

EarACE: Empowering Versatile Acoustic Sensing via Earable Active Noise Cancellation Platform

YETONG CAO, Beijing Institute of Technology, China

CHAO CAI, Huazhong University of Science and Technology, China

ANBO YU, HT Acoustics Technology, China

FAN LI, Beijing Institute of Technology, China

JUN LUO, Nanyang Technological University, Singapore

In recent years, particular attention has been devoted to *earable acoustic sensing* due to its numerous applications. However, the lack of a common platform for accessing raw audio samples has forced researchers/developers to pay great efforts to the trifles of prototyping often irrelevant to the core sensing functions. Meanwhile, the growing popularity of *active noise cancellation* (ANC) has endowed common earphones with high standard acoustic capability yet to be explored by sensing. To this end, we propose EarACE to be the first acoustic sensing platform exploiting the native acoustics of commercial ANC earphones, significantly improving upon self-crafted earphone sensing devices. EarACE takes a compact design to handle hardware heterogeneity and to deliver flexible control on audio facilities. Leveraging a systematic study on in-ear acoustic signals, EarACE gains abilities to combat performance sensitivity to device wearing states and to eliminate body motion interference. We further implement three major acoustic sensing applications to showcase the efficacy and adaptability of EarACE; the results evidently demonstrate EarACE's promising future in facilitating earable acoustic sensing research.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: Earable computing, acoustic sensing, platform development, ANC earphones, vital signs monitoring, user authentication.

ACM Reference Format:

Yetong Cao, Chao Cai, Anbo Yu, Fan Li, and Jun Luo. 2023. EarACE: Empowering Versatile Acoustic Sensing via Earable Active Noise Cancellation Platform. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 2, Article 47 (June 2023), 23 pages. <https://doi.org/10.1145/3596242>

1 INTRODUCTION

Enabling a new modality of hands-free sensing and low-intrusive information exchange, *earable computing* [17] has ushered in a revolution in human wearable computing [33, 49]. Though promising progress has been achieved for HCI (human-computer interaction) [4, 6, 35, 36] and context-aware computations [9, 28, 29, 37, 39], relying on dedicated earable sensors renders them expensive and inflexible, as opposed to leveraging *earphones* that attract significant attention recently due to their low-cost and ubiquity. In particular, a recent proposal [22] exploits earphones to sense the difference in air pressure between two ears for profiling human states; it shows successes in recognizing heart rate, touch gestures, identities, and even voices. However, leveraging differential signal only captures part of the original excitation sources, so this design may not be applicable to scenarios demanding fine-grained information.

Authors' addresses: [Yetong Cao](mailto:yetongcao@bit.edu.cn), yetongcao@bit.edu.cn, Beijing Institute of Technology, China; [Chao Cai](mailto:chriscai@hust.edu.cn), chriscai@hust.edu.cn, Huazhong University of Science and Technology, China; [Anbo Yu](mailto:yuanbo@ht-acousitcs.com), yuanbo@ht-acousitcs.com, HT Acoustics Technology, China; [Fan Li](mailto:fli@bit.edu.cn), fli@bit.edu.cn, Beijing Institute of Technology, China; [Jun Luo](mailto:junluo@ntu.edu.sg), junluo@ntu.edu.sg, Nanyang Technological University, Singapore.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2023 Copyright held by the owner/author(s).

2474-9567/2023/6-ART47

<https://doi.org/10.1145/3596242>

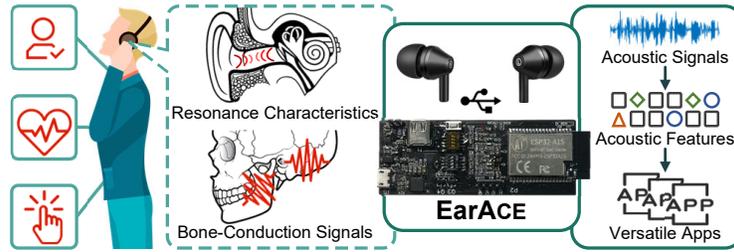


Fig. 1. The conceptual construction of EarACE.

Meanwhile, as the demand for a more natural communication experience and the awareness of noise-induced hearing loss continue to grow, the market share of *active noise cancellation* (ANC) earphones increases drastically. Given their high-standard in-ear acoustic facilities (e.g., microphones), they are expected to improve existing earable sensing capabilities significantly. However, proposals claiming to be designed for ANC earphones, such as behavioral analysis [3, 14, 32], daily health monitoring [34], and user authentication [25–27, 59], are often implemented and tested on self-built prototypes; yet lacking a *well-engineered construction of native acoustics* has made the practical effectiveness of these prototypes very questionable. Essentially, the missing of flexible access to the native acoustics offered by ANC earphones is holding back the progress of earable acoustic sensing, as researchers/developers have to make repetitive efforts in (re-)building hardware prototypes often incompetent and irrelevant to core sensing algorithms. In a nutshell, whether ANC earphones may enhance wearable sensing or simply stop at being a daily fashion accessory *strongly depends on the development of a common sensing platform that directly exploits the native acoustics offered by ANC*.

To this end, we present EarACE, the first earable ANC platform to empower versatile acoustic sensing exploiting the native acoustics of commercial ANC earphones, aiming to supplant self-crafted earphone sensing devices. As shown in Fig. 1, signals induced by human vitals and activities propagate to their ears through bones and cause unique ear canal resonance characteristics. EarACE captures them with ANC in-ear microphones to enable versatile sensing functionalities. Specifically, EarACE aims to

- *Enable Convenient and Unobtrusive Sensing*: A compact plug-in peripheral as the sensing platform, readily attachable to arbitrary ANC earphones.
- *Allow Flexible Control and Reconfiguration on Acoustic Facilities*: A user-friendly GUI to select microphones or speakers available to sense or transmit, tune sensor configuration, and enable/disable the ANC function.
- *Ensure Reliable Measurements*: Detecting fine-grained *device wearing state* (DWS) to obtain desired signal quality, while eliminating body motion interference to ensure accurate sensing under practical scenarios.

In order to meet the above requirements, we propose several design strategies for implementing EarACE. First of all, we identify common features of the audio front-ends to determine the basic platform design, minimizing the adaptations to accommodate hardware heterogeneity. Secondly, we develop an Android-based control interface in order to mask the nuance of the underlying hardware; it enables users to readily configure platform parameters and access sensing data. Thirdly, as DWSs may significantly affect the sensing quality, we carefully study the acoustic characteristics under different DWSs. This study allows us to quantitatively profile their impacts by particular features of an impedance indicator, which in turn enables EarACE to detect fine-grained DWSs. Finally, to cope with body motions heavily affecting earable sensing quality, we investigate the different acoustic characteristics between left and right ears: by associating path sensitivity of signals with non-periodicity of interference, EarACE can be endowed with an interference reference that further warrants an elimination algorithm to ensure reliable sensing in real-life scenarios.

Given the above strategies, EarACE exhibits three main technical advantages over existing earable platforms, mainly due to our novel design strategies and direct integration with ANC native acoustics. First, it empowers versatile acoustic sensing upon earables with a much wider frequency range than existing developments. Second, it provides utility functions to ensure reliable sensing measurements, allowing researchers and developers to focus on core sensing algorithms without being troubled by reliability issues. Third, EarACE supports a broad range of applications, including both active and passive sensing. In summary, our paper makes the following major contributions:

- We propose EarACE as the first acoustic sensing platform leveraging commercial ANC earphones; it aims to enable flexible access to the native acoustics of the increasingly popular ANC facilities for enhancing in-ear acoustic sensing.
- We develop a plug-in platform and an Android-based control interface to greatly simplify the development of acoustic sensing applications, by making laborious hardware-related details transparent to developers.
- Enabled by EarACE, we are the first to extract fine-grained heartbeat waveforms via earables; it is widely understood that fine-grained heartbeat monitoring is highly non-trivial by wearable sensing.
- As sensing applications often call for the awareness of DWSs, we develop a novel algorithm to quantify fine-grained DWSs, so as to improve sensing quality under diversified circumstances.
- We propose an interference elimination algorithm leveraging, for the first time, dual-ear asymmetry; it associates path sensitivity of signals with non-periodicity of motion-induced interference, in order to extract clean measurements for ensuring reliable sensing.
- We demonstrate the efficacy and adaptability of EarACE by implementing three representative applications, namely heartbeat monitoring, hand-face interaction analysis, and user authentication; the promising results firmly indicate that EarACE is fully capable of meeting diversified acoustic sensing requirements.

Note that, in addition to delivering a versatile earable sensing platform for supporting common research interests, EarACE enables every application developed upon it to be readily convertible into commercial products, exactly because of its seamless integration with ANC native acoustics. This second aspect further makes EarACE significantly superior to existing counterparts.

The rest of this paper is organized as follows. In Section 2, we present the background on in-ear acoustic sensing, along with our initial studies to motivate our design; the common literature review has been merged into this section for brevity. In Section 3, we give an overview of the EarACE platform and then elaborate on its hardware and software components. We further report our implementation of three applications along with their evaluations in Section 4. Finally, we discuss the limitations in Section 5 and conclude the paper in Section 6. Since EarACE aims to facilitate the transition from an experiment platform to fully integrated earable-computing-ready ANC earphones, we elaborate on its construction details in the Appendix to supplement the brief overview in Section 3.

2 UNDERSTANDING IN-EAR ACOUSTICS

Latest earable sensing platforms often have limited applicability: they either lose track of critical information (e.g., HeadFi [22] captures only the differential signals between two ears) or are restricted to certain frequency ranges (e.g., OESense [32] handles signals only below 50Hz). Therefore, in order to gain versatile capabilities in a wider range of acoustic sensing, EarACE aims to fully leverage the well-engineered native acoustics offered by ANC earphones. To this end, we first investigate the opportunities and challenges of exploiting ANC earphone acoustics (e.g., in-ear microphones), by exploring three major categories of human sensing:

Physiological Signal Monitoring. As one of the “inlets” into human body, the ear canal offers unique opportunities for acoustic sensing to capture physiological signals, such as respiration rate [38, 57, 58], heart rate [34], and coughing [53], which can be leveraged to detect physiological or pathological states (e.g., sleep stages and

diseases). Herein, we explore a challenging case of acquiring the heart sounds produced by the closure of the heart valves, a.k.a. *phonocardiogram* (or PCG) [2].

Activity and Behavior Analysis. Given their proximity to the human body, earphones allow for close track of various body activities and behaviors. Though several earable computing applications [3, 4, 41] have succeeded in facial expression recognition, head movement detection, and step counting, we believe earable acoustic sensing can further advance activity and behavior analysis. In particular, considering that *hand-face interactions* [32] such as tapping and sliding could generate unique vibrations, we investigate a practically usable task of recognizing hand-face interactions, which could deliver a new input experience to users.

Security and Authentication. As earphones are increasingly pervasive in our daily life, they may provide an additional line of security defense for mobile devices by, for example, delivering a stand-alone authentication channel. Recently, the research interest has focused on ear canal biometrics, such as the static geometry of the ear canal [27] and ear canal deformation due to articulation activities [59]. Based on the facts that the human ear canal is a closed space consisting of different reflective surfaces and that the acoustic signals reflected by these surfaces encode unique features [52], we study the practical application of using earphones for *user authentication* via *active sensing*, i.e., playing random contents (e.g., music and voice messages) as acoustic excitation and sensing the signal reflections as “earprint”.

2.1 Understanding Feasibility

We first investigate whether ANC native acoustics are sufficiently sensitive to support the three typical applications. Specifically, we conduct the experiment in a quiet lab and record i) PCG signal, ii) the sound of sliding and tapping on the cheek, and iii) the in-ear reflection of a piece of music from three volunteers (denoted as V_1, V_2, V_3); we derive a transfer function [27] from the music reflections as biometrics. Fig. 2(a) shows examples of the measurements for three applications in respective rows.

One period of PCG is known to consist of two sounds: S_1 indicates atrioventricular valves closing and S_2 indicates the semilunar valves closing [1]. From the top panel, we can clearly observe that the recorded signals successfully capture both S_1 and S_2 , showing that in-ear microphones are highly sensitive to such subtle excitation signals. We can also observe, from the middle panel, distinctive signal fluctuations when sliding (denoted by W_s) and tapping (denoted by W_t) behaviors occur, demonstrating the feasibility of behavior analysis. Finally, the bottom panel confirms that the transfer functions of three volunteers exhibit evidently discernible distinctions,

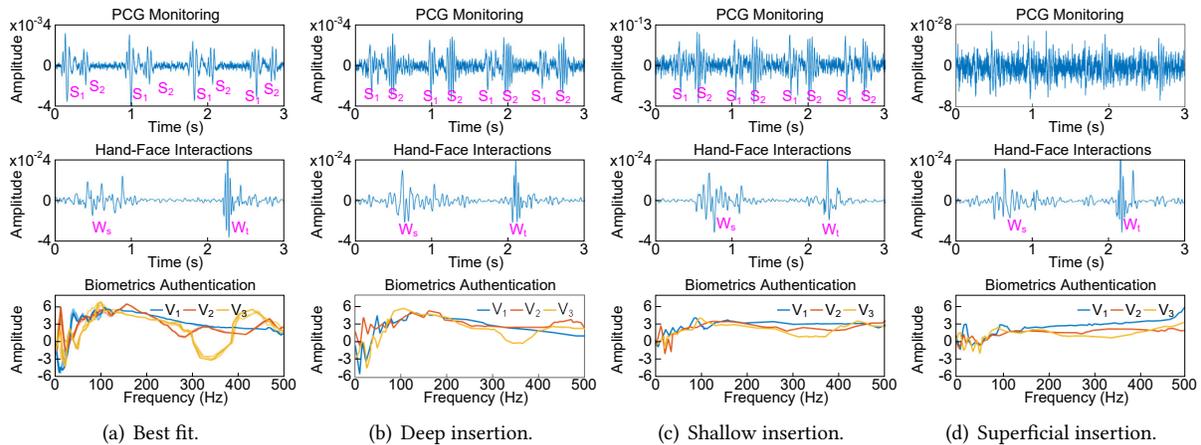


Fig. 2. Impact of device wearing states (DWSs) on three acoustic sensing applications.

proving that acoustic characteristics are sensitive to the ear canal resonance and can hence be adopted as biometrics. In short, this preliminary study confirms the feasibility of both passively and actively using in-ear acoustics to enable various sensing applications.

2.2 Impact of Device Wearing States (DWSs)

When placing an object in an ear canal (a small S-shape tube roughly 30 mm in length), the in-ear acoustic characteristics are affected by i) *occlusion*, ii) *occlusion effect*, and iii) *ear resonance* [48]. Occlusion denotes the closure of the ear canal (e.g., by earphone as an earplug), causing attenuation of external acoustic signals. Occlusion effect, not to be confused with occlusion, is the self-generated intensification of low-frequency components (< 50Hz) when an ear canal is occluded; a shallow insertion of an earpiece strengthens this effect. Ear resonance refers to the high-frequency (> 100 Hz) amplification associated with an open (or partially open) ear canal, which can be greatly reduced by a deep earpiece insertion. In short, the in-ear acoustic characteristics may vary significantly with the DWS.

We explore three DWSs in this preliminary study (leaving detailed explorations on fine-grained DWS to Section 3.3), namely deep insertion (completely occluded), shallow insertion (with a small vent), and superficial insertion (with a large vent); the results demonstrating that changes in DWS have different impacts for these applications are shown in Fig. 2(b), 2(c), and 2(d), respectively. Firstly, for PCG monitoring (the top panels), the two separated heart sounds are evident under both deep and shallow insertion conditions, but they become barely recognizable under the superficial insertion condition. Further measurements indicate that the interfering sources come from the external environment. Secondly, for activity analysis (the middle panels), we find that the on-face sliding and tapping sounds can always be differentiated regardless of the DWS, demonstrating that the hand-face interactions collected from ears are less sensitive to the DWS. Thirdly, for biometric authentication (the bottom panels), we can observe that changing the DWS has a significant impact on the transfer function of the ear canal, and the impact varies from subject to subject. Completely occluding the ear canal allows the transfer functions of different subjects to exhibit discernible differences over the range of [0, 50]Hz and [300, 500]Hz, whereas the other two conditions render the transfer functions of all subjects insufficient for the identification purpose.

In summary, quantifying fine-grained DWS is often crucial to improve sensing quality. However, current earable acoustics sensing solutions only support the binary classification of whether the wearer removes the headphones or not [25], whereas quantitative analysis of DWS remains to be an understudied area, which motivates our new designs by exploiting the electro-acoustic model [15] in Section 3.3.

2.3 Interferences to Acoustic Sensing

Among the diversified interferences to our in-ear acoustic sensing applications, we select three typical ones, namely nodding (affecting ear canal resonances), walking (generating bone-conduction noise), and speaking (causing mixed interferences). As a baseline, we also consider a quasi-static sitting position; the results are reported in Fig. 3 following the same convention as Fig. 2. Firstly, PCG signal monitoring appears immune to nodding (comparable to sitting), but speaking and walking largely overwhelm the heart sounds. Secondly, the hand-face interactions are similar to PCG, as nodding has minor impacts on the recorded signals but the other two activities can substantially contaminate the sensing results. Fortunately, these interferences and on-face interactions do not necessarily occur at the same time, hence they could be removed by straightforward time-domain filtering. Finally, biometric authentication would better be conducted under quasi-static (i.e., sitting) state, because speaking significantly alters the acoustic spectrum below 200Hz, while walking produces a wider range of interference from 10Hz to 300Hz.

We tentatively conclude that active sensing (i.e., biometric authentication) is sensitive to ear canal resonance noise while passive sensing (i.e., PCG monitoring and hand-face interaction analysis) may not. More prominently,

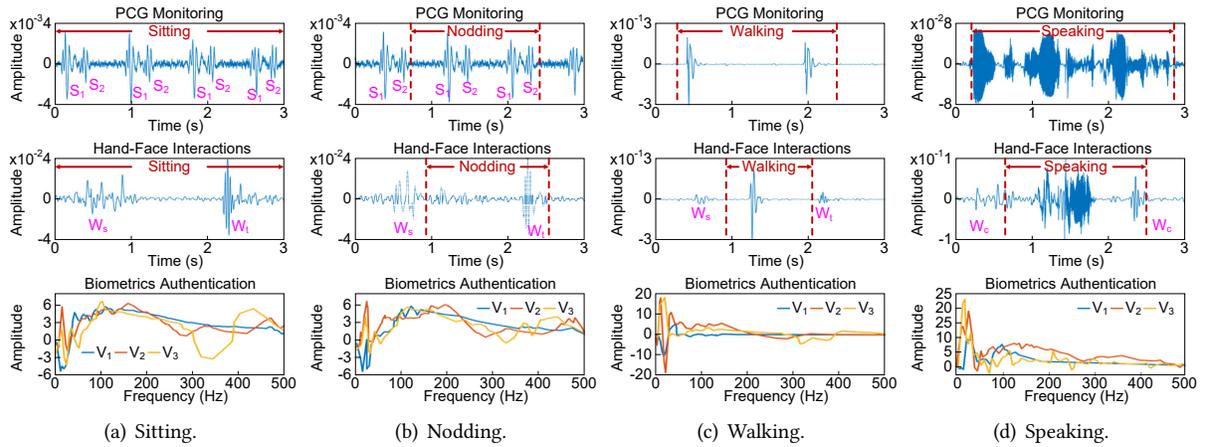


Fig. 3. Diversified interferences to three acoustic sensing applications.

as both sensing modes are sensitive to bone-conduction noise, addressing this interference is very critical for improving the sensing performance, but it is also very challenging as bone-conduction noise often affects different applications in distinctive ways. For excitation signals occurring intermittently (e.g., hand-face interactions), noise might be eliminated by time-domain trimming, but continuous excitation signals (e.g., PCG) cannot be handled similarly. Although many earable acoustic sensing systems have respectively designed denoising methods, they only focus on eliminating noise in specific frequency bands relevant to their respective applications [22, 25], which failed to cover the wide frequency range of real-life bone-conduction noise (covering from 0.5Hz upward and caused by a vast number behaviors such as walking and speaking). Therefore, it is imperative to design advanced interference elimination methods to handle the complex nature of bone-conduction noises.

2.4 Differential Signals Between Two Ears

Due to the asymmetry of human bodies (e.g., bones, muscles, organs, and especially two ears), the signals collected in two ears are normally different. On the contrary, earphones are artificially made to be symmetric, so differential signals collected from both ears may remove system errors (or unknown system parameters) possessed by earphones [22]. Nonetheless, signal differences may not retain the features of the raw signals, so we hereby study whether and how we may explore two-ear signal difference in EarACE given our three typical

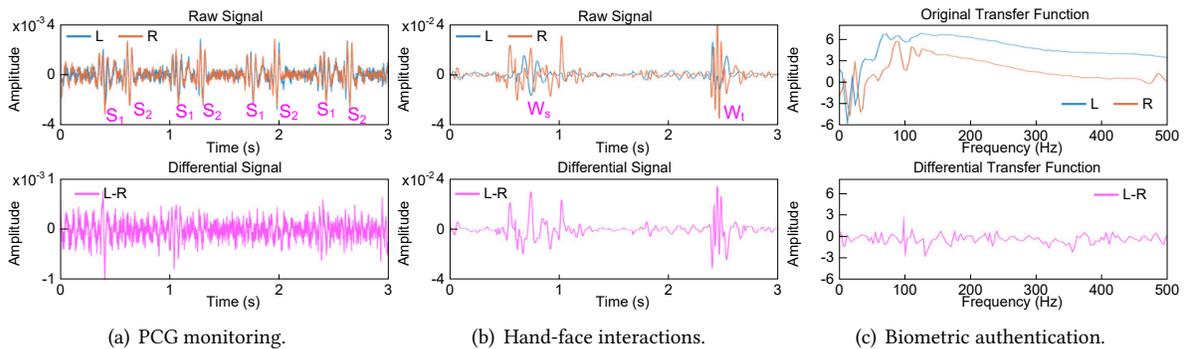


Fig. 4. Demonstrating acoustic signal differentiation between two ears.

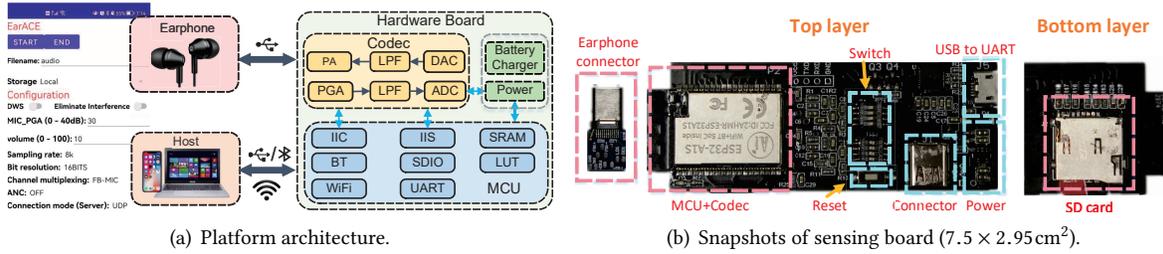


Fig. 5. The overview of EarACE: (a) overall system construction and (b) snapshot of its hardware.

applications. In Fig. 4, the top panels show signals collected from two ears (under the sitting state), while the differential signals are depicted on the bottom panels. One may clearly observe that, whereas differentiating both PCGs and transfer functions results in a white-noise-like outcome, the hand-face interactions do seem to admit signal differentiation: the results share prominent features with the signals collected from the right ear when it was the right-hand tapping/sliding on the face. As EarACE directly uses commodity ANC microphones with well-engineered acoustic capabilities to collect acoustic signals, it does not need to exploit signal differentiation to remove unknown parameters. However, the positive case with hand-face interactions does offer us an inspiration to exploit the dual-ear asymmetry so as to address the interference issue raised in Section 2.3.

3 SYSTEM DESIGN

As shown in Fig. 5(a), EarACE contains a sensing board and a platform control interface on an arbitrary host (e.g., an Android phone). It also involves two novel algorithms for ensuring robust acoustic sensing. While the sensing board can be readily attached to most ANC earphones in a plug-in manner, the control interface provides flexible tuning of hardware and sensing parameters. The sensing board streams data to the host in real time, which are processed by the sensing algorithms for profiling the DWSs and eliminating the body motion interference. In the following, we elaborate on the individual components of EarACE.

3.1 Hardware Design

Given the design requirements raised in Section 1, EarACE's sensing board has a compact design but offers powerful processing capabilities and versatile compatibility with major ANC earphones. As shown in Fig. 5, EarACE is built upon ES8388 [21] stereo codec with a configurable sampling rate ranging from 8 to 96 kHz and up to 24-bit resolution. The internal PGA (programmable gain amplifier) has a maximum of 24 dB gain, adequate to accommodate microphones with heterogeneous sensitivity. The acoustic front-end, including multi-channel microphones and speakers, is routed to this codec via a USB-C socket. We specifically engineer this USB connector to accommodate major analog microphones, including Electret Condenser Microphones (ECM) and MEMS (micro-electromechanical systems) ones. Additionally, we design the (signal) pre-processing circuit to be hardware-adaptive to microphone heterogeneity, entailing only minor changes on resistors or capacitors. This codec can be configured by an MCU via a serial interface. We refer interested readers to the Appendix for more details on wiring ANC earphones up to EarACE.¹

We use ESP32 [20], a feature-rich MCU with versatile connectivity and plentiful memory spaces as the core processor. This powerful MCU is meant to not only enable flexible configurations on the codec but also deploy stand-alone real-time processing algorithms. To obtain convenient usability and accessibility, we leverage this MCU to also provide wireless connectivity via either Wi-Fi or Bluetooth for remote processing; sensing results,

¹Note that EarACE does not change the default capability of ANC earphones in simultaneous sensing (by microphones) and communication (by speakers); it only piggybacks on the acoustic facilities to further expand their sensing capability.

raw audio samples, or configurations can all be streamed between the host and the MCU via wireless links. To allow stand-alone continuous sampling of a huge volume of data, we specifically put an auxiliary SD card slot on the hardware board to enable local data storage. We reserve the UART port for debugging and for experienced researchers to update or develop prioritized firmware. Other onboard resources include a battery charger, a user button, several reserved IOs for hardware extensions such as IMUs, and switches for audio channel multiplexing. This design allows EarACE to emulate the full functionality of a potential sensing-ready ANC earphone, capable of offering insights for a future integrated product design.

3.2 Platform Control

EarACE offers multiple approaches for platform control. One can flexibly choose the way to access raw audio samples and playback arbitrary sounds in either stand-alone or remote mode; both modes offer full functionality for configurations and sensing. In particular, the recorded channels can be manually configured by switches shown in Fig. 5(b) under both modes. The stand-alone mode requires no host support and is therefore suitable for continuous sensing with negligible user interventions. One can start or stop the recording or playback processes via a user button (on the bottom layer hence not shown in the figure due to space limit) under this mode. Given the wireless connectivity discussed in Section 3.1, we specifically write a MATLAB script for PCs and an Android application for smartphones to enable remote configurations. As shown in Fig. 5(a), the Android app allows for flexibly tuning microphone gain (via the MIC_PGA item), adjusting the speaker volume, changing the sampling rate or bit resolution, etc. In addition to these essential settings, we also provide application-level configurations such as DWS profiling and ANC on-off switch. For the MATLAB script running on PCs, these settings are listed in a readable JSON file and are synchronized to the sensing board once a connection is established.

3.3 Profiling Device Wearing States (DWSs)

As studied in Section 2.2, profiling DWS across different users and earphones is crucial, yet it is inherently challenging as both ear canals and earphones have diverse structures.

3.3.1 Elector-Acoustic Model. To qualitatively explain the effect of DWS on the acoustic characteristics of an ear canal, we introduce two configurations of an electro-acoustic model [15]: namely *perfect* and *partial occlusions* distinguished by the absence and presence of a *vent* at the ear canal entrance. Specifically, the ear canal is considered a cylindrical tube with varying radii in both cases.

For the perfect occlusion shown in Fig. 6(a), the acoustic behavior can be modeled in the frequency domain as:

$$\frac{P_{EC}}{Z_S U_S} = \frac{Z_{EC}}{Z_S + Z_{EC}}, \quad (1)$$

where P_{EC} and Z_{EC} are respectively the sound pressure and the corresponding impedance at the ear canal, U_S and Z_S are the source sound volume velocity and source impedance, related to the Thevenin equivalent source sound pressure by $P_S = U_S Z_S$. Eqn. (1) indicates that the Thevenin pressure P_{EC} is affected by both Z_S and Z_{EC} , where Z_{EC} is a varying factor under all perfectly occluded cases; its value changes with the insertion depth of an occlusion device.

For the partial occlusion shown in Fig. 6(b), the presence of a vent makes the earlier model insufficient. We simplify the vent as a high-pass filter with impedance Z_V and sound pressure P_V , so that the total acoustic impedance of the ear canal becomes the parallel of Z_V and Z_{EC} :

$$Z'_{EC} = \frac{Z_V Z_{EC}}{Z_V + Z_{EC}}. \quad (2)$$

Though the actual model for this case is far more complicated and less well studied, a simplified version can be obtained by replacing Z_{EC} in Eqn. (1) with Z'_{EC} ; this allows us to safely conclude that the Thevenin pressure P_{EC}

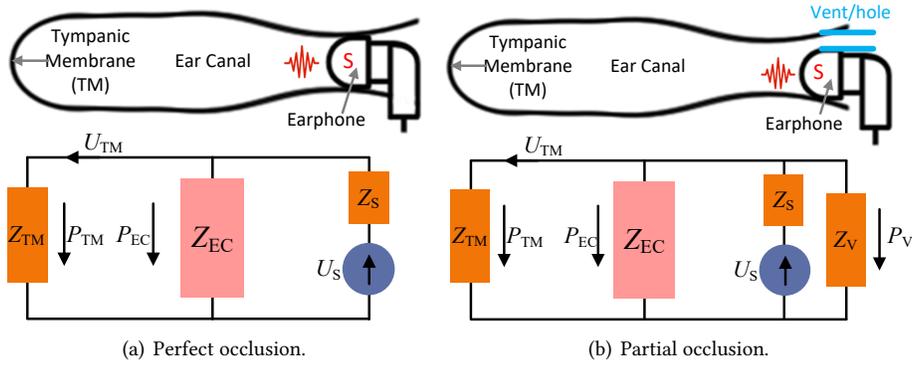


Fig. 6. Two electro-acoustic models of ear canal.

is a function of Z_S , Z_{EC} , and Z_V . Since Z_V can be approximated as a function of the vent radius and the distance from the vent to the tympanic membrane [15], the vent size and device insertion depth both affect P_{EC} .

Remarks: In short, the degree of occlusion (vent size) and the earphone insertion depth both affect the acoustic characteristics of an ear canal significantly. This also highlights the necessity to profile DWS for all in-era acoustic sensing applications.

3.3.2 Understanding Impedance Factor. Profiling DWS demands quantitative characteristics, so we need an indicator correlated with the changes in DWS. As discussed in Section 3.3.1, sound pressure and impedance are all affected by DWS variations, but their quantities (and their respective impacts from DWS) are physically very different. We decide to target the impedance because it varies to a much greater extent than the pressure, even in a uniform canal. Because the energy loss of acoustic signals due to propagation in the ear canal is very small, the sound pressures measured at the acoustic source, the eardrum, and the earphone (after reflected back) differ only by a barely detectable phase factor determined by the length of the canal.

Prior research and clinical audiometry [30, 54] have studied the ear canal impedance via many modalities (e.g., dedicated multi-cavity-multi-resistor sensors), yet they are mostly inapplicable due to heavy user involvements or special hardware requirements. Therefore, we develop a simple yet effective solution. Consider a sine sweep (with bandwidth ranging from 0Hz to 10kHz) sent into the ear canal as the excitation signal; the simultaneously recorded reflection signal can be used to figure out the frequency response within a wideband. The sound pressure P_S at the earphone can be obtained by the earphone and accounted by the sum of excitation sound pressure P_S^+ and reflection sound pressure P_S^- ; a complex reflectance coefficient is obtained as $\Gamma = P_S^-/P_S^+$. Assuming the canal area is of uniform area at the measurement point, the impedance Z_{EC} can be approximated by [43, 50]:

$$Z_{EC} = z_0 \frac{1 + \Gamma}{1 - \Gamma}. \quad (3)$$

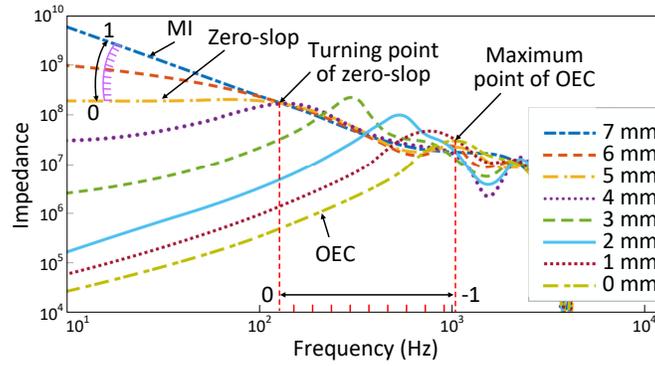
The characteristic acoustic impedance z_0 is defined as $z_0 = c\rho/A$, where ρ is the density of air, c is the speed of sound, and A is the cross-sectional area at the measuring position. Variations due to the use of a slightly deviated area A to define z_0 are shown to have a minor effect on impedance measurements [30, 54], so using the average cross-sectional area A_0 of the adult instead of the true area is adequate for meaningful modeling. We summarize the involved parameters in Table 1.

We measure the Z_{EC} of an adult based on Eqn. (3) with earphones inserted at different depths in the ear canal and illustrate the results in Fig. 7. We observe that, under perfect occlusion (6 and 7 mm insert depth), the impedance is inversely proportional to frequency, since the residual volume of the middle ear dominates the impedance [66]. When insertion depth decreases and the ear canal is only partially occluded by the earphone, a

Table 1. Constants used for calculation.

Name	Constant	Value	Units
Speed of sound	c	33480	cm/s
Density of air	ρ	0.001223	g/cm^3
Average cross-sectional area of adult ear canal [51]	A_0	0.442	cm^2

vent appears and the impedance-frequency relation changes drastically. First, the amplitude drops significantly below around 100Hz. Moreover, below a certain frequency, impedance becomes proportional to frequency, hence leading to a local maximum at higher frequencies, which explains the acoustic enhancement by the occlusion effect on such frequencies [32]. For the open ear canal (0mm insertion), the impedance seems to be proportional to the frequency component of up to around 1kHz. In conclusion, the impedance under different device insertion depths and the degree of the ear canal occlusion clearly exhibit observable distinctions; it is hence proven to be sufficiently indicative of DWS.

Fig. 7. An example of Z_{EC} values under various depths of insertion; axes are scaled for better illustrations.

3.3.3 Profiling DWSs. The acoustic characteristics of an ear canal differ from person to person [25] and also differ for the left and right ears of the same person [13]. Based on the impedance-depth relation, we aim to offer a systematic calibration procedure to capture the DWS for two ears of individual users. Specifically, a new user to EarACE should slowly insert the earphone into his/her ear canal until no further insertion is possible (maximum distance). Meanwhile, EarACE sends the swept signal repeatedly via the earphones and records the reflections. Consequently, multiple impedance-frequency trajectories are obtained, where those i) at the maximum insertion depth, ii) when the device is barely inserted, and iii) with a zero slope between 0 and 100Hz (indicating the sudden turning from perfect to partial occlusion), are treated as three references and denoted by MI, OEC, and Zero-slop, respectively.

When in use for profiling, we first calculate the impedance-frequency trajectory for a given DWS. Then the instantaneous slope of the impedance trace at 100Hz is used to characterize whether the ear canal is completely occluded: a negative slope indicates a perfect occlusion, and a positive slope indicates a partial occlusion. For perfect occlusions, we calculate the ratio of the slope of the current trajectory to the slope of the maximum insertion depth, both in the range 0-100Hz, shown as the purple arc with scales in Fig. 7. A ratio closer to 1 indicates a deeper insertion depth. For partial occlusions, mapping the current insertion distance to a slope ratio may not be sufficiently robust based on our experience. Therefore, we obtain a new indicator η whose scales are

shown between red (vertical) dashed lines in Fig. 7:

$$\eta = (f_c - f_{\text{Zero-slop}}) / (f_{\text{Zero-slop}} - f_{\text{OEC}}), \quad (4)$$

where f_c and f_{OEC} refer to the frequencies that respectively maximize the current impedance trajectory and the OEC trajectory, while $f_{\text{Zero-slop}}$ denotes the turning point of the Zero-slop trajectory. This DWS profiling method is performed in real time and involves less than 3.5s of sweeping sound playing time and data processing time. The resulting DWS profile allows researchers to develop signal process algorithms adaptively and wearers to adjust their DWS in real time to match particular application requirements.

3.4 Eliminating Body Motion Interference

As shown in Section 2.3, the acoustic characteristics exploited by EarACE can be particularly susceptible to interference from body motions. Though removing such interference can be hard for conventional methods, we propose to leverage the different acoustic characteristics between left and right ears, as inspired by the studies in Section 2.4.

3.4.1 Interference Demixing. As the attenuation of sound propagation through bones and tissues is sensitive to propagation path, the body motion (e.g., walking and exercising) interference acts differently on the left and right ears due to body asymmetry. To demonstrate this asymmetry, we give an example of body motion interference on PCG signals collected from the two ears in Fig. 8, where the interference always appears differently in the left and right ears, and the interference intensity, as expected, can be much higher than PCG. We consider the signals collected from each ear as mixtures of *sensing target* and *body motion interference*, and our goal is to extract a clean sensing target out of a noisy audio mixture. Moreover, many applications involve random body motions that are often non-periodic, so periodic or quasi-periodic sensing targets (e.g., physiological signals) can also be distinguished.

Therefore, we propose to exploit this contrast along with two-ear difference (see Section 2.4) to extract the sensing target. Let $y_L(t)$ and $y_R(t)$ represent the in-ear audio signal collected from the left and right ears, which compose a signal vector $y(t) = \{y_L(t), y_R(t)\}$. Assuming a linear propagation channel for the source signal vector $s(t)$ through bones and tissues to an ear canal (to be lifted later), i.e., $y(t) = As(t)$ with A being a mixing matrix. In our case, $s(t) = \{s_T(t), s_I(t)\}$ with $s_T(t)$ and $s_I(t)$ respectively representing the sensing target and body motion interference; we aim to find a demixing matrix W such that $s(t) = W^T y(t)$ separates the sensing target from the interference. Since the sensing target is temporally correlated, but the body motion interference is not, they satisfy the following relations for a nominal period ξ^* :

$$\begin{cases} \mathbb{E}[s_T(t)s_T(t - \xi^*)] > 0 \\ \mathbb{E}[s_T(t)s_I(t - \xi^*)] = \mathbb{E}[s_I(t)s_I(t - \xi^*)] = 0 \end{cases}, \quad (5)$$

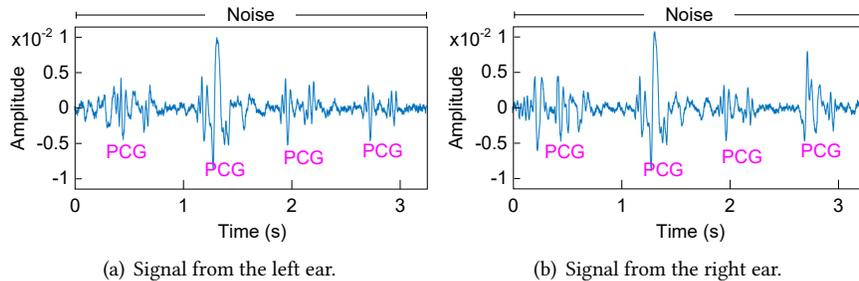


Fig. 8. Examples of body motion interference contaminating PCG signals collected from both ears. The two signals exhibit evident differences.

where \mathbb{E} is the expectation operator. Let us define $C(W) = \mathbb{E}[s(t)s(t - \xi^*)] = W^\top \mathbb{E}[y(t)y(t - \xi^*)]W$, where $^\top$ refers to matrix transpose. Thanks to the symmetry of $C(W)$, it can be further decomposed as:

$$C(W) = \frac{1}{2}C(W) + \frac{1}{2}C(W)^\top = \frac{1}{2}W^\top (H_y(\xi^*) + H_y(\xi^*)^\top)W, \quad (6)$$

where $H_y(\xi^*) = \mathbb{E}[y(t)y(t - \xi^*)^\top]$. Apparently, the closer W projects $y(t)$ on $s(t)$, the larger the resulted value is, as projecting on $s_1(t)$ leads to zero. Therefore, extracting $s_T(t)$ implies maximizing $C(W)$. Under the constraint that $\|W\|_2 = 1$, the maximizing problem can be solved by finding the eigenvectors (denoted by $\mathbb{E}\mathbb{V}$ operator) of $H_y(\xi^*) + H_y(\xi^*)^\top$:

$$W = \mathbb{E}\mathbb{V} (H_y(\xi^*) + H_y(\xi^*)^\top). \quad (7)$$

Then $s_T(t) = w_1^\top y(t)$ where w_1 is the eigenvector corresponding to the largest eigenvalue. This perfectly recovers $s_T(t)$ if the nominal period is exact.

3.4.2 Clean Desired Signals Extraction. In reality, three practical issues should be considered. First, the nominal period ξ^* generally does not exist. For example, the interbeat interval of heartbeats varies over time. To approximate ξ^* , we use the autocorrelation method [5] to obtain a mean period within a specified interval of interest. Second, the expectation of finite samples becomes a numerical average, leading to non-zero values for the uncorrelated signals in Eqn. (5). To avoid “leaking” $s_T(t)$ onto other eigenvalues under the above imperfect conditions, we turn to consider the eigenvectors corresponding to the top- K eigenvalues in W . We set $K = 3$ as the default value for our experiments, but one can modify it according to specific applications.

The last issue is that the above interference demixing solution is based on the unrealistic assumption of a linear signal mixture model, which may not enable accurate extraction of desired clean signals. Therefore, we first perform interference demixing to separate the interference signal, then we use it as the reference signal to input into an adaptive filter to finally obtain the clean sensing target. Fig. 9(a) shows an example of the demixed interference signal, referred to as the *interference reference*. Specifically, we apply ASLMS (adaptive step-size least mean squares) [42] to eliminate the interference of body motions. ASLMS can effectively handle the dynamic nature of most biological signals (i.e., to be nonstationary and change substantially in their properties over time). Moreover, by updating the step size, it can work well at various interference levels. Consequently, by combing the interference reference and the mixed signal at either ear, we can extract a clean sensing target. Fig. 9(b) illustrates the extracted PCGs from both ears, clearly demonstrating the effectiveness of our overall method.

4 DEMONSTRATIVE APPLICATIONS

Due to its flexibility and robustness, EarACE supports a wide range of potential applications covering both passive and active sensing. To demonstrate its efficacy and adaptability, we conduct full scale experiments on the three representative applications discussed in Section 2. These applications do not necessarily impair the

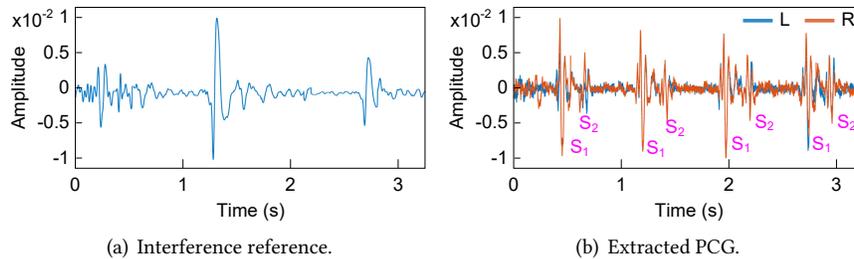


Fig. 9. An example of body motion interference elimination; note the scale difference between (a) and (b).

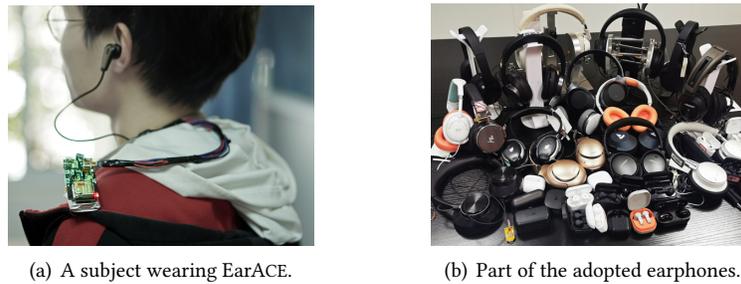


Fig. 10. EarACE (with ANC earphones) in action.

common usage of earphones and we demonstrate the success of playing music while authenticating a valid user in Section 4.3. In particular, we evaluate their performance in realistic settings, where 35 subjects (20 male and 15 female, aging between 20 to 50) are involved. To ensure device diversity, we have tried 50 pairs of ANC earphones (including in-ear, on-ear, and over-ear) with varying prices of around \$100, but we choose only 18 pairs of in-ear ANC earphones for the following experiments, since they outperform other types according to our empirical studies. Fig. 10(a) illustrates a subject wearing our EarACE prototype, and Fig. 10(b) shows the part of the involved earphones (those not under NDA). Note that the placement of EarACE shown in Fig. 10(a) is only meant to clearly showcase the hardware; the actual placements during usage would certainly be at more stable positions. Also, we want to stress again that the eventual target of EarACE is to be integrated into ANC earphones. The captured data are transmitted via Wi-Fi to a PC and processed in real-time there.

4.1 PCG Monitoring

We first demonstrate that EarACE can effectively collect PCG in the face of common body motions. Valvular and heart-related diseases have been a significant public health challenge [45, 56]. Efficient disease evaluation and prevention call for daily PCG monitoring: an effective way of capturing clinically valued heart rate variability (HRV) features such as systolic period, diastolic period, and the standard deviation of interbeat intervals (SDNN) [44]. Therefore, we particularly showcase the extraction of these three parameters never being achieved by existing earable sensing solutions. In addition, the pilot study presented in Section 2 suggests that the performance of PCG monitoring is highly affected by changes in DWS and body motion interference. We specifically conduct experiments under different DWSs and interference conditions, aiming to further emphasize EarACE’s novelties and contributions in improving in-ear acoustic sensing.

4.1.1 Experiment Setup and Signal Processing. We let each subject wear EarACE prototype with different earphones; each PCG signal recording lasts for 5 minutes. To study the system performance in practical scenarios, we ask the subjects to rewear their earphones for each recording, while performing daily activities (e.g., walking, making the bed, and cooking) during recording; this allows the collected audio data to include both DWS variations and body motion interference. Meanwhile, the ground truth signals are obtained by an ECG monitor [47]. In practical operations, EarACE first profiles the DWS to determine whether the ear canal is occluded to block external environment noise (the major noise identified in Section 2), before applying the demixing algorithm in Section 3.4 to retrieve PCG. In the experiment, we clearly mark the time period with body motions in order to conduct comparative studies on the effectiveness of EarACE. Moreover, to demonstrate EarACE’s advantages over existing earphone-based platforms, namely HeadFi [22] and OESense [32], we apply their principles to EarACE by respectively extracting the differential signal from the two ears and filtering out components above 50Hz, so as to set up two comparative baselines. Doing these actually favor these two platforms, as they originally rely on

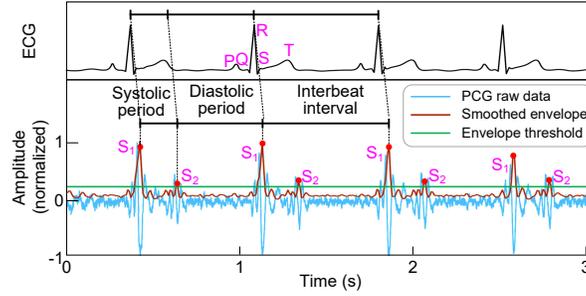


Fig. 11. Detecting S_1 and S_2 using threshold and envelope, and aligning PCG with the ground truth ECG.

self-crafted prototypes for sensing, whereas these baselines are implemented upon EarACE’s well-engineered ANC native acoustics.

Prior to HRV features extraction, EarACE identifies the major heart sounds S_1 and S_2 (see Section 2.1). Since the periods of S_1 and S_2 exhibit significant fluctuation in the demixed signal while non-heart-sound periods are more stationary and lower in amplitude, EarACE detects them based on whether the demixed signal exceeds a certain threshold. Specifically, an average Shannon Energy envelope is computed as follows:

$$\mathcal{E}_x = \frac{1}{N} \sum_{t=1}^N x(t)^2 \cdot \log x(t)^2 \quad (8)$$

where $x(t)$ is the normalized signal in the range of $[-1,1]$, and N is a window size usually set to $0.02s$ with $0.01s$ overlapping [31]. Then a smoothed envelope is obtained by applying zero-phase filtering on the envelope, and the peaks of S_1 and S_2 are identified as \mathcal{E}_x exceeding an empirically set threshold value of 0.2 (we leave an adaptive threshold for future explorations). The systolic period, diastolic period, and interbeat intervals are extracted based on these peaks as shown in Fig. 11; it also illustrates the alignment between PCG and its corresponding ECG signal. It can be observed that the QRS complex and T wave of ECG are temporally aligned with features of PCG. Therefore, we use the ECG to derive the ground truth for the HRV features as suggested by [18], so as to validate the quality of the PCG retrieved by EarACE.

4.1.2 Results. As discussed earlier, PCGs can be affected by DWS and body motion interference. To demonstrate the significant advantage of the proposed DWS profiling method and the body motion interference elimination method, we group the data according to whether the earphones completely occlude (i.e., $CO \in \{0, 1\}$) the ear canal (indicated by DWS profile) and whether our body motion interference elimination method is deployed

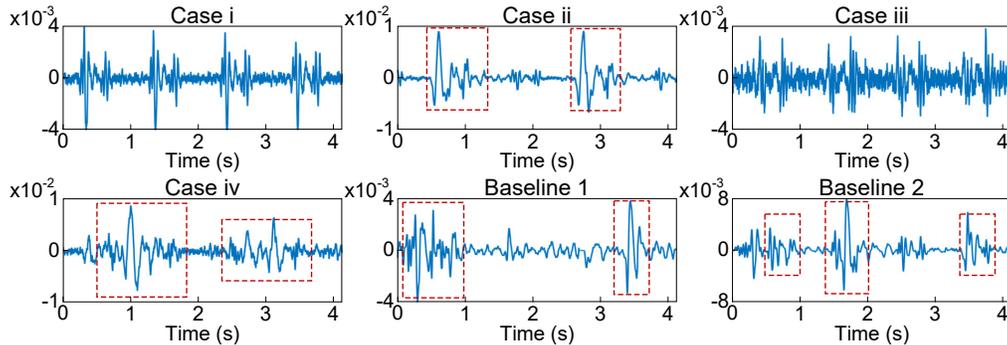


Fig. 12. Snapshots of PCG under the four cases (left ear) and two baselines, with interferences of walking highlighted with red dotted lines.

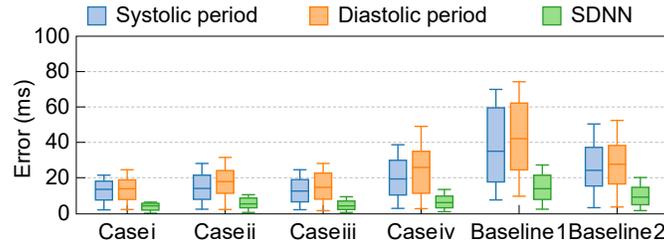


Fig. 13. Performance comparisons with respect to PCG monitoring and HRV feature estimation.

(i.e., $IE \in \{0, 1\}$). Specifically, we study four cases: Case i) $CO = 1, IE = 1$, Case ii) $CO = 1, IE = 0$, Case iii) $CO = 0, IE = 1$, and Case iv) $CO = 0, IE = 0$. Besides, we denote the upgraded implementations of HeadFi [22] and OESense [32] on EarACE by Baseline 1 and Baseline 2, respectively. Snapshots of PCG under the four cases and two baselines are demonstrated in Fig. 12.

The experiment results, as shown in Fig. 13, are depicted by box plot of the systolic period, diastolic period, and SDNN in the four cases and the baselines. One can readily observe that all four cases surpass the two baselines. This indeed conforms to the observation made in Section 2 that EarACE is beneficial for sensing applications by retaining original information, rather than deducing differential signals from two ears or being confined to a limited frequency range. Moreover, Case i achieves the lowest error for all HRV features as expected. Specifically, the three features receive respective median errors of 14.31 ms (4.77% of the 300ms nominal systolic period), 14.74ms (2.95% of the 500ms nominal diastolic period), and 5.35ms. The decent performance strongly confirms the efficacy of EarACE in obtaining accurate HRV estimations. One can also observe that cases with proper DWS ($CO=1$) yield lower error, further confirming that DWS profiling method advances HRV feature extraction. Moreover, Case ii exhibits higher error compared with Case i, confirming that our interference elimination is necessary and effective. Finally, the error of Case iii being a bit smaller than that of Case iv suggests that our proposed body motion interference elimination algorithm may handle a broader range of interference (e.g., from the external environment) beyond body motions. Overall, we can confidently conclude that EarACE gains enhanced in-ear acoustic sensing due to its effective body motion interference elimination and DWS profiling.

4.2 Hand-Face Interaction Analysis

We showcase the usefulness of EarACE in this section, by considering an application for detecting multiple hand-face interactive gestures. Note that EarACE is readily applicable to other behavior analysis applications such as eating habit analysis [7, 55] and facial expression monitoring [3]; we consider this one because it appears to be one of the most popular earable sensing tasks in recent publications [32, 60].

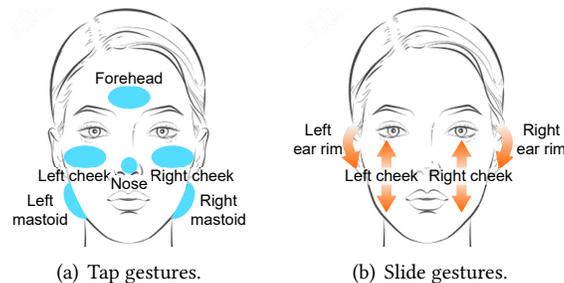


Fig. 14. Hand-face interactive gesture design.

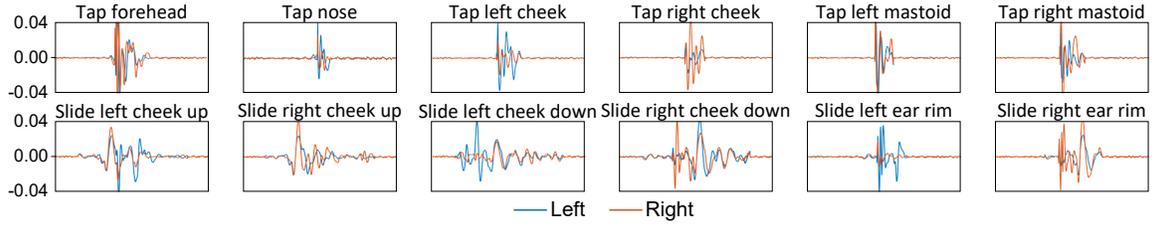


Fig. 15. Snapshots of gesture-induced waveform.

4.2.1 Experiment Setup and Signal Processing. We consider six tapping gestures and six sliding gestures associating different face positions and actions, as shown in Fig. 14. The gesture set studied by EarACE, matching that considered in [32], is one of the largest of its kind considered by earphone-based platforms; we use Fig. 15 to showcase examples signals for all gestures. We ask the subjects to respectively perform these gestures while wearing the EarACE prototype. Each subject repeats every gesture multiple times using a different pair of earphones each time to include small variations in devices. In total, we collect 21,000 gestures for evaluation. Observations from Section 2.2 suggest that changes in DWS barely affect the features of the acquired signals, and the body motion interference should usually be avoided for applications using hand-face interaction as a control interface. Therefore, we focus on evaluating the discriminability of the gesture-induced signals collected by EarACE, under a normal DWS and absent of other unnecessary body motions (e.g., walking or exercising).

Interaction gesture analysis involves detection and classification. EarACE first performs gesture detection to check the presence of a gesture, based on the observation that the *short-term energy* (STE) introduced by a hand-face gesture is significantly higher than that of system noise. Specifically, the STE of sensed signal $x(m)$ is retrieved as follows:

$$\text{STE}_n = \sum_{m=n-N+1}^n [x(m)\omega(n-m)]^2, \quad n - N + 1 \leq m \leq n, \quad (9)$$

where $\omega(n-m)$ is a window, N is the window length, and n represents the last sample that the window targets. EarACE adopts a threshold to judge the presence of a gesture by whether STE_n goes beyond the threshold. We empirically set the window to be rectangular with a 30ms length and the threshold to be twice the average STE_n computed adaptively from the past 100ms.

Upon detecting a gesture, features from both time and frequency domains, including gesture duration, the main frequency of the signal envelope, the root-mean-square energy, spectral centroid, zero-crossing rate, the first 13 components of MFCC (Mel-frequency cepstral coefficients), and the temporal differential of the MFCC, are extracted for further classification. The features from both left and right channels are combined into a feature vector and input into a classifier. In addition to the twelve gesture classes, there is a NULL class to avoid classifying interference, such as unexpected body motions, as gestures. In experiments, we split the collected data into 80% training data and 20% testing data. Given the diverse methods available for classification, we choose to evaluate the performance of recognizing hand-face interaction gestures with four commonly used classifiers: random forest (RF), decision tree (DT), k-nearest neighbors (kNN), and support vector machine (SVM) with RBF kernel. All classifiers are implemented under default settings.

4.2.2 Results. We conduct five-fold cross-validation on the collected data. Fig. 16 compares the performance of different classifiers in terms of precision and recall. We can observe that all classifiers achieve precision and recall over 80% with our handpicked features. In particular, the RF classifier achieves 96.39% precision and 94.69% recall, rendering a trial with popular deep neural networks redundant. Overall, the results evidently confirm the effectiveness of EarACE in collecting fine-grained and hence highly discriminable gesture-induced signals suitable for virtually any classifiers, mostly credited to the use of ANC native acoustics.

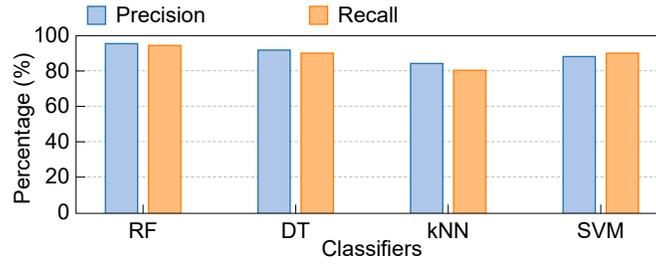


Fig. 16. Performance of hand-face gesture recognition with commonly used classifiers.

4.3 User Authentication

EarACE is not only capable of performing passive sensing to acquire physiological signals and body activities, but can also actively probe the ear canal structure. In this section, we study the performance of EarACE in continuous authentication, leveraging a piece of music reflected by the ear canal structure unique to an individual.

4.3.1 Experiment Setup and Signal Processing. We let each subject wear EarACE while listening to a clip of *Canon* [40] for 5 sessions, each lasting for thirty seconds and divided into fifteen non-overlapping recordings each of two seconds long. This excitation signal is arbitrarily chosen as any clips without silent intervals can serve as candidates. Based on our observation in Section 2.2 that the ear canal biometrics are highly sensitive to the DWS, we in particular ask the subjects to rewear the earphones with different DWS to further study EarACE’s performance in authenticating users. Similar to the previous application, unnecessary body motions are minimized to emulate a realistic application scenario. In total, we obtain 2,625 two-second recording samples for evaluation. Our experiments aim to demonstrate that user authentication can be achieved by the transfer function (see Section 2.1) that uniquely characterizes the ear canal structure of a certain user.

Basically, the earphone, ear canal, and eardrum together are considered as a black box system, and the transfer function characterizing it may not be sufficiently unique if obtained assuming linearity and time-invariance. Therefore, we consider a more complex yet practical system model involving nonlinearity and memorylessness, as illustrated in Fig. 17. Inside the system, the excitation signal $s_e(t)$ in Fig. 17(a) first passes through a memoryless nonlinear distorter characterized by a n -th order Volterra kernel $k_n(t)$ [23], and then reverberated through a linear filter $f(t)$. In the meantime, certain noise $n(t)$ can be involved and added to the measured signal $s_m(t)$ in Fig. 17(a):

$$\begin{aligned} s_m(t) &= n(t) + s_e(t) \otimes k_1(t) \otimes l(t) + \dots + s_e^n(t) \otimes k_n(t) \otimes l(t), \\ &= n(t) + s_e(t) \otimes \ell(t) + \dots + s_e^n(t) \otimes \ell_n(t), \end{aligned} \quad (10)$$

where the convolution \otimes between nonlinear distortion $k_n(t)$ and linear reverberation $l(t)$ is replaced by $\ell_n(t)$. This essentially describes the transfer process by a set of impulse responses, each of them being convolved with a distinct power of the excitation signal [23].

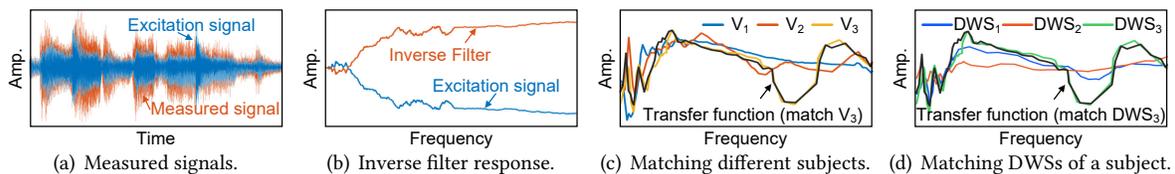


Fig. 17. The procedure of obtaining transfer function and subsequent authentication.

We first resolve the noise $n(t)$ by taking a number of synchronous averages to $s_m(t)$. Since deconvolving using Fast Fourier Transformation usually leads to time aliasing artifacts, we generate an inverse filtered signal $g(t)$ [24], which inverts the response of the excitation signal in Fig. 17(b). This process allows us to derive a transfer function by integrating the averaged output $\hat{s}_m(t)$ and inverse filtered signals $g(t)$ in the frequency domain. As discussed in Section 2, the transfer function trajectories of different subjects have unique and stable variation patterns and exhibit discernible differences mostly over the range of [0, 50]Hz and [300, 500]Hz. Therefore, we employ a DTW (dynamic time warping) [8, 63] based similarity matching scheme to classify a given user based on the transfer function in the range of [0-500]Hz. Specifically, in the registration phase, we record a user's transfer function as his/her template. Upon authenticating a user, we look for the best similarity matching between the current user's transfer function and the stored templates: the user is authenticated if the matched template corresponds to whom he/she claims to be, as shown by Fig. 17(c); otherwise the authentication fails. Furthermore, the transfer function varies with DWS even for the same user as claimed earlier, so EarACE profiles DWS with indicator η (see Section 3.3.3) to accurately identify a user, as shown in Fig. 17(d).

4.3.2 Results. We evaluate EarACE's performance in authentication, and in particular, we look into the influence of the DWS. We group the measured signals into four datasets according to the DWS indicator η , namely [-1,-0.5), [-0.5,0), [0,0.5), and [0.5,1]. The template set and test set are disjointly selected from our dataset with 2,625 samples, with the former accounting for 60% and the latter 40%. Although prior studies have explored acoustic sensing to identify users based on the ear canal structure [27, 59], they failed to handle changes in DWS. To further study how profiling of DWS advances earlier proposals, we consider the case where all collected data are blended without DWS differentiation as the baseline. Fig. 18 compares the *false accept rate* (the probability that a different user is wrongly authenticated as a given user) and the *false reject rate* (the probability that a legitimate user is not authenticated) of the four groups and baseline. We can observe that all four groups significantly outperform the baseline, firmly demonstrating that our DWS profiling method is very critical to ensure a successful user authentication with reasonable performance. Moreover, we note that all four groups achieve reasonable results, where the false accept rates and false reject rates of [0,0.5) and [0.5,1] groups are all below 10%. The results not only evidently confirm the profile of DWS is indeed accurate so that the authentication template and tested data are matched, but also validate the adequate capability of EarACE in authenticating users under various DWS.

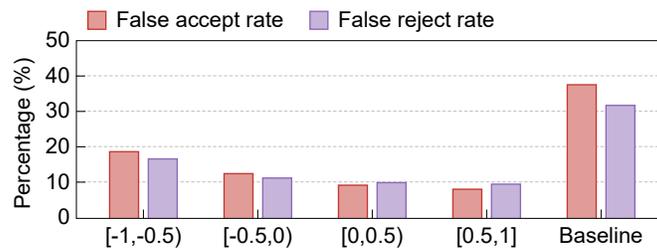


Fig. 18. Performance of user authentication.

5 DISCUSSIONS AND FUTURE WORK

We discuss a few open issues on EarACE that can inspire future studies.

- We propose a fast and effective method to eliminate non-periodic motion interference for a periodic sensing target. If the periodicity of the target is inconspicuous, motion interference elimination may be achieved using the novel blind source separation method [16, 64] leveraging signals collected from both ears.

- We have not thoroughly evaluated EarACE under different levels of ambient noise. Fortunately, ANC earphones integrate inverse filters to perform noise cancellation, which should greatly reduce environment noise by adding to the acquired signals. However, fully integrating the sensing ability of EarACE with noise cancellation may still need a careful engineering process.
- Profiling the DWS can help developers to build more practical applications and it is crucial for ensuring reliable measurements, so it is worth investigating how multi-dimensional DWS indicators can be adapted to heterogeneous ear structures. In fact, we plan to elaborate on a study of device heterogeneity in an extended report.
- It would be interesting to explore how EarACE can support multiple applications concurrently, potentially with specifically designed signal waveforms. Moreover, combining earable sensing with contact-free sensing techniques with much longer sensing ranges (e.g., those based on either radio-frequency [16, 19, 61, 65] or acoustics [10–12]) may help extend to a ubiquitous sensing environment.
- One may be concerned about whether the audio quality of music playback is affected. Since EarACE’s plug-in design does not affect the internal structure and circuits of the earphone, it should not affect the music quality, and all our 35 test subjects agree that no difference from normal earphones is perceived when attaching EarACE.
- EarACE is so far tested in rather arbitrary attaching manners, because the eventual goal of EarACE is to be integrated into ANC earphones to deliver a better user experience; we are discussing with local industries to accelerate the process.
- EarACE plays a sweep signal ranging from 0-10kHz as the excitation signal to profile DWS, which potentially interrupts the played music. Therefore, our follow-up study would investigate the possibility of using higher frequencies (i.e., above 16kHz that are hardly audible to most adults) for this purpose.
- The prototype has two levels of system voltage, including 3.3V and 1.8V. The power composition of EarACE depends on many aspects, such as sampling rate, data transmission mode, and whether ANC algorithm is on or not. During our experiments, EarACE can support up to 10 hours of continuous operation with a 450mAh battery and 3.3V voltage. We expect a fully integrated earable sensing product to substantially cut down power consumption.

6 CONCLUSION

In this paper, we have proposed EarACE, the first acoustic sensing platform leveraging commercial ANC earphones for enabling versatile sensing applications. We systematically study the characteristics and affecting factors of the acoustic signals across three major human sensing categories, in order to motivate our platform design. Consequently, EarACE provides flexible yet powerful hardware control to the native acoustics offered by ANC earphones and two critical functions to ensure reliable measurements, which has maximized the sensing capability for diversified applications and lowered the development barrier for novice researchers. We have tested three representative applications on the EarACE and evaluated their performance with 35 subjects and 50 pairs of ANC earphones. Our results have evidently demonstrated the promising capability of EarACE in sensing abundant information across human physiological signals, behaviors, and biometrics.

REFERENCES

- [1] Abbas K Abbas and Rasha Bassam. 2009. Phonocardiography Signal Processing. *Synthesis Lectures on Biomedical Engineering* 4, 1 (2009), 1–194.
- [2] John Allen. 2007. Photoplethysmography and Its Application in Clinical Physiological Measurement. *Physiological Measurement* 28, 3 (2007), R1.
- [3] Takashi Amesaka, Hiroki Watanabe, and Masanori Sugimoto. 2019. Facial Expression Recognition Using Ear Canal Transfer Function. In *Proc. of ACM ISWC 2019*. 1–9.

- [4] Toshiyuki Ando, Yuki Kubo, Buntarou Shizuki, and Shin Takahashi. 2017. CanalSense: Face-Related Movement Recognition System Based on Sensing Air Pressure in Ear Canals. In *Proc. of the 30th ACM UIST*. 679–689.
- [5] Allan Kardec Barros and Andrzej Cichocki. 2001. Extraction of Specific Signals With Temporal Structure. *Neural computation* 13, 9 (2001), 1995–2003.
- [6] 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 *Proc. of ACM UbiComp/ISWC Adjunct*. 1333–1338.
- [7] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments. *Proc. of ACM IMWUT 2017* 1, 3 (2017), 1–20.
- [8] Donald J Berndt and James Clifford. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. In *Proc. of the 3rd ACM KDD-AAAIWS*. Seattle, WA, USA:, 359–370.
- [9] Nam Bui, Nhat Pham, Jessica Jacqueline Barnitz, Zhanan Zou, Phuc Nguyen, Hoang Truong, Taeho Kim, Nicholas Farrow, Anh Nguyen, Jianliang Xiao, Robin Deterding, Thang Dinh, and Tam Vu. 2019. eBP: A Wearable System For Frequent and Comfortable Blood Pressure Monitoring From User’s Ear. In *Proc. of the 25th ACM MobiCom*. 1–17.
- [10] Chao Cai, Zhe Chen, Henglin Pu, Liyuan Ye, Menglan Hu, and Jun Luo. 2020. AcuTe: Acoustic Thermometer Empowered by a Single Smartphone. In *Proc. of the 18th ACM SenSys*. 28–41.
- [11] Chao Cai, Henglin Pu, Peng Wang, Zhe Chen, and Jun Luo. 2021. We Hear Your PACE: Passive Acoustic Localization of Multiple Walking Persons. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 55 (2021), 24 pages.
- [12] Chao Cai, Rong Zheng, and Jun Luo. 2022. Ubiquitous Acoustic Sensing on Commodity IoT Devices: A Survey. *IEEE Communications Surveys & Tutorials* 24, 1 (2022), 432–454.
- [13] Yetong Cao, Chao Cai, Fan Li, Zhe Chen, and Luo Jun. 2023. HeartPrint: Passive Heart Sounds Authentication Exploiting In-Ear Microphones. In *Proc. of IEEE INFOCOM 2023*. 1–10.
- [14] Yetong Cao, Huijie Chen, Fan Li, and Yu Wang. 2021. CanalScan: Tongue-Jaw Movement Recognition via Ear Canal Deformation Sensing. In *Proc. of IEEE INFOCOM 2021*. 1–10.
- [15] Kévin Carillo, Olivier Doutres, and Franck Sgard. 2021. On the Removal of the Open Earcanal High-Pass Filter Effect Due to Its Occlusion: A Bone-Conduction Occlusion Effect Theory. *Acta Acustica* 5 (2021), 36.
- [16] Zhe Chen, Tianyue Zheng, Chao Cai, and Jun Luo. 2021. MoVi-Fi: Motion-robust Vital Signs Waveform Recovery via Deep Interpreted RF Sensing. In *Proc. of the 27th ACM MobiCom*. 392–405.
- [17] Romit Roy Choudhury. 2021. Eearable Computing: A New Area to Think About. In *Proc. of the 22nd ACM HotMobile*. 147–153.
- [18] Gari D Clifford, Francisco Azuaje, Patrick McSharry, et al. 2006. *Advanced Methods and Tools for Ecg Data Analysis*. Vol. 10. Artech House Boston.
- [19] Shuya Ding, Zhe Chen, Tianyue Zheng, and Jun Luo. 2020. RF-Net: A Unified Meta-Learning Framework for RF-enabled One-Shot Human Activity Recognition. In *Proc. of the 18th ACM SenSys*. 517–530.
- [20] Espressif. 2022. ESP32 Technical Reference Manual. https://www.espressif.com/sites/default/files/documentation/esp32_technical_reference_manual_en.pdf
- [21] Everest. 2022. ES8388: Low Power Stereo Audio CODEC With Headphone Amplifier. <http://www.everest-semi.com/pdf/ES8388%20DS.pdf>
- [22] Xiaoran Fan, Longfei Shangguan, Siddharth Rupavatharam, Yanyong Zhang, Jie Xiong, Yunfei Ma, and Richard Howard. 2021. HeadFi: Bringing Intelligence to All Headphones. In *Proc. of the 27th ACM MobiCom*. 147–159.
- [23] Angelo Farina. 2000. Simultaneous Measurement of Impulse Response and Distortion With a Swept-Sine Technique. *Journal of the Audio Engineering Society* (february 2000), 1–24.
- [24] Angelo Farina and Fabrizio Righini. 1997. Software implementation of an MLS analyzer with tools for convolution, auralization and inverse filtering. In *Journal of the Audio Engineering Society*.
- [25] Andrea Ferlini, Dong Ma, Robert Harle, and Cecilia Mascolo. 2021. EarGate: Gait-Based User Identification With In-Ear Microphones. In *Proc. of the 27th ACM MobiCom*. 337–349.
- [26] 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. *Proc. of ACM IMUWT 2021* 5, 1 (2021), 1–25.
- [27] Yang Gao, Wei Wang, Vir V. Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using Ear Canal Echo for Wearable Authentication. *Proc. of ACM IMWUT 2019* 3, 3 (2019), 1–24.
- [28] Valentin Goverdovsky, David Looney, Preben Kidmose, and Danilo P. Mandic. 2016. In-Ear EEG From Viscoelastic Generic Earpieces: Robust and Unobtrusive 24/7 Monitoring. *IEEE Sensors Journal* 16, 1 (2016), 271–277.
- [29] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Allesandro Montanari. 2018. Earables for Personal-Scale Behavior Analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89.
- [30] Douglas H Keefe, Robert Ling, and Jay C Bulen. 1992. Method to Measure Acoustic Impedance and Reflection Coefficient. *The Journal of the Acoustical Society of America* 91, 1 (1992), 470–485.

- [31] H Liang, Sakari Lukkarinen, and Iiro Hartimo. 1997. Heart Sound Segmentation Algorithm Based on Heart Sound Envelopogram. In *Computers in Cardiology 1997*. 105–108.
- [32] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. 2021. OESense: Employing Occlusion Effect for in-Ear Human Sensing. In *Proc. of the 19th ACM MobiSys*. 175–187.
- [33] Steve Mann. 1997. Wearable Computing: A First Step Toward Personal Imaging. *Computer* 30, 2 (1997), 25–32.
- [34] Alexis Martin and Jérémie Voix. 2017. In-Ear Audio Wearable: Measurement of Heart and Breathing Rates for Health and Safety Monitoring. *IEEE Transactions on Biomedical Engineering* 65, 6 (2017), 1256–1263.
- [35] Denys JC Matthies, Bernhard A Strecker, and Bodo Urban. 2017. Earfieldsensing: A Novel In-Ear Electric Field Sensing to Enrich Wearable Gesture Input Through Facial Expressions. In *Proc. of the 35th ACM CHI*. 1911–1922.
- [36] Alessandro Montanari, Chulhong Min, Akhil Mathur, Utku Acer, Marc Van den Broeck, and Fahim Kawsar. 2022. ESense. <https://www.esense.io/>
- [37] 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.
- [38] Yunyoung Nam, Bersain A Reyes, and Ki H Chon. 2015. Estimation of Respiratory Rates Using the Built-in Microphone of a Smartphone or Headset. *IEEE Journal of Biomedical and Health Informatics* 20, 6 (2015), 1493–1501.
- [39] Anh Nguyen, Raghda Alqurashi, Zohreh Raghebi, Farnoush Banaei-Kashani, Ann C Halbower, Thang Dinh, and Tam Vu. 2016. In-Ear Biosignal Recording System: A Wearable for Automatic Whole-Night Sleep Staging. In *Proc. of the 14th ACM WearSys*. 19–24.
- [40] Johann Pachelbel. 1680. Cannon in D Major.
- [41] Jay Prakash, Zhijian Yang, Yu-Lin Wei, and Romit Roy Choudhury. 2019. STEAR: Robust Step Counting from Earables. In *Proc. of the 1st ACM EarComp*. 36–41.
- [42] M Raghu Ram, K Venu Madhav, E Hari Krishna, Nagarjuna Reddy Komalla, and K Ashoka Reddy. 2011. A Novel Approach for Motion Artifact Reduction in Ppg Signals Based on AS-LMS Adaptive Filter. *IEEE Transactions on Instrumentation and Measurement* 61, 5 (2011), 1445–1457.
- [43] Sarah R Robinson. 2017. *Effects of Ear-Canal Geometry and Middle-Ear Pressure on Wideband Acoustic Reflectance*. Ph.D. Dissertation. University of Illinois at Urbana-Champaign.
- [44] Fred Shaffer and J.P. Ginsberg. 2017. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health* 5 (2017), 258.
- [45] Patricia A Sharpe, Noreen M Clark, and Nancy K Janz. 1991. Differences in the Impact and Management of Heart Disease Between Older Women and Men. *Women & Health* 17, 2 (1991), 25–43.
- [46] Ying Song, Yu Gong, and S.M. Kuo. 2005. A Robust Hybrid Feedback Active Noise Cancellation Headset. *IEEE Transactions on Speech and Audio Processing* 13, 4 (2005), 607–617.
- [47] SparkFun. 2014. SparkFun Single Lead Heart Rate Monitor - AD8232. <https://www.sparkfun.com/products/12650>. Accessed: 2022-02-28.
- [48] Wayne Staab. 2022. Human Ear Canal. <https://hearinghealthmatters.org/waynesworld/2014/human-ear-canal-viii/>
- [49] Thad Starner. 1996. Human-Powered Wearable Computing. *IBM systems Journal* 35, 3.4 (1996), 618–629.
- [50] Stefan Stenfelt and Sabine Reinfeldt. 2007. A Model of the Occlusion Effect With Bone-Conducted Stimulation. *International Journal of Audiology* 46, 10 (2007), 595–608.
- [51] Michael R Stinson and Gilles A Daigle. 2005. Comparison of an Analytic Horn Equation Approach and a Boundary Element Method for the Calculation of Sound Fields in the Human Ear Canal. *The Journal of the Acoustical Society of America* 118, 4 (2005), 2405–2411.
- [52] Michael R Stinson and BW Lawton. 1989. Specification of the Geometry of the Human Ear Canal for the Prediction of Sound-Pressure Level Distribution. *The Journal of the Acoustical Society of America* 85, 6 (1989), 2492–2503.
- [53] Yasutaka Umayahara, Zu Soh, Kiyokazu Sekikawa, Toshihiro Kawae, Akira Otsuka, and Toshio Tsuji. 2018. Estimation of Cough Peak Flow Using Cough Sounds. *Sensors* 18, 7 (2018), 2381.
- [54] Susan E Voss and Jont B Allen. 1994. Measurement of Acoustic Impedance and Reflectance in the Human Ear Canal. *The Journal of the Acoustical Society of America* 95, 1 (1994), 372–384.
- [55] Tri Vu, Feng Lin, Nabil Alshurafa, and Wenyao Xu. 2017. Wearable Food Intake Monitoring Technologies: A Comprehensive Review. *Computers* 6, 1 (2017), 4.
- [56] Chuyu Wang, Lei Xie, Wei Wang, Yingying Chen, Yanling Bu, and Sanglu Lu. 2018. RF-ECG: Heart Rate Variability Assessment Based on Cots RFID Tag Array. *Proc. of ACM IMWUT 2018 2*, 2 (2018), 1–26.
- [57] Tianben Wang, Daqing Zhang, Leye Wang, Yuanqing Zheng, Tao Gu, Bernadette Dorizzi, and Xingshe Zhou. 2018. Contactless Respiration Monitoring Using Ultrasound Signal With Off-The-Shelf Audio Devices. *IEEE Internet of Things Journal* 6, 2 (2018), 2959–2973.
- [58] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW Based Contactless Respiration Detection Using Acoustic Signal. *Proc. of ACM IMWUT 2018 1*, 4 (2018), 1–20.
- [59] 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. *Proc. of ACM IMWUT 2020 5*, 1 (2021), 1–27.
- [60] Xuhai Xu, Haitian Shi, Xin Yi, WenJia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K Dey. 2020. Earbuddy: Enabling On-Face Interaction via Wireless Earbuds. In *Proc. of CHI 2020*. 1–14.

- [61] Shujie Zhang, Tianyue Zheng, Hongbo Wang, Zhe Chen, and Jun Luo. 2023. Quantifying the Physical Separability of RF-Based Multi-Person Respiration Monitoring via SINR. In *Proc. of the 20th ACM SenSys*. 47–60.
- [62] Jihui Zhangg, Thushara D. Abhayapala, Prasanga N. Samarasinghe, Wen Zhang, and Shouda Jiang. 2016. Sparse Complex FxLMS for Active Noise Cancellation over Spatial Regions. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 524–528.
- [63] Tianming Zhao, Jian Liu, Yan Wang, Hongbo Liu, and Yingying Chen. 2018. PPG-Based Finger-Level Gesture Recognition Leveraging Wearables. In *Proc. of IEEE INFOCOM 2018*. IEEE, 1457–1465.
- [64] Tianyue Zheng, Zhe Chen, Shujie Zhang, Chao Cai, and Jun Luo. 2021. MoRe-Fi: Motion-Robust and Fine-Grained Respiration Monitoring via Deep-Learning UWB Radar. In *Proc. of the 19th ACM SenSys*. 111–124.
- [65] Tianyue Zheng, Zhe Chen, Shujie Zhang, and Jun Luo. 2022. Catch Your Breath: Simultaneous RF Tracking and Respiration Monitoring with Radar Pairs. *IEEE Trans. on Mobile Computing* (2022), 1–14.
- [66] T Zurbrügg, A Stirnemann, Martin Kuster, and Hervé Lissek. 2014. Investigations on the Physical Factors Influencing the Ear Canal Occlusion Effect Caused by Hearing Aids. *Acta Acustica United with Acustica* 100, 3 (2014), 527–536.

APPENDIX

A typical in-ear ANC earphone has two microphones and a speaker inside each earbud.² As shown in Fig. 19, the Feed Back (FB) microphone (a.k.a. in-ear microphone) is placed inside the ear cup to monitor in-ear sounds [46, 62], while the Feed Forward (FF) microphone is placed in the outer space of the earbud to collect outside noise. Whereas the data delivered by these acoustic front-ends are crucial for versatile sensing applications, currently no earphone vendors offer corresponding SDKs for accessing them. Consequently, we have to directly tap into raw analog signals of microphones and speakers, and apply our own codec to digitize or generate these signals, in order to exploit these well-engineered ANC acoustic front-ends. To this end, we rewire the earphone to bypass any circuits preventing our codec from directly interfacing with these acoustic front-ends. On one hand, our method is advantageous to smartphone back-ended solutions as smartphones have very limited access to ANC earphones: they can only feed data to speakers but have no access to the in-ear microphones. On the other hand, it is also superior to self-built prototypes integrating in-ear microphones with low-end earphones, as these mock-ups are often built without engineering experience. The total cost of EarACE’s sensing board is around \$12, which could be further reduced upon mass production.



Fig. 19. The structure of an ANC earbud.

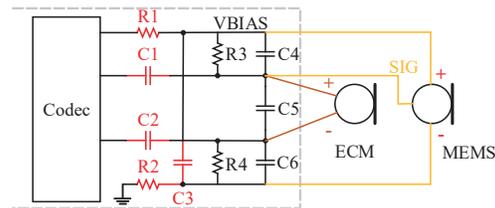


Fig. 20. Connections for ECM and MEMS microphones. For MEMS microphones, we only need R1, C1, and C3 but short out C2 and R2, leaving other components vacant.

The major challenge faced by wiring ANC earphones to EarACE is the heterogeneous acoustic front-ends equipped by various products; it requires sophisticated knowledge of analog circuit design and a deep understanding of their working principles. Basically, the microphone technology can be roughly categorized into two types: Electret condenser microphone (ECM) and micro-electromechanical system (MEMS). For ECM-based earphones,

²Other structures are also available but often have less competitive noise cancellation effect and hence are less common; we hence omit the discussions on them.

each earbud has six wires, two for a speaker and four for two microphones. In contrast, MEMS-based ones can have more wires as each microphone may use up to three wires, respectively corresponding to VCC, GND, and signals, though most manufacturers would merge the two GNDs or even VCCs. We illustrate the schematics for connecting both microphone types to our codec in Fig. 20. Because speakers always have two wires without polarity regardless of their constructions, they can be directly connected to EarACE without any trouble.

Since the above ANC acoustic front-ends entail many wires (up to 12 for ECM-based earphones), so we decide to use a USB Type-C (USB-C) socket to gather these wires and then connect them to our board, as shown in Fig. 5. We choose USB-C for its compactness in size and richness in pins. Note that we use USB-C to carry only (raw) analog signals from/to microphones and speakers. This is in stark contrast to normal mobile devices (e.g., smartphones) using USB-C only as a digital data interface, so our USB-C cannot be directly plugged into an arbitrary USB-C compatible device, eliminating the possibility of using such a device to serve as the back-end processing unit. The last challenge is the non-trivial task of distinguishing the wires out of earbuds, which we achieve through iterative trials leveraging a function generator and an oscilloscope. Essentially, given an incorrect connection, the acoustic front-ends would behave abnormally; for example, the observed signals would encounter severe distortions. After cracking dozens of earphones, we have learned that, for non-True Wireless Stereo (TWS) ANC earphones, there exist visible wires out of an earbud and the codecs are detached from the earbuds, as shown in Fig. 21(a). As a result, directly tapping onto the identified wires and connecting them to EarACE would incur little risk of breaking an earbud's inner structures. For TWS earphones, things get more complicated as we have to open an earbud and search for possible connections inside, as depicted in Fig. 21(b). In general, we would recommend users sticking to non-TWS earphones for prototype testing, until we may get EarACE fully integrated with ANC earphones.

To summarize, in order to customize EarACE for a certain pair of ANC earphones, one can resort to the following steps. First, bypassing unnecessary circuitry (e.g., codec) of an ANC earbud but retaining all acoustic sensors, namely the microphones and speakers. Second, determining the type (ECM or MEMS) of onboard microphones and correspondingly wiring the acoustic sensors to the earphone connector of EarACE shown in Fig. 5(b) via the labels printed on the PCB board. Third, accommodating the circuit configurations of the EarACE's mother board for connection according to Fig. 20. Finally, plugging the earphone connector to the mother board. At this time, an EarACE prototype is fully ready to be used.



Fig. 21. Non-TWS earphones (a) have visible wires outside. TWS earphones (b) have to rely on manual connections inside the earbud so as to bypass internal circuits.