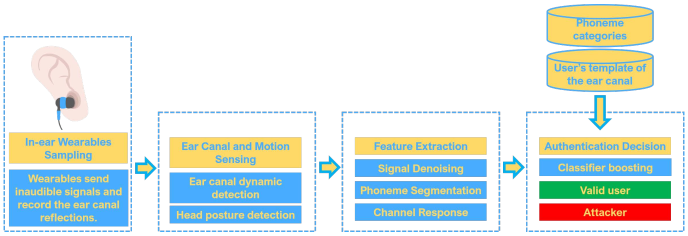

Figure 2: System flow of the system.

includes the geometry changes that characterizes the behavioral properties of the user while speaking. Recent prior work shows that the static geometry of the ear canal is unique for every individual [3]. We find that the ear canal deformation due to articulation activities includes more dynamic information and could provide better and more secure user authentication while the user is speaking.

Although we cannot directly measure the canal dynamic motions, we can infer such motions based on the articulatory movements. However, measuring the articulatory movements requires specialized sensors attached to the articulators, which are impractical. We solve such a challenge by looking at the phonemes in the user’s speech. Specifically, each phoneme pronunciation corresponds to unique and consistent articulatory movements. The canal dynamic motions thus could be identified by recognizing each phoneme and the corresponding articulatory movements. Consequently, the ear canal dynamic motions for the phonemes that are invoked by similar jaw and tongue movements will share high similarity and could be categorized into the same group. Such a categorization can also reduce the calculation complexity and shorten the authentication time.

To perform user authentication, our system extracts fine-grained acoustic features that correspond to the ear canal deformation and compares these features against the user enrolled profile. To evaluate EarDynamic, we conduct experiments with 24 participants in various noisy environments. The results show EarDynamic achieves high accuracy and maintain comparable performance in different noisy environments and under various daily activities.

2 SYSTEM DESIGN

Approach Overview. The key idea underlying our user authentication system is to leverage the advanced acoustic sensing capabilities of the ear wearable device to sense the dynamic deformation of the user’s ear canal. As illustrated in Fig. 1, when the user is wearing EarDynamic, the earbud will emit an inaudible chirp signal to probe the ear canal. Then the signal reflected from the ear canal will be captured by the inward-facing microphone that can be further utilized to extract the dynamic deformation of the ear canal.

Our system has the ability to work under both static and dynamic scenarios of the ear canal. When there is no head movements or articulatory motions detected, the user is under a static scenario, where the ear canal geometry remains the same throughout the authentication process. Thus, the captured signal reflections represent the physiological characteristics of the user’s ear canal. Then the extracted features that correspond to static geometry of the ear canal are utilized to compare against the user enrolled profiles to determine if it is the legit user.

Different from the static scenario, dynamic deformation represents the combination of both the physiological and behavioral characteristics of the ear canal. To better leverage the ear canal deformation, we categorize various deformation motions into different groups based on phoneme pronunciation, such that each group shares similar jaw and tongue movements. Such an approach has the benefit of improving system usability by simplifying the profile enrollment process. For both static and dynamic scenarios, we leverage the channel response to measure the ear canal dynamic deformation and its geometry information. Specifically, the channel response of the ear canal is the ratio of the reflected signal to the incident probe signals. The channel response depicts how the ear canal deformation (i.e., wireless channel). To further enhance our system and make authentication more accurate, we consider each phoneme as one classifier and boost them into one stronger classifier. An adaptive boosting classifier is trained by iteratively adding the primary classifier of each phoneme. The acoustic features of both the static and dynamic of the ear canal are fed into the boosting classifier for authentication.

System Flow. Our system consists of four major components: ear wearable sampling, ear canal, and motion sensing, dynamic feature extraction, and user authentication, shown in Fig. 2. The authentication process could be triggered on-demand or continuously depending on the applications. Once triggered, our system will first send the probe signals and record the signal reflections. Meanwhile, the audible signal that contains phoneme information of the user’s speech will also be recorded. For feature extraction, we need to process the captured signals that consist of the reflected inaudible signals and the audible signals. Our system then segments the separated audible signals into a sequence of phonemes and maps them to the corresponding inaudible components for capturing the ear canal deformation. Lastly, our system will authenticate the user based on the extracted information from previous steps. We utilize a sequence of phoneme-based classifiers that can be combined as one stronger classifier to improve classification accuracy. If a positive decision is given, then the user is considered as legit, otherwise unauthorized.

Ear Canal Deformation Categorization. To better leverage the ear canal deformation and improve the usability of our system, we categorize various dynamic motions into different groups based on the phoneme pronunciations. The underlying principle of ear canal deformation categorization is that similar articulatory gestures, i.e., jaw and tongue motions, will have a similar impact on the

Table 1: Deformation Categories Based on Phoneme

Deformation (Articulator) Category	Phonemes
Tongue Forward, Jaw Open Slightly	[iː] [ɪ] [ɪə] [eɪ] [ə] [eə] [ɜː]
Tongue Lower, Jaw Open Widely	[æ] [ai] [ɒ] [ɑ] [ɔː] [au]
Tongue Back, Jaw Open Slightly	[ʊ] [uː] [ʊə]
Tongue Back, Jaw Open Moderately	[oʊ] [ɔɪ] [e] [ʌ]
Tongue Raised, Jaw Open Widely	[tʃ] [tr] [ts] [dʒ] [dr] [dz]
Tongue Raised, Jaw Open Slightly	[f] [s] [ʃ] [h] [v] [z] [ʒ] [r]
Tongue Fricative, Jaw Open Slightly	[θ] [ð] [l]

geometry of the ear canal. In particular, each phoneme is produced by a sequence of coordinated movements of several articulators. In this work, we mainly focus on two articulators (i.e., jaw and tongue) that contribute the most to the ear canal deformation. For example, the phoneme sound of [ɔː] and [ɑ] both have a lower and backward position with an open jaw. Thus, these two phonemes result in a similar impact on the ear canal deformation and are categorized into the same group. We also eliminate several phonemes due to the fact that they have minimal usage of articulators, which leads to almost no impact on the ear canal deformation. For instance, when the user pronounced the phoneme [p], no ear canal deformation was detected. The categorization results of commonly used vowels and consonants are summarized in Table 1.

As each phoneme contains unique formats (e.g., frequencies), we thus could segment and identify each phoneme by analyzing the spectrogram of the audible signal. In particular, we first leverage the automatic speech recognition protocol to identify each word in the sample speech [10]. Then, we utilize MAUS as the primary way of phoneme segmentation and labeling [4]. It is done by transferring the samples into expected pronunciation and searching for the highest probability in a Hidden Markov Model [10]. The segmented and labeled phonemes will be categorized according to Table 1 for further analysis.

3 PERFORMANCE EVALUATION

Experimental Setup. The authentication process could be happening in various environment under everyday use scenarios. Thus, to evaluate our system's performance in real-world environments, we choose various locations including home, office, grocery store, vehicle, and parks to conduct experiments and ask the participants to wear our system in their natural habits. Existing earbuds that equipped with inward-facing microphones on the market such as Apple Airpods Pro [2] are less desirable due to accessibility of the raw data . In this work, we built our prototype system utilizing only off-the-shelf hardware to demonstrate its practicability and compatibility. The microphone is attached to the earbud where located in front of the speaker, and kept in the center of the cross-section area. Such design can mitigate the impact of wearing the earbuds in different angles. The total cost of this prototype is very low which is more affordable to a wider range of customers.

Data Collection and Metrics. We recruit 24 participants for the experiments including 12 females and 12 males with an age range from 20 to 40. The participants are informed about the goal of our experiments and asked to talk in their natural way of speaking. Each participant is asked to speak 10 sentences with length varies from 2 to 20 words under different environment. The sentences include some commonly used voice commands like "Hey Google", "Alexa, play some music" as well as other short daily conversation pieces.

Overall Performance. We evaluate our system's overall performance against the mimic attack. To launch a mimic attack, the adversary will wear the earables and issue the same voice command by mimicking the victim's way of speaking. For such an attack, the adversary tries to spoof the system by performing similar articulator gestures with respect to the victim. We can observe that EarDynamic can achieve overall accuracy of 93.04%, recall of 97.38%, the precision of 95.02%, and an F1 score of 96.84% across different environments and participants. Furthermore, the median accuracy, recall, precision, and F1 score are 93.97%, 98.78%, 95.40%, and 96.85%, respectively.

4 CONCLUSIONS

In this work, we propose EarDynamic, a continuous and passive user authentication system that leverages the ear canal deformation sensed by the ear wearable. Our study shows that the ear canal deformation due to articulation activities is unique for each individual and contains both the static geometry and dynamic motion of the ear canal when the user is speaking. We sense the ear canal deformation with an acoustic based approach that utilizes the microphone and speaker on the earables. We also build a prototype of EarDynamic with off-the-shelf accessories by embedding an inward facing microphone inside an earbud. Experiment results show that EarDynamic is highly accurate in authenticating users under different noisy environments with various daily activities.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful feedback. This work was partially supported by the NSF Grants CNS-2131143; CNS-1910519; and DGE-1565215.

REFERENCES

[1] T. Amesaka, H. Watanabe, and M. Sugimoto. 2019. Facial expression recognition using ear canal transfer function. In *Proceedings of the 23rd International Symposium on Wearable Computers.* 1–9.

[2] Apple. 2020. Apple AirPods Pro. https://www.apple.com/airpods-pro/

[3] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin. 2019. EarEcho: Using Ear Canal Echo for Wearable Authentication. *Proceedings of the ACM on IMWUT* 3, 3 (2019), 1–24.

[4] T. Kisler, F. Schiel, and H. Sloetjes. 2012. Signal processing via web services: the use case WebMAUS. In *Digital Humanities Conference 2012.*

[5] E. Marasco and A. Ross. 2014. A survey on antispoofing schemes for fingerprint recognition systems. *ACM Computing Surveys (CSUR)* 47, 2 (2014), 1–36.

[6] M. K. Ravishankar. 1996. *Efficient Algorithms for Speech Recognition.* Technical Report. Carnegie-Mellon Univ Pittsburgh pa Dept of Computer Science.

[7] Z. Wang, S. Tan, L. Zhang, and J. Yang. 2018. ObstacleWatch: Acoustic-based obstacle collision detection for pedestrian using smartphone. *Proceedings of the ACM on IMWUT* 2, 4 (2018), 1–22.

[8] L. Zhang, S. Tan, Z. Wang, Y. Ren, Z. Wang, and J. Yang. 2020. VibLive: A Continuous Liveness Detection for Secure Voice User Interface in IoT Environment. In *Annual Computer Security Applications Conference.* 884–896.

[9] L. Zhang, S. Tan, and J. Yang. 2017. Hearing your voice is not enough: An articulatory gesture based liveness detection for voice authentication. In *Proceedings of the 2017 ACM SIGSAC Conference on CCS.* 57–71.

[10] L. Zhang, S. Tan, J. Yang, and Y. Chen. 2016. Voicelive: A phoneme localization based liveness detection for voice authentication on smartphones. In *Proceedings of the 2016 ACM SIGSAC Conference on CCS.* 1080–1091.