

HeartPrint: Passive Heart Sounds Authentication Exploiting In-Ear Microphones

Yetong Cao* Chao Cai† Fan Li* Zhe Chen‡ Jun Luo§

* School of Computer Science and Technology, Beijing Institute of Technology, China

† College of Life Science and Technology, Huazhong University of Science and Technology, China

‡ China-Singapore International Joint Research Institute, China

§ School of Computer Science and Engineering, Nanyang Technological University, Singapore

Email: yetongcao@bit.edu.cn, chriscai@hust.edu.cn, fli@bit.edu.cn, nick.chitchan@gmail.com, junluo@ntu.edu.sg

Abstract—Biometrics has been increasingly integrated into wearable devices to enhance data privacy and security in recent years. Meanwhile, the popularity of wearables in turn creates a unique opportunity for capturing novel biometrics leveraging various embedded sensing modalities. In this paper, we study a new intracorporal biometrics combining the uniqueness of i) heart motion, ii) bone conduction, and iii) body asymmetry. Specifically, we design HeartPrint as a passive yet secure user authentication system: it exploits the bone-conducted heart sounds captured by (widely available) dual *in-ear microphones* (IEMs) to authenticate users, while neatly leveraging IEMs renders itself transparent to users without impairing the normal functions of earphones. To suppress the interference from other body sounds and audio produced by the earphones, we develop a novel interference elimination method using modified non-negative matrix factorization to separate clean heart sounds from background interference. We further explore the uniqueness of IEM-recorded heart sounds in three aspects to extract a novel biometric representation, based on which HeartPrint leverages a convolutional neural model equipped with a continual learning method to achieve accurate authentication under drifting body conditions. Extensive experiments with 18 pairs of commercial earphones on 45 participants confirm that HeartPrint can achieve 1.6% FAR and 1.8% FRR, while effectively coping with major attacks, complicated interference, and hardware diversity.

I. INTRODUCTION

As the key component of ANC (active noise cancellation) earphones for enhancing listening experience, *in-ear microphones* (IEMs) have gained increasing popularity and availability. Thanks to their proximity to (inner) human body, IEMs may enable a great number of applications including human-computer interactions [1], [2] and context-aware computations [3], [4], among which extracting unique biometrics for authentication is an important one. These biometrics include, for instance, ear canal geometry [5], dynamic ear canal deformation [6], and unique body sounds (e.g., voice [7], [8], tooth clicking [9], [10], and walking [11]). However, these approaches heavily involve *active* participation of earphone speakers [5], [6] or users [7]–[11], which may hamper the

Fan Li is the corresponding author. The work of Fan Li is partially supported by the National Natural Science Foundation of China (NSFC) under Grant No.62072040. The work of Chao Cai is partially supported by NSFC under Grant No.6220011218. The work of Jun Luo is partially supported by National Research Foundation (NRF) Future Communications Research & Development Programme (FCP) grant FCP-NTU-RG-2022-015.

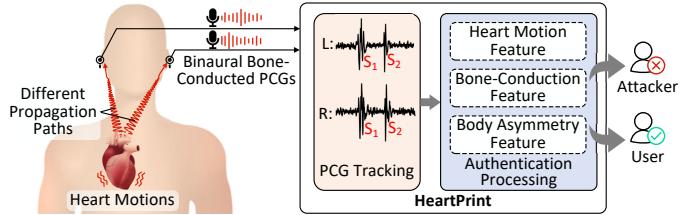


Fig. 1. The conceptual construction of HeartPrint.

normal function of earphones or cause inconvenience to users. Given the growing sensitivity of personal data, there is an urgent need to protect these data from infringement via an authentication solution demanding *zero involvement*.

The heart-based biometrics has rich research history and are known to be unique (i.e., distinguishable across subjects), measurable (i.e., hard to hide), nonvolitional (i.e., unknown to the user), secure (i.e., difficult to counterfeit), and present in all living individuals (i.e., intrinsic liveness) [12]. Consequently, a number of authentication proposals have explored these unique biometrics, including ECG (electrocardiogram) [13], PPG (photoplethysmogram) [14], SCG (seismocardiogram) [15], and BCG (ballistocardiography) [16]. Nonetheless, existing approaches usually rely on special hardware unavailable to common users and are viable only when involving particular user activities (e.g., holding the sensor against the chest). A few proposals have explored earphone-based heart rate tracking [17], demonstrating that heart sounds may be captured by IEMs. Inspired by them, we take one step forward by reusing the IEMs of ANC earphones to capture the unique and fine-grained heart sounds (a.k.a., *phonocardiogram* or PCG) as biometrics to enable *pure passive* user authentication.

To this end, we explore the bone-conducted PCGs collected from both ears as a novel biometrics holding uniqueness in three aspects. First, generated from heart motions (e.g., closure of a valve), PCG is determined by highly individual-dependent heart properties such as geometric shape, volume, and moving dynamics [12]. Second, PCG undergoes complex absorption and dispersion as it propagates toward both ears via bones and tissues. Studies have revealed that individuals hold unique bone-conduction patterns due to their distinct bone structures,

bone densities, and bone-tissue ratios [9]. Third, the human heart, slightly on the left side of the body, causes different propagation paths for PCG to travel toward both ears, while different individuals show distinct body asymmetry patterns as confirmed by various studies [8], [18]. In short, each individual produces unique binaural bone-conducted PCGs.

In this paper, we propose a passive authentication system named HeartPrint, leveraging the unique binaural bone-conducted PCGs captured using dual IEMs widely available in ubiquitous earables (e.g., ANC earphones, hearing aids). As shown in Fig. 1, heart motions produce PCGs that propagate through bones and tissues and finally reach the left and right ears by different paths. HeartPrint captures the binaural bone-conducted PCGs (typically consisting of two major components S_1 and S_2) from IEMs in both ears and extracts unique biometrics for differentiating valid users and attackers. As earables have already well accepted in society, PCG-based authentication on earables offers multiple advantages, including wide availability, ubiquitous deployability, and transparent operations without bothering either users or earables.

Realizing this passive and secure authentication scheme faces several challenges in practice. First of all, PCGs are weak and can easily be buried under stronger background interference (e.g., earphone audio and other body sounds). Therefore, these weak signals must be distilled from stronger interference without impairing the general usage of earphones. Second, though the collected binaural bone-conducted PCGs show distinctions between different users, extracting unique and secure biometric information from them is still an unexplored area. Last but not least, the changes in the user's body over time (e.g., losing weight, lacking sleep) can lead to variations in the data distribution of the extracted biometrics [19]. However, how to thoroughly capture such distributions within a limited registration period can be highly non-trivial in practice.

To address the above challenges, we develop an interference elimination method utilizing the modified *non-negative matrix factorization* (NMF) to separate bone-conducted PCGs from background interference. Then, by studying the characteristics of binaural bone-conducted PCGs, we design a biometric representation that captures uniqueness in heart motion, bone conduction, and body asymmetry. Last, we propose a CNN-based continual learning scheme utilizing *elastic weight consolidation* (EWC), which quickly updates the model parameters using limited data and adaptively accommodates biometric drifts over time. To summarize, we make the following major contributions in HeartPrint:

- We propose HeartPrint, the first earable-based passive authentication system, to leverage both binaural bone-conducted PCGs and IEMs. HeartPrint holds the potential to enable earables with a more anti-attack, disclosure-resistant, and user-friendly authentication experience.
- We develop a novel interference elimination method that effectively reduces the impact of other body sounds and the background audio input of earphones.
- We design a novel biometric representation that captures the uniqueness of the user's heart motion, bone conduction, and body asymmetry information.

- We propose a novel CNN-based continual learning scheme that enables secure authentication using limited registration data and overcomes biometric drifts.
- We conduct extensive evaluations using 18 commercial earphones and 45 participants. The decent results validate the promising practical usability of HeartPrint.

II. PRELIMINARIES

A. Attack Model

Our attack model assumes that an attacker with full knowledge of HeartPrint's working principle has obtained access to the earable device. Since PCGs cannot be controlled consciously, the attacker cannot mimic the victim's PCG on a living human body, but it may launch attacks only in the following ways:

- i) Random Attack: the attacker wears the victim's device and expects to fool HeartPrint using his/her own bone-conducted PCG.
- ii) Replay attacks: Gaining the access to the victim's medical records (in particular the PCG gathered via electronic stethoscopes), the attacker may replay the PCG to HeartPrint for authentication.

Since the attacker is unable to directly record PCGs of a victim, he/she has to steal from the medical records, which makes the replay attack very difficult to launch. Moreover, converting the stolen PCG to the bone-conducted PCG acceptable to HeartPrint can be extremely hard (if not impossible) without the full knowledge of the victim's body construction. Therefore, we omit the security analysis in designing HeartPrint, but only perform experiments to validate its impossibility.

B. Uniqueness Analysis

Sound wave propagating through a homogeneous medium can be modeled as [20]:

$$p = p_0 e^{-\theta d}, \quad (1)$$

where p is the decayed sound pressure, p_0 is the source sound pressure, θ is a decay property of the propagation medium, and d is the propagation distance. Therefore, the bone-conducted PCGs in ears are determined by the sound source (i.e., heartbeat), decay property of bone and tissue, and the relative distance between the heart and the ear.

As shown in Fig. 2, a human heart contains two atria, two ventricles, and four valves. The atria and ventricles

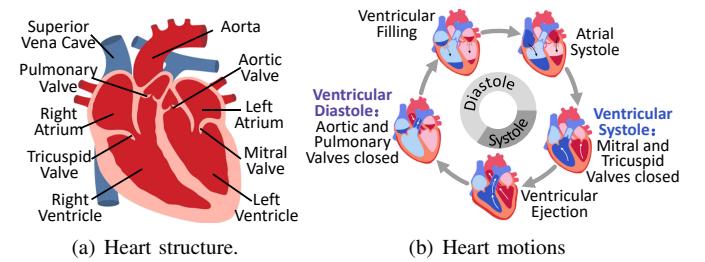


Fig. 2. Heart structure (a) and heart motion cycle (b).

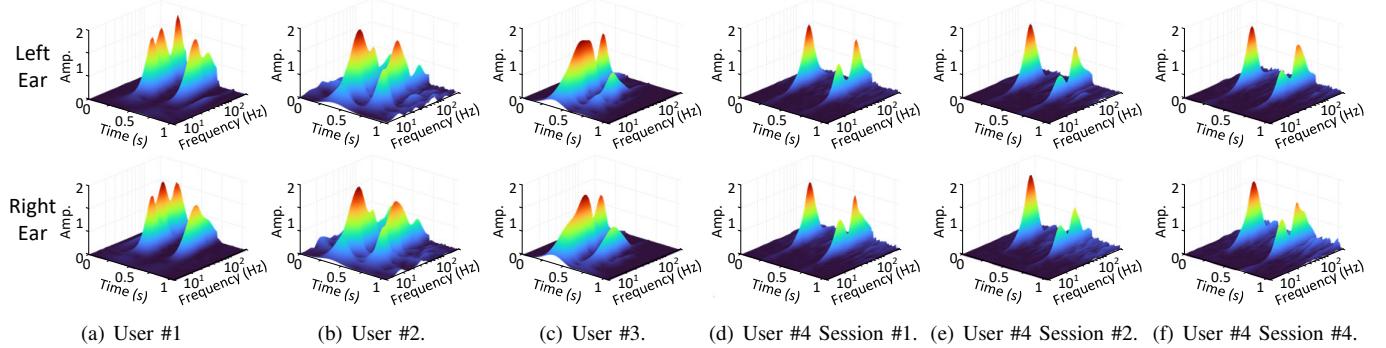


Fig. 3. The CWT of PCG samples collected from different users and the same user across different sessions.

alternatively contract (systole) and relax (diastole) to circulate the blood through lungs and body tissue, which generate PCGs as valves open and close [21]. Previous studies confirmed that different individuals' PCGs hold clear distinctions [22], [23], because their hearts have different volumes, geometric shapes, and moving dynamics (e.g., speed, acceleration). Moreover, the heart sound propagates through bone and tissue and finally reaches the ears. Individual human bodies possess unique properties (i.e., bone structure, bone density, and bone-tissue ratio, thus further adding cross-user distinctions in the resulting bone-conducted PCG. Finally, the heart is at different distances from the left and right ears because the human heart is slightly on the left side of the body; this asymmetry (again very much unique to individual bodies) in turn contributes to the uniqueness of the IEM-recorded PCGs. In short, the pattern of binaural bone-conducted PCGs should uniquely represent individual users.

C. Feasibility Studies

Benefit from the amplification boost induced by the occlusion effect [24], IEMs can capture bone-conducted PCG with improved SNR (signal-to-noise-ratio). To validate the feasibility of using binaural bone-conducted PCGs for authentication, we conduct a preliminary study with four users to answer two fundamental questions.

Do Different Users Produce Unique Binaural Bone-Conducted PCGs?: We ask each user to sit down and collect data using a pair of commercial earphones with IEMs; the continuous wavelet transform (CWT) analysis on users' PCGs are shown in Fig. 3(a)-3(d). In each figure, the top panel shows the results from the left ear and the bottom panel shows those from the right ear. We can readily observe that i) different users exhibit clear distinctions in bone-conducted PCGs, and ii) individual users also exhibit discernible body asymmetry patterns: i.e., differences in frequency and amplitude of PCGs between their right and left ears.

Are the Patterns of Binaural Bone-Conducted PCGs Consistent over Time for the Same User?: We collect data during several sessions at different times, and we let users virtually maintain the same (normal) earphone wearing states across these sessions. We showcase the CWT results from two

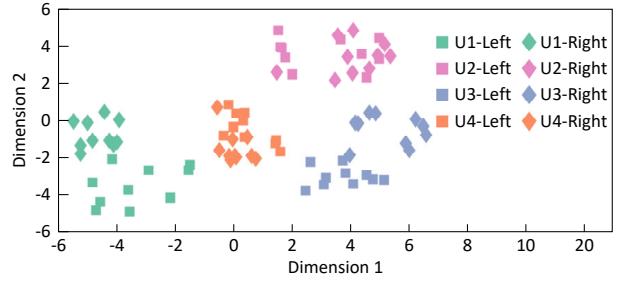


Fig. 4. t-SNE projection of CWT for all users across ten sessions. U1, U2, U3, and U4 represent each user.

additional sessions for user #4 in Fig. 3(e) and 3(f): the results clearly hold consistent patterns across three sessions.

In summary, the binaural bone-conducted PCGs contain sufficient biometrics that is unique across users and remains consistent over time. To reconfirm this, Fig. 4 shows the t-SNE (t-distributed stochastic neighbor embedding) projections of features described in Section III-D across four users, with each user recording bone-conducted PCGs ten times.

III. SYSTEM DESIGN

A. Overview of System Architecture

The architecture of HeartPrint is shown in Fig. 5. The system constantly collects bone-conducted PCGs from the user's left and right ears using earphones with IEMs. In practice, the in-ear sounds inevitably contain interference caused by other body sounds and the background audio input of the earphones. Therefore, HeartPrint first determines whether interference presents in the IEM-recorded sounds based on the periodicity signatures. If interference is detected, the collected data will be served to *Interference Elimination* to separate clean binaural bone-conducted PCGs. Specifically, the *Interference Elimination* generates a signal representation using STFT (short-time Fourier transform) and then decomposes it using the modified NMF; clean bone-conducted PCGs can then be reconstructed based on the clustered sub-components of NMF result. After that, *Representation Extraction* kicks in to derive a novel biometric representation from the binaural bone-conducted PCGs, by combining the uniqueness of heart motion, bone conduction, and body asymmetry. Finally, HeartPrint adopts a CNN-based deep learning model to make *Authentication*

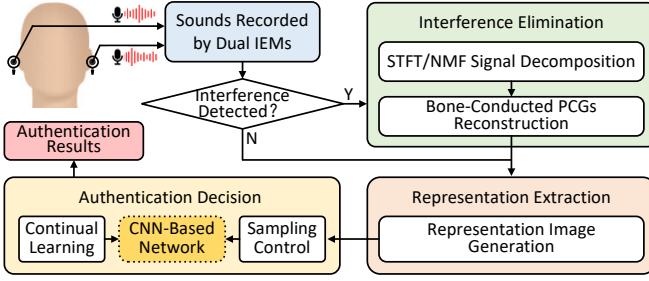


Fig. 5. The architecture of HeartPrint.

Decision. A novel continual learning method is developed to enable efficient training given limited data and overcome biometric drifts over time, with a sampling method to achieve energy-efficient on resource-constrained earphones.

B. Interference Detection

To avoid running expensive interference elimination algorithms during interference-free periods, we build a lightweight method to detect interference. A widely applied solution for interference detection in PCGs is to use the periodicity signatures [25], which first searches for a 3-second interference-free reference PCG signal, then correlates the reference sounds with the collected signal and uses temporal and spectral similarity to discriminate between interference-contaminated periods and interference-free periods. HeartPrint adopts this interference detection method, but it takes the interference-free bone-conducted PCG (obtained during registration) of the user (whom the wearer claims to be) as the reference, saving the additional computation overhead of searching for the reference signal. This modified method not only retains good performance, but also reduces computational costs.

C. Interference Elimination

To achieve reliable and continuous bone-conducted PCG tracking, we need to remove the interference in the collected data. Since bone-conducted PCG mostly impacts the IEM below 160Hz [26], we first employ a low-pass filter with a cut-off frequency of 160Hz to remove high-frequency interference. However, there still left significant interference coming from two sources, including (i) the background audio input of the earphones, and (ii) other body sounds such as walking sounds. Intuitively, we can implement adaptive filtering, EMD (empirical mode decomposition), and wavelet-based approaches to extract clean PCGs. However, these approaches either require additional sensors to obtain interference reference or rely on the assumption of a low aliasing degree of PCG and interference, which are not feasible in our case. Therefore, we build an innovative interference elimination method based on NMF. Although NMF has been utilized for physiological signal denoising [27], [28], to the best of our knowledge, it has never been used for our application scope.

The collected in ear sound $x(n)$ is a mixture of bone-conducted PCG $s(n)$, audio input of the earable $a(n)$, and other body sound $b(n)$, which can be expressed as $x(n) = s(n) + a(n) + b(n)$. The proposed NMF-based interference

elimination framework takes $x(n)$ as input and aims to extract $s(n)$. In the following, we detail the two steps of our design.

1) *STFT/NMF Signal Decomposition:* The input audio signal is first processed by STFT to obtain the time-frequency spectators $x(f, t) = s(f, t) + a(f, t) + b(f, t)$ with $f \in [1, F]$ and $t \in [1, T]$ as the signal representation. Based on that, we obtain the magnitude spectrogram matrix of $x(f, t)$ to be $X = [x_1, \dots, x_T] \in \mathbb{R}^{F \times T}$ and $x_t = [|x(1, t)|, \dots, |x(F, t)|]^T$. Similarly, we define $S \in \mathbb{R}^{F \times T}$, $A \in \mathbb{R}^{F \times T}$, and $B \in \mathbb{R}^{F \times T}$ representing the magnitude spectrogram matrices of $s(f, t)$, $a(f, t)$, and $b(f, t)$, respectively.

The traditional NMF decomposes the magnitude spectrogram matrix as:

$$X \approx \tilde{X} = WH, \quad (2)$$

where $\tilde{X} \in \mathbb{R}^{F \times T}$ is the estimated matrix; $W \in \mathbb{R}^{F \times K}$ is the basis matrix that represents the spectral pattern of different types of signal sources; $H \in \mathbb{R}^{K \times T}$ is the activation matrix that represents when each source occurs during a time frame.

Ideally, by estimating \tilde{X} and clustering the sub-components, we can obtain clean bone-conducted PCGs ($\tilde{S} = W_S H_S$), audio inputs ($\tilde{A} = W_A H_A$), and body sounds ($\tilde{B} = W_B H_B$), where $H = [H_S, H_A, H_B]^T$, and $W = [W_S, W_A, W_B]$. However, applying the traditional NMF method can not handle the wide spectrum of interference of audio inputs and body sounds, thus it is not suitable for our application. Instead, we modify the NMF method by adding an unsupervised component $\tilde{X}_{un} = W_{un} H_{un}$ to handle the large variety of interference not covered in the training phase, avoiding such interference being assigned to the PCGs [29]. Therefore, we estimate \tilde{X} by minimizing the cost function formulated as:

$$\begin{aligned} & \min_{W, H} (\eta_S D(S|\tilde{S}) + \eta_A D(A|\tilde{A}) + \eta_B D(B|\tilde{B}) \\ & + D(X_{un}|\tilde{X}_{un}) + \frac{1}{T} \sum D(X|\tilde{X})), \end{aligned} \quad (3)$$

where η_S , η_A , and η_B are the weights representing the level of co-factorization. D is defined as the sum of the Kullback–Leibler divergence and the sparsity penalty on H :

$$D(X|\tilde{X}) = D_{KL}(X|\tilde{X}) + \mu \|H\|_1, \quad (4)$$

and μ defines the importance of the sparsity constraint.

We solve the optimization problem (3) using a simple yet effective gradient descent method. Instead of the traditional NMF, we adopt a modified cost function, hence our updating rule in the following is also different:

$$\begin{aligned} W & \leftarrow W \otimes \frac{(\frac{X}{WH})H^T + W \otimes (W \otimes OH^T)}{OH^T + W \otimes (W \otimes (\frac{X}{WH})H^T)}, \\ H & \leftarrow H \otimes \frac{W^T \frac{X}{WH}}{W^T O + \mu}, \end{aligned} \quad (5)$$

where W and H are initialized to be random non-negative matrices; \otimes and division denote the element-wise product and division, respectively; $O \in \mathbb{R}^{F \times T}$ is an all-ones matrix composed of F rows and T columns.

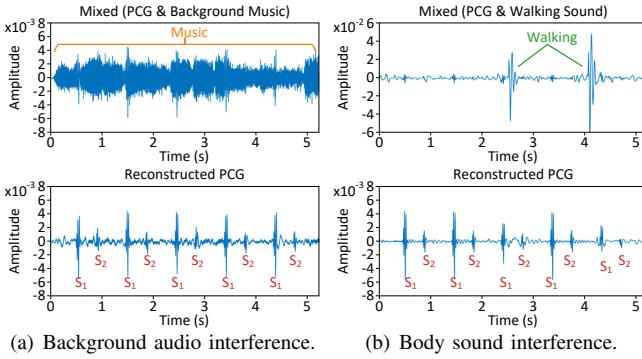


Fig. 6. Examples of interference elimination.

2) *Bone-Conducted PCGs Reconstruction*: To extract clean PCGs, we examine the spectral correlation between the user's registered PCGs and each sub-component of the estimated matrix. Specifically, we adopt the cosine distance to measure the spectral correlation. Having the cosine distance close to 1 suggests that the sub-component is highly likely to belong to the PCG. In practice, we cluster sub-components with cosine distance greater than ψ together, where ψ is experimentally set to be $0.7 \sim 0.9$ depending on the individual. Based on the clustered sub-components, we apply inverse STFT to the extracted clean PCG, *i.e.*, $s(n)$. Fig. 6(a) and Fig. 6(b) show the examples of extracting PCGs from background audio and other body sounds, respectively. We can observe that the impacts of these interference are significantly reduced after applying our proposed method. We also conduct experiments in Section IV-D to further demonstrate the effectiveness of our interference elimination method.

D. Representation Extraction

Our goal is to develop a set of person-distinguishable representations from the binaural bone-conducted PCGs. As statistical features could not cope with the wide spectrum of human biological characteristics and directly processing PCGs using deep learning technology would produce biased biometrics, we design a novel biometric representation that captures the uniqueness of heart motion, body conduction, and body asymmetry between two ears:

a) *Heart Motion Feature*: The *Mel frequency cepstral coefficients* (MFCCs) describe the dynamics among different frequency bands of sounds generated from heart motion, combining linear and nonlinear characteristics [30]. We extract

MFCC from the collected PCGs, which has been confirmed effective in various PCG pattern recognition tasks [31], [32]. Specifically, we extract 16 MFCC features from PCGs of each ear (16×2 features in total).

b) *Body Conduction Feature*: the bone-conducted PCG contains unique information about the conduction properties of bone and tissue. To characterize the sound propagation channel, we use the commonly used acoustic feature for sound resonance analysis, *linear prediction coefficients* (LPCs) [33], [34]. LPC abstracts the complex propagation paths from the heart to an ear as a channel and characterizes the rendering of PCG in the channel. Specifically, we extract 16-order LPCs from data of each ear (16×2 features in total).

c) *Body Asymmetry Feature*: Due to body asymmetry, PCGs propagate to the left and right ears via different paths. Subsequently, the bone-conducted PCG collected from two ears exhibit different degrees of enhancement or attenuation at different frequencies and time frames. To accurately model the body asymmetry information between the two ears, we first divide the bone-conducted PCG of each ear into 32 equal parts according to the frequency range scale between 0 Hz to 160 Hz (*i.e.*, each part has a bandwidth of 5 Hz). Then, we compute the *euclidean distances* (EDs) between the left and right in-ear sounds of each part as the body asymmetry features, which forms a 32-dimensional vector.

To avoid building multiple classifiers for heterogeneous features, we combine them into an integral biometric representation. Specifically, we first segment the binaural bone-conducted PCGs into 130ms non-overlapping windows, which is chosen because human heart cycle typically does not exceed 130 ms [35]. Then we apply Hamming window and segment the data into 32 8ms frames with overlap and extract the three sets of features from each frame. These features are grouped into three matrices (as shown in Fig. 7(a)) corresponding to the MFCCs and LPCs of the left ear, MFCCs and LPCs of the right ear, and EDs in both ears, respectively. Then we combine them into an RGB image by scaling their values to integers between 0 and 255 (as shown in Fig. 7(b)), where the three matrices correspond to the red, green, and blue channels.

E. Authentication Decision

We design a CNN-based deep learning framework to distinguish between attackers and valid users, as shown in Fig. 8. Except for the CNN-based user classifier, we construct a sampling controller to avoid quickly draining the earable's

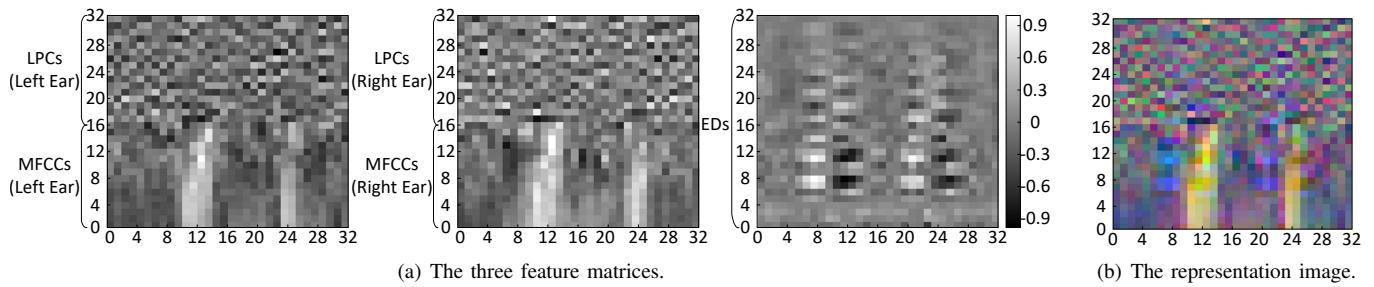


Fig. 7. Our proposed representation extraction scheme.

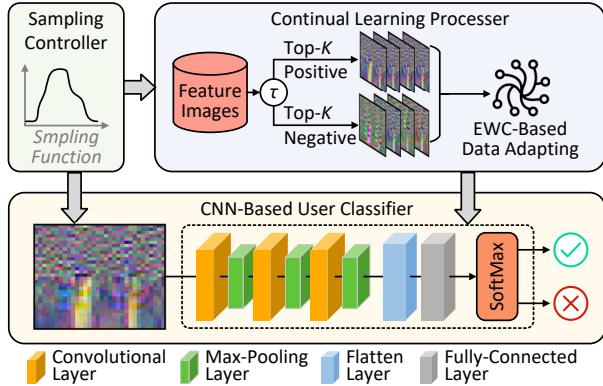


Fig. 8. The CNN-based deep learning framework.

battery. Recall the challenge that biometrics changes over time and users prefer to provide very limited amounts of training data during registration. We build a continual learning method to quickly and adaptively update the parameters of the CNN-based user classifier without retraining a new model.

1) *CNN-Based User Classifier*: Three convolution blocks are used to extract latent information from the input biometric representation; each block involves a 2D-convolution layer using ReLU (rectified linear unit) followed by a max-pooling layer to learn the locally salient information. The three convolution layers use 32, 64, and 64 kernels, respectively, and the kernel sizes are 3×3 . Then, we build a flatten layer following the convolution blocks to make the extracted multidimensional representation into a one-dimensional vector. Through a fully connected layer with 64 units and an output layer using the *Softmax* function, the vector is compressed and finally classified as a valid user or an attacker.

2) *Continual Learning Processor*: Continuously updating the biometric information during usage would be useful to enrich the training dataset and learn the changing data distributions over time. We tackle this by unsupervised enriching the training datasets and adapting to new data using EWC [36]. By carefully studying plenty of incremental learning and transfer learning paradigms, we choose EWC as it is one of the most prominent approaches to cope with catastrophic forgetting issue [37]. Our basic idea is to protect the performance of the existing CNN-based model while modifying the EWC method to adapt the CNN-based model to new training datasets.

Specifically, we set a threshold τ as the minimal probability of recognizing a valid user. When a new sample is fed into the CNN-based user classifier, we calculate the probability p that the user is valid and compare it with τ . If p is larger than τ , we pseudo-label the sample as positive (valid user); otherwise pseudo-label the sample as negative (attacker). Then, we sort the samples based on their probabilities of being classified as a valid user and construct a new training dataset by selecting the top- K positive/negative samples with the highest/lowest probabilities. Specifically, K is empirically set as 70% in our proof-of-concept study but can be adjusted depending on the particular situation.

Then we update the CNN-based model using the recently

built training dataset. Considering learning of the old dataset and the new dataset are two tasks M and N, the basic idea is to utilize EWC to constrain the change of the most important parameters for the old task. The standard EWC can be implemented by adding a quadratic penalty on the difference between the parameters (i.e., loss function):

$$L = L_N(\theta_{MN}) + \sum_i \frac{\lambda}{2} \cdot \Gamma_{M,i}(\theta_{MN,i} - \theta_{M,i}^*)^2, \quad (6)$$

where θ_{MN} is the weights and biases to optimize performance, $L_N(\theta_{MN})$ is the loss for task N without that penalty, λ sets the importance of task M compared with task N. Γ is the diagonal of Fisher information matrix with i labeling each parameter.

We expect this approach can enable CNN-based models to cope with limited registration data and biometric drifts. However, inaccurate parameters of the initial CNN-based model (constructed at registration) and the typically overestimated the approximated parameter importances λ could hinder the continual learning [38]. Therefore, the importance of a parameter θ_k is modified to include how much it contributed to a drop in the loss w_k and how far it moved compared to its initial value Δ_k . Subsequently, the total loss function is modified by adding a surrogate loss which approximates the summed loss functions of previously trained datasets as:

$$L^* = L + c \sum_k \Psi_k(\tilde{\theta}_k - \theta_k)^2, \quad (7)$$

where c is a constant representing the weight of the pre-parameters $\tilde{\theta}_k$, which is empirically set as 0.6. The regularization strength $\Psi_k = \sum w_k / (\Delta_k)^2$ is summed over all training datasets. Experiment demonstrates that the proposed method significantly benefits the system performance, which is presented in Section IV-E.

3) *Sampling Controller*: Continuously running authentication algorithm can easily drain the small battery of earables. Thus, we design a variable authentication frequency control scheme to trade off continuity against energy-efficiency. We adjust the authentication frequency to be high when intense interference presents and low when there is little interference. Therefore, we define a variable frequency function $\mu(k)$ proportional to the change of energy of the contaminated PCGs $x(n)$ and the separated clean bone-conducted PCGs $s(n)$:

$$\mu(k) = \begin{cases} \mu_0 & \text{if } 0 \leq \omega(k) < \mu_0 \\ \omega(k) & \text{if } \mu_0 \leq \omega(k) < \beta, \\ 1 & \text{if } \beta \leq \omega(k) \end{cases} \quad (8)$$

where $\omega(k) = \alpha \cdot \mathbb{E}(x_k(n)) / \mathbb{E}(s_k(n))$ with k labels each 130 ms window, α , μ_0 , and β are experimentally determined constants in our default model. α sets $\mu(k)$ between 0 and 1, and μ_0 and β can be manually adjusted for systems with different security requirements. Using this method, the transient interference generated when the wearer changes can alert the system to increase the authentication frequency. Whereas the authentication frequency keeps low when the user keeps a constant state to save energy.

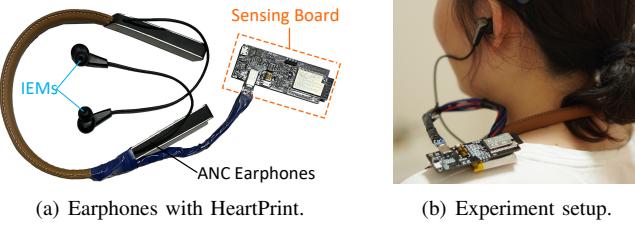


Fig. 9. HeartPrint (on ANC earphones) in action.

IV. EVALUATION

A. Experimental Methodology

1) *Implementation:* Although IEMs are getting embedded into many commercial ANC earphones, we notice that existing manufacturers do not provide direct access to raw IEMs measurements. Therefore, we build a sensing board (as shown in Fig. 9(a)) with a dedicated codec to collect IEM data in real time, which bypasses the earphone’s circuits and does not affect the acoustic front-ends. We equip the sensing board with an ESP32 [39] module, which provides wireless connectivity via Wi-Fi and Bluetooth, to send IEM-recorded sounds to a remote laptop for processing. The sensing board can accommodate major ANC earphones, and we test 18 pairs of commercial ANC earphones with varying prices around \$100. All involved devices have silicone earbud tips with user-specific sizes to ensure good ambient noise isolation.

2) *Data Collection:* We recruit 45 participants, including 26 males and 19 females aging between 20 to 50, for evaluation. All data are collected following an experimental protocol approved by the ethical review board of the institute. The entire data collection process took three months, during which the participants are asked to respectively wear each pair of earphones in their natural habits and record the binaural bone-conducted PCGs for about 130 to 200 minutes, as shown in Fig. 9(b). To better simulate the practical usage of HeartPrint, we particularly choose various locations, including a controlled lab room, home, and a park to conduct experiments. Meanwhile, a trained expert uses a digital stethoscope to record the PCG over four main heart valve areas as the source for replay attacks. Scenarios without and with interference are adopted to evaluate HeartPrint. We particularly conduct a 24-hour experiment to validate HeartPrint’s effectiveness for continuous monitoring. Besides, we conduct a number of experiments to validate HeartPrint’s key algorithms. Finally, we evaluate HeartPrint with different earphones wearing states.

3) *Evaluation Metrics:* We use two metrics to evaluate HeartPrint. *False accepted rate* (FAR) is the percentage of unauthorized users accepted by the system; *False reject rate* (FRR) is the percentage of a valid user incorrectly rejected by the system; The smaller these metrics are, the better the performance is.

B. Overall Performance and Security Study

We first evaluate the overall performance of HeartPrint by examining the FAR and FRR in the static scenario. We conduct a 5-fold-cross-validation where data are randomly shuffled

and use 20% for testing and the remaining 80% for training. During the experiment, we consider each participant (e.g., P_1) as a valid user and remaining participants (e.g., $P_2 \sim P_{45}$) as attackers. The experiment simulates the random attack where attackers expect their data can fool the system. Overall, the average FAR and FRR over 45 participants are 1.6% and 1.8%, respectively. Fig. 10 and Fig. 11 show the detailed FAR and FRR results of each participant. We notice that HeartPrint achieves good performance and gives accurate classification labels based on the captured biometrics.

Moreover, we study the replay attack scenario that attackers attempt to forge the binaural bone-conducted PCGs using chest PCG gathered via electronic stethoscopes. Specifically, we attach a loudspeaker in the heart position to a medical human silicone simulator (have a similar texture to the human skin) and use earphones to record the replayed PCGs that propagate to the simulator’s ears. We find that all participants have FAR below 3.6% and obtain the average FAR of 2.3%, as shown in Fig. 12. The reason is that although the replayed chest PCG greatly correlate with binaural bone-conducted PCG, they fail to present the unique body asymmetry information. Overall, the successful thwarting of random attack and replay attack attempts confirms that HeartPrint provides a high level of resistance to major attacks.

C. 24-hour Study

We ask five participants to continuously record their binaural bone-conducted PCGs for 24 hours (contain daily activities) and show the FAR and FRR averaged over all five participants in Fig. 13. We note that FAR and FRR fluctuate slightly over the day. This is because heart rhythm changes all the time, causing small variations in the IEM-recorded PCGs. Indeed, such a phenomenon conforms to the common sense that the human state affects heart rhythms. Overall, the FAR results are all below 3.2% with an average value of 2.0%. Besides, the FRR results are all below 5.0% with an average value of 3.8%. The low FAR and FRR over 24 hours confirm the robust usability of HeartPrint for continuous authentication.

D. Effectiveness of Interference Elimination

To further study the effectiveness of our proposed interference elimination method, we use the data collected in the static scenario for training and data collected under interference (i.e., background audio input and other body sounds) for testing. We conduct experiments according to the interference source (i.e., $IS \in \{\text{audio, motion}\}$), and whether HeartPrint is implemented with the interference elimination method (i.e., $IE \in \{\text{with, without}\}$). We particularly study four cases: Case i : $IS = \text{audio}, IE = \text{without}$; Case ii : $IS = \text{motion}, IE = \text{without}$; Case iii : $IS = \text{audio}, IE = \text{with}$; and Case iv : $IS = \text{motion}, IE = \text{with}$. Fig. 14 shows the authentication results in view of FAR and FRR of four cases. The FAR seems to be immune to interference and remains stable. If HeartPrint is implemented without applying interference elimination method, the FRR reaches over 35.0%, which confirms the challenge that interference greatly degrades

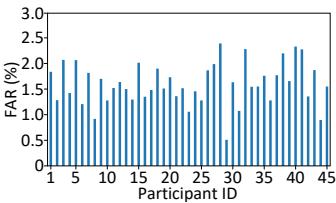


Fig. 10. Overall FAR.

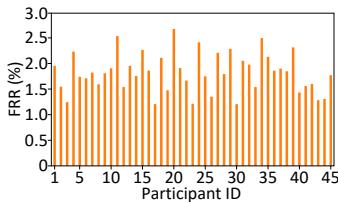


Fig. 11. Overall FRR.

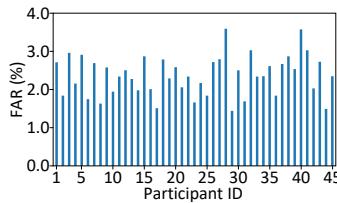


Fig. 12. FAR of replay attacks.

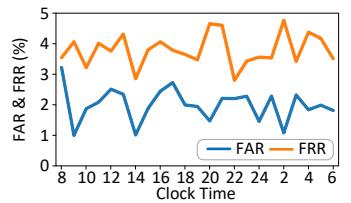


Fig. 13. 24-hour performance.

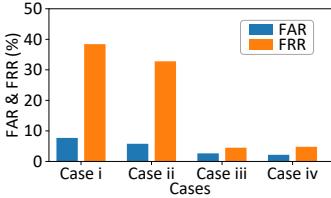


Fig. 14. Performance with & without interference elimination.

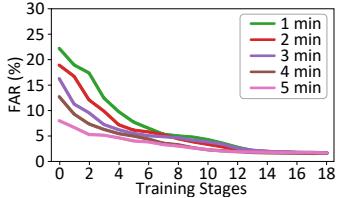


Fig. 15. FAR of different training data sizes.

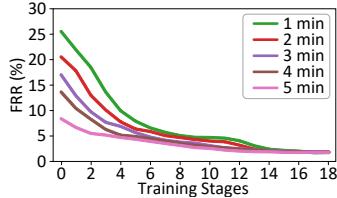


Fig. 16. FRR of different training data sizes.

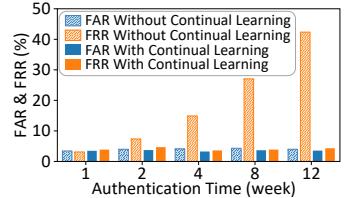


Fig. 17. Long-term performance.

the authentication performance. After applying interference elimination, the FRR drops by 30.6% to 4.4%. The decent performance not only verifies the effectiveness of our interference elimination method, but also shows that HeartPrint is a practical approach promising to be applied in real daily usage.

E. Effectiveness of Continual Learning

The continual learning method aims to relieve the burden of providing a large amount of registration data and achieve robustness against biometric changes over time. We investigate the effectiveness of our continual learning method in terms of registration data size and long-term performance.

1) Amount of Registered Data for Training: Since the training data size influences the CNN-based deep learning network's performance and time for data collection, we particularly evaluate 1, 2, 3, 4, and 5 minutes static binaural bone-conducted PCGs of each participant for training, respectively. Fig. 15 and Fig. 16 respectively show the FAR and FRR with different continual training stages (5 minutes of data are added for each stage). At stage 0, the training data size contributes to improving the model accuracy. Using 1-minute data for training has the worst FAR of 22.1% and the FRR of 25.5%, which is insufficient for secure authentication. After applying the continual learning method, we can notice the overall trend of FAR and FRR of each case drop significantly. In particular, the case of using 1-minute training data has FAR below 2.0% and FRR below 2.5% when the training stage increases to 14. The results show the effectiveness of our proposed method, that is, it has high confidence using limited registration data.

2) Long-Term Performance: The continual learning method helps the system adapt to more data; thus, the system per-

formance changes accordingly. We study HeartPrint's long-term stability over three months and particularly test each participant's data after the periods of 1, 2, 4, 8, and 12 weeks, respectively. We show the averaged FAR and FFR of HeartPrint with and without applying the continual learning method in Fig.17. Without using the continual learning method, we can observe that the FAR and FRR stay stable within 4 weeks, but increase rapidly after 5 weeks. After using the continual learning method, HeartPrint can maintain FAR and FRR below 2.5% in 12 weeks. The results not only confirm the effectiveness of our continual learning method, but also indicate our system has good robustness in long-term stability.

F. Impact of Earphone Wearing States

The IEM-recorded sounds of the same participant could be different with different earphone insertion depths and rotation angles, arising from different wearing habits and earphone shapes. We conduct experiments by asking the participants to collect data with insertion depth $\gamma = \{2mm, 4mm, 6mm\}$ and rotation angle $\rho = \{0^\circ, 45^\circ, 90^\circ\}$. Table I summarizes the FAR and FRR quantified by the above two factors. Overall, HeartPrint obtains stable results in different wearing states. Besides, we notice that larger insertion depths result in lower FAR and FRR. This may be because deeper insertion of earphones provides better ear canal sealing and a more solid position fixation.

V. RELATED WORK

User authentication on mobile and wearable devices is indispensable to prevent users' privacy from leakage. Recently, heart-based biometrics has emerged as a popular authentication method since they offer the advantages of liveness detection, thereby providing strong tolerance to spoofing and replay attacks [41]. Existing solutions mainly utilize the EEG [13], SCG [15], BCG [16], and PPG [14], [42]. However, obtaining ECG signals requires attaching additional hardware to the mobile devices and wearables, which incur a high cost and narrow deployment. Although SCG and BCG can be recorded using widely integrated motion sensors, they only work in specific

TABLE I
IMPACT OF DIFFERENT DEVICE WEARING STATES.

| $\rho \backslash \gamma$ | 2 mm | 4 mm | 6 mm | 8 mm | 10 mm |
|--------------------------|----------|----------|----------|----------|-----------------|
| 0° | FAR=1.9% | FAR=1.6% | FAR=1.4% | FAR=1.4% | FAR=1.3% |
| | FRR=2.2% | FRR=1.9% | FRR=1.7% | FRR=1.6% | FRR=1.6% |
| 45° | FAR=2.1% | FAR=1.6% | FAR=1.4% | FAR=1.4% | FAR=1.4% |
| | FRR=2.5% | FRR=1.9% | FRR=1.8% | FRR=1.6% | FRR=1.6% |
| 90° | FAR=2.0% | FAR=1.6% | FAR=1.4% | FAR=1.4% | FAR=1.4% |
| | FRR=2.2% | FAR=1.9% | FRR=1.7% | FRR=1.7% | FRR=1.6% |

TABLE II
COMPARISON WITH EARABLE-BASED AUTHENTICATION APPROACHES.

| Work | Nakamura et al. [40] | EarDynamic [6] | EarEcho [5] | EarGate [11] | TeethPass [9] | Vocal Resonance [7] | EarPrint [8] | HeartPrint |
|-----------------------------------|-----------------------|------------------|---------------------|---------------|-------------------------|---------------------|--------------|----------------------|
| Modality | EEG sensors | IEMs | IEMs | IEMs | IEMs | IEMs | IEMs | IEMs |
| Feature | EEG rhythms | Ear canal motion | Ear canal structure | Walking sound | Teeth occultation sound | Voice sound | Voice sound | PCG |
| Unrestricted Scenario | Yes | No | Yes | No | No | No | No | Yes |
| Mode | Passive | Active | Active | Semi-active | Semi-active | Semi-active | Semi-active | Passive |
| Well-engineered commercial device | No | No | No | No | No | No | No | Yes |
| Subject | 15 | 24 | 20 | 31 | 22 | 29 | 23 | 45 |
| Result | FAR<2.3% FRR<32.2% | Acc.=93.04% | BAC>94.5% | BAC>80% | FRR=1.6% | Acc. 94.2%-96.1% | Acc.=96.36% | FAR=1.6% FRR=1.8% |

Acc. represents accuracy; BAC represents the balanced accuracy.

recording conditions such as holding the sensor against the chest, which involve cumbersome efforts for frequent use. PPG is a popular approach widely used in commercial smartwatches and fitness trackers, but it is sensitive to the dynamic gestures that compress the arterial geometry [43], making PPG-based methods not applicable in scenarios that involve hand gestures. In our work, we only use the IEMs widely integrated into commercial earphones to capture the heart sounds propagating to both sides of the ears, introducing a novel passive heart-based biometric for authentication.

The proliferation of earables has created unique opportunists for authentication systems. Nakamura *et al.* [40] propose to identify users with the in-ear EEG sensor, but the requirement of dedicated sensors unavailable on commercial devices makes it unsuitable to be adopted widely. Our method only employs the readily available IEMs of ANC earphones, thus is promising for widespread deployment. EarEcho [5] and EarDynamic [6] emit audible/inaudible ultrasonic sounds to probe the ear canal and identify users. However, occupying the earphone's speaker hinders the earphone's audio functionality, resulting in a poor user experience. There have been active works that extract body sounds for authentication [7]–[11]. They rely on users' participation in performing particular activities, such as walking, teeth clicking, and speaking. Our method transparently and passively tracks the binaural bone-conducted PCGs to authenticate users without impairing the common usage of the earphones. Moreover, we particularly propose an interference elimination method to ensure secure authentication in various scenarios.

Table II compares HeartPrint with existing typical works on earable-based authentication in terms of the used modality, feature, usage scenario restriction, sensing mode, the used device, involved subjects, and the reported result. To the best of our knowledge, we are the first passive and continuous authentication solution that leverages binaural bone-conducted PCGs and IEMs. Moreover, all the prior approaches are implemented using 1 to 3 pairs of self-build prototypes (e.g., manually attach microphones into non-smart earphones) with a small number of participants. In contrast, HeartPrint uses 18 pairs of commercial ANC earphones and the evaluation involves 45 subjects, which is so far the largest device and participant set that has been used for earable-based authentication solutions.

VI. DISCUSSION AND FUTURE WORK

As a novel biometric authentication system, HeartPrint may still exhibit a few limitations in the current stage. First, the energy computation of HeartPrint with different resource-constrained earphones is yet to be explored. It is worth investigating how the proposed sampling control method can help HeartPrint operate at a non-trivial energy floor. Second, we have not seriously studied the impact of hardware differences between earphones. However, our experiments are conducted using all 18 pairs of different commercial earphones, where hardware differences exist and do not seem to affect HeartPrint. Third, we have not explored the impact of ambient noise. As HeartPrint is designed for ANC earphones that integrate IEMs, the impact of ambient noise could be readily reduced by the phase inversion function [44]. Last but not least, the propagation pattern of PCGs during speaking could change due to changes in oral morphology. We are planning to combine HeartPrint with the existing bone-conducted voice-based authentication method [8] to further improve the practical usability of HeartPrint.

VII. CONCLUSION

In this paper, we propose a new passive authentication system HeartPrint; it exploits the unique binaural bone-conducted PCGs and IEMs widely available in commercial earphones. Notably, HeartPrint works continuously and is transparent to users without impairing the common usage of earphones. To render HeartPrint practical, we propose an interference elimination method to eliminate the interference caused by other body sounds and the background audio input of the earphones. Leveraging the uniqueness of heart motion, bone conduction, and body asymmetry, HeartPrint extracts a novel biometric representation from the binaural bone-conducted PCGs. Based on this representation, a CNN-based neural model employing an innovatively designed continual learning method is developed to enable accurate authentication while coping with limited training data and biometric drifts. Our experiments with 45 participants using 18 pairs of commercial ANC earphones demonstrate that HeartPrint has very low FAR and FRR under various practical conditions. We envision HeartPrint to be an appealing line of defense for earphones privacy (stand alone or companion with smartphone) thanks to its security, unobtrusiveness, and user-friendliness.

REFERENCES

- [1] T. Amesaka, H. Watanabe, and M. Sugimoto, "Facial Expression Recognition Using Ear Canal Transfer Function," in *Proc. of the 23rd ACM ISWC*, 2019, pp. 1–9.
- [2] X. Xu, H. Shi, X. Yi, W. Liu, Y. Yan, Y. Shi, A. Mariakakis, J. Mankoff, and A. K. Dey, "EarBuddy: Enabling On-Face Interaction via Wireless Earbuds," in *Proc. of ACM CHI 2020*, 2020, pp. 1–14.
- [3] D. Ma, A. Ferlini, and C. Mascolo, "OESense: Employing Occlusion Effect for In-Ear Human Sensing," in *Proc. of the 19th ACM MobiSys*, 2021, pp. 175–187.
- [4] Q. Yang and Y. Zheng, "DeepEar: Sound Localization with Binaural Microphones," in *Proc. of IEEE INFOCOM 2022*, 2022, pp. 1–10.
- [5] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin, "EarEcho: Using Ear Canal Echo for Wearable Authentication," *Proc. of ACM IMWUT 2019*, vol. 3, no. 3, pp. 1–24, 2019.
- [6] Z. Wang, S. Tan, L. Zhang, Y. Ren, Z. Wang, and J. Yang, "EarDynamic: An Ear Canal Deformation Based Continuous User Authentication Using In-Ear Wearables," *Proc. of ACM IMWUT 2021*, vol. 5, no. 1, pp. 1–27, 2021.
- [7] R. Liu, C. Cornelius, R. Rawassizadeh, R. Peterson, and D. Kotz, "Vocal Resonance: Using Internal Body Voice For Wearable Authentication," *Proc. of the ACM IMWUT*, vol. 2, no. 1, pp. 1–23, 2018.
- [8] Y. Gao, Y. Jin, J. Chauhan, S. Choi, J. Li, and Z. Jin, "Voice in Ear: Spoofing-Resistant and Passphrase-Independent Body Sound Authentication," *Proc. of ACM IMWUT 2021*, vol. 5, no. 1, pp. 1–25, 2021.
- [9] Y. Xie, F. Li, Y. Wu, H. Chen, Z. Zhao, and Y. Wang, "TeethPass: Dental Occlusion-based User Authentication via In-ear Acoustic Sensing," in *Proc. of IEEE INFOCOM 2022*, 2022, pp. 1–10.
- [10] Z. Wang, Y. Ren, Y. Chen, and J. Yang, "Toothsonic: Earable authentication via acoustic toothprint," *Proc. of ACM IMWUT 2022*, vol. 6, no. 2, pp. 1–24, 2022.
- [11] A. Ferlini, D. Ma, R. Harle, and C. Mascolo, "EarGate: Gait-Based User Identification With In-Ear Microphones," in *Proc. of The 27th ACM MobiCom*, 2021, pp. 337–349.
- [12] F. Lin, C. Song, Y. Zhuang, W. Xu, C. Li, and K. Ren, "Cardiac Scan: A Non-Contact and Continuous Heart-Based User Authentication System," in *Proc. of the 23rd ACM MobiCom*, 2017, pp. 315–328.
- [13] J. S. Arteaga-Falconi, H. Al Osman, and A. El Saddik, "Ecg authentication for mobile devices," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 3, pp. 591–600, 2015.
- [14] Y. Cao, Q. Zhang, F. Li, S. Yang, and Y. Wang, "PPGPass: Nonintrusive and Secure Mobile Two-Factor Authentication via Wearables," in *Proc. of IEEE INFOCOM 2020*, 2020, pp. 1917–1926.
- [15] L. Wang, K. Huang, K. Sun, W. Wang, C. Tian, L. Xie, and Q. Gu, "Unlock With Your Heart: Heartbeat-Based Authentication on Commercial Mobile Phones," *Proc. of ACM IMWUT 2018*, vol. 2, no. 3, pp. 1–22, 2018.
- [16] K. Jiekeng, G. Jakllari, and A.-L. Beylot, "I Want to Know Your Hand: Authentication on Commodity Mobile Phones Based on Your Hand's Vibrations," *Proc. of ACM IMWUT*, vol. 6, no. 2, p. 58, 2022.
- [17] S. Nirjon, R. F. Dickerson, Q. Li, P. Asare, J. A. Stankovic, D. Hong, B. Zhang, X. Jiang, G. Shen, and F. Zhao, "MusicalHeart: A Hearty Way of Listening to Music," in *Proc. of the 10th ACM SenSys*, 2012, pp. 43–56.
- [18] L. Kompanets, "Biometrics of Asymmetrical Face," in *Proc. of ICBA*, 2004, pp. 67–73.
- [19] R. Asplund and H. Aberg, "Sleep and Cardiac Symptoms Amongst Women Aged 40-64 Years," *Journal of Internal Medicine*, vol. 243, no. 3, pp. 209–213, 1998.
- [20] T. L. Szabo, "Chapter 12 - Nonlinear Acoustics and Imaging," in *Diagnostic Ultrasound Imaging: Inside Out (Second Edition)*, T. L. Szabo, Ed. Academic Press, 2014, pp. 501–563.
- [21] T. R. Reed, N. E. Reed, and P. Fritzson, "Heart Sound Analysis For Symptom Detection and Computer-Aided Diagnosis," *Simulation Modelling Practice and Theory*, vol. 12, no. 2, pp. 129–146, 2004.
- [22] K. Phua, J. Chen, T. H. Dat, and L. Shue, "Heart Sound As A Biometric," *Pattern Recognition*, vol. 41, no. 3, pp. 906–919, 2008.
- [23] N. El-Bendary, H. Al-Qaheri, H. M. Zawbaa, M. Hamed, A. E. Hassanien, Q. Zhao, and A. Abraham, "HSAS: Heart Sound Authentication System," in *Proc. of IEEE NaBIC 2010*, 2010, pp. 351–356.
- [24] K. Carillo, O. Doutres, and F. Sgard, "Theoretical Investigation of The Low Frequency Fundamental Mechanism of The Objective Occlusion Effect Induced By Bone-Conducted Stimulation," *The Journal of The Acoustical Society of America*, vol. 147, no. 5, pp. 3476–3489, 2020.
- [25] D. Kumar, P. Carvalho, M. Antunes, R. Paiva, and J. Henriques, "Noise Detection During Heart Sound Recording Using Periodicity Signatures," *Physiological Measurement*, vol. 32, no. 5, pp. 599–618, 2011.
- [26] P. Arnott, G. Pfeiffer, and M. Tavel, "Spectral Analysis of Heart Sounds: Relationships Between Some Physical Characteristics and Frequency Spectra of First and Second Heart Sounds In Normals and Hypertensives," *Journal of Biomedical Engineering*, vol. 6, no. 2, pp. 121–128, 1984.
- [27] M. Niegowski and M. Zivanovic, "ECG-EMG Separation by Using Enhanced Non-Negative Matrix Factorization," in *Proc. of the 36th IEEE EMBC*, 2014, pp. 4212–4215.
- [28] C. Ye, K. Toyoda, and T. Ohtsuki, "Blind Source Separation on Non-Contact Heartbeat Detection by Non-Negative Matrix Factorization Algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 482–494, 2019.
- [29] E. Grooby, C. Sitaula, D. Fattah, R. Sameni, K. Tan, L. Zhou, A. King, A. Ramanathan, A. Malhotra, G. A. Dumont, and Others, "Noisy Neonatal Chest Sound Separation For High-Quality Heart and Lung Sounds," *Arxiv Preprint Arxiv:2201.03211*, pp. 1–12, 2022.
- [30] J. Li, K. Fawaz, and Y. Kim, "Velody: Nonlinear Vibration Challenge-Response For Resilient User Authentication," in *Proc. of ACM CCS 2019*, 2019, pp. 1201–1213.
- [31] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of Feature-Based and Deep Learning-Based Classifiers For Detection of Abnormal Heart Sounds," in *2016 Computing In Cardiology Conference*, 2016, pp. 621–624.
- [32] W. Fu, X. Yang, and Y. Wang, "Heart Sound Diagnosis Based on DTW and MFCC," in *Proc. of the 3rd IEEE CISPL*, vol. 6, 2010, pp. 2920–2923.
- [33] H. Wakita, "Estimation of Vocal-Tract Shapes From Acoustical Analysis of The Speech Wave: The State of The Art," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 3, pp. 281–285, 1979.
- [34] O. O. Akande and P. J. Murphy, "Estimation of The Vocal Tract Transfer Function With Application To Glottal Wave Analysis," *Speech Communication*, vol. 46, no. 1, pp. 15–36, 2005.
- [35] A. Leal, D. Nunes, R. Couceiro, J. Henriques, P. Carvalho, I. Quintal, and C. Teixeira, "Noise Detection in Phonocardiograms by Exploring Similarities in Spectral Features," *Biomedical Signal Processing and Control*, vol. 44, pp. 154–167, 2018.
- [36] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, and Others, "Overcoming Catastrophic Forgetting In Neural Networks," *Proc. of The National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.
- [37] R. M. French, "Catastrophic Forgetting In Connectionist Networks," *Trends In Cognitive Sciences*, vol. 3, no. 4, pp. 128–135, 1999.
- [38] F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," in *International Conference on Machine Learning*, 2017, pp. 3987–3995.
- [39] Espressif, "ESP32 Technical Reference Manual," 2022. [Online]. Available: https://www.espressif.com/sites/default/files/documentation/esp32_technical_reference_manual_en.pdf
- [40] T. Nakamura, V. Goverdovsky, and D. P. Mandic, "In-Ear EEG Biometrics for Feasible and Readily Collectable Real-World Person Authentication," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 648–661, 2017.
- [41] A. S. Rathore, Z. Li, W. Zhu, Z. Jin, and W. Xu, "A Survey on Heart Biometrics," *ACM Computing Surveys*, vol. 53, no. 6, pp. 1–38, 2020.
- [42] T. Zhao, Y. Wang, J. Liu, Y. Chen, J. Cheng, and J. Yu, "TrueHeart: Continuous Authentication on Wrist-Worn Wearables Using PPG-Based Biometrics," in *Proc. of IEEE INFOCOM 2020*, 2020, pp. 30–39.
- [43] T. Zhao, J. Liu, Y. Wang, H. Liu, and Y. Chen, "PPG-Based Finger-Level Gesture Recognition Leveraging Wearables," in *Proc. of IEEE INFOCOM 2018*, 2018, pp. 1457–1465.
- [44] B. Benoit, C. Camasta, M. Kenny, K. Li, R. Romanowski, and S. Kevin, "Engineering Silence: Active Noise Cancellation," *North Carolina State University*, pp. 1–12, 2012.