



BreathPro: Monitoring Breathing Mode during Running with Earables

CHANGSHUO HU, Singapore Management University, Singapore

THIVYA KANDAPPU, Singapore Management University, Singapore

YANG LIU, University of Cambridge, United Kingdom

CECILIA MASCOLO, University of Cambridge, United Kingdom

DONG MA*, Singapore Management University, Singapore

Running is a popular and accessible form of aerobic exercise, significantly benefiting our health and wellness. By monitoring a range of running parameters with wearable devices, runners can gain a deep understanding of their running behavior, facilitating performance improvement in future runs. Among these parameters, breathing, which fuels our bodies with oxygen and expels carbon dioxide, is crucial to improving the efficiency of running. While previous studies have made substantial progress in measuring breathing rate, exploration of additional breathing monitoring during running is still lacking. In this work, we fill this gap by presenting BreathPro, the first breathing mode monitoring system for running. It leverages the in-ear microphone on earables to record breathing sounds and combines the out-ear microphone on the same device to mitigate external noises, thereby enhancing the clarity of in-ear breathing sounds. BreathPro incorporates a suite of well-designed signal processing and machine learning techniques to enable breathing mode detection with superior accuracy. We implemented BreathPro as a smartphone application and demonstrated its energy-efficient and real-time execution.

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

Additional Key Words and Phrases: Running Monitoring, Breathing Analysis, Breathing Mode, Earables, Audio Processing

ACM Reference Format:

Changshuo Hu, Thivya Kandappu, Yang Liu, Cecilia Mascolo, and Dong Ma. 2024. BreathPro: Monitoring Breathing Mode during Running with Earables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 2, Article 71 (June 2024), 25 pages. <https://doi.org/10.1145/3659607>

1 INTRODUCTION

Aerobic exercise is an essential component of a well-rounded fitness routine, offering a positive impact on cardiovascular health, stamina, endurance, mental well-being, and overall physical fitness [36, 61]. Nowadays, running has gained popularity as a preferred choice of aerobic exercise among individuals due to its simplicity, accessibility, and affordability. Unlike other aerobic exercises that necessitate specialized equipment or facilities, running only requires running shoes and a place to run. Statista Research Department's 2022 report indicates that approximately 50 million Americans (or 15% of the U.S. population) participate in some form of running or jogging, and this figure is still increasing [5]. Recently, advancements in wearable technology, such as wristbands

*Corresponding author: Dong Ma, dongma@smu.edu.sg, Singapore Management University

Authors' Contact Information: [Changshuo Hu](mailto:cs.hu.2023@phdcs.smu.edu.sg), cs.hu.2023@phdcs.smu.edu.sg, Singapore Management University, Singapore; [Thivya Kandappu](mailto:thivya.kandappu@smu.edu.sg), thivya.kandappu@smu.edu.sg, Singapore Management University, Singapore; [Yang Liu](mailto:yl868@cam.ac.uk), yl868@cam.ac.uk, University of Cambridge, United Kingdom; [Cecilia Mascolo](mailto:cm542@cam.ac.uk), cm542@cam.ac.uk, University of Cambridge, United Kingdom; [Dong Ma](mailto:dongma@smu.edu.sg), dongma@smu.edu.sg, Singapore Management University, Singapore.

Permission to make digital or hard copies of all or part 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 components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, or post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2474-9567/2024/5-ART71

<https://doi.org/10.1145/3659607>

and wireless earbuds, have facilitated the automatic monitoring of various running parameters (e.g., pace, cadence, intensity, heart rate (HR), and so on [12]). This technology enables a comprehensive assessment of running activities, providing valuable insights for potential performance improvement [35, 38].

Existing studies have primarily measured the following three factors while running: 1) Location-related factors such as velocity, orientation, position, step counts, stride frequency, and stride length, have been extensively studied by leveraging in-the-pocket sensors [11, 13, 16, 24, 42, 50, 60], e.g., inertial measurement units (IMUs) and the Global Positioning System (GPS); 2) Strike-related factors such as foot placement and focus, ground reaction force, as well as force distribution, are mainly measured using multiple IMUs worn on the body [74, 75], sensors in shoes [32, 71], and force platforms on grounds [58, 67]; 3) Physiology-related factors such as HR and respiration rate (RR), can be monitored with chest-worn straps, like Polar [6] for HR monitoring, and Zephyr [7] for both HR and RR monitoring. Although these methods are effective, they are invasive in nature. To this end, hEARt [15] proposes to monitor HR while running using in-ear microphone audio from earables. RunBuddy [31] employs both smartphones and headphones to monitor Locomotor Respiratory Coupling (LRC) during running, which can then be utilized for respiration rate monitoring.

Although previous studies have achieved remarkable progress in running assessments through monitoring various factors, breathing type (i.e., breathing using mouth or nose), which is highly correlated with the volume of oxygen exchange and therefore affects the running performance [23, 40, 47], has not been investigated yet. In this paper, we delve into additional breathing monitoring during running, and for the first time, introduce the concept of *breathing mode* within the context of the human respiratory system, namely, *the combination of breathing type (including nose or mouth breathing) and breathing phase (including inspiration or expiration)*. In detail, we identify four breathing modes: nasal inhalation (NI), nasal exhalation (NE), oral inhalation (OI), and oral exhalation (OE). While individuals are capable of self-identifying their breathing mode, constant self-monitoring can divert their attention from crucial factors such as surrounding pedestrians, traffic, posture, and pace. On the contrary, automatic and continuous monitoring of breathing type not only alleviates the need for self-monitoring but also offers an opportunity to guide runners in adopting proper breathing techniques for improved running performance.

We investigate the feasibility of breathing mode detection by utilizing earables, a promising candidate for this task due to the following considerations: 1) Earables are widely chosen companions by runners as they provide entertainment and interaction features; 2) Their on-body location situates them in close proximity to respiratory and cardiovascular systems, offering an ideal position for breath monitoring; 3) Previous studies have demonstrated their ability to capture a variety of accurate and motion-resilient physiological parameters using the onboard in-ear and out-ear microphones, such as HR [15, 17] and RR [62, 69]. Specifically, in-ear microphones have been proven to reliably capture heart sounds [15, 17] and breathing sounds [30, 44] due to the occlusion effect [63] that amplifies the low-frequency component of bone-conducted sounds. Out-ear microphones have been shown to record breath sounds roughly during running, but they are also highly sensitive to capturing ambient noise [31].

To this end, this paper proposes BreathPro, a portable and lightweight earable-based system for high-fidelity breathing mode monitoring during running. We initially leverage in-ear microphones on earables as they can capture both air- and bone-conducted sounds generated by respiration during running, as investigated in Section 3.2. However, the following technical challenges need to be carefully addressed:

- The development of a machine learning (ML) based breathing mode classification model requires precisely annotated data, namely, accurately segmented in-ear signals between inspiration and expiration. However, the breathing signals collected from the in-ear microphone are considerably weak and span across wide frequency ranges, making it challenging to discern the transitions of breathing phases in both time and frequency domains. To resolve this issue, we leverage the ground truth signal collected near the nose to

guide the segmentation of the in-ear signal for model training. During inference, the segmentation process is eliminated through our innovative design of frame-based classification, and the performance is boosted with the proposed high-quality frame selection scheme and post-processing.

- Our breathing mode classification model, which is trained using clean breathing sounds captured by in-ear microphones in quiet environments, encounters considerable performance degradation when the user runs in a noisy environment. To tackle this issue, we observe that external noise is first picked up by the out-ear microphone, then modulated by the occluded ear canal, before finally being captured by the in-ear microphone. Consequently, we propose to utilize the out-ear signal to diminish the in-ear noise. This is achieved by estimating the residual noise inside the ear using the out-ear signal and subsequently subtracting it from the noisy in-ear signal, yielding cleaner in-ear breathing sounds.

In summary, this paper makes the following contributions:

- To the best of our knowledge, BreathPro is the *first wearable system that tracks runners' breathing modes*. We contribute to the research community by introducing a new sensing application and encouraging future research for performance improvement from both algorithmic and modality perspectives.
- We meticulously design a comprehensive suite of signal processing and machine learning techniques to create BreathPro. Specifically, we introduce an innovative noise reduction method for in-ear audio sensing by leveraging out-ear signals to attenuate in-ear noise, thereby improving the quality of in-ear breathing sounds during running.
- Utilizing data collected from 25 participants, our results demonstrate that BreathPro achieves an accuracy of 90.70% and 98.52% in classifying four breathing modes at the frame level and phase level, respectively. This performance is sound even under a range of challenging conditions, such as varying intensities and types of ambient noise and different inhalation and exhalation sequences.
- Our measurements on energy consumption and latency demonstrate the minimal overhead of BreathPro, suggesting its efficient and real-time execution on smartphones accompanied by earables.

2 RELATED WORK

In this section, we overview related research in three categories: running monitoring systems, respiration monitoring systems, and in-ear sensing applications.

2.1 Running Monitoring Systems

While there are many running monitoring systems, we focus this review on those that perform physiological monitoring during running, such as HR [6, 15] and RR [7, 31, 73]. These have been tracked using chest-worn straps, like Polar for HR monitoring [6] and Zephyr for both HR and RR monitoring [7]. While these methods are efficient, they are also invasive. In response to this, hEARt [15] proposed monitoring HR during running with in-ear microphone through leveraging the occlusion effect to enhance low-frequency bone-conducted sounds in the ear canal and applying a deep learning-based motion artifact mitigation framework. RunBuddy [31] employed a combination of a smartphone's IMU and a headphone's microphone to monitor LRC during running, which indicates possible ratios between the stride and breathing frequencies and thus can be used for RR monitoring. ER-Rhythm [73] also developed a method for monitoring LRC using RFID while running on a treadmill.

Our work aligns closely with the above works while delving deeper into the exploration of additional breathing information, i.e., breathing mode, which provides valuable insights for managing a runner's breathing, thereby enhancing overall running performance.

2.2 Respiration Monitoring Systems

2.2.1 Non-earables based Systems. Several works in the past have focused on using various on-body and off-body sensors to monitor various aspects of human respiratory system. Acoustic signals are an intuitive method for detecting respiration. By using microphones on masks [59] and smartphones[49], respiratory signals can be captured for respiratory rate estimation. However, these methods require opportunistic placement (close to mouth/nose) of the audio sensor, making it challenging to achieve seamless monitoring anytime and anywhere. Chu et al. [21] developed a disposable respiration sensor that can be placed on the ribcage and abdomen to measure the expansion and contraction during respiration. Zephyr [7] utilized a pressure sensor to measure the chest movements during respiration to monitor RR and respiration volume. However, these on-body sensors can be intrusive and uncomfortable for runners. This is particularly true for specialized chest bands that need to be tightly fastened around the chest, potentially leading to discomfort, restricted free breathing, and limited movement during running. Existing studies also employ body-worn photoplethysmography (PPG) and electrocardiography (ECG) to capture heart signals and utilize respiratory sinus arrhythmia (RSA) for monitoring RR [19], which have been leveraged by commercial products, such as Apple Watch [2] and Garmin [3].

Transitioning to off-body devices, a range of sensing modalities has been utilized for respiration monitoring. Hu et al. [34] introduced a dual-mode imaging system that operates on both visible and long-wave infrared wavelengths from RGB and thermal videos respectively, to non-invasively and unobtrusively measure RR and respiration volume by capturing the temperature changes caused by respiration using a thermal camera. Massaroni et al. [45] utilized a laptop's integrated RGB camera and proposed a post-processing algorithm to capture the chest movements incurred by respiration to monitor RR. Similar to capturing the minute movements caused by respiration, Liu et al. [41] proposed to track RR during sleep by using off-the-shelf WiFi, through analyzing the channel information in both time and frequency domains. However, camera-based solutions raise privacy concerns and lack portability, while wireless-signals-based solutions also face limitations in portability. Additionally, these approaches are specifically designed for stationary situations and are not suited for use during running.

2.2.2 Earable based Systems. Due to the close and fixed distance between human ears and mouth/nose, earable is considered an advantageous commodity wearable for respiratory activity sensing. We group earable-based breathing monitoring systems into two categories. The first category involves IMU-based breath activity detection in a static environment. Rahman et al. [55] have introduced algorithms for estimating respiration rate during resting positions, utilizing inertial sensors embedded in common earbuds to capture breath-incurred motions and mitigate errors caused by passive and active head motion. The low-power accelerometer in the earbuds is also employed to generate a set of breathing biomarkers, encompassing breathing rate, depth, and symmetry, facilitating guided breathing exercises for users [54]. Similarly, Röddiger et al. [57] estimated respiration rate by filtering respiration-related body motions at the ear using accelerometer and gyroscope data. However, this inertial sensing approach heavily relies on underlying motion and is applicable only when the user is at rest. The second category involves acoustic-based breathing rate estimation. Kumar et al. [39] collected data from the microphone of a near-field headphone before, during, and after strenuous exercise, deploying a multi-task LSTM network to estimate respiratory rate in varying background noise conditions. Ren et al. [56] utilize earbuds' microphones for fine-grained sleep monitoring, with a focus on scenarios such as heavy breathing after exercise or specific environments like sleep.

In summary, previous studies on respiration monitoring have predominantly concentrated on estimating respiration rate and volume. The majority of these approaches are confined to stationary settings or involve proxy sensor placement. Our work stands out as the first to explore the feasibility of breathing mode detection under active scenarios using wearables.

2.3 Other Audio-based In-ear Sensing Systems

Earphones, inherently equipped with speakers and microphones, have proven to be advantageous for various in-ear sensing applications, including human-computer interaction (HCI) [43, 51], health monitoring [15, 20, 37], and authentication procedures [27, 64, 70, 72]. For instance, Dong et al. [43] implemented OESense to facilitate hand-to-face gesture interactions by harnessing an in-ear microphone to capture bone-conducted sounds, enabling the recognition of various tapping positions on the face. Prakash et al. [51] proposed an approach for HCI by using teeth actions, specifically tapping and sliding, which create detectable vibrations in the jaw and skull, and produce vibratory signals detectable by earphones. Christofferson et al. [20] leveraged the internal and external microphones presented in active noise-canceling earbuds to distinguish sounds associated with poor or disordered sleep such as snoring, teeth grinding, and restless movements. Truong et al. [66] estimated blood pressure by leveraging the distinct propagation times of sound and blood within the human body. They employed both in-ear microphones and PPG to measure the vascular transit time (VTT), calculated as the time difference between the S1 heart sound and the PPG upstroke in one pulse cycle. Jin et al. [37] developed a non-invasive system for monitoring ear conditions by emitting a probing chirp and analyzing the recorded echoes evoked by the chirp sound stimulus, which can detect three major hearing health conditions: ruptured eardrum, earwax buildup and blockage, and otitis media. Ferlini et al. [27] proposed a gait-based authentication technique by leveraging the low-frequency in-ear sounds generated by stepping motions while walking.

Compared to conventional out-ear microphones, the newly emerged in-ear microphone offers two advantages for earable sensing: (1) by positioning the microphone inside the ear canal, the earbud shell naturally attenuates external noises and therefore enhances the signal-to-noise ratio of the in-ear sensing signal, and (2) the sealing of the ear canal opening creates the occlusion effect [63] that amplifies the low-frequency bone-conducted sounds, thereby enabling the detection of various body sounds. All of the above in-ear sensing systems operate independently yet intersect with our work. Our research serves as a valuable addition to the field of earable sensing, further broadening the scope of in-ear sensing applications.

3 PRELIMINARY

3.1 Effect of Breathing Modes on Running

Respiration, the process of breathing to fuel our bodies with oxygen and getting rid of carbon dioxide [22], is crucial for running as such intense activity consumes oxygen and produces wastes faster. During running, people can breathe using their nose (refers to nasal breathing) or mouth (refers to oral breathing), or an alternation of both. Nasal and oral breathing can lead to different physiological responses in the body [23, 47]. Specifically, oral breathing during exercises at 60% of one's aerobic capacity allows a greater volume of air to be utilized. However, nasal breathing filters pollutants from the air and facilitates bronchial nitric oxide production, which may positively influence exercise. Generally, nasal breathing is comfortable at lower exercise intensity, but at 35-41 L/min total ventilation (VE), individuals usually switch from breathing nasally to oral breathing [29, 40].

The selection of the best breathing mode during running also depends on factors such as exercise intensity, environment temperature, and humidity [22, 25]. During intense running sessions, such as sprints or interval training, oral breathing is preferred to maximize airflow and oxygen intake, supporting the increased energy demands. On the other hand, nasal breathing can be beneficial during long-distance runs at a comfortable pace, facilitating the regulation of airflow and ensuring a steady and controlled intake of oxygen. In hot and humid environments, oral breathing aids in dissipating heat efficiently by facilitating greater airflow and faster cooling. Conversely, nasal breathing helps warm and humidify the inhaled air in the nasal cavity, which is particularly useful in cold weather [52].

The automatic recognition of a runner’s breathing mode using earables can significantly enhance the effectiveness of coaching during running¹, leading to improved endurance and reduced fatigue. For instance, based on factors such as run intensity, stage, and environmental conditions, real-time suggestions for optimal breathing mode can be provided, without distracting the runner’s engagement by requiring manual focus on their breathing. Post-run reports can analyze the breathing modes employed during the session, offering insights into potential improvements. Additionally, a running assistant can automatically design a standard or personalized breathing mode management scheme before each run and continuously track whether the runner adheres to it in real time.

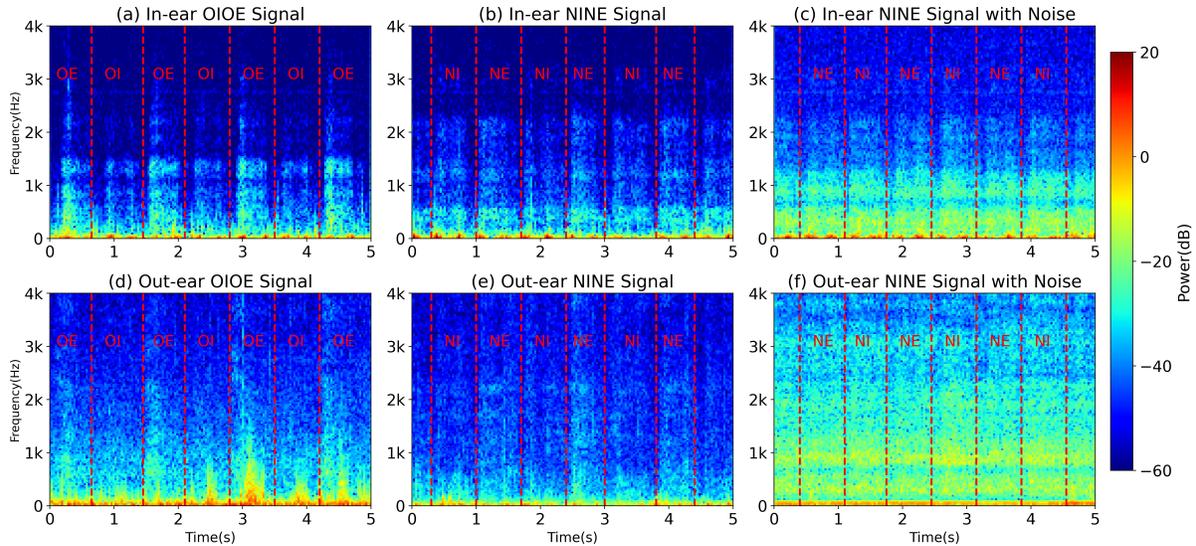


Fig. 1. In-ear and out-ear spectrograms of two breathing phase combinations collected in quiet environments (a, b, d, e), and a NINE combination from noisy environments (c, f). The breathing signals captured by in-ear microphones are more pronounced, and more resilient to ambient noise.

3.2 Feasibility Exploration: In-ear vs. Out-ear Microphone

The above discussions have clearly shown the importance and benefits of monitoring runners’ breathing modes during running. Unfortunately, there is no existing common wearable that can support such operations. In this paper, we focus on earables (e.g., earphones), a highly-adopted wearable device in recent years, particularly during exercises (e.g., running). Conventional earphones typically feature microphones positioned near the outer shell, known as out-ear microphones, which are designed to capture air-conducted human voices for communication purposes. Recent earphones equip additional microphones in the earphone housing, known as in-ear microphones, that face inward to capture residual sounds in the ear canal for the purpose of active noise cancellation. Next, we explore the feasibility of using the in-ear microphone to detect runners’ breathing modes, motivated by recent works that demonstrate the capability of in-ear microphones to capture human-generated vibrations/sounds through bone-conduction and occlusion effect [63].

We developed a prototype (details presented in Section 5.1) and collected some data with both the in-ear and out-ear (for comparison purposes) microphones when a subject undergoes different breathing modes during running. Figure 1 shows the spectrograms of the collected signals, which infer three promising findings. First, we

¹Note that we mainly focus on the detection of breathing mode, while the subsequent coaching for running is out of the scope of this work.

can observe from the upper row that different breathing modes (OI, OE, NI, NE) exhibit distinct characteristics across frequencies (mainly from 500 Hz to 3000 Hz), showing the potential to classify them using the in-ear signals. Second, comparing the first and second rows, the in-ear signal yields a higher signal-to-noise ratio (SNR) as it captures bone-conducted breathing sounds, while the breathing cycles can be barely noticed in the out-ear signal due to the high air attenuation. Third, from the last column, when external noise (around 60 dB) exists, both in-ear and out-ear signals are polluted, although the out-ear signal encounters a stronger impact in high frequencies (>2000 Hz). This implies that breathing mode detection with the in-ear microphone is still vulnerable to external noise and proper noise removal techniques should be applied.

The above studies demonstrated the feasibility of using the in-ear microphones to capture and distinguish runners' breathing modes, and identified potential factors that might affect the detection performance. Next, we present our design to realize the idea and optimize the performance under different scenarios.

4 SYSTEM DESIGN

4.1 Overview

The proposed BreathPro achieves accurate and robust detection of breathing modes under various running scenarios by opportunistically tapping on both in-ear and out-ear microphones. While the in-ear microphone mainly captures bone-conducted sounds generated by respiration, the out-ear microphone can be utilized to record ambient sound for subsequent noise reduction. Figure 2 shows the BreathPro methodological pipeline, consisting of the following phases:

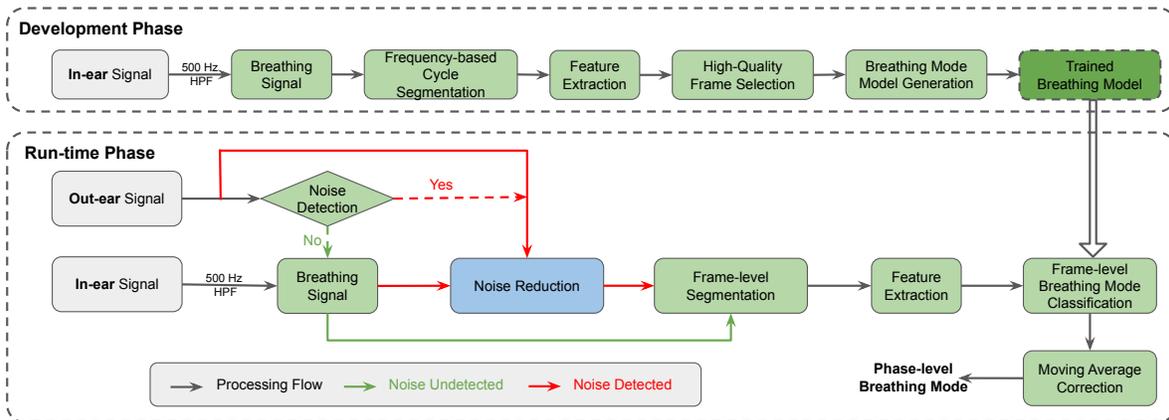


Fig. 2. The methodological pipeline of BreathPro.

- Development Phase:** The goal of this phase is to develop trained machine learning (ML) models to detect breathing modes. We start off by extracting frequencies that contain breathing-related information. As illustrated in Figure 1, the high-frequency component (high pass filtered with a cutoff frequency of 500 Hz) of the in-ear signal shows distinct characteristics across different breathing modes. As the breathing sounds recorded in the ear canal are considerably weak, we first adopt a frequency-based approach to segment the raw signal into different breathing phases. To be more specific, the ground truth data recorded by an external microphone placed on the philtrum is initially segmented by identifying the transitions between breathing phases, which are represented by the valleys of the frequency-energy curve. Subsequently, the identified transitions are used to segment the in-ear breathing signals recorded concurrently with the ground truth signal. We then extract

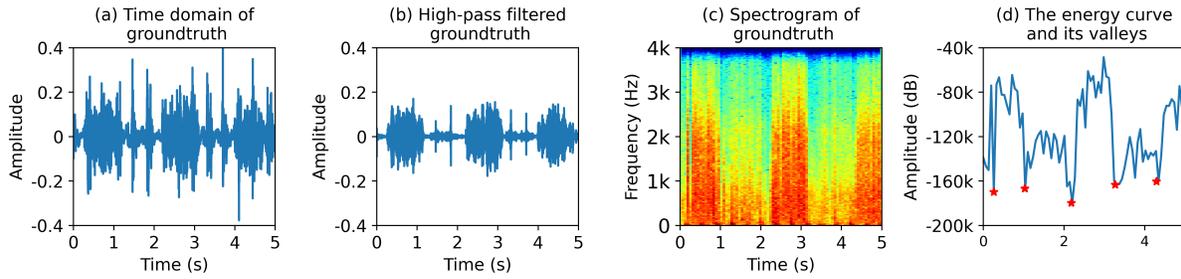


Fig. 3. Spectrogram-based breathing phase segmentation process. (a) the raw ground truth breathing signal, (b) the high-pass filtered ground truth breathing signal, (c) the spectrogram of filtered ground truth breathing signal, and (d) the energy curve derived from the spectrogram, in which the valleys indicate transition intervals.

40 MFCC features over each frame of 100ms and slide through the entire segmented breathing phase (with the hop length of 50ms). By carefully applying a series of techniques (explained in Section 4.2), the frames with low quality (that are either polluted by spontaneous environmental noise or located at the transition between adjacent phases) are eliminated. Afterward, the MFCC features of the remaining frames are labeled and fed into an ML classifier for training, resulting in a model that can recognize different breathing modes.

- Run-time Phase:** This phase focuses on classifying input signals into different breathing modes using the pre-trained model, with techniques to minimize the influence of background noise such as traffic sounds. Specifically, we first leverage the out-ear microphone to estimate the ambient noise level. If the noise level is above a certain threshold, the noise reduction module (described in Section 4.3) will be activated for in-ear noise cancellation with the aid of the out-ear microphone, where the core idea is to estimate the residual noise in the ear and subtract it from the in-ear signal. Otherwise, the module will be bypassed and the filtered breathing signal will be directly fed to the subsequent modules for frame-level segmentation, feature extraction, and classification with the pre-trained model. Finally, as a breathing phase spans across multiple frames, we propose a moving average scheme to smooth the frame-level classification results, obtaining a higher phase-level accuracy.

4.2 Classification Model Development

4.2.1 Spectrogram-based Breathing Phase Segmentation: Although the breathing patterns are roughly seen from the spectrograms presented in Figure 1, applying segmentation directly on the spectrograms of the in-ear signal yields poor performance because (1) some breathing modes (nasal inhaling and nasal exhaling) produce weaker breathing sounds and can be easily disrupted by environmental noise, and (2) the boundaries of the adjacent breathing phases are vague and would result in inaccurate segmentation. Thus, in the model development phase, we concurrently record the signals from another microphone placed near the nose to obtain the ground truth of breathing phase. Figure 3 (a) and Figure 3 (b) compare the original ground truth signal and the high-pass filtered breathing signal with a cutoff frequency of 500 Hz. As shown in Figure 3(b) and Figure 3(c), the inhaling and exhaling phases present distinctive patterns and clear transitions (the change of frequency components and the associated energy). The high-pass filter also eliminates other body-generated sounds such as footstep or heart beat sounds as they mainly reside in frequencies lower than 100 Hz [15, 43]. Note that the ground truth data is dedicated to breathing phase segmentation and annotation of the in-ear signal, and will not be employed as input for model training.

To segment the spectrogram of the ground truth signal, we conducted a Short-Time Fourier Transform (STFT) on each five-second window to acquire its spectrogram expressed in decibels. We then sum up the energy across

all frequency bands to construct a frequency energy curve, as shown in Figure 3(d). Based on the observation that there exists a silence period (no breathing) during the transition, we extract the valleys of the frequency energy curve as the transitions between adjacent breathing phases. Then, we manually label the segmented phases as inhalation or exhalation. Specifically, since the two phases occur alternately, we only need to identify the first phase manually and the subsequent phases can be labeled based on the alternation. Finally, we apply the index extracted from the ground truth signal to the in-ear breathing signal (extracted with a >500 Hz high-pass filter) and segment it².

4.2.2 Framing and Feature Extraction: To capture the fine-grained breathing mode information from the breathing sounds, we extract the Mel-Frequency Cepstral Coefficients (MFCC) features that are designed to approximate the human auditory system’s response at varying frequencies and has been widely used in diverse human-generated sounds including speech, music, and breathing sounds [8]. Specifically, prior research has successfully used MFCC for breathing event detection during running [31], COVID-19 detection using breathing sounds [10, 53], as well as respiratory disease classification [48]. In detail, each segmented breathing phase is further split into 100 ms frames with 50 ms hop length using the slide window technique. 40 MFCC features are extracted for each frame. We further compare the performance of MFCC against other features in Section 6.1.3.

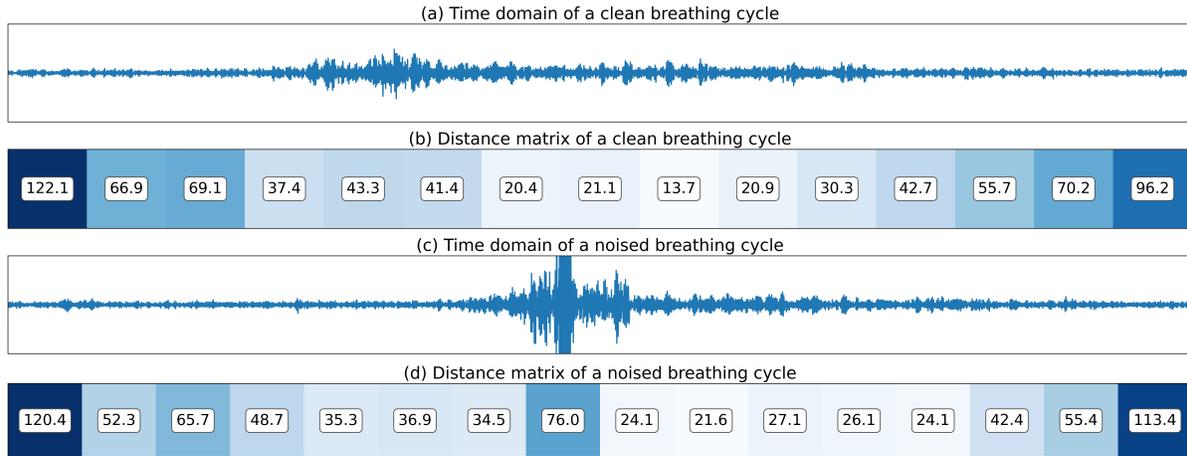


Fig. 4. Distance matrix between the template and each frame of (a) a clean breathing phase and (c) a breathing phase with a sudden noise. Frames with a distance greater than 50 from the template will be excluded.

4.2.3 High-quality Frame Selection: As mentioned earlier, there is a transition between inhaling and exhaling phases, where the human does not perform any breathing or generate noticeable breathing sounds. Since our current segmentation technique cannot exclude these transitions, some frames at the beginning and end of each segmented phase may not represent the corresponding breathing modes, leading to low-quality features for model training. Moreover, as shown in Figure 4(c), there might be spontaneous environmental noise occurring in the middle of breathing.

To mitigate those low-quality and noisy frames, we propose a frame selection scheme. In detail, we first generate a *feature template* by computing the average of MFCC features from all the frames in the current

²Note that the ground truth signal is only used during training to segment the in-ear breathing sounds, while is not needed during run-time as we do not perform breathing phase segmentation and only use frames for classification.

segmented phase. Specifically, with a frame length of 100 ms and 50 ms overlap, a single breathing phase (around 0.7 second) can be split into 15 frames for template computation. Then, we calculate the Minkowski distance ($p=2$) of MFCC features between each frame i and the template, i.e., $d_i (i = 1, 2, 3 \dots)$. As exemplified in Figure 4, frames with pure breathing signals have shorter distances to the template, while the transition frames (two ends of Figure 4 (b) and (d)) and noisy frames (middle of Figure 4 (d)) show relatively larger distances. Afterward, we compute the average distance (d_{avg}) of all the frames in current segmented phase. If the ratio $|d_i - d_{avg}|/d_{avg}$ is higher than a certain threshold α , frame i is considered low-quality and will be excluded for model training. We empirically set α as 2 to retain high-quality frames and this design is able to handle the varying signal strength (i.e. volume) of breathing sounds.

4.2.4 Model Training: The extracted and selected breathing features and their corresponding labels are utilized to train a machine-learning model. Four commonly used ML classifiers, namely Support Vector Machine (SVM), K Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF), are considered, and the evaluation conducted in Section 6.1.1 will help identify the classifier with the best performance.

4.3 Run-time Breathing Mode Classification

As our breathing mode classification model is trained with clean breathing sounds and the in-ear breathing signal is vulnerable to environmental noise, we introduce the out-ear microphone to deal with this issue at run-time. Specifically, the out-ear microphone can be used for two purposes: noise detection and noise reduction.

4.3.1 Noise Detection: The out-ear signal is a combination of the runner's breathing sound, the associated wind noise, and other environmental sounds. Since breathing sounds are weak and the running-induced wind noise exists all the time, the out-ear signal can be considered an approximation of the environment noise. Thus, we compute the overall energy of the out-ear signal across different frequency bands and convert it to sound volume in decibels (dB). If the noise level exceeds a certain threshold (e.g., 55 dB), BreathPro will activate the noise reduction module to clean the noisy in-ear signal before segmentation. Otherwise, the noise reduction module will be bypassed and the unprocessed breathing signal will be split into frames directly.

4.3.2 Noise Reduction: We observe a phenomenon that *any external noise is firstly captured by the out-ear microphone, subsequently modulated by the occluded ear canal cavity between the earbud and eardrum, before finally being captured by the in-ear microphone*. This indicates that there exists a correlation between the in-ear and out-ear noise signals and inspires us to leverage the out-ear signal for in-ear noise reduction. Specifically, our idea is to estimate the residual noise inside the ear using the out-ear signal and then subtract it from the noisy in-ear signal to obtain clean in-ear breathing sounds. It is worth noting that, aside from external noise, the out-ear signal may also contain faint air-conducted breathing sounds. As a result, the in-ear signal comprises three components - bone-conducted breathing sounds, external noise attenuated by the ear, and air-conducted breathing sound attenuated by the ear. During noise removal, BreathPro treats the two components in the out-ear signal as a whole and removes both, which ensures that the in-ear signal still contains bone-conducted breathing sounds for breathing mode classification.

To realize this idea, we first need to obtain the correlation between the in-ear and out-ear signals. A common way is to develop a machine learning or deep learning model to learn such correlation automatically. However, this method requires a large volume of data for model training and incurs high computation overhead during inference. From Figure 1, we discovered that the two signals mainly differ in the energy distributions across different frequencies, namely, the in-ear signal shows stronger energy at low frequencies and weaker energy at high frequencies, compared to the out-ear signal³. Thus, we transform the out-ear and in-ear signals into the

³This is also supported by the occlusion effect that the low-frequency part of the in-ear signal is amplified and its high-frequency part is suppressed [18].

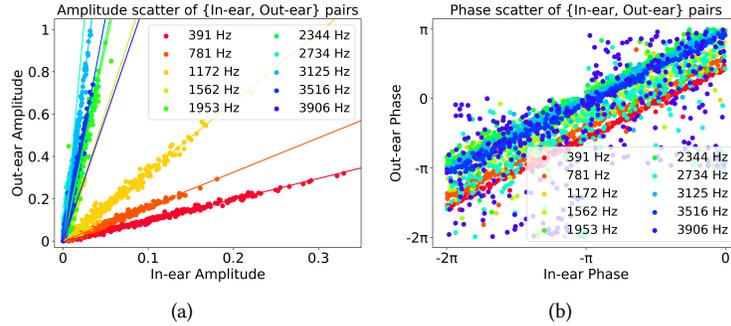


Fig. 5. The scatter plot of (a) amplitudes and (b) phases for {out-ear, in-ear} pairs at 10 different frequency bands.

frequency domain using Short-Time Fourier Transform (STFT), and compute the amplitude and phase at each frequency band.

Figure 5(a) and Figure 5(b) show the scatter plot of amplitudes and phases for {in-ear, out-ear} pairs at 10 different frequency bands. We observe linear correlations between the in-ear and out-ear signals in both amplitude and phase plots at a single frequency band. Specifically, when the out-ear signal is stronger, the in-ear signal also exhibits greater strength. This observation aligns with the fundamental sound propagation model, i.e., inverse-square law [26], which indicates that given the fixed distance between the in-ear and out-ear microphone, the attenuation ratio is deterministic. As such, we leveraged a *lightweight linear regression model* to establish mappings of {out-ear, in-ear} for each frequency band⁴. Specifically, we formulated a linear relationship between the amplitude and phase of out-ear and in-ear sounds as represented by:

$$S_{in}^i = a^i * S_{out}^i + b^i, i \in [1, 1024], \quad (1)$$

where the superscript i denotes the frequency band, and S_{out} and S_{in} denote the amplitude or phase of the out-ear and in-ear sounds, respectively. We set the `n_fft` of STFT as 2048, therefore resulting 1024 frequency bands. The right part of Figure 6 summarizes the flow to derive the in-ear and out-ear correlations, referred to as *Offline Coefficients Learning*. Specifically, we first play a background sound with various frequencies and record the in-ear and out-ear signals (named as template). Then, STFT is applied to both signals to obtain the amplitudes and phases at different frequency bands, which are used to derive the coefficients with linear regression. Since the coefficients derivation is based on the fundamental mechanism of sound propagation, it remains unaffected by different breathing modes and noise conditions.

The pipeline for run-time noise removal is illustrated in the left part of Figure 6, referred to as *Online Noise Reduction*. Figure 7 shows an example of the noise reduction performance at different stages. In detail, as exemplified in Figure 7, when external noise is detected, we first apply STFT to the out-ear signal to obtain its amplitude and phase values at different frequency bands (Figure 7(a)). Then, for each band, we computed the transformed amplitude and phase using the derived coefficients with Equation (1) (Figure 7(c)). Afterward, we subtract the transformed amplitude and phase from the corresponding in-ear noisy amplitude and phase (Figure 7(b)), resulting in a clean in-ear breathing pattern (Figure 7(d)). Finally, we apply iSTFT to recover the denoised in-ear breathing sounds⁵.

⁴Note that speaking will break the linear correlation as it influences the two microphones differently (i.e., through bone conduction for the in-ear microphone and air conduction for the out-ear microphone). In our paper, we only consider the external noise that affects both microphones via air conduction, as it is rare for a runner to engage in conversation during a run.

⁵Note that our noise removal method differs from existing active noise cancellation (ANC) technique in two aspects: (1) ANC aims to cancel the noise heard by humans which usually needs to consider the primary path (from noise source to human ear drum) and secondary path

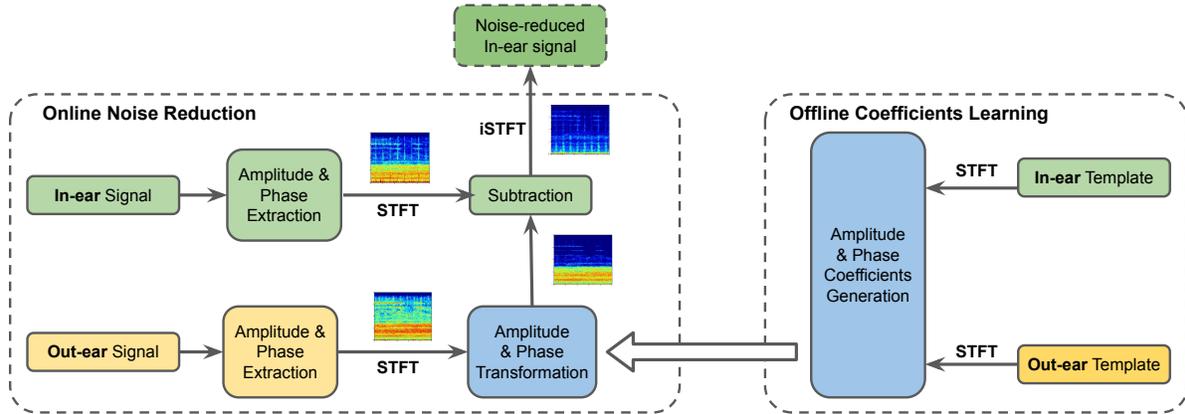


Fig. 6. The flowchart of noise reduction module, including offline coefficients learning (right part) and online noise reduction (left part).

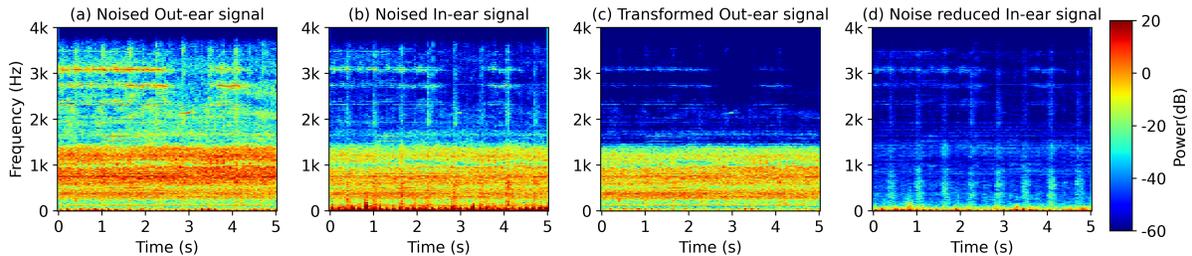


Fig. 7. The noise reduction performance at different stages.

4.3.3 Signal Segmentation, Feature Extraction, and Classification: In the development phase, we segment the breathing signal into different phases (inhaling or exhaling) with a strong reference signal from the nose. However, no reference signal is available during run-time. Thus, we segment the breathing signal into frames with a length of 100 ms using a sliding window (no overlap). For each frame, we extract the MFCC features and feed them to the trained ML model for classification. As such, we can obtain a breathing mode estimation for every frame.

4.3.4 Moving Average Correction: In our run-time pipeline, multiple frame-level estimations might belong to a single breathing phase. Based on the fact that the breathing mode (i.e., nasal or oral) remains unchanged within a single phase, there is a chance to correct some wrongly-classified frames by considering adjacent frames. Thus, we propose to leverage the moving average algorithm (with a window length of 10 frames) to smooth the frame-level estimations. Specifically, as shown in Figure 8 (b), we found that applying the moving average algorithm once can correct most of the spontaneous errors while failing to deal with consecutive errors. Thus, we apply another moving average operation to the corrected estimations and this successfully corrects consecutive errors, as compared between Figure 8 (c) and (d). Since breathing mode is defined for a single inhalation or

(from speaker to human ear drum) [68], while our method aims to eliminate the impact of external noise on a body-generated sound and we only need to consider the path between out-ear and in-ear microphones; (2) existing ANC techniques usually applies different filters to produce the anti-noise, while our solution leverage the mapping between in-ear and out-ear signals on different frequencies. Since the mapping can be easily derived by recording both microphones' signal simultaneously, our approach can adapt to personal characteristics (e.g., geometry of ear) for improved noise removal performance.

exhalation phase, reporting phase-level (i.e., aggregated breathing mode for a breathing phase) estimation is sufficient and reasonable from the user perspective.

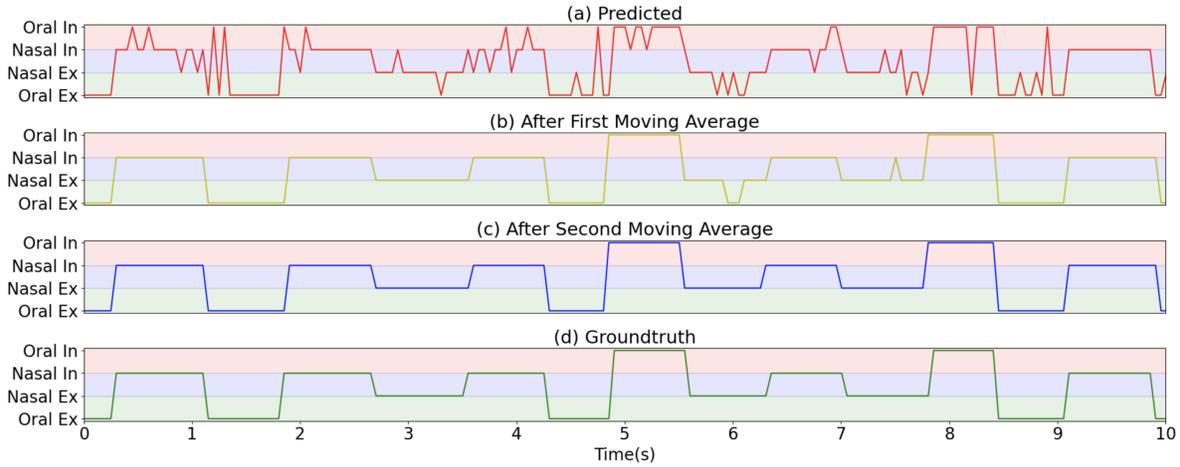


Fig. 8. The breathing mode of (a) predicted by BreathPro, (b) after first moving average, (c) after second moving average, and (d) the ground truth.

5 PROTOTYPING AND DATA COLLECTION

5.1 Prototype

BreathPro only requires an in-ear microphone and an out-ear microphone, which have been equipped in many commercial wireless earbuds such as Apple AirPods [1] and Huawei Freebuds [4]. Particularly, the in-ear microphone is primarily intended for noise cancellation purposes, where the algorithms are executed on the onboard audio chip for extremely small latency. Thus, the manufacturers usually do not release the API for accessing the raw in-ear signal.

To address this limitation, we designed and constructed a pair of earbuds to evaluate our proposed system. As shown in Figure 9(a), our customized earbuds comprise a 3D-printed shell and two analog microphones (CMC-4015-40L100 by CUI Devices), with one placed near the earbud tip to capture in-ear sound (referred to as the in-ear mic) and the other one integrated into the end of the handle to pick up out-ear sound (referred to as the out-ear mic). Both microphones are connected to a Bela Mini development board through a 3.5mm audio jack, which features an integrated development environment (IDE) for data recording. The Bela Mini board is then connected to a Raspberry Pi so that the data can be collected remotely when the subjects are running. We set the sampling rate to 8 kHz as human-produced sounds are usually below 4 kHz [46]. The participant wears the earbuds and the rest the prototype components are powered by a power bank and placed in a small bag to be carried during running, as of in Figure 9(b). Additionally, to ensure participant comfort during running, as well as good sealing quality, we provide three different sizes of foam ear tips to fit various ear canal sizes. Since both the in-ear and out-ear microphones struggle to capture a strong breathing signal with clear inhale and exhale transitions, we attached another microphone (philtrum microphone) under each participant’s nose. Due to its close proximity and direct contact with the breathing airflow, the collected audio can serve as ground truth for breathing cycle segmentation.

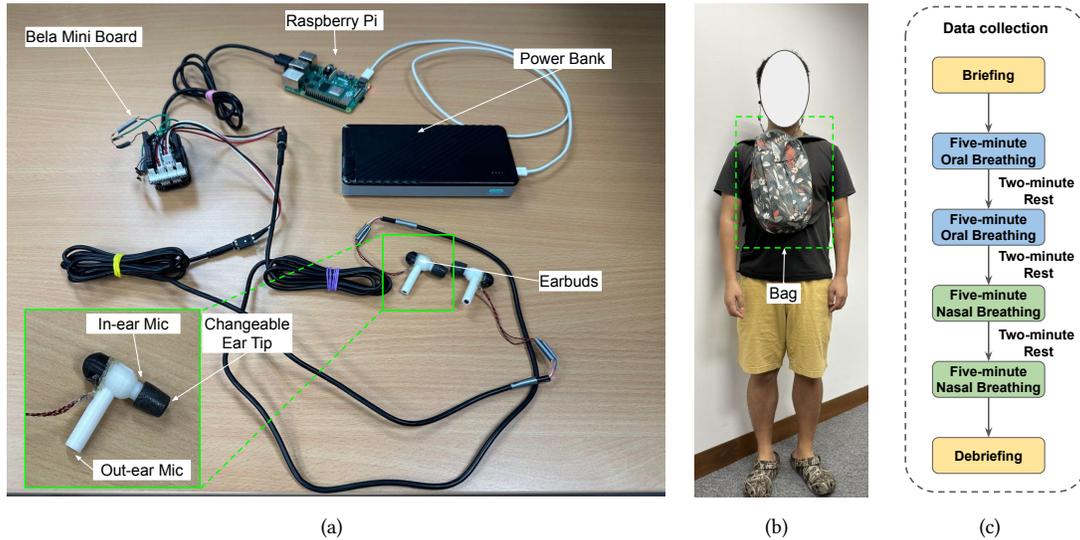


Fig. 9. (a) The customized earbuds prototype and accompanying data recording device, (b) illustration of a participant wearing the device, and (c) flowcharts of breathing data collection.

5.2 Data collection

Upon receiving approval from the Institutional Review Board (IRB), we recruited a total of 25 participants, comprising 16 males and 9 females, for large-scale data collection. The age range of the participants was between 23 and 31 years (mean = 26.72, standard deviation = 2.35). More than half of the participants engaged in regular exercise, including fitness training or aerobic activities at least twice a week, while the rest engaged in exercise less frequently. To account for real-world running conditions such as natural airflow, we conducted the data collection on an outdoor rubber running track. The total duration of the data collection process was approximately 40 minutes. Figure 9(c) illustrates the flow of data collection and we provide a detailed description as follows.

Initially, each participant received a briefing about the study procedure and provided signed consent. Subsequently, the participant wore the earbud along with other hardware components of the prototype in a bag, with the assistance of the investigator. The formal data collection was comprised of four five-minute running sessions, interspersed with two-minute break sessions to allow for rest and recovery. Participants were instructed to perform pure oral breathing (i.e., OIOE) in the first two running sessions and pure nasal breathing (i.e., NINE) in the last two sessions. Given that humans naturally adopt different breathing modes in various stages of the run, participants were considered to have mastered each breathing mode and capable of performing them accurately in the experiment. In addition, to label oral or nasal breathing, we assume the participants adhered to the provided instructions and maintained a consistent breathing mode throughout each running session. To mitigate the risk of discrepancies, we also collected feedback from the participants after each session. If they reported a potential deviation from the instructed breathing mode, we manually examined the data and excluded incorrect segments (this is achievable as oral and nasal breathing show distinct characteristics across different frequencies as illustrated in Figure 1).

6 EVALUATION

Next, we evaluate BreathPro with the collected data, aiming to answer the following questions: (1) what accuracy can BreathPro achieve in detecting the breathing mode? (Section 6.1.1, Section 6.1.2 and Section 6.1.3); (2) will different combinations of inhalation and exhalation modes affect the performance? (Section 6.1.4); (3) how

effective are the proposed techniques such as noise removal and error correction? (Section 6.1.5, Section 6.1.6); (4) what performance of can BreathPro obtain in more realistic running scenarios? (Section 6.1.7); and (5) what is the run-time overhead of BreathPro on smartphones? (Section 6.2).

6.1 Breathing Mode Classification Performance

6.1.1 Individual Model Performance: We make predictions for each respiratory frame, resulting in a series of initial predicted labels, and further apply the moving average (MA) technique twice to mitigate spontaneous errors. We refer to these as “frame-level classification”, and the accuracy after the second MA will be reported. In practical applications, however, users may only need to know their breathing mode at a phase-level. Therefore, we utilize the ground truth data to determine each breathing cycle and subsequently perform a majority voting to determine the breathing mode for a particular breathing phase. We evaluate the performance with four typical ML classifiers (KNN, SVM, RF, DT) and report the results for KNN only as it achieves the best performance.

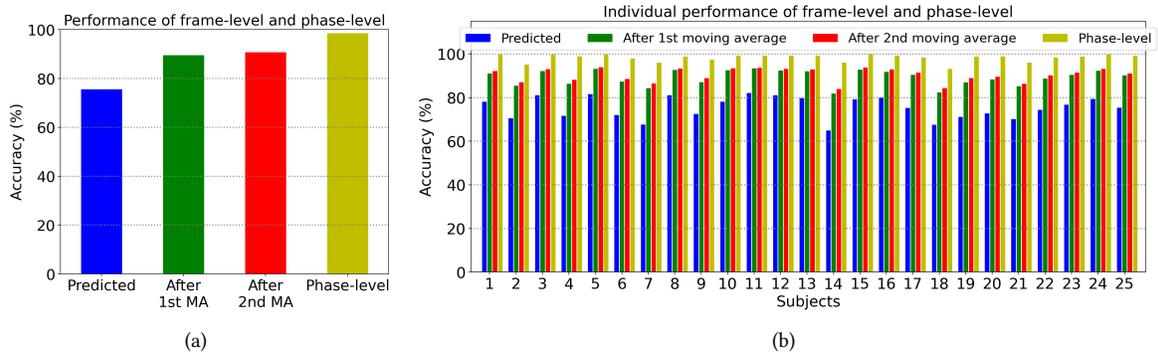


Fig. 10. (a) Overall breathing mode recognition performance of frame-level and phase-level classification, and (b) individual performance of frame-level and phase-level classification.

As shown in Figure 10(a), the overall accuracy of the initial frame-level predictions stands at 75.59%, which is sharply improved to 89.49% and 90.70% after applying the first and second MA, respectively. Additionally, the phase-level accuracy is further improved and reaches an impressive value of 98.52%. Figure 10(b) presents the individual performance of 25 subjects, considering both frame-level and phase-level classification. By combining the insights from Figure 8 and Figure 10(b), several observations can be made. Firstly, it is evident that both the first and second moving average effectively correct the majority of spontaneous errors, although they may introduce slight shifts in transition points. Secondly, non-regular runners (e.g., subjects 3, 12, and 16) tend to demonstrate better performance compared to regular runners. This can be attributed to the fact that non-regular runners may not have developed the same level of muscle coordination and breathing techniques as regular runners. Consequently, their breathing sounds are characterized by higher intensity, rapidity, and shallowness, resulting in more noticeable and distinguishable patterns.

6.1.2 Performance of Leave-one-out Test. Having witnessed the superior performance of per-individual model, in this section, we aim to focus on understanding the impact of individual differences (shape of the ear canal, bone/tissue structure of human body, and respiratory sounds) in the prediction accuracy. To this end, in this section, we use an user-agnostic model. More specifically, we employ the leave-one-out test, where 24 subjects are used for training and the remaining one subject is used for testing, and iterate across all subjects. Figure 11 reveals variations in performance across subjects. Despite an initial prediction accuracy of 69.9%, which drops approximately 6% when compared to using individual models, the utilization of the first and second MA significantly lifts the

accuracy up to 85.25% and 86.90%, respectively. Furthermore, the phase-level accuracy also reaches a remarkable value of 96.44%, with only 2% accuracy drop. Although the breathing sounds from male and female subjects are quite different (with females' breathing sounds tending to contain more high-frequency components than males'), their breathing sounds collected by in-ear microphones are less different as the occluded ear canal suppresses high-frequency components and amplifies the low-frequency components. This may also suggest that breathing signals among individuals recorded by our approach are more similar, allowing for the utilization of a general model for analysis and prediction.

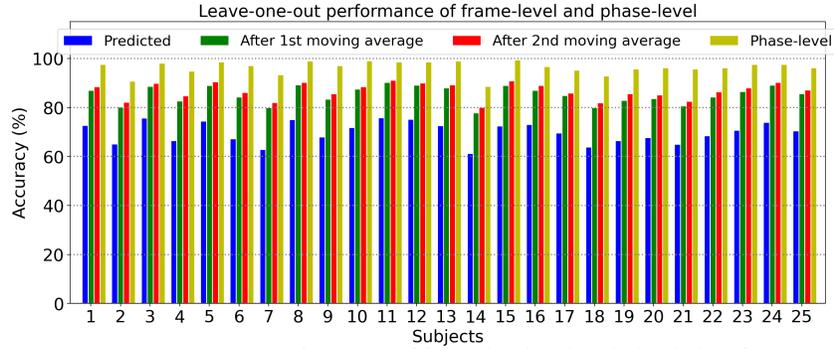


Fig. 11. Leave-one-out performance of frame-level and cycle-level classification.

6.1.3 Impact of Different Features. To substantiate the effectiveness of MFCC for fine-grained breathing mode information extraction, we retrained the breathing mode classification model for each individual with different features including (i) statistical features such as mean, median, root-meansquare, maximum, minimum, 1st and 3rd quartile, interquartile range, standard deviation, skewness, and kurtosis [14]; (ii) Linear-Frequency Cepstral Coefficients (LFCC) [76]; and (iii) Constant-Q Cepstral Coefficients (CQCC) [65]. Specifically, the MFCC employs a Mel scale designed to mimic the human perception of pitch, offering higher resolution at lower frequencies and coarser resolution at higher frequencies. In contrast, CQCC is grounded in the constant-Q scale, closely mirroring the human cochlea's response to various frequencies, as the cochlea is sensitive to frequencies in a non-linear manner. Finally, LFCC utilizes a linear frequency scale, where frequency bins are evenly spaced. This characteristic might be beneficial in specific applications that prioritize linear frequency features. Table 1 presents the average classification accuracy across 25 subjects based on these features at different stages of the pipeline. We can observe that MFCC consistently outperforms other features in different stages, demonstrating the effectiveness of our feature design.

Table 1. Performance comparison of different features at different stages.

| | Statistical | LFCC | CQCC | MFCC |
|--------------|-------------|--------|--------|--------|
| Predicted | 42.91% | 72.75% | 69.94% | 75.59% |
| After 1st MA | 55.73% | 88.57% | 86.73% | 89.49% |
| After 2nd MA | 57.1% | 88.47% | 87.70% | 90.70% |
| Phase-level | 59.35% | 94.20% | 92.79% | 98.52% |

6.1.4 Impact of Different Inhalation and Exhalation Sequence. Due to constraints in terms of time and the participants' physical stamina, we were only able to collect data on two combinations of breathing patterns: oral inhalation and exhalation (OIOE), as well as nasal inhalation and exhalation (NINE). However, there are two additional possible combinations of breathing patterns, namely oral inhalation with nasal exhalation (OINE) and nasal inhalation with oral exhalation (NIOE). The acoustic features of respiratory sounds may be influenced by different combinations of breathing patterns. For example, OIOE can maintain an open mouth position, while OINE may result in repeated opening and closing of the mouth, leading to different oral inhalation sounds.

Table 2. Phase-level performance of different inhalation and exhalation combinations.

| Training | OIOE+NINE | | OINE+NIOE | | OIOE+NINE+OINE+NIOE | |
|----------|-----------|-----------|-----------|-----------|---------------------|-----------|
| Testing | OIOE+NINE | OINE+NIOE | OIOE+NINE | OINE+NIOE | OIOE+NINE | OINE+NIOE |
| S1 | 95.53% | 88.19% | 87.24% | 96.57% | 96.72% | 94.13% |
| S2 | 93.10% | 86.67% | 86.97% | 92.84% | 93.27% | 92.31% |
| S3 | 95.32% | 88.83% | 89.46% | 93.6% | 95.7% | 95.47% |
| S1+S2+S3 | 90.81% | 84.51% | 83.83% | 90.48% | 91.8% | 91.77% |

To examine how different combinations of breathing patterns affect classification performance, we conducted an experiment in which three participants applied all four possible breathing combinations while running. Table 2 presents the classification performance achieved by extracting oral inhalation (OI), oral exhalation (OE), nasal inhalation (NI), and nasal exhalation (NE) from the OIOE and NINE combinations, the OINE and NIOE combinations, as well as all four breathing combinations as separate training data. The performance was then tested on OIOE and NINE, as well as OINE and NIOE as test data. Table 2 shows that respiratory features extracted from only two combinations cannot be directly applied to the other two combinations. The accuracy decreases by more than 6% when classifying data from the untrained combinations. However, such degraded accuracy of >85% is already good enough for real-world deployment. Furthermore, if all four combinations are included in the training data, negligible accuracy loss is presented for all test data, as illustrated in the rightmost column.

6.1.5 Performance with different ambient noise levels. Although earbuds can attenuate some external noise naturally, making in-ear signals more resistant to external noise than out-ear signals, excessive environmental noise may still interfere with respiratory sounds. In order to examine the influence of different levels of environmental noise on respiratory classification performance, we utilized a speaker to play pre-recorded traffic noise at various volumes, simulating a scenario in which a runner is exercising outdoors along a road.

Table 3 presents the performance results for three subjects in both the noisy and noise-reduced conditions. From the first column, it can be observed that despite the use of earplugs to attenuate external noise, the noise still dominates compared to the faint breathing sounds. Consequently, this leads to low accuracy in the initial frame-level predictions, with over half of the classifications being incorrect at higher noisy levels (e.g., 80dB). Additionally, the application of moving average not only fails to correct these errors but may actually increase the number of misclassified frames. Therefore, noise reduction is deemed necessary in such a scenario. By comparing the third (*original*) and fifth (*noise-reduced (individual template)*) column of Table 3, we can observe that our noise-reduction scheme can significantly improve the performance by up to around 40% depending on the noise level.

Our noise-reduction approach relies on an occluded ear canal to obtain the correlation between the in-ear and out-ear signals. However, as mentioned by [28, 33], the shape of the ear canal of each individual is unique and this distinction can even be utilized for authentication purposes. To investigate whether these inter-individual distinctions would impact the noise-reduction effectiveness, we further conducted the following experiments.

Table 3. Phase-level performance with different ambient noise levels.

| | | original | noise-reduced (general template) | noise-reduced (individual template) |
|----|-------|----------|-------------------------------------|--|
| S1 | quiet | 95.53% | NA | NA |
| | 60dB | 79.72% | 83.63% | 91.32% |
| | 70dB | 65.20% | 75.35% | 87.44% |
| | 80dB | 42.13% | 71.81% | 81.94% |
| S2 | quiet | 93.10% | NA | NA |
| | 60dB | 70.80% | 79.82% | 89.18% |
| | 70dB | 57.65% | 73.16% | 88.50% |
| | 80dB | 39.40% | 67.33% | 85.12% |
| S3 | quiet | 95.32% | NA | NA |
| | 60dB | 78.12% | 81.45% | 91.96% |
| | 70dB | 61.41% | 75.49% | 87.42% |
| | 80dB | 51.90% | 68.80% | 85.27% |

Specifically, three subjects are examined under different noise levels using their own personalized templates (previously generated) and a general template (averaged from the templates of the three individuals) for noise reduction. From the fourth (*noise-reduced (general template)*) and fifth (*noise-reduced (individual template)*) column of Table 3, it is evident that using the individualized template for noise reduction significantly improves the accuracy. In detail, under the 60 dB, 70 dB, and 80 dB noise levels, the average accuracies reach 90.82%, 87.79%, and 84.11%, respectively. In contrast, using the general template yields lower accuracies of 81.63%, 74.67%, and 69.31%, respectively. Therefore, we recommend users generate their own personalized noise reduction templates for better breathing classification performance. The overhead of generating the individual template is minimal as it only requires playing a simulated sound of around 10 seconds and performing linear regression.

6.1.6 Performance with different ambient noise types. To evaluate the robustness of our noise reduction scheme in various ambient noise environments, we conducted additional experiments with three individuals. Specifically, each participant wore the prototype and ran while holding a smartphone nearby. The smartphone was playing different types of noise on YouTube including traffic noise, crowd conversation, and popular music from the Billboard Top 50 of 2023, each at a noise level of approximately 70dB. In Table 4, we show both the frame and phase-level accuracy averaged across the three participants. We observe that BreathPro can successfully remove various ambient noises and achieved an overall phase-level accuracy exceeding 85%, in classifying breathing modes. Our noise reduction scheme demonstrates a notable improvement, enhancing phase-level accuracy by more than 20% compared to the case without noise removal.

Table 4. Performance comparison of different ambient noise types at different stages.

| | Traffic | | Crowd | | Music | |
|--------------|----------|---------------|----------|---------------|----------|---------------|
| | original | noise-reduced | original | noise-reduced | original | noise-reduced |
| Predicted | 43.17% | 62.53% | 42.93% | 63.90% | 41.25% | 60.27% |
| After 1st MA | 53.23% | 77.80% | 52.77% | 79.67% | 50.01% | 75.27% |
| After 2nd MA | 55.60% | 79.50% | 55.30% | 81.60% | 52.50% | 77.60% |
| Phase-level | 65.20% | 87.44% | 64.02% | 89.25% | 61.27% | 85.51% |

6.1.7 In-the-wild Test: To evaluate the performance of BreathPro in more realistic running scenarios, including uncontrolled breathing modes and prolonged running times, we conducted in-the-wild experiments with three regular runners in two distinct scenarios. In the first scenario, a single participant wore our prototype and ran freely and continuously for 40 minutes along a city sidewalk (with both hard and rubber surfaces) surrounded by conversations of crowds, traffic sounds, and other environmental noises, at around 60 to 70 dB. The subject jogged at a speed of 10km/h in the first 30 minutes, then started to accelerate at 30-33 minutes, and finally stabilized at 15km/h after 33 minutes. During the first 30 minutes, the subject adopted the breathing mode of NINE due to the moderate running intensity. In the acceleration phase, the subject gradually began to exhale through the mouth (i.e., NIOE) and finally breathed completely through the mouth (i.e., OIOE) at the sprint stage. The subject is a regular runner and is able to maintain the instructed breathing mode at different stages, ensuring the reliability of labels. In the second scenario, another two subjects jogged at a self-selected pace (subject 2: 9.67-11 km/h; subject 3: 7.69-9.27 km/h) and employed any breathing modes they preferred throughout a 30-minute running session. An additional philtrum microphone was attached to assist ground truth annotation, and a running application was utilized for monitoring the running speed.

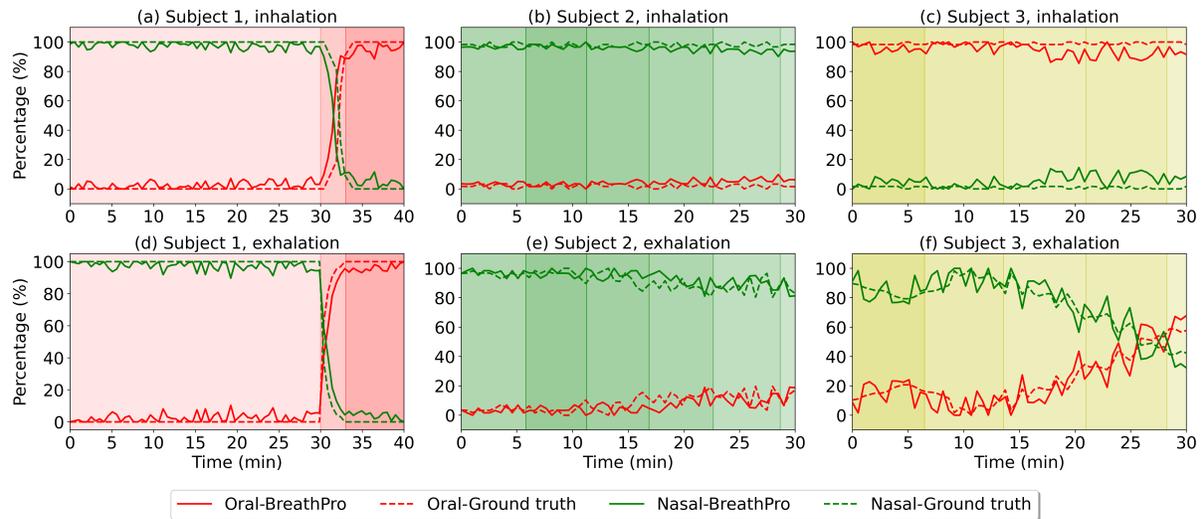


Fig. 12. In-the-wild longitudinal breathing mode monitoring of three subjects. Colored boxes represent different running speeds over time, with a deeper color indicating a higher running speed (subject 1: 10-15 km/h; subject 2: 9.67-11 km/h; subject 3: 7.69-9.27 km/h). The percentages in sub-figures represent the portion of breathing phases classified as oral (red lines) or nasal (green lines) during inhalation (a, b, c) and exhalation (d, e, f) within a 30-second window.

Figure 12 illustrates the breathing mode tracking performance of BreathPro in the in-the-wild longitudinal test of three subjects. Figure 12 (a) and (d) correspond to the first mentioned scenario, and the remaining sub-figures depict the second scenario. The red and green solid lines in the sub-figures represent the percentages of breathing phases classified by BreathPro as oral and nasal within a 30-second window, while the red and green dotted lines indicate the ground truth of the breathing mode. It is evident that BreathPro accurately tracks the transitions of breathing mode in both inhalation and exhalation with the pre-trained model. Notably, compromised tracking performance is observed during the acceleration stage for subject 1 in Figure 12 (a) and (d), and the reason is that the breathing combination of NIOE is not included in the model training, as explained in Section 6.1.4. We believe the performance would be improved once all the breathing combinations are involved during training. Furthermore, while each participant has their preferred breathing style, such as NINE for subject 2 and OINE for

subject 3, it may shift with increased running distance due to the demands of gas exchange or the accumulation of fatigue. In the case of these two subjects, the prediction accuracy in the latter half of the session is lower than in the first half. This could be attributed to two factors: 1) a gradual shift in breathing modes, potentially even involving two types (oral and nasal) of exhalation simultaneously, and 2) as stamina diminishes, maintaining clear, singular breathing patterns close to the training data becomes challenging, leading to degraded classification results. In summary, the average phase-level accuracy of these three subjects is 89.93%, indicating the effectiveness and robustness of BreathPro in realistic scenarios.

6.2 System Performance

Since all the existing earphones are paired with a smartphone using Bluetooth, we implement BreathPro as an Android application and evaluate its system performance based on Xiaomi 13 which is equipped with a Snapdragon 8 Gen 2 processor and a battery with 4500 mAh capacity. The breathing mode pipeline is segmented into five stages. Firstly, the pre-processing stage applies a high-pass filter. The subsequent steps, namely noise detection and noise reduction, are responsible for detecting whether the external noise exceeds a threshold and utilizing pre-generated {in-ear, out-ear} amplitude and phase templates to remove external noise from the in-ear microphone. The feature extraction stage generates MFCC features at the frame level. Lastly, the prediction of breathing mode and the double-moving average constitute the post-processing stage.

We evaluate the CPU load and latency by executing a specific portion of the code looping 1000 times, repeating this process five times, and calculating the average value of the five iterations. To assess battery usage, we measure the power consumption of a specific code segment by running it in the background for one hour while the screen is turned off. We then subtract the power consumption during this period from the power consumption during one hour of idle standby, yielding the power consumption specifically attributed to the code segment. Table 5 presents the system performance of BreathPro based on the processing of a one-second data. We can observe that the proposed breathing mode recognition pipeline can be completed within 84.44 ms respectively, guaranteeing real-time detection. The CPU load is low at around 5%. In terms of power consumption, we convert the measurements to battery usage (%/hour), and our results indicate that the task consumes about 1.1% battery capacity separately when processing one-hour data, which is similar to typical smartphone applications such as music player (2%/hour). All the results reveal the lightweight design of our classification pipeline, ensuring that the system does not burden the mobile device and can operate for extended periods without draining excessive battery power.

Table 5. System performance of *RunCoach*, executed on Xiaomi 13 (battery capacity of 4500mAh) to process one-second data.

| | Pre-processing | Noise Detection | Noise Reduction | Feature Extraction | Post-processing | Overall |
|-----------|------------------------|-----------------|-----------------|--------------------|-----------------|---------|
| Breathing | CPU (%) | 3.3 | 3.7 | 5.6 | 6.2 | - |
| | Latency (ms) | 0.34 | 0.19 | 64.12 | 17.22 | 84.44 |
| | Energy (mAh) | 0.0001 | 0.0001 | 0.0321 | 0.0013 | 0.0444 |
| | Battery Usage (%/hour) | 3 | 3 | 4 | 5 | 1.13 |

7 DISCUSSION AND FUTURE WORK

Generalizability to other scenarios: The fundamental idea of BreathPro is to identify breathing modes using the in-ear microphone on earables. Although we showcased its feasibility and performance during running, the concept and effectiveness can be extended to other vigorous-intensity exercises that generate distinct breathing

sounds, such as cycling, racewalking, hiking uphill, etc. It is crucial to note that if an activity generates slow and weak breathing sounds, the performance of our approach may be considerably compromised and may even fail.

Breathing rate estimation: As shown in Figure 8, after applying the second moving average, we can clearly identify the start, end, and duration of each breathing phase. By counting the number of breaths within a specified period of time, the respiration rate can also be derived. Note that two breathing phases constitute one breathing cycle. Thus, we compute the respiration rate for each participant, obtaining an average mean absolute error of 1.88 breath per minute (BPM), which is comparable with existing approaches [9]. The result demonstrates the feasibility and superior performance of deriving breathing rate based on breathing mode estimation.

Impact of running speed: Different individuals or the same runner at different running stages can run at different speeds, and the running speed is often positively correlated with the breathing rate. As shown in Figure 12, running speed has minimal impact on our breathing mode detection, which is attributed to the frame-level classification of BreathPro. More specifically, the breathing speed only affects the number of frames in a single phase (more frames at lower speeds, and vice versa), while the phase-level accuracy relies on a majority vote of all the frames. Even at a high breathing rate of one cycle per second, there are nine frames within a phase, allowing spontaneous classification errors to be easily corrected.

Impact of music playback: While earphones are commonly used during running, their primary function is music playback. This raises concerns about simultaneous music playback and breathing mode detection, as the captured breathing signals may be overwhelmed by the music. One potential solution is to employ source separation algorithms to extract the breathing sounds from the distorted signal, given that the delivered sound is known from the earphone system. However, this approach requires careful consideration of the frequency responses of the speaker and microphones, necessitating further engineering exploration such as automatic gain control. Additionally, in the presence of environmental noise, the impact of music on the in-ear and out-ear microphones may differ, warranting further investigation into their correlation.

Development of a user-invariant model. Because the shape of the ear canal, bone/tissue structure, and breathing sounds of different users vary, the frame-level and phase-level accuracy of the leave-one-out test (i.e., user-invariant model) drops 6% and 2% respectively, compared to those of individual models. In the current BreathPro, we only extract basic MFCC features and classify breathing modes with a KNN model. In the future, it would be promising to develop a more accurate user-invariant model by augmenting the collected breathing data and training a deep neural network for classification.

Feedback delivery methods: We delve into the future prospects of our system, particularly focusing on the user-friendly delivery of feedback regarding breathing modes during running. We are considering innovative approaches that prioritize the convenience and preferences of runners. One such approach is the provision of post-run summaries, which offer comprehensive reports after each running session. These summaries allow runners to analyze their breathing patterns in detail, without any interruption during their activity. Another promising direction is the development of customizable notifications. This feature would empower users to personalize how and when they receive feedback, thereby minimizing any potential disturbance to their running experience. These potential delivery methods represent key areas of our future work, aiming to enhance the practicality and effectiveness of our system for runners.

Tidal volume estimation. Tidal volume refers to the amount of air that is inhaled or exhaled during a breathing cycle. It is of great importance in analyzing the pulmonary function, which would in turn be used to understand the running efficiency and refine the breathing strategy during running. However, tidal volume is usually measured with a special equipment at controlled conditions and accurately measuring it in the wild seems to be impossible. Based on the observation that breathing sounds show varying characteristics, such as intensity, frequency, and duration, at different breathing modes, an interesting future direction is to estimate the tidal volume using in-ear breathing signals.

Transplantation to commercial earbuds. We developed our own prototype because commercial manufacturers typically do not release APIs for accessing the raw in-ear signal. The customized prototype does not contain any special signal processing circuit, while simply connecting a microphone to the audio codec for recording. Thus, we believe the fundamental concept of BreathPro is likely to be applied on commercial earbuds. Moreover, our device is wired, which may introduce additional noise during running (e.g., rubbing with clothes or human skin), implying a potential performance improvement on wireless earbuds. However, different earbuds may feature microphones with varied physical layouts, diverse frequency responses and amplification ratio, potentially influencing our noise reduction scheme. Therefore, we recognize that it might require more investigation before deploying our system on commercial earbuds.

8 CONCLUSION

In this paper, we present a novel earbuds-based system, named BreathPro, for breathing mode monitoring during running. Employing a microphone to record the in-ear sounds, BreathPro first extracts signals related to human breathing with a high-pass filter. Then, BreathPro adopts a well-designed signal processing pipeline and ML-based classification scheme to recognize runners' breathing modes. In addition, BreathPro exploits the out-ear microphone to mitigate the noise in the breathing sounds captured by the in-ear microphone, thereby improving the performance and robustness of breathing mode detection. With data collected from 25 subjects, we experimentally demonstrated the superior performance of BreathPro, namely, 98.52% phase-level accuracy for breathing mode recognition. We also implemented BreathPro as a smartphone application and demonstrated its lightweight property with power and latency measurement.

9 ACKNOWLEDGMENTS

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

REFERENCES

- [1] Online. AirPods Pro (2nd generation). <https://www.apple.com/sg/airpods-pro/>.
- [2] Online. Apple Watch. <https://www.apple.com/watch/>.
- [3] Online. Garmin. <https://support.garmin.com/en-GB/?faq=>.
- [4] Online. Huawei Freebuds 5. <https://consumer.huawei.com/en/headphones/freebuds5/>.
- [5] Online. Number of running and jogging participants in the United States from 2010 to 2021. <https://www.statista.com/statistics/190303/running-participants-in-the-us-since-2006/>.
- [6] Online. Polar H10 Heart Rate Sensor. <https://www.polar.com/sg-en/sensors/h10-heart-rate-sensor>.
- [7] Online. Zephyr HxM Heart rate Monitor. <https://www.zephyranywhere.com/system/hxm>.
- [8] Zrar Kh Abdul and Abdulbasit K Al-Talabani. 2022. Mel Frequency Cepstral Coefficient and its applications: A Review. *IEEE Access* (2022).
- [9] Mohamed Ali, Ali Elsayed, Arnaldo Mendez, Yvon Savaria, and Mohamad Sawan. 2021. Contact and remote breathing rate monitoring techniques: A review. *IEEE Sensors Journal* 21, 13 (2021), 14569–14586.
- [10] Mohanad Alkhodari and Ahsan H Khandoker. 2022. Detection of COVID-19 in smartphone-based breathing recordings: A pre-screening deep learning tool. *PLoS one* 17, 1 (2022), e0262448.
- [11] Moustafa Alzantot and Moustafa Youssef. 2012. UPTIME: Ubiquitous pedestrian tracking using mobile phones. In *2012 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 3204–3209.
- [12] Kyle R Barnes and Andrew E Kilding. 2015. Running economy: measurement, norms, and determining factors. *Sports medicine-open* 1, 1 (2015), 1–15.
- [13] Mattia Bertschi, Patrick Celka, Ricard Delgado-Gonzalo, Mathieu Lemay, Enric M Calvo, Olivier Grossenbacher, and Ph Renevey. 2015. Accurate walking and running speed estimation using wrist inertial data. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 8083–8086.
- [14] Chloë Brown, Jagmohan Chauhan, Andreas Grammenos, Jing Han, Apinan Hasthanasombat, Dimitris Spathis, Tong Xia, Pietro Cicuta, and Cecilia Mascolo. 2020. Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. *arXiv preprint*

- arXiv:2006.05919* (2020).
- [15] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, and Cecilia Mascolo. 2021. hEARt: Motion-resilient Heart Rate Monitoring with In-ear Microphones. *arXiv preprint arXiv:2108.09393* (2021).
 - [16] Inge Bylemans, Maarten Weyn, and Martin Klepal. 2009. Mobile phone-based displacement estimation for opportunistic localisation systems. In *2009 Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*. IEEE, 113–118.
 - [17] Yetong Cao, Chao Cai, Fan Li, Zhe Chen, and Jun Luo. 2023. HeartPrint: Passive Heart Sounds Authentication Exploiting In-Ear Microphones. *Heart* 50, S1 (2023), S2.
 - [18] Kévin Carillo, Olivier Doutres, and Franck Sgard. 2020. 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* 147, 5 (2020), 3476–3489.
 - [19] Peter H Charlton, Drew A Birrenkott, Timothy Bonnici, Marco AF Pimentel, Alistair EW Johnson, Jordi Alastruey, Lionel Tarassenko, Peter J Watkinson, Richard Beale, and David A Clifton. 2017. Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review. *IEEE reviews in biomedical engineering* 11 (2017), 2–20.
 - [20] Kenneth Christofferson, Xuyang Chen, Zeyu Wang, Alex Mariakakis, and Yuntao Wang. 2022. Sleep Sound Classification Using ANC-Enabled Earbuds. In *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 397–402.
 - [21] Michael Chu, Thao Nguyen, Vaibhav Pandey, Yongxiao Zhou, Hoang N Pham, Ronen Bar-Yoseph, Shlomit Radom-Aizik, Ramesh Jain, Dan M Cooper, and Michelle Khine. 2019. Respiration rate and volume measurements using wearable strain sensors. *NPJ digital medicine* 2, 1 (2019), 8.
 - [22] Julius H Comroe. 1965. Physiology of respiration. *Academic Medicine* 40, 9 (1965), 887.
 - [23] George Dallam, Bethany Kies, et al. 2020. The effect of nasal breathing versus oral and oronasal breathing during exercise: a review. *J. Sports Res* 7 (2020), 1–10.
 - [24] Luis Enrique Díez, Alfonso Bahillo, Jon Otegui, and Timothy Otim. 2018. Step length estimation methods based on inertial sensors: A review. *IEEE Sensors Journal* 18, 17 (2018), 6908–6926.
 - [25] David Elad, Michael Wolf, and Tilman Keck. 2008. Air-conditioning in the human nasal cavity. *Respiratory physiology & neurobiology* 163, 1-3 (2008), 121–127.
 - [26] LB Evans and LC Sutherland. 1970. Absorption of sound in air. *Wyle Labs Research Staff Report WR* (1970), 70–14.
 - [27] Andrea Ferlini, Dong Ma, Robert Harle, and Cecilia Mascolo. 2021. EarGate: gait-based user identification with in-ear microphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 337–349.
 - [28] Yang Gao, Wei Wang, Vir V Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using ear canal echo for wearable authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.
 - [29] Roberta L Hall. 2005. Energetics of nose and mouth breathing, body size, body composition, and nose volume in young adult males and females. *American Journal of Human Biology: The Official Journal of the Human Biology Association* 17, 3 (2005), 321–330.
 - [30] Feiyu Han, Panlong Yang, Shaojie Yan, Haohua Du, and Yuanhao Feng. 2023. Breathsing: Transparent and continuous in-ear authentication using bone-conducted breathing biometrics. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE, 1–10.
 - [31] Tian Hao, Guoliang Xing, and Gang Zhou. 2015. RunBuddy: a smartphone system for running rhythm monitoring. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 133–144.
 - [32] Mahmoud Hassan, Florian Daiber, Frederik Wiehr, Felix Kosmalla, and Antonio Krüger. 2017. Footstriker: An EMS-based foot strike assistant for running. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 1 (2017), 1–18.
 - [33] Changshuo Hu, Xiao Ma, Dong Ma, and Ting Dang. 2023. Lightweight and Non-Invasive User Authentication on Earables. In *Proceedings of the 24th International Workshop on Mobile Computing Systems and Applications*. 36–41.
 - [34] Meng-Han Hu, Guang-Tao Zhai, Duo Li, Ye-Zhao Fan, Xiao-Hui Chen, and Xiao-Kang Yang. 2017. Synergetic use of thermal and visible imaging techniques for contactless and unobtrusive breathing measurement. *Journal of biomedical optics* 22, 3 (2017), 036006–036006.
 - [35] Mark Janssen, Ruben Walravens, Erik Thibaut, Jeroen Scheerder, Aarnout Brombacher, and Steven Vos. 2020. Understanding different types of recreational runners and how they use running-related technology. *International Journal of Environmental Research and Public Health* 17, 7 (2020), 2276.
 - [36] David Jiménez-Pavón, Ana Carbonell-Baeza, and Carl J Lavie. 2020. Physical exercise as therapy to fight against the mental and physical consequences of COVID-19 quarantine: Special focus in older people. *Progress in cardiovascular diseases* 63, 3 (2020), 386.
 - [37] Yincheng Jin, Yang Gao, Xiaotao Guo, Jun Wen, Zhengxiong Li, and Zhanpeng Jin. 2022. EarHealth: an earphone-based acoustic otoscope for detection of multiple ear diseases in daily life. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*. 397–408.
 - [38] Armağan Karahanoğlu, Rúben Gouveia, Jasper Reenalda, and Geke Ludden. 2021. How are sports-trackers used by runners? Running-related data, personal goals, and self-tracking in running. *Sensors* 21, 11 (2021), 3687.
 - [39] Agni Kumar, Vikramjit Mitra, Carolyn Oliver, Adeeti Ullal, Matt Biddulph, and Irida Mance. 2021. Estimating respiratory rate from breath audio obtained through wearable microphones. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine*

- & *Biology Society (EMBC)*. IEEE, 7310–7315.
- [40] Chase O LaComb, Richard D Tandy, Szu Ping Lee, John C Young, and James W Navalta. 2017. Oral versus nasal breathing during moderate to high intensity submaximal aerobic exercise. *International Journal of Kinesiology and Sports Science* 5, 1 (2017), 8–16.
- [41] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. 2015. Tracking vital signs during sleep leveraging off-the-shelf wifi. In *Proceedings of the 16th ACM international symposium on mobile ad hoc networking and computing*. 267–276.
- [42] Yang Liu, Yonghang Jiang, Zhenjiang Li, and Jianping Wang. 2019. Rulers on Our Arms: Waving to Measure Object Size through Contactless Sensing. *ACM Transactions on Sensor Networks (TOSN)* 15, 1 (2019), 1–25.
- [43] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. 2021. OESense: employing occlusion effect for in-ear human sensing. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*. 175–187.
- [44] 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.
- [45] Carlo Massaroni, Daniela Lo Presti, Domenico Formica, Sergio Silvestri, and Emiliano Schena. 2019. Non-contact monitoring of breathing pattern and respiratory rate via RGB signal measurement. *Sensors* 19, 12 (2019), 2758.
- [46] Elvira Mendoza, Nieves Valencia, Juana Muñoz, and Humberto Trujillo. 1996. Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS). *Journal of voice* 10, 1 (1996), 59–66.
- [47] AR Morton, K King, S Papalia, Carmel Goodman, KR Turley, and JH Wilmore. 1995. Comparison of maximal oxygen consumption with oral and nasal breathing. *Australian journal of science and medicine in sport* 27, 3 (1995), 51–55.
- [48] Krishna Mridha, Shakil Sarkar, and Dinesh Kumar. 2021. Respiratory disease classification by CNN using MFCC. In *2021 IEEE 6th International Conference on Computing, Communication and Automation (ICCCA)*. IEEE, 517–523.
- [49] 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.
- [50] Shahriar Nirjon, Jie Liu, Gerald DeJean, Bodhi Priyantha, Yuzhe Jin, and Ted Hart. 2014. COIN-GPS: Indoor localization from direct GPS receiving. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. 301–314.
- [51] Jay Prakash, Zhijian Yang, Yu-Lin Wei, Haitham Hassanieh, and Romit Roy Choudhury. 2020. EarSense: earphones as a teeth activity sensor. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–13.
- [52] Donald F Proctor and Ib Harald Peder Andersen. 1982. *The nose: upper airway physiology and the atmospheric environment*. Elsevier Biomedical Press Amsterdam.
- [53] Ariana Tulus Purnomo, Ding-Bing Lin, Tjahjo Adiprabowo, and Willy Fitra Hendria. 2021. Non-contact monitoring and classification of breathing pattern for the supervision of people infected by COVID-19. *Sensors* 21, 9 (2021), 3172.
- [54] Md Mahbubur Rahman, Tousif Ahmed, Mohsin Yusuf Ahmed, Minh Dinh, Ebrahim Nemati, Jilong Kuang, and Jun Alex Gao. 2022. Breathe buddy: Tracking real-time breathing exercises for automated biofeedback using commodity earbuds. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (2022), 1–18.
- [55] Md Mahbubur Rahman, Tousif Ahmed, Mohsin Yusuf Ahmed, Ebrahim Nemati, Minh Dinh, Nathan Folkman, Md Mehedi Hasan, Jilong Kuang, and Jun Alex Gao. 2021. Towards motion-aware passive resting respiratory rate monitoring using earbuds. In *2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 1–4.
- [56] Yanzhi Ren, Chen Wang, Jie Yang, and Yingying Chen. 2015. Fine-grained sleep monitoring: Hearing your breathing with smartphones. In *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 1194–1202.
- [57] Tobias Röddiger, Daniel Wolfram, David Laubenstein, Matthias Budde, and Michael Beigl. 2019. Towards respiration rate monitoring using an in-ear headphone inertial measurement unit. In *Proceedings of the 1st International Workshop on Earable Computing*. 48–53.
- [58] Melvyn Roerdink, Claudine JC Lamothe, Peter J Beek, et al. 2008. Online gait event detection using a large force platform embedded in a treadmill. *Journal of biomechanics* 41, 12 (2008), 2628–2632.
- [59] Chiara Romano, Andrea Nicolò, Lorenzo Innocenti, Marco Bravi, Sandra Miccinilli, Silvia Sterzi, Massimo Sacchetti, Emiliano Schena, and Carlo Massaroni. 2023. Respiratory Rate Estimation during Walking and Running Using Breathing Sounds Recorded with a Microphone. *Biosensors* 13, 6 (2023), 637.
- [60] Nirupam Roy, He Wang, and Romit Roy Choudhury. 2014. I am a smartphone and i can tell my user’s walking direction. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. 329–342.
- [61] Edward P Sarafino and Timothy W Smith. 2014. *Health psychology: Biopsychosocial interactions*. John Wiley & Sons.
- [62] Chen-Hsuan Shih, Naofumi Tomita, Yanick X Lukic, Álvaro Hernández Reguera, Elgar Fleisch, and Tobias Kowatsch. 2019. Breeze: Smartphone-based acoustic real-time detection of breathing phases for a gamified biofeedback breathing training. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 4 (2019), 1–30.
- [63] 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.
- [64] Xue Sun, Jie Xiong, Chao Feng, Wenwen Deng, Xudong Wei, Dingyi Fang, and Xiaojiang Chen. 2023. Earmonitor: In-ear Motion-resilient Acoustic Sensing Using Commodity Earphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–22.

- [65] Massimiliano Todisco, Héctor Delgado, and Nicholas Evans. 2017. Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification. *Computer Speech & Language* 45 (2017), 516–535.
- [66] Hoang Truong, Alessandro Montanari, and Fahim Kawsar. 2022. Non-invasive blood pressure monitoring with multi-modal in-ear sensing. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6–10.
- [67] Gijsbertus Jacob Verkerke, AL Hof, W Zijlstra, W Ament, and G Rakhorst. 2005. Determining the centre of pressure during walking and running using an instrumented treadmill. *Journal of biomechanics* 38, 9 (2005), 1881–1885.
- [68] Hong-Son Vu, Kuan-Hung Chen, and Tien-Mau Fong. 2015. Active noise control for in-ear headphones: Implementation and evaluation. In *2015 IEEE International Conference on Consumer Electronics-Taiwan*. IEEE, 264–265.
- [69] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW based contactless respiration detection using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–20.
- [70] Zi Wang, Yili Ren, Yingying Chen, and Jie Yang. 2022. Toothsonic: Earable authentication via acoustic toothprint. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–24.
- [71] Paweł W Woźniak, Monika Zbytniewska, Francisco Kiss, and Jasmin Niess. 2021. Making Sense of Complex Running Metrics Using a Modified Running Shoe. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–11.
- [72] Yadong Xie, Fan Li, Yue Wu, Huijie Chen, Zhiyuan Zhao, and Yu Wang. 2022. TeethPass: Dental Occlusion-based User Authentication via In-ear Acoustic Sensing. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 1789–1798.
- [73] Yanni Yang, Jiannong Cao, and Xiulong Liu. 2019. ER-rhythm: Coupling exercise and respiration rhythm using lightweight COTS RFID. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–24.
- [74] Xiaoyi Zhang, Ming-Chun Huang, Fengbo Ren, Wenyao Xu, Nan Guan, and Wang Yi. 2013. Proper running posture guide: a wearable biomechanics capture system. In *Proceedings of the 8th International Conference on Body Area Networks*. 83–89.
- [75] Hongyu Zhao, Zhelong Wang, Sen Qiu, Jiabin Wang, Fang Xu, Zhengyu Wang, and Yanming Shen. 2019. Adaptive gait detection based on foot-mounted inertial sensors and multi-sensor fusion. *Information Fusion* 52 (2019), 157–166.
- [76] Xinhui Zhou, Daniel Garcia-Romero, Ramani Duraiswami, Carol Espy-Wilson, and Shihab Shamma. 2011. Linear versus mel frequency cepstral coefficients for speaker recognition. In *2011 IEEE workshop on automatic speech recognition & understanding*. IEEE, 559–564.