

TITLE OF THE INVENTION: Method And System Of Identifying Breathing Mode And Gait

REFERENCE TO RELATED APPLICATIONS

- 5 [0001] The present application claims priority to Singapore patent application number 10202400691T with a filing date of 13 March 2024 and titled “RunCoach: Earbud-assisted Running System for Gait and Breathing Monitoring” and is incorporated herein by reference.

10 FIELD OF THE INVENTION

[0002] The present application relates to computer implemented methods identifying a breathing mode and gait of a user.

BACKGROUND OF THE INVENTION

- 15 [0003] 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. 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
- 20 specialized equipment or facilities, running only requires a pair of running shoes and a secure 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. Prioritizing safety and efficiency during running is thus crucial to minimize the risk of injuries and
- 25 maximize performance potential. However, a majority of runners, particularly those untrained, often lack the knowledge to control or manage their running parameters, such as pace, cadence, intensity, heart rate (HR), respiration, and so on. Hence, it is important to automatically monitor various running parameters and provide real-time feedback to runners while running to enhance the running experience, reduce
- 30 the risk of injuries, and improve overall efficiency and performance.

[0004] Existing studies have primarily measured the following three kinds of factors while running to enable automatic monitoring and provide prompt feedback to runners: 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, e.g., inertial measurement units (IMUs) and the Global Positioning System (GPS). These solutions are reliable and portable, leading to successful and widespread development in commercial smartphones and smartwatches; 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, sensors in shoes, and force platforms on grounds. These approaches are either bulky to users resulting in lower acceptance, or costly to implement with limited monitoring coverage; 3) Physiology-related factors such as HR and respiration rate (RR), can be monitored with chest-worn straps for HR or/and RR monitoring. Although these methods are effective, they are invasive in nature. To this end, hEARt (Butkow et al., 2021, arXiv preprint arXiv:2108.09393) proposes to monitor HR while running using in-ear microphone audio from earables. RunBuddy (Hao et al., Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 133–144) employs both smartphones and headphones to monitor Locomotor Respiratory Coupling (LRC) during running, which can then be utilised for RR monitoring. ER-Rhythm (Yang et al., 2019, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies ,3, 4 (2019), 1–24) also monitors LRC using RFID while running on a treadmill.

[0005] Although previous studies have achieved remarkable progress in running assistance through monitoring various factors, certain areas remain untapped and warrant further investigation. Primarily, there exists a demand for a portable and lightweight solution to measure factors related to foot strikes. Furthermore, current studies have yet to reveal fine-grained physiological factors that could provide valuable insights. For instance, in addition to the breathing rate, how the runner is breathing (i.e., using mouth or nose) is highly correlated with the volume of oxygen exchange and therefore affects the running performance and efficiency.

SUMMARY OF THE INVENTION

[0006] In a first aspect, there is provided a computer implemented method of identifying a breathing mode of a user. The method comprises obtaining an in-ear signal data of a user; filtering the in-ear signal data to obtain a breathing signal data; segmenting the breathing signal data into a plurality of frames with a window sliding algorithm; extracting a plurality of features from each frame in the plurality of frames; feeding the plurality of features from each frame to a trained machine learning model; obtaining a breathing mode estimation for each frame from the trained machine learning model; applying a moving average algorithm to the breathing mode estimation for the plurality of frames to correct an error if present in the breathing mode estimation to provide a plurality of corrected frames; and identifying a breathing mode for each frame of the plurality of corrected frames. If there are no errors in the frame, the breathing mode will be the same in the corrected frame and uncorrected frame (i.e. before applying the moving average algorithm).

[0007] Preferably, the method further comprises obtaining an out-ear signal data; determining from the out-ear signal data if an ambient noise level is above a threshold level; if the ambient noise level is at or above the threshold level, estimating a residual noise level from the out-ear signal data and subtracting the residual noise level from the breathing signal data to denoise the breathing signal data before segmenting the breathing signal data; and if the ambient noise level is below the threshold level, proceeding to segmenting the breathing signal data.

[0008] Preferably, estimating the residual noise level from the out-ear signal data comprises obtaining a plurality of amplitude and phase values of the out-ear signal data at different frequency bands; obtaining a corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands; computing a transformed plurality of amplitude and phase values of the out-ear signal data at different frequency bands based on a linear correlation between the corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands and the plurality of amplitude and phase values of the out-ear signal data at different frequency bands; and subtracting the residual noise level

from the breathing signal data comprises subtracting the transformed plurality of amplitude and phase values of the out-ear signal data at different frequency bands from the corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands; and converting the subtracted plurality of amplitude and phase values to the breathing signal data to denoise the breathing signal data.

[0009] Preferably, filtering the in-ear signal data comprises applying a high pass filter to exclude signals of less than 500 Hz.

[0010] Preferably, each frame in the plurality of frames comprises a fixed length, and preferably there is no overlap of frames in the sliding window algorithm.

[0011] In an embodiment, a machine learning model to be trained is selected from K-Nearest Neighbours (KNN), SVM (Support Vector Machine), RF (Random Forest), and DT (Decision Tree). Preferably the machine learning model is K-Nearest Neighbours (KNN).

[0012] In an embodiment, the training method of the trained machine learning model comprises segmenting a training in-ear signal data with reference to a reference breathing signal to provide a plurality of segmented training in-ear signal data; splitting each of the plurality of segmented training in-ear signal data into a plurality of training frames; extracting a plurality of features from each training frame in the plurality of training frames; applying a high-quality frame selection algorithm to the plurality of training frames to select high quality training frames in the plurality of training frames; and inputting the selected high quality training frames into the machine learning model. Preferably, the reference breathing signal indicates transitions between adjacent breathing phases of inhalation and exhalation and is preferably obtained from an external microphone, more preferably the external microphone is placed proximal to a nose.

[0013] Preferably, the high-quality frame selection algorithm comprises computing an average of the plurality of features from the plurality of training frames in each segmented training in-ear signal data; generating a feature template from the computed average of the plurality of features from the plurality of training frames; determining the Minkowski distance between the plurality of features from each

training frame and the feature template; and selecting high-quality frames based on the Minkowski distance.

[0014] Preferably, the training method comprises collecting the training in-ear signal data at least in part from the user.

5 [0015] Preferably, applying the moving average algorithm to the breathing mode estimation is performed twice.

[0016] Preferably, the plurality of features comprises Mel-frequency cepstral coefficients (MFCC) features.

[0017] Preferably, obtaining a breathing mode estimation for each frame from the
10 trained machine learning model comprises classifying the plurality of features from each frame to determine the breathing mode estimation.

[0018] In an embodiment, the trained machine learning model is remotely located. In an embodiment, the breathing mode estimation and/or the breathing mode is selected from (the group consisting of) oral inhalation, oral exhalation, nasal
15 inhalation, and nasal exhalation.

[0019] Preferably, the method further comprises determining a respiration rate.

[0020] In a second aspect, there is provided a computer implemented method of identifying a gait of a user. The method comprises obtaining an in-ear signal data of a user; filtering the in-ear signal data to obtain a gait signal data; segmenting the
20 gait signal data with an envelope-based algorithm to obtain a segmented peak; extracting a plurality of features in the segmented peak; feeding the plurality of features from the segmented peak to a trained machine learning gait model; and obtaining a gait of the user for the segmented peak from the trained machine learning model. The segmented peak may be used to identify a foot step of the user
25 and optionally determine cadence of the user. The gait ground truth is determined based on the instructed gait type for each session, assuming that the user consistently follows the instructed gait.

[0021] Preferably, filtering the in-ear signal data comprises applying a low pass filter to exclude signals of more than 100 Hz.

30 [0022] Preferably, the envelope-based algorithm comprises applying a Hibbert transform to generate an envelope signal of the gait signal; smoothing the envelope

signal; detecting a peak and its adjacent region of the envelope signal to obtain the segmented peak.

[0023] Preferably, extracting the plurality of features in the segmented peak comprises performing a Fast Fourier transform (FFT) method or a Mel-frequency cepstral coefficients (MFCC) method, preferably performing the FFT method.

[0024] In an embodiment, a machine learning gait model to be trained is selected from K-Nearest Neighbours (KNN), SVM (Support Vector Machine), RF (Random Forest), and DT (Decision Tree). Preferably, the machine learning gait model is KNN.

10 [0025] Preferably, segmenting the gait signal data with an envelope-based algorithm comprises obtaining a plurality of segmented peaks, and the method further comprises generating a count of the plurality of segmented peaks. Advantageously, this allows the step count or cadence to be determined.

[0026] Preferably, obtaining a gait of the user for the segmented peak from the
15 trained machine learning model comprises classifying the plurality of features from the segmented peak to obtain the gait of the user.

[0027] In an embodiment, the trained machine learning model is remotely located.

[0028] In an embodiment, the gait is selected from a forefoot strike, an overstride,
20 and a heel strike.

[0029] In a third aspect, there is provided a computer implemented method of identifying a cadence of a user. The method comprises obtaining an in-ear signal data of a user; filtering the in-ear signal data to obtain a gait signal data; segmenting the gait signal data with an envelope-based algorithm to obtain a segmented peak;
25 and determining the cadence of the user. Preferably, segmenting the gait signal data with an envelope-based algorithm comprises obtaining a plurality of segmented peaks, and the method further comprises generating a count of the plurality of segmented peaks.

[0030] In a fourth aspect, there is provided a method of cancelling noise in an in-ear microphone. The method comprises obtaining an in-ear signal data of a user;
30 obtaining an out-ear signal data; determining from the out-ear signal data if an

ambient noise level is above a threshold level; if the ambient noise level is at or above the threshold level, estimating a residual noise level from the out-ear signal data; and subtracting the residual noise level from the in-ear signal data to denoise the in-ear signal data; and if the ambient noise level is below the threshold level,
5 determine that the in-ear signal data need not be denoised.

[0031] Preferably, estimating the residual noise level from the out-ear signal data comprises obtaining a plurality of amplitude and phase values of the out-ear signal data at different frequency bands; obtaining a corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands; computing
10 a transformed plurality of amplitude and phase values of the out-ear signal data at different frequency bands based on a linear correlation between the corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands and the plurality of amplitude and phase values of the out-ear signal data at different frequency bands; and subtracting the residual noise level
15 from the in-ear signal data comprises subtracting the transformed plurality of amplitude and phase values of the out-ear signal data at different frequency bands from the corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands; and converting the subtracted plurality of amplitude and phase values to the in-ear signal data to denoise the in-ear signal
20 data.

The linear correlation between the amplitude and phase values of the in-ear and out-ear signal at different frequency bands can be determined by playing a set of noise signals within a frequency range and analyzing the corresponding signal responses. Experimental analysis has shown that different types of ambient noise
25 do not affect the computed linear correlation values, indicating the robustness of the correlation across various noise environments. Different users exhibit unique linear correlation profiles. A general linear correlation template can be applied across users, obtaining a user-specific linear correlation profile can further enhance noise cancellation performance.

[0032] In an embodiment, a data processing apparatus comprises means for carrying out the method according to any one of the first aspect, second aspect,

third aspect, or fourth aspect. In an embodiment, a system comprises a pair of earbuds; and the data processing apparatus.

[0033] In an embodiment, a computer program or computer-readable medium comprises instructions which, when the program is executed by a computer, cause the computer to carry out the method according to any one of the first aspect, second aspect, third aspect, or fourth aspect.

[0034] Advantageously, the methods and systems described allows a user to monitor their breathing and gait while running allowing the use to optimize their running performance and efficiency.

BRIEF DESCRIPTION OF THE DRAWINGS

[0035] Figure (FIG.) 1 shows an illustration of three gaits and close up of their foot.

[0036] FIG. 2 shows In-ear (upper row) and Out-ear (lower row) signals of three running gaits and a noisy Heel Strike gait in the time domain. In-ear signals exhibit a more pronounced waveform and clearer inter-class differences compared to out-ear signals.

[0037] FIG. 3 shows in-ear and out-ear signals of two breathing phase combinations (panels a, b, d, e) and a noisy NINE combination (panels c, f) in the frequency domain. The breathing signals captured by in-ear microphones are more pronounced, and more resilient to ambient noise. The panel of each may be also be termed as the FIG. number and panel alphabet, for example FIG. 3a, FIG. 3b.

[0038] FIG. 4 shows the methodological pipeline of RunCoach.

[0039] FIG. 5 shows the envelope-based cycle segmentation of gait signal. Panel (a) shows the raw heel strike signal, panel (b) shows the low-pass filtered heel strike signal, panel (c) shows the upper envelopes of low-pass filtered heel strike signal, and panel (d) shows the peaks of upper envelopes and 200 ms samples indicating the initial contacts.

[0040] FIG. 6 shows the frequency-based breathing motion segmentation of the ground truth breathing signal collected by wired earpods. Panel (a) shows the raw ground truth breathing signal, panel (b) shows the high-pass filtered ground truth

breathing signal, panel (c) shows the spectrogram of ground truth breathing signal, and panel (d) shows the energy curve derived from the spectrogram and its valleys indicating transition intervals.

[0041] FIG. 7 shows the time domain and distance matrix between the template and each frame of a clean breathing phase (panels (a) and (b)) and a breathing phase with a sudden noise (panels (c) and (d)). Frames with a distance greater than 50 from the template will be excluded.

[0042] FIG. 8 shows the scatter plot of (a) amplitudes and (b) phases for {out-ear, in-ear} pairs at 10 different frequency bands.

[0043] FIG. 9 shows the flowchart of noise reduction module, including offline coefficients learning (right part) and online noise reduction (left part).

[0044] FIG. 10 shows the noise reduction performance at different stages.

[0045] FIG. 11 shows the breathing mode of (a) predicted by RunCoach, (b) after first moving average, (c) after second moving average, (d) after low-pass filter, and (e) the ground truth.

[0046] FIG. 12 shows in panel (a) the customized earbuds prototype and accompanying data recording device, panel (b) picture of a participant wearing the device, and panel (c) flowcharts of gait and breathing data collection.

[0047] FIG. 13 shows in panel (a) the overall gait type recognition performance of four typical classifiers, and in panel (b) individual performance of employing MFCC and FFT.

[0048] FIG. 14 shows in panel (a) the overall confusion matrix of each gait type, and in panel (b) individual performance of each gait type.

[0049] FIG. 15 shows in panel (a) individual gait type recognition performance for leave-one-out test, and in panel (b) performance comparison between employing general model and individual model.

[0050] FIG. 16 shows in panel (a) the overall breathing mode recognition performance of frame-level and cycle-level classification, and in panel (b) individual performance of frame-level and cycle-level classification.

[0051] FIG. 17 shows the Leave-one-out performance of frame-level and cycle-level breathing mode classification.

[0052] FIG. 18 shows in panel (a) time domain signals from left and right ear during a forefoot strike running, and (b) the RMSE of left and right ear signals.

[0053] FIGs. 19 to 21 each shows a flowchart of an embodiment.

5 DETAILED DESCRIPTION OF THE INVENTION

[0054] In the following description, numerous specific details are set forth in order to provide a thorough understanding of various illustrative embodiments of the invention. It will be understood, however, to one skilled in the art, that embodiments of the invention may be practiced without some or all of these specific details. Embodiments described in the
10 context of one of the methods or devices are analogously valid for the other methods or devices. Similarly, embodiments described in the context of a method are analogously valid for a device, and vice versa.

[0055] Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs.

15 [0056] The phrase “at least one of A and B” means it requires only A alone, B alone, or A and B, i.e. only one of A or B is required.

[0057] The systems and methods described above may be implemented in many different ways in many different combinations of hardware, software, firmware, or any combination thereof. In one example, the systems and methods may be implemented with
20 a computer having a processor and a memory communicably coupled thereto, where the memory stores instructions, which when executed by the processor, cause the processor to perform the aforementioned methods. The processor may include any type of circuit such as, but not limited to, a microprocessor, a microcontroller, a graphics processor, a digital signal processor, or another processor. The processor may also be implemented with
25 discrete logic or components, or a combination of other types of analogue or digital circuitry, combined on a single integrated circuit or distributed among multiple integrated circuits. All or part of the logic described above may be implemented as instructions for execution by the processor, controller, or other processing device and may be stored in a tangible or non-transitory machine-readable or computer-readable medium such as flash memory,
30 random access memory (RAM) or read only memory (ROM), erasable programmable read only memory (EPROM) or other machine-readable medium such as a compact disc read only memory (CDROM), or magnetic or optical disk. A product, such as a computer program product, may include a storage medium and computer readable instructions stored on the

medium, which when executed in an endpoint, computer system, or other device, cause the device to perform operations according to any of the description above. The memory can be implemented with one or more hard drives, and/or one or more drives that handle removable media, such as diskettes, compact disks (CDs), digital video disks (DVDs), flash memory keys, and other removable media.

[0058] The systems and methods can also include a display device, an audio output and a controller, such as a keyboard, mouse, trackball, game controller, microphone, voice-recognition device, or any other device that inputs information. The processing capability of the system may be distributed among multiple system components, such as among multiple processors and memories, optionally including multiple distributed processing systems. Parameters, databases, and other data structures may be separately stored and managed, may be incorporated into a single memory or database, may be logically and physically organized in many different ways, and may implemented in many ways, including data structures such as linked lists, hash tables, or implicit storage mechanisms. Programs may be parts (e.g., subroutines) of a single program, separate programs, distributed across several memories and processors, or implemented in many different ways, such as in a library, such as a shared library (e.g., a dynamic link library (DLL)). The DLL, for example, may store code that performs any of the system processing described above. The systems and methods can be implemented over a cloud.

[0059] Unless specifically stated otherwise, discussions herein using words such as “processing,” “computing,” “calculating,” “determining,” “presenting,” “displaying,” or the like may refer to actions or processes of a machine (e.g., a computer) that manipulates or transforms data represented as physical (e.g., electronic, magnetic, or optical) quantities within one or more memories (e.g., volatile memory, non-volatile memory, or any suitable combination thereof), registers, or other machine components that receive, store, transmit, or display information.

[0060] 1. Introduction

[0061] Running is a popular and accessible form of aerobic exercise, significantly benefiting our health and wellness. Despite its popularity, potential risks are associated with unoptimized running parameters, and there is a general lack of knowledge among runners on how to manage these parameters. This underlines the need for innovative, portable, and user-friendly running assistant systems. While previous studies have made substantial progress, there are still gaps in the portable

and lightweight monitoring of foot strike parameters, as well as in the exploration of fine-grained breathing-related factors. To address these needs, described herein is an earable-based running assistant system 100 (termed RunCoach) designed to provide continuous, high-fidelity monitoring of diverse gait types and breathing modes during running. RunCoach utilizes the unique properties of in-ear and out-ear microphones from earables to overcome technical challenges with developing such a system 100. Notably, a novel noise reduction approach that leverages out-ear signals to mitigate in-ear noise was developed, thereby enhancing the clarity of in-ear breathing sounds. RunCoach has been extensively evaluated under various challenging conditions and demonstrates high accuracy in identifying three gait types and classifying four breathing modes. With its portability and operational efficiency, and its ability to deliver real-time detection results through a user's smartphone, RunCoach is a significant advancement in running assistance. The described methods and systems 100 fill the existing gaps in the field, facilitating safer and more efficient running practices.

[0062] Described herein is a further step towards running assistance by investigating portable and lightweight solutions for these yet-to-be-explored aspects. Specifically, the methods and systems 100 delve into the foot placement and focus on parameter analysis for each strike, particularly identifying different gait types when the user's foot makes contact with the ground, including overstriding, forefoot striking, and heel striking (detailed in Section 3.2). A aim is to provide runners with real-time detection results that are critical to guide their foot strikes and help them optimize their running form, thereby reducing the risk of injuries. Additionally, fine-grained breathing mode analysis was incorporated, particularly distinguishing between nasal inhalation (NI), nasal exhalation (NE), oral inhalation (OI), and oral exhalation (OE) for breath monitoring (detailed in Section 3.3), which make it possible to offer personalized respiratory coaching to runners and help them maintain optimal breathing patterns throughout their runs for enhanced running efficiency.

[0063] These were investigated 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 locations situate 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, motion-resilient physical and physiological parameters using microphones (including in-ear and out-ear) and IMUs, such as step count, pronation, HR, and RR. Specifically, IMUs on headphones demonstrate resilience to motion artifacts, thanks to their positioning near the relatively stable structures of the human head. In-ear microphones have been proven to reliably capture heart and foot strike sounds due to the occlusion effect. Out-ear microphones have been shown to record breath sounds roughly during running, and they are also highly sensitive to capturing ambient noise.

[0064] RunCoach, a portable and lightweight earable-based running assistant system for continuous, high-fidelity monitoring of three gait types and four breathing modes during running is described herein. RunCoach provides real-time information that could be key to offer insightful feedback to runners, leading to mitigated risk of injuries and enhanced running efficiency. This was achieved by initially leveraging in-ear microphones on earables since it is capable of capturing tissue- and bone-conducted sounds generated by both footsteps and respiration during running as investigated in Section 4.

[0065] However, the following technical challenges need to be carefully addressed:

[0066] i) For both gait type and breathing mode classification, accurately segmented in-ear signals representing single events are essential for precise labeling to train the classification models, and inappropriate segmentation can lead to degraded model performance. To resolve this issue, a series of signal processing methods were meticulously designed to automate event segmentation for gait type classification. As the breathing signals are substantially weaker and span across wider frequency ranges, they are challenging to discern in both time and frequency domains. Consequently, the ground truth signal was leveraged to guide the segmentation of the in-ear signal for model training merely, which is eliminated

during inference through our innovative design of a high-quality frame selection method and frame-based classification.

[0067] ii) The 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, it was observed 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, it was proposed 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.

[0068] Some of the key features and advantages of the methods and systems herein include:

[0069] (i) To the best of our knowledge, RunCoach is the first earable-based running assistant for continuous, hifidelity monitoring of various gait types and breathing modes, which have not been investigated in previous studies. By addressing this gap, RunCoach has the potential to significantly enhance the quality and safety of running practices, making it a noteworthy contribution to the field.

[0070] (ii) A comprehensive suite of signal processing and machine learning techniques was carefully designed to develop RunCoach. Notably, a novel noise reduction method for in-ear audio sensing was introduced by leveraging out-ear signals to diminish in-ear noise, thus enhancing the quality of in-ear breathing sounds during running.

[0071] (iii) Utilizing data collected from 25 participants, our results demonstrate that RunCoach achieves an accuracy of 87.80% in identifying three gait types, and 98.52% in classifying four breathing modes at the phase level. This performance is consistent even under a range of challenging conditions, such as varying intensities of ambient noise, different ground surfaces (rubber vs cement), and a variety of shoe types.

[0072] (iv) RunCoach is a portable and lightweight running assistant system developed by earables. It can be conveniently operated directly on a user's

smartphone (RunCoach was implemented as a smartphone application), providing real-time information that could be key to offer insightful feedback to runners while running. Specifically, processing one-second data, the power consumption measured on a Xiaomi smartphone are 0.0055mAh and 0.0444mAh, which is similar to that of typical smartphone applications like a music player.

[0073] In addition, for both gait type and breathing mode classification models, RunCoach could achieve real-time response, i.e., 11.01ms (milliseconds) and 84.44ms execution time on a Xiaomi 13 smartphone for one-second data processing.

[0074] FIG. 19 shows a flow chart of a computer implemented method 1900 of identifying a breathing mode of a user. The method 1900 includes obtaining an in-ear signal data of a user in block 1902; filtering the in-ear signal data to obtain a breathing signal data in block 1904; segmenting the breathing signal data into a plurality of frames with a window sliding algorithm in block 1906; extracting a plurality of features from each frame in the plurality of frames in block 1908; feeding the plurality of features from each frame to a trained machine learning model in block 1910; obtaining a breathing mode estimation for each frame from the trained machine learning model in block 1912; applying a moving average algorithm to the breathing mode estimation for the plurality of frames to correct an error if present in the breathing mode estimation to provide a plurality of corrected frames in block 1914; and identifying a breathing mode for each frame of the plurality of corrected frames in block 1916. If there are no errors in the frame, the breathing mode will be the same in the corrected frame and uncorrected frame (i.e. before applying the moving average algorithm).

[Claim 1] FIG. 20 shows a flow chart of a computer implemented method 2000 of identifying a gait of a user. The method 2000 includes obtaining an in-ear signal data of a user in block 2002; filtering the in-ear signal data to obtain a gait signal data in block 2004; segmenting the gait signal data with an envelope-based algorithm to obtain a segmented peak in block 2006; extracting a plurality of features in the segmented peak in block 2008; feeding the plurality of features from the segmented peak to a trained machine learning gait model in block 2010; and

obtaining a gait of the user for the segmented peak from the trained machine learning model in block 2012.

[0075] FIG. 21 shows a flow chart of a method 2100 of cancelling noise in an in-ear microphone. The method 2100 includes obtaining an in-ear signal data of a user in block 2102; obtaining an out-ear signal data in block 2104; determining from the out-ear signal data if an ambient noise level is above a threshold level in block 2106; if the ambient noise level is at or above the threshold level, estimating a residual noise level from the out-ear signal data and subtracting the residual noise level from the in-ear signal data to denoise the in-ear signal data in block 2108; and if the ambient noise level is below the threshold level, determine that the in-ear signal data need not be denoised in block 2110.

[0076] 2. Related Work

[0077] Research in running assistant systems, respiration monitoring systems, and audio-based in-ear sensing applications have been previously reported. Running assistant systems typically focus on location-related, strike-related, and physiology-related aspects.

[0078] Location factors such as velocity, orientation, position, step counts, stride frequency, and stride length were measured using in-the-pocket sensors like IMUs and GPS. Strike-related factors, such as foot placement and focus, ground reaction force, and force distribution have primarily been measured using multiple IMUs worn on the body [79, 80], integrating sensors in shoes, and employing force platforms on the ground. Physiology-related factors such as heart rate and respiratory rate have been tracked using chest-worn straps. While these methods are efficient, they are also invasive. However, the present application identifies a need for portable, lightweight solutions to accurately capture foot strike factors and the exploration of more nuanced physiological factors to gain deeper insights. To this end, RunCoach dives deeper by developing portable and lightweight solutions that enable continuous, high-fidelity monitoring across various gait types and breathing modes. This not only addresses the existing gaps in research but also sets the stage for new frontiers in comprehensive running assistance.

[0079] Existing research uses a range of sensor types to monitor human respiration. But previous studies on respiration monitoring have primarily focused on estimating RR and respiration volume, and most of these approaches are limited to stationary settings. Particularly, there is currently no portable and non-invasive solution available for monitoring detailed respiration factors during running scenarios. The present application aims to propose a portable and non-invasive solution specifically designed for accurately classifying fine-grained breathing modes while running.

[0080] Earphones, inherently equipped with speakers and microphones, have proven to be advantageous for various in-ear sensing applications, including human-computer interaction (HCI), health monitoring, and authentication procedures. In particular, Ferlini et al. (Proceedings of the 27th Annual International Conference on Mobile Computing and Networking. 337–349) proposed a gait-based authentication technique by leveraging the low-frequency in-ear sounds generated by stepping motions while walking, which travel uniquely through each individual's distinct musculoskeletal structure and can be amplified by the occluded ear canal.

[0081] 3. Description

[0082] An earable-based portable running assistant system 100 to assist runners with

appropriate running gaits and breathing types has been developed. The most common running injuries (i.e., overuse injuries) occur mainly in the lower limbs of the body due to the prolonged habits of using improper gestures such as overstride, and subsequently result in a reduction/cessation of training or medical consequences. Similarly, adapting appropriate breathing during running, i.e., nasal inhalation and oral exhalation, helps to increase ventricular efficiency and regulate body temperature . The present running assistant system 100 aims to improve the user's experience with these. The different running parameters in general are first explained and detail the parameters focused upon to develop the system 100.

[0083] 3.1 Running Parameters

[0084] Running provides a variety of health benefits, including the strengthening of limb and core muscles, the improvement of cardiovascular endurance and health,

and the promotion of mental health such as the release of stress, anxiety, and depression. Despite its benefits, running also carries the risk of injuries. The most common running-related injuries arise in the lower leg, including ankle, shin, and foot, that can result from various factors, such as overuse, poor running gaits, or wearing unsuitable footwear. Additionally, using incorrect breathing techniques while running can result in a lack of oxygen supply to the body, causing increased fatigue and decreased running efficiency. With the increasing engagement of people in running, there is a demand for tools and technologies that can facilitate the monitoring of running behaviors, relieve risks of injuries, and optimize runners' proficiency, by means of tracking various running-related parameters. For example, many commercial wristbands or smartwatches (e.g., FitBit, Apple Watch, etc.) can monitor the distance and speed of the run, as well as the heart rate of the runners. A comprehensive literature search was done and typical parameters that are important for runners to track their running status, reduce the risk of injury, and improve their running efficiency were identified. These parameters were categorized into stride-related, strike-related, physiology related, and statistic-related, and the definition of each parameter presented in Table 1.

[0085] Some of the parameters (such as stride length, stride frequency, heart rate, etc.) are relatively easy to be measured and have been incorporated into existing wearable products, e.g., Apple Watch, Galaxy Watch, and FitBit. The present application focus on two parameters - gait type and breathing mode, that have not been captured in widely-adopted and portable wearable devices yet. The rationale to investigate these parameters is twofold. Firstly, these parameters play a crucial role in mitigating the risk of injuries (gait type) and enhancing running efficiency (breathing mode). Secondly, most runners (especially untrained ones) lack awareness of controlling/managing their strike and breathing, thereby benefiting a large volume of potential users.

[0086] Table 1. Running-related parameters may affect risk of injury and running efficiency.

Stride-related	Stride Length	The distance covered in a single stride
----------------	---------------	---

	Stride Frequency	The number of strides per minute, also known as cadence.
	Step Time	The duration each foot spends on the ground.
Strike-related	Foot Strike Type	The part a foot makes initial contact with the ground.
	Impact Force Gs	The impact forces experienced by the body during a footstrike.
	Vertical Oscillation	The up-and-down movement of the body's center of mass.
Physiology-related	Respiration Rate	The number of breaths taken per minute during running.
	Breathing Mode	Breathing through nose, mouth, or a combination of both.
	Heart rate	The heart rate during or after running.
Statistic-related	Distance and Velocity	The total distance and the speed during a running session.

[0087] 3.2 Gait Type

[0088] Stride vs. Strike: Usually, stride and strike are used interchangeably to describe the movement of the human foot during running. However, there are small differences between them. Specifically, stride depicts the process of moving the foot from one location to another (i.e., the swing phase), while strike refers to the part of the foot that contacts the ground first (i.e., the landing phase). As shown in Table 2, the stride type could be normal or over, while the strike pattern could be forefoot or heel. In the present application, gait type is defined as the combination of different stride and strike types and mainly focus on forefoot strike, heel strike, and overstride, as illustrated in Table 2 and FIG. 1. Note that the combination of overstride with forefoot strike is ignored as it is unlikely to happen in practice. Next, the three gait types are described in detail.

[0089] Overstride: Overstride occurs when the front leg extends too far in front of the runner's center of mass during running. Due to the vertical misalignment between the center of mass and contact position, the human body (especially the leg and knees) tends to absorb the energy during landing. Therefore, overstride is usually considered an improper and dangerous gait type that might lead to discomfort or even injuries. In addition, overstride might also decrease energy efficiency as it increases the braking force exerted on the body.

[0090] Forefoot Strike: Forefoot strike refers to the case where the runner performs a normal stride and lands on the ball of the foot or the toes. It is usually considered effective for sprinting and short bursts of speed, while landing too far forward on the toes is not recommended for longer distances as it might lead to shin splints or other injuries.

[0091] Heel Strike: Heel strike refers to the case where the runner performs a normal stride and lands with the heel. It is a preferred strike pattern for distance runners as it is more energy-efficient at slow to medium speeds. However, heel strikes can exert additional strain on the knees, leading to knee and hip pain.

[0092] Note that there is no uniformly-agreed strike pattern for runners and the selection of strike depends on (1) the stage of the run (e.g., forefoot strike for sprinting and heel stride for long-distance running), (2) the proficiency level of runners (e.g, forefoot for beginning runners), and (3) runners' preference.

[0093] Table 2. Three running gaits, classified by the stride-strike matrix

Stride \ Strike	Forefoot	Heel
	NA	Overstride
Over		
Normal	Forefoot Strike	Heel Strike

[0094] 3.3 Breathing Modes

[0095] Respiration, the process of breathing to fuel our bodies with oxygen and getting rid of carbon dioxide, is crucial for running as such intense activity consumes oxygen and produces waste faster. While the breathing rhythm has been widely investigated and demonstrated to be critical for running, existing literature ignores

another important feature of breathing - the breathing mode (which is defined herein as the fine-grained segmentation of oral vs nasal and inspiration vs. expiration). During running, people can breathe using their nose (refers to nasal breathing) or mouth (refers to oral breathing), or a combination of both. The selection of the best breathing mode during running depends on factors such as exercise intensity, environment temperature, and humidity. 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. For example, Hall (American Journal of Human Biology: The Official Journal of the Human Biology Association 17, 3 (2005), 321–330) demonstrated that at 60% of maximal oxygen consumption, oral breathing leads to significantly higher oxygen consumption, ventilation, and respiration rate compared to nasal breathing. 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. In the present application, inhaling and exhaling are treated as independent breathing stages and therefore there are four breathing modes: nasal inhaling (NI), nasal exhaling (NO), oral inhaling (OI), and oral exhaling (OE).

[0096] 4 FEASIBILITY EXPLORATION: IN-EAR VS. OUT-EAR MICROPHONE

[0097] The above discussions have clearly shown the importance and benefits of monitoring the gait types and breathing modes during running. Unfortunately, there is no existing common wearable that can support such operations. Hence, an aim was to design an advanced running behavior monitoring system 100 using high-adherence wearables. Specifically, a focus was placed on earphones, a highly-adopted wearable device in recent years, particularly during exercises (e.g., running) for music playback. Compared to smartphones or smartwatches with relatively large form factors, earphones can only embed a few sensors due to their small size and limited battery capacity. 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, the feasibility of using the in-ear microphone to detect runner's gait type and breathing modes was explored. This idea was motivated by recent works that demonstrate the capability of in-ear microphones to capture human-generated vibrations/sounds through bone conduction and occlusion effect (Stenfelt and Reinfeldt, 2007, International journal of audiology 46, 10 (2007), 595–608). A prototype was developed (details presented in Section 6.1) and data was collected with both the in-ear and out-ear (for comparison purposes) microphones when a subject is running. FIG. 2 shows the waveform of the signals from different gait types captured with both microphones. It can be observed from the left three columns that (1) the pattern of the waveform is distinct for different gait types due to the varying ways of hitting the ground and different propagating paths (i.e., the human skeleton and bone structure). In detail, the forefoot contains more tissues than the heel, resulting in larger signal absorption in forefoot strike than those in heel strike and overstride. Due to the angle difference between the human body and the foot, heel strike and overstride experience slightly different propagation paths and therefore different patterns. (2) Although both in-ear and out-ear microphones are positioned at distances approximating the user's height from the sound source (foot strike), the in-ear microphone yields stronger gait signals as it possesses two distinct advantages. Firstly, in-ear microphones receive vibrations that directly propagate through the lower limbs and body to the ear canal through bone conduction. In contrast, out-ear microphones capture air-conducted sound generated between the contact of the shoe and the ground, resulting in greater signal attenuation. Secondly, the earphone seals the ear canal and therefore forms the occlusion effect, which amplifies the low-frequency foot strike-induced sounds (<100 Hz). The right two columns in FIG. 2 compare the impact of external noise on both microphones. It can be seen from the raw signals that both microphones are affected. However, after lowpass filtering (<100 Hz), the in-ear

signal (through bone conduction thus resides in low frequencies) shows clear gait patterns while the out-ear signal (through air conduction thus resides in high frequencies) is completely polluted by the noise.

[0098] FIG. 3 shows the spectrogram of the signals collected under different breathing modes during running. First, it can be observed 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, 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.

[0099] The above studies demonstrated the feasibility of using the in-ear microphones to capture and distinguish runner's gait types and breathing modes, and identified potential factors that might affect the detection performance. Next, the design was developed to realize the idea and optimize the performance under different scenarios.

[00100] 5 SYSTEM DESIGN

[00101] 5.1 Overview

[00102] The system design of RunCoach is described. RunCoach is an earbuds-based running assistance system that monitors and identifies different types of running gaits and breathing modes by opportunistically tapping on both in-ear microphone and outear microphones. While the in-ear microphone captures tissue and bone-conducted sounds generated by both footsteps and respiration, the out-ear microphone can be used to record ambient sound for subsequent noise reduction. FIG. 4 shows the RunCoach methodological pipeline, consisting of the following phases:

[00103] • Development Phase: The goal of this phase is to develop trained machine learning (ML) models to detect gait types and breathing modes. The phase starts off by decomposing a single in-ear signal into two streams with different frequency ranges: (i) The low pass filtered (with a cutoff frequency of 100 Hz) component, called the gait signal, captures the gait-related signals as the vibrations propagated to the ear canal via bone conduction are at low frequency. To segment this signal into corresponding footsteps, an envelope-based approach is applied to the gait signal, where the peaks of the smoothened envelope are used to localize each footstrike. The episode that consists of neighboring regions (width of 100ms) of both sides of the detected peaks is segmented as a single footstep. The energy spectrum of each segment at different frequency bands is then extracted as discriminative features and trained an ML classifier to recognize the three gait types (explained in detail in Section 5.2).

[00104] (ii) As illustrated in FIG. 3, the high frequency (high pass filter with a cutoff frequency of 500 Hz) component of the in-ear signal shows distinct characteristics of different breathing modes (i.e., NI, NE, OI, and OE). As the breathing sounds recorded in the ear canal are considerably weaker, a frequency-based approach was first adopted to segment different breathing modes. To be more specific, the ground truth data recorded by an external microphone placed on the philtrum is first segmented by identifying the transitions between breathing motions that are represented by the valleys of the frequency-energy curve, and subsequently used to segment the in-ear recorded breathing signals. 40 MFCC features are extracted over each frame of 100ms and slide through the entire segmented breathing motion (with the hop length of 50ms). By carefully applying a series of techniques (explained in Section 5.3) 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 remaining MFCC features within a segmented motion are averaged and fed into an ML classifier for training, resulting in a model that can recognize different breathing modes.

[00105] • Run-time Phase: This phase focuses on classifying gait and breathing signals into appropriate classes using the pre-trained models generated in the

Development Phase. For gait type classification, a similar pipeline detailed in the development phase is followed to extract FFT features and then utilize the pre-trained ML model for online classification. Meanwhile, the count of the segmented episodes is regarded as an observable proxy of the number of footsteps, which can be used to calculate the cadence, i.e., steps taken per minute. For the breathing phase classification, a slightly different pipeline was used – specifically, since the breathing signals captured by the in-ear microphone are considerably weaker and vulnerable to environmental noise (>500 Hz) such as traffic sounds, and music, in the run-time phase, the ambient noise level is first checked using the out-ear microphone. If the noise level is above a certain threshold, the noise reduction module (described in Section 5.3.2) is first used to reduce the impact of the external noise with the aid of the out-ear microphone. To be specific, a phenomenon was uncovered where the ambient noise (recorded by the out-ear microphone) is modulated by the occluded ear canal before being picked up by the in-ear microphone. It is observed that the modulation index between the two signals at different frequencies can be represented by linear correlation coefficients, and subsequently used to cancel the noise in the in-ear breathing signal by subtracting the estimated modulated ambient noise in the ear canal. It was next proceeded to segment the breathing signals into small frames and extract the MFCC features for breathing phase classification. As a breathing phase spans across multiple frames, a moving average scheme is proposed to smooth the frame-level recognition results, obtaining a higher phase-level accuracy. Finally, the respiration rate can be estimated based on the low-pass filtered phase-level breathing mode classification.

[00106] 5.2 Gait Type Classification

[00107] 5.2.1 Classification Model Development. To enable continuous monitoring of users' gait types during running, a machine-learning model that can classify the target gait types was first developed. After collecting the various gait type data from multiple users, the pipeline shown in Figure 4 is followed to train the recognition model.

[00108] **Envelope-based Cycle Segmentation:** As shown in FIG. 5(a), the footstrike-induced and the patterns are clear. To eliminate possible environmental

noise (usually higher than 100 Hz), a low-pass filter with a cut-off frequency of 100 Hz is applied to the in-ear signal. The resulting signal refers to the gait signal (FIG. 5(b)). Then, the envelope of the gait signal is extracted using the Hilbert transform. Afterward, a low-pass filter with a cut-off frequency of 5 Hz is applied to smooth the envelope signal ((FIG. 5(c)). Finally, peak detection on the filtered envelope was performed, where the peaks correspond to the time when the human foot hits the ground. Therefore, the samples that are located 100 ms before and 100 ms after each peak are regarded as the gait signal for a single foot strike (200 ms in total) (FIG. 5(d)).

[00109] **Feature Extraction and Model Training:** MFCC is a commonly used feature extraction technique in speech and audio processing. However, its effectiveness in capturing low-frequency signals like gait sounds (<100Hz) may be limited as it is designed based on human ear sensitivity to higher frequencies. In detail, MFCC decomposes the signal into Mel spectrograms with varying bandwidth, i.e., higher resolution in the human audible frequency range while extremely low resolution for signals below 100 Hz. Thus, it is proposed to apply FFT, which divides the spectrum into bands with equal widths, to the segmented signals to obtain the energy features. Concretely, after removing the DC component, a 65536-point FFT is applied to each 200 ms gait signal, yielding a frequency resolution of 0.12 Hz. The frequency range below 100 Hz is divided into 40 sub-frequency bands, each spanning 2.5 Hz. By calculating the average energy in each sub-frequency band, 40 energy values are obtained as the features. Then, an ML-based approach for classifying different gait types is considered due to its superior performance and relatively low computation. Thus, the extracted features are then fed into an ML model for training. Four common machine learning classifiers were considered, namely Support Vector Machine (SVM), K Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF), and select the best one as evaluated in Section 7.1.1. These classification models are implemented using the Python scikit-learn package (Pedregosa et al.. Journal of Machine Learning Research 12 (2011), 2825–2830)

[00110] 5.2.2 Run-time Gait Monitoring. In the run-time stage, the same pipeline is followed to process the in-ear signal (including filtering and segmentation) and extract the features. The resulting features are passed to the pre-trained ML model for real-time gait-type classification. Furthermore, since each peak corresponds to a step, the peaks may be counted to derive the runner's stride rate or cadence.

[00111] 5.3 Breathing Mode Classification

[00112] 5.3.1 Classification Model Development. Similar to gait type monitoring, a breathing mode classification model needs to be developed before RunCoach can monitor a runner's breathing in run-time. The pipeline is shown in FIG. 4.

[00113] **Spectrogram-based Breathing Phase Segmentation:** Compared to the gait signal, breathing signals are much weaker and span across wider frequency ranges. Thus, it can be barely observed from the time domain signals and thus it is extracted from the frequency domain. However, although the breathing patterns are roughly seen from the spectrograms presented in FIG. 3, purely applying segmentation 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 affected 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. FIG. 6(a) and FIG. 6(b) compare the original ground truth signal and the high-pass filtered breathing signal with a cutoff frequency of 500 Hz. As shown in FIG. 6(c), the inhaling and exhaling motions present distinctive patterns and clear transitions (the change of frequency components and the associated energy). Note that the ground truth data is dedicated to breathing phase segmentation and labeling and will not be employed as input for training models.

[00114] To segment the spectrogram of the ground truth signal, a Short-Time Fourier Transform (STFT) is conducted on each five-second window to acquire its spectrogram expressed in decibels. The energy across all frequency bands is summed up to construct a frequency energy curve, as shown in FIG. 6(d). Based

on the observation that there exists a silence period (no breathing) during the transition, the valleys of the frequency energy curve are extracted as the transitions between adjacent breathing phases. Finally, the index extracted from the ground truth signal is applied to the breathing signal (extracted from the in-ear signal with a >500 Hz high-pass filter) and segment it.

[00115] **Framing and Feature Extraction:** Since breathing sounds contain energy at higher frequencies, the MFCC features are simply extracted for breathing event detection. Specifically, 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.

[00116] **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 the 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 FIG. 7(c), there might be spontaneous environmental noise occurring in the middle of breathing.

[00117] To mitigate those low-quality and noisy frames, a high-quality frame selection scheme is proposed. In detail, a feature template is first generated by computing the average of MFCC features from all the frames in the current segmented phase. Then, the Minkowski distance ($p=2$) between the MFCC features from each frame and the template is calculated. As exemplified in FIG. 7, frames with pure breathing signals have short distances to the template, while the transition frames and noisy frames show relatively larger distances. Thus, a distance threshold of 50 is empirically set to retain high-quality frames for more accurate model development.

[00118] **Model Training:** Similar to the gait model, 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 7.2.1 will help identify the classifier with the best performance.

[00119] 5.3.2 Run-time Breathing Monitoring. As our breathing mode classification model is trained with clean breathing sounds and the in-ear breathing signal is vulnerable to environmental noise, the out-ear microphone is introduced to deal with this issue at run-time. Specifically, the out-ear microphone can be used for two purposes: noise detection and noise reduction.

[00120] **Noise Detection:** The out-ear signal is a combination of 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, the overall energy of the out-ear signal across different frequency bands is computed and converted to sound volume in decibels (dB). If the noise level exceeds a certain threshold (e.g., 55 dB), RunCoach 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 to frames directly.

[00121] **Noise Reduction:** A phenomenon was observed that any external noise is firstly captured by the out-ear microphone 120, subsequently modulated by the occluded ear canal cavity between the earbud 105 and eardrum, before finally being captured by the in-ear microphone 110. This indicates that there exists a correlation between the in-ear and out-ear noise signals and was used to leverage the out-ear signal for in-ear noise reduction. Specifically, the idea is to estimate the residual noise inside the ear using the out-ear signal⁴ and then subtract it from the noisy in-ear signal to obtain clean in-ear breathing sounds.

[00122] To realize this idea, the correlation between the in-ear and out-ear signals is first obtained. 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 FIG. 3, it was 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 signal. Thus, the out-ear and in-ear signals were transformed into the frequency domain using Short-Time Fourier Transform (STFT), and compute the amplitude and phase at each frequency band. FIG. 8(a) and FIG. 8(b) show the scatter plot of amplitudes and phases for {out-ear, in-ear} pairs at 10 different frequency bands. Linear correlations between the in-ear and out-ear signals were observed in both the amplitude and phase plots. As such, a lightweight linear regression model is leveraged to establish mappings of {out-ear, in-ear} for each frequency band. Specifically, a linear relationship was formulated between the amplitude and phase of out-ear and in-ear sounds as represented by:

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

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 in 1024 frequency bands. The right part of FIG. 9 summarizes the flow to derive the in-ear and out-ear correlations, referred to as

Offline Coefficients Learning. Specifically, a background sound with various frequencies is first played and the in-ear and out-ear signals are recorded (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.

The pipeline for run-time noise removal is illustrated in the left part of FIG. 9, referred to as Online Noise Reduction. FIG. 10 shows an example of the noise reduction performance at different stages. In detail, when external noise is detected, STFT is first applied to the out-ear signal to obtain its amplitude and phase values at different frequency bands (FIG. 10(a)). Then, for each band, the transformed amplitude and phase is computed using the derived coefficients with Equation (1) (FIG. 10(c)). Afterward, the transformed amplitude and phase is subtracted from the corresponding in-ear noisy amplitude and phase (FIG. 10(b)), resulting in a clean in-ear breathing pattern (FIG. 10(d)). Finally, iSTFT is applied to recover the denoised in-ear breathing sounds.

[00126] **Signal Segmentation, Feature Extraction, and Classification:** In the development phase, the breathing signal is segmented into different phases (inhaling or exhaling) with a strong reference signal from the nose. However, no reference signal is available during run-time. Thus, the breathing signal is segmented into frames with a length of 100 ms using a sliding window (no overlap). For each frame, the MFCC features are extracted and feed them to the trained ML model for classification. As such, a breathing mode estimation is obtained for every frame.

[00127] **Moving Average Correction:** From the application perspective, runners only require phase-level (either inhalation or exhalation) updates about their breathing mode. In the 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, it is proposed to leverage the moving average algorithm (with a window length of 10 frames) to smooth the frame-level estimations. Specifically, as shown in FIG. 11(b), it was found that applying the moving average algorithm once can correct most of the spontaneous errors while failing to deal with consecutive errors. Thus, another moving average operation is applied to the corrected estimations and this successfully corrects consecutive errors, as shown in FIG. 11(c). Moreover, from the final smoothed breathing mode estimation curve, illustrated in Figure 11 (c), the length of each breathing phase or the count of breathing within a certain period can be clearly identified (circle marker in FIG. 11(d)), thereby obtaining the respiration rate. Given that the respiration rate of average individuals typically falls within the range of 40 to 60 per minute when engaged in physical activity, a low-pass filter with a cutoff frequency of 2 Hz is employed to obtain the desired respiration rate.

[00128] 6 PROTOTYPING AND DATA COLLECTION

[00129] 6.1 Prototype

[00130] RunCoach only requires an in-ear microphone and an out-ear microphone, which have been equipped in many commercial wireless earbuds such as Apple AirPods and Huawei Freebuds. 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.

[00131] To address this limitation, a pair of earbuds 105 were designed and constructed to evaluate the proposed system 100. As shown in FIG. 12(a), the customized earbuds 105 comprise a 3D-printed shell and two analog microphones 110, 120 (CMC-4015-40L100 by CUI Devices), with one placed near the earbud tip to capture in-ear sound (referred to as the in-ear microphone 110) and the other one integrated into the end of the handle to pick up out-ear sound (referred to as the out-ear microphone 120). Both microphones 110, 120 are connected to a Bela Mini development board 125 through a 3.5mm audio jack, which feature an integrated development environment (IDE) for data recording. The Bela Mini board 125 is then connected to a Raspberry Pi 130 so that the data can be collected remotely when the subjects are running. The sampling rate is set to 8 kHz as human-produced sounds are usually below 4 kHz. All prototype components are powered by a power bank 135 and placed in a small bag 140, as in FIG. 12(b). Additionally, to ensure participant comfort during running, as well as good sealing quality, three different sizes of foam ear tips 115 were provided to fit various ear canal sizes.

[00132] 6.2 Data collection

[00133] Upon receiving approval from the Institutional Review Board (IRB), a total of 25 participants was recruited, 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), and their height ranged from 1.58 m to 1.87 m (mean = 1.71 m, standard deviation = 0.09 m). Their body mass index (BMI) ranged from 18.82 to 27.05 (mean = 22.67, standard deviation = 2.36). Among them, 22 individuals fell within the healthy weight range, while the remaining 3 individuals were classified as overweight. 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. Each participant was instructed to perform specific gait types or breathing modes while running on an

outdoor rubber track. The total duration of the data collection process was approximately 1 hour. FIG. 12(c) illustrates the flow of data collection and a detailed description of the gait session and breathing session is provided as follows.

[00134] Prior to the gait section, participants are provided with detailed instructions on forefoot strike, heel strike, and overstride to ensure the proper execution of specific gait types. The participants are instructed to perform these gait types during each running session, with each gait type being recorded for a duration of two minutes and repeated twice. A one-minute break is scheduled between each two-minute session to mitigate potential movement deviations caused by participant fatigue. There are no specific speed requirements throughout the entire data collection process, allowing participants to choose their preferred running pace based on their own comfort. It is assumed that participants would ideally demonstrate each gait type, representing the ground truth for the corresponding session. To measure the ground truth of running cadence, a separate individual manually counts the steps while the participants were running. Each participant completed six running sessions involving three types of gait, resulting in a total of 49,886 steps across all subjects.

[00135] The procedure for collecting breathing data in the study followed a similar approach to the gait collection. Before the formal collection of breathing data, participants are instructed to attempt breathing solely through the nose or mouth to ensure smooth breathing during running. Subsequently, each participant is asked to perform specific breathing types (nasal breathing or oral breathing) during each running session, which lasts for five minutes and is repeated twice. A two-minute inter-session break is provided to allow for rest and recovery. Since both the in-ear and out-ear microphones cannot capture a strong breathing signal with clear inhale and exhale transitions, another microphone was attached under the nose of each participant. Due to the close distance and direct contact with the breathing airflow, the collected audio can serve as the ground truth for breathing cycle segmentation as well as the breathing rate.

[00136] 7 EVALUATION

[00137] Based on the collected data, the performance of RunCoach is evaluated in this section. Specifically, the aim is to answer the following questions. For gait type detection, (1) how accurately can RunCoach achieve and whether the user needs to collect individual data and train a personal model? (Section 7.1.3 and
5 Section 7.1.4); (2) what is the impact of different design and practical factors, such as the use of FFT features, different shoes and ground conditions? (Section 7.1.5 and Section 7.1.6). For breathing mode detection, (1) what accuracy can RunCoach
[00138] achieve in detecting the breathing mode? (Section 7.2.1 and Section 7.2.2) ; (2) will different combinations of inhalation and exhalation modes affect the
10 performance? (Section 7.2.3); (3) how effective are the proposed techniques such as noise removal and error correction? (Section 7.2.4). And finally, what is the run-time overhead of RunCoach on smartphones? (Section 7.3)

[00139] 7.1 Gait Type and Cadence Estimation

[00140] 7.1.1 Comparison of Classifiers. As mentioned in Section 5.2.1,
15 conventional machine learning (ML) based classifiers was adopted to recognize different gait types due to the simplicity of the task and the lightweight nature of ML
[00141] classifiers. Specifically, for each subject, the collected data was split into a training set (75%) and a testing set (25%) and implemented four typical classifiers, namely, KNN, SVM, DT, and RF for training and testing. FIG. 13(a) compares the
20 average recognition accuracy over the 25 subjects based on the classifiers. It can be observed that all the four classifiers achieve more than 80% recognition accuracy, with KNN attaining the highest accuracy at 87.80%. Consequently, the subsequent results are reported with KNN as the classifier.

[00142] 7.1.2 Comparison of Features: MFCC vs. FFT. As stated in Section 5.2.1,
25 instead of extracting commonly-used MFCC features for audio signals, it was proposed to use the FFT features for gait type classification due to the extremely low-frequency property of the gait signals. To demonstrate the effectiveness of the proposal, the MFCC features were also extracted and applied the same classification pipeline to the collected data. FIG. 13(b) compares the recognition
30 accuracy of MFCC and FFT features across different subjects, where it can be seen that FFT features consistently outperform MFCC features. Averaged over all the

subjects, MFCC features obtain an accuracy of 80.01%, while FFT features yield an accuracy of 87.80% with around 8% improvement.

[00143] 7.1.3 Detailed Performance. FIG. 14(a) illustrates the confusion matrix of the three gait types over all the subjects using KNN and FFT features. It can be
5 observe that forefoot strike achieves the best accuracy, while the model mainly confuses with heel strike and overstride. The main reason is that heel strike and overstride are both landed with the heel, which might create similar patterns in the in-ear waveforms. In contrast, forefoot strike is landed with the fore part of the foot, where there are more bone and tissue compositions, thereby leading to more
10 discriminative patterns.

[00144] In FIG. 14(b), the accuracy of each gait type for each subject is plotted. First, it can be seen there is no significant performance variation across different subjects, indicating that the approach can be applied to different individuals. Second, interestingly, it was observed that the forefoot strike does not always
15 achieve the highest accuracy as analyzed above, while heel strike and overstride are recognized more accurately for some subjects (e.g., subjects 2, 9, 13 for heel strike, and subject 19 for overstride). After an in-depth investigation, we found that the recognition performance is also related to the preferred running gait type of the user. For instance, subject 2 usually runs with heel strike. Therefore, during data
20 collection, subject 2 can perform more consistent heel strikes compared to the other two gait types, resulting in higher-quality heel strike data for model training and in turn higher accuracy during testing.

[00145] 7.1.4 Performance of Leave-one-out Test. The above results are obtained based on the individual models, which require users to collect personal data for
25 model training. To assess whether a pre-trained model can be applied to unseen users, the leave-one-out test was conducted. In detail, the data from one subject was iteratively select as the testing set, while the data from the remaining 24 subjects are used to train a model. The testing accuracies are shown in FIG. 15(a), where a performance degradation can be observed, with an average accuracy drop
30 from 87.70% to 78.10%. The result is rooted in two reasons: 1) gait signals captured inside the ear canal are propagated through bone conduction across the whole

body, while different people might have distinct skeletons, leading to a unique modulation on the gait signals; 2) different subjects might perform the three gait types in slightly different ways, resulting in pattern variations. Therefore, directly applying a general model suffers from a relatively poor performance.

5 [00146] To solve this issue, the user may provide some personal samples for model training. Here, two different strategies were investigated: 1) pure individual model training and 2) model fine-tuning with personal data. For the first strategy, individual models were trained with different numbers of personal samples (from 10 to 300). As shown in the lighter curve in FIG. 15(b), with only 10 samples, the
10 individual model yields even poorer performance compared to the general model. This is because the general model can benefit from a larger volume of training data. By increasing the number of personal samples, the accuracy gradually increases until it saturates at 87.67% with 140 samples. For the second strategy, other subjects' samples were combined with personal samples for model retraining. As
15 shown in the darker curve, the accuracy consistently increases with more personal samples. In specific, with 70 samples, it achieves the same performance as the individual models. While with more samples, the performance is lower because the variations from other subjects might interfere with the current subject.

[00147] Thus, it may be concluded from the data that 1) without any personal data,
20 our approach can offer 78.1% gait type recognition accuracy; 2) if the user provides limited samples (<140), the model should be trained together with other subjects' data; 3) if more personal data is provided (>140), the model should be trained solely with the individual data.

[00148] 7.1.5 Impact of Different Shoes. The approach relies on the
25 sounds/vibrations generated during the contact between human foot and ground to distinguish different gait types. One possible factor that can affect the generated sounds is the running shoes, which might have different sole materials or air cushioning thickness. To investigate such an impact, a small-scale study with three subjects was conducted. Specifically, each subject wore two different shoes and
30 performed the three gait types respectively. Then, different training and testing schemes were studied and the results are presented in Table 3.

[00149] Table 3. Classification Performance of Different Shoes

Training	Shoe1		Shoe2		Shoe1+Shoe2	
Testing	Shoe1	Shoe2	Shoe1	Shoe2	Shoe1	Shoe2
S1	90.83%	85.27%	85.74%	87.91%	89.30%	86.82%
S2	87.29%	83.1%	81.40%	84.67%	85.50%	84.33%
S3	90.08%	82.98%	82.47%	86.32%	86.78%	84.83%

[00150] It can be observed that 1) training and testing on the same type of shoes yields the best performance; 2) when training and testing with different shoes, i.e.,
5 unseen shoes, there is around a 4% accuracy drop as the material and geometry of the shoes introduce variations on the generated gait signals; 3) if the data from both shoes are used for training, the individual accuracy can be improved compared to the case of unseen shoes. The impact of shoes is already considered in the leave-one-out test, as each subject wore their own shoes for data collection. In summary,
10 the impact of running shoes is limited and the described approach can be applied to unseen shoes with acceptable performance.

[00151] 7.1.6 Impact of Ground Conditions. Similar to wearing different shoes, runners may run on different ground materials, such as rubber tracks or cement roads. Each of these grounds possesses unique properties. For instance, rubber
15 tracks have a greater capacity for shock absorption and longer contact time between the foot and the track, which can influence the sound generated by foot strikes and consequently affect classification performance. To investigate the impact, three subjects were recruited to run on both a rubber track and a cement road, which represent the softest and hardest common ground materials during running.
20 Similarly, different training and testing schemes were applied to understand the impact. As shown in Table 4, it can be observed that 1) running on hard ground (i.e., cement) yields a better performance (e.g., 93.40% versus 90.83% for S1) because soft ground can absorb/attenuate some of the vibrations generated during foot strikes; 2) when model training and testing are performed on different ground
25 conditions, i.e., unseen scenarios, the approach suffers from around a 7% accuracy drop due to the pattern shifts of the gait waveforms; 3) when the training data

contain both ground conditions, the recognition accuracies are improved. Since the ground conditions are relatively limited, users can collect data from each condition for model training. In addition, runners usually have individual preferences on the ground condition, so the ground condition-specific model can be trained to obtain optimal performance.

[00152] Table 4. Gait classification performance of different ground conditions.

Training	Rubber		Cement		Rubber+Cement	
Testing	Rubber	Cement	Rubber	Cement	Rubber	Cement
S1	90.83%	85.35%	83.61%	93.40%	86.47%	90.74%
S2	87.29%	82.06%	81.53%	89.77%	84.19%	87.13%
S3	90.08%	84.17%	84.49%	92.46%	87.46%	90.13%

[00153] 7.1.7 Cadence Calculation. After gait cycle segmentation, each cycle corresponds to a single foot strike. By counting the number of cycles within a certain period, the cadence of the runners may be easily obtained, without relying on other sensors or algorithms. Remarkably, RunCoach achieves an overall accuracy of 99.69% in cadence calculation across all 25 subjects.

[00154] 7.2 Breathing Mode and Respiration Rate Estimation

[00155] 7.2.1 Individual Model Performance: Predictions for each respiratory frame are made, resulting in a series of initial predicted labels, and the moving average (MA) technique is further applied twice to mitigate spontaneous errors. These are referred to 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, the ground truth data is utilized to determine each breathing cycle and subsequently perform a majority voting to determine the breathing mode for a particular breathing phase. The performance was evaluated with four typical ML classifiers (KNN, SVM, RF, DT) and report the results for KNN only as it achieves the best performance.

[00156] As shown in FIG. 16(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%. FIG. 16(b) presents the individual performance of 25 subjects, considering both frame-level and phase-level classification. By combining the insights from FIG. 11 and FIG. 16(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, their breathing sounds are characterized by higher intensity, rapidity, and shallowness, resulting in more noticeable and distinguishable patterns.

[00157] 7.2.2 Performance of Leave-one-out Test. The leave-one-out test is also conducted for the breathing mode recognition. FIG. 17 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 male's), 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 compared to gait signals, where there are significant inter-individual distinctions, breathing signals among individuals are more similar, allowing for the utilization of a general model for analysis and prediction.

[00158] 7.2.3 Impact of Different Inhalation and Exhalation Sequence. Due to constraints in terms of time and the participants' physical stamina, only data on two combinations of breathing patterns were collected: 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.

[00159] To examine how different combinations of breathing patterns affect classification performance, an experiment was conducted in which three participants utilized all four possible breathing combinations while running.

[00160] Table 5 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 5 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, when all four combinations are included in the training data, good classification is achieved for all test data, as illustrated in the rightmost column. These findings suggest that different breathing combinations can affect classification performance, but this issue can be overcome by including all four combinations in the training data.

[00161] Table 5. 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%	93.36%	95.7%	95.47%
S1+S2+S3	90.81%	84.51%	83.83%	90.48%	91.8%	91.77%

[00162] 7.2.4 Performance with different ambient noise levels. Although earbuds
105 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
5 levels of environmental noise on respiratory classification performance, a speaker
was utilized to play pre-recorded traffic noise at various volumes, simulating a
scenario in which a runner is exercising outdoors along a road.

[00163] Table 6 presents the performance results for three subjects in both the
noisy and noise-reduced conditions. From the first column, it can be observed that
10 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
15 of misclassified frames. Therefore, noise reduction is deemed necessary in such a
scenario. By comparing the first (original) and third (noise-reduced (general
template)) rows of Table 6, it can be observed that the noise-reduction scheme can
significantly improve the performance by up to around 40% depending on the noise
[00164] level.

[00165] The noise-reduction approach relies on an occluded ear canal to obtain
the correlation between the in-ear and out-ear signals. However, 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 template and further impact effectiveness of noise
25 reduction in practical applications, the following experiments were conducted.
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
second (noise-reduced (general template)) and third (noise-reduced (individual
30 template)) rows of Table 6, 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 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, it is recommended for users to 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 to play a simulated sound and perform linear regression.

[00166] Table 6. Phase-level performance with different ambient noise level.

		original	noise-reduced (general template)	noise-reduced (individual template)
S1	quiet	95.53%	NA	NA
	60dB	79.7%	83.63%	91.32%
	70dB	65.2%	75.35%	87.44%
	80dB	42.1%	71.81%	81.94%
S2	quiet	92.1%	NA	NA
	60dB	70.8%	79.82%	89.18%
	70dB	57.6%	73.16%	88.5%
	80dB	39.4%	67.33%	85.12%
S3	quiet	94.2%	NA	NA
	60dB	78.1%	81.45%	91.96%
	70dB	61.4%	75.49%	87.42%
	80dB	51.9%	68.8%	85.27%

[00167] 7.2.5 Respiration Rate Calculation: Following the application of the second moving average, a low-pass filter

[00168] with a cutoff frequency of 2 Hz is applied. Subsequently, the peaks and valleys (represented by orange dots)

[00169] depicted in Figure 11 (d) serve as markers for each breathing phase. The respiration rate can be computed by

[00170] counting the peaks or valleys in a given period. Across all subjects, the overall accuracy for respiration rate

[00171] calculation is 95.08%. The performance varies among individuals, with the lowest accuracy recorded at 89.92%

[00172] and the highest accuracy achieved at 98.8%.

[00173] 7.3 System Performance

[00174] RunCoach was implemented as an Android application and its system performance evaluated on a Xiaomi 13 phone device which is equipped with a Snapdragon 8 Gen 2 processor and a battery with 4500mAh capacity. The gait type classification process is divided into three distinct stages. The pre-processing stage includes tasks such as applying a low-pass filter, detecting peaks, and calculating cadence. Subsequently, the feature extraction part generates FFT features for each individual gait cycle. Finally, the post-processing phase entails inferring a label for each gait. Additionally, the breathing mode pipeline is segmented into five parts. Firstly, the pre-processing stage applies a high-pass filter. The subsequent sections, 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 section generates MFCC features in the frame level. Lastly, the prediction of breathing mode and the twice-moving average, constitute the post-processing stage.

[00175] The CPU load and latency was evaluated by executing a specific portion of the code looping 1000 times, repeating this process five times, and calculating the average value of five iterations. To assess battery usage, the power consumption of a specific code segment was measured by running it in the background for one hour while the screen is turned off. The difference between the power consumption during this period and the power consumption during one hour of idle standby yields the power consumption specifically attributed to the code segment. Table 7 presents the system performance of RunCoach based on the processing of a one-second data. It can be observed that gait type and breathing mode recognition can be completed within 11.01 ms and 84.44 ms respectively, guaranteeing real-time detection. The CPU load for both tasks are low at around 5%. In terms of power consumption, the measurements to battery usage are converted to %/hour, and the results indicate that both tasks consumes about 1.1%/hour 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 the classification pipelines, which ensures

that the system does not burden the mobile device and can operate for extended periods without draining excessive battery power.

[00176] Table 7. 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
Gait	CPU (%)	3.4	-	-	4.8	6	-
	Latency (ms)	0.38	-	-	5.91	4.72	11.01
	Energy (mAh)	0.0001	-	-	0.0030	0.0024	0.0055
	Battery Usage (%/hour)	3	-	-	4	4	1.10
Breathing	CPU (%)	3.3	3.7	5.6	5.2	6.2	-
	Latency (ms)	0.34	0.19	64.12	2.57	17.22	84.44
	Energy (mAh)	0.0001	0.0001	0.0321	0.0013	0.0108	0.0444
	Battery Usage (%/hour)	3	3	4	4	5	1.13

5

[00177] 8 DISCUSSION AND FUTURE WORK

[00178] Impact of different running speeds. Different runners or the same runner at different phases of a running session can have different speeds. The adjustment of running speeds is mainly based on the stride length, while the runners tend to maintain a similar running cadence. This implies that each gait cycle is likely to have the same time span at different speeds, thereby having a limited impact on the cycle

10

[00179] length. However, to adapt the stride length, runners need to adjust the contacting force between the foot and ground, thereby increasing the magnitude of the gait signal. To understand how the running speed affects gait type recognition, one subject was recruited to run at three different speeds (slow-7.5 km/h, normal-10 km/h, and fast-12 km/h, respectively) and performed the gait type recognition.

[00180] Table 8 presents gait type recognition performance with different training and testing data. It can be observed that when training and testing are performed on data with different speeds, there is an average accuracy drop of around 5%. This implies that there is a slight difference between the gait patterns under different speeds, which may be attributed to the inherent biomechanical and muscle activation variations associated with running. However, if we combine the data at all running speeds for training, the accuracy can be improved significantly and close to the case where training data and testing data are from the same speed.

[00181] Table 8. Gait type recognition performance at different running speeds.

Training Testing	Slow	Normal	Fast	Slow+Normal+Fast
Slow	87.81%	83.32%	81.08%	87.58%
Normal	84.55%	89.15%	80.52%	87.36%
Fast	83.54%	80.74%	86.13%	86.02%

[00182] Symmetry: earbuds & feet. RunCoach utilizes a single pair of earbuds to detect user's gait type, where each gait cycle is defined as a single foot strike. One limitation of this design is that it cannot identify the foot performing currently detected gait type. For example, some runners might perform different gait types at different feet during running due to a prior injury or just a wrong running gait. Thanks to the symmetric design of earbuds (left and right), it was found that this issue can be easily resolved by comparing the energy of signals from both earbuds. Specifically, if a foot strike is produced by the left foot, the signals will propagate to both earbuds, but the path to the left earbud is shorter than that to the right earbud, and vice versa. As a result, more attenuation is experienced for the right earbud and thus has a lower signal amplitude. FIG. 18(a) shows the

original waveform of left ear and right ear signals and FIG. 18(b) illustrates the root-mean-square energy (RMSE) curve calculated from the waveform. It can be clearly observed alternate peaks for the left and right ear signals, which can be mapped to the left and right foot strikes. Moreover, due to the variations in the left and right propagation paths, it is possible to combine both earbuds' signals to improve the gait type classification performance.

[00183] 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 under 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.

[00184] 9 CONCLUSION

[00185] Described herein is a earbuds-based system 100, named RunCoach, for running gait and breathing monitoring. Employing a microphone to record the in-ear sounds during running, RunCoach first extracts signals related to human gait and breathing respectively with frequency-based filtering. Then, for each signal, RunCoach adopts a well-designed signal processing pipeline and ML-based classification scheme to recognize runners' gait type or breathing modes. In addition, RunCoach 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, the superior performance of RunCoach was experimentally demonstrated, namely, 87.80% and 98.52% accuracy for gait type and breathing mode (phase-level) recognition, respectively. RunCoach was implemented on a smartphone and its run-time overhead measured.

[00186] The machine learning models described herein are trained with the data from 25 users and it is likely that with a greater number of users to generate the

data to train the machine learning model, the methods described herein may improve its accuracy and applicability.

[00187] The computer implemented method of identifying a breathing mode of a user is described in detail herein with respect to a user running. This is because the breathing signal from the in-ear microphone is inherently weaker as the inhalation and exhalation sound is further from the ear. Hence, whilst running the inhalation and exhalation of a user is generally louder which increases the signal quality. This makes the method of identifying the breathing mode easier. It may be possible for the method to be employed to the walking of a user. However, to improve the accuracy, additional breathing signal data of users walking may be needed to train the machine learning model. Similarly, the method may be applied to determine the breathing mode of a user at rest or sleeping and the machine learning model may be similarly trained, for example whether the user is breathing or snoring. The limitations of determining the breathing mode when a user is resting may be due to the sensitivity of the in-ear microphone rather than the method itself, and improvements in the microphone sensitivity may increase the applications of the described methods.

[00188] The computer implemented method of identifying a gait of a user is described in detail herein with respect to a user running and may be also applicable to the walking of a user. However, to improve the accuracy, additional gait signal data of users walking may be needed to train the machine learning model.

[00189] The method of cancelling noise in an in-ear microphone may have different applications and may be used to remove the effects of the ambient noise in the in-ear microphone signal and allow more relevant and pertinent information and data to be extracted, analysed and applied. For example, the noise cancelling method may be applied to determine the breathing mode of a user, the gait of a user, and the heart rate of a user. Other relevant applications may also use the noise cancelling method.

CLAIMS

[Claim 1] A computer implemented method of identifying a breathing mode of a user, the method comprising:

obtaining an in-ear signal data of a user;

5 filtering the in-ear signal data to obtain a breathing signal data;

segmenting the breathing signal data into a plurality of frames with a window sliding algorithm;

extracting a plurality of features from each frame in the plurality of frames;

10 feeding the plurality of features from each frame to a trained machine learning model;

obtaining a breathing mode estimation for each frame from the trained machine learning model;

applying a moving average algorithm to the breathing mode estimation for the plurality of frames to correct an error if present in the breathing mode estimation to

15 provide a plurality of corrected frames; and

identifying a breathing mode for each frame of the plurality of corrected frames.

[Claim 2] The method according to claim 1, further comprising obtaining an out-ear signal data;

20 determining from the out-ear signal data if an ambient noise level is above a threshold level;

if the ambient noise level is at or above the threshold level, estimating a residual noise level from the out-ear signal data and subtracting the residual noise level from the breathing signal data to denoise the breathing signal data before segmenting

25 the breathing signal data; and

if the ambient noise level is below the threshold level, proceeding to segmenting the breathing signal data.

[Claim 3] The method according to claim 2, wherein estimating the residual
30 noise level from the out-ear signal data comprises

obtaining a plurality of amplitude and phase values of the out-ear signal data at different frequency bands;

obtaining a corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands;

- 5 computing a transformed plurality of amplitude and phase values of the out-ear signal data at different frequency bands based on a linear correlation between the corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands and the plurality of amplitude and phase values of the out-ear signal data at different frequency bands; and
- 10 subtracting the residual noise level from the breathing signal data comprises subtracting the transformed plurality of amplitude and phase values of the out-ear signal data at different frequency bands from the corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands; and converting the subtracted plurality of amplitude and phase values to the
- 15 breathing signal data to denoise the breathing signal data.

[Claim 4] The method according to any one of claims 1 to 3, wherein filtering the in-ear signal data comprises applying a high pass filter to exclude signals of less than 500 Hz.

20

[Claim 5] The method according to any one of claims 1 to 4, wherein each frame in the plurality of frames comprises a fixed length, and preferably there is no overlap of frames in the sliding window algorithm.

- 25 [Claim 6] The method according to any one of claims 1 to 5, wherein a machine learning model to be trained is selected from K-Nearest Neighbours (KNN), SVM (Support Vector Machine), RF (Random Forest), and DT (Decision Tree), preferably the machine learning model is KNN.

- 30 [Claim 7] The method according to claim 6, wherein a training method of the trained machine learning model comprises

segmenting a training in-ear signal data with reference to a reference breathing signal to provide a plurality of segmented training in-ear signal data;
splitting each of the plurality of segmented training in-ear signal data into a plurality of training frames;

5 extracting a plurality of features from each training frame in the plurality of training frames;

applying a high-quality frame selection algorithm to the plurality of training frames to select high quality training frames in the plurality of training frames; and
inputting the selected high quality training frames into the machine learning model.

10

[Claim 8] The method according to claim 7, wherein the reference breathing signal indicates transitions between adjacent breathing phases of inhalation and exhalation and is preferably obtained from an external microphone, more preferably the external microphone is placed proximal to a nose.

15

[Claim 9] The method according to claim 7 or claim 8, wherein the high-quality frame selection algorithm comprises computing an average of the plurality of features from the plurality of training frames in each segmented training in-ear signal data; generating a feature template from the computed average of the plurality of
20 features from the plurality of training frames; determining the Minkowski distance between the plurality of features from each training frame and the feature template; and selecting high-quality frames based on the Minkowski distance.

[Claim 10] The method according to any one of claims 7 to 9, wherein the training
25 method comprises collecting the training in-ear signal data at least in part from the user.

[Claim 11] The method according to any one of claims 1 to 10, wherein applying the moving average algorithm to the breathing mode estimation is performed twice.

30

[Claim 12] The method according to any one of claims 1 to 11, wherein the plurality of features comprises Mel-frequency cepstral coefficients (MFCC) features.

5 [Claim 13] The method according to any one of claims 1 to 12, obtaining a breathing mode estimation for each frame from the trained machine learning model comprises classifying the plurality of features from each frame to determine the breathing mode estimation.

10 [Claim 14] The method according to any one of claims 1 to 13, wherein the trained machine learning model is remotely located.

[Claim 15] The method according to any one of claims 1 to 14, wherein the breathing mode estimation and/or the breathing mode is selected from oral inhalation, oral exhalation, nasal inhalation, and nasal exhalation.

15

[Claim 16] The method according to any one of claims 1 to 15, further comprising determining a respiration rate.

20 [Claim 17] A computer implemented method of identifying a gait of a user, the method comprising:

obtaining an in-ear signal data of a user;

filtering the in-ear signal data to obtain a gait signal data;

segmenting the gait signal data with an envelope-based algorithm to obtain a segmented peak;

25 extracting a plurality of features in the segmented peak;

feeding the plurality of features from the segmented peak to a trained machine learning gait model; and

obtaining a gait of the user for the segmented peak from the trained machine learning model.

30

[Claim 18] The method according to claim 17, wherein filtering the in-ear signal data comprises applying a low pass filter to exclude signals of more than 100 Hz.

5 [Claim 19] The method according to claim 17 or claim 18, wherein the envelope-based algorithm comprises applying a Hibbert transform to generate an envelope signal of the gait signal; smoothing the envelope signal; detecting a peak and its adjacent region of the envelope signal to obtain the segmented peak.

10 [Claim 20] The method according to any one of claims 17 to 19, wherein extracting the plurality of features in the segmented peak comprises performing a Fast Fourier transform (FFT) method or a Mel-frequency cepstral coefficients (MFCC) method, preferably performing the FFT method.

15 [Claim 21] The method according to any one of claims 17 to 20, wherein a machine learning gait model to be trained is selected from K-Nearest Neighbours (KNN), SVM (Support Vector Machine), RF (Random Forest), and DT (Decision Tree), preferably the machine learning gait model is KNN.

20 [Claim 22] The method according to claim 21 wherein segmenting the gait signal data with an envelope-based algorithm comprises obtaining a plurality of segmented peaks, and the method further comprises generating a count of the plurality of segmented peaks.

25 [Claim 23] The method according to any one of claims 17 to 22, wherein obtaining a gait of the user for the segmented peak from the trained machine learning model comprises classifying the plurality of features from the segmented peak to obtain the gait of the user.

30 [Claim 24] The method according to any one of claims 17 to 23, wherein the trained machine learning model is remotely located.

[Claim 25] The method according to any one of claims 17 to 24, wherein the gait is selected from a forefoot strike, an overstride, and a heel strike.

[Claim 26] A method of cancelling noise in an in-ear microphone, the method comprising:

obtaining an in-ear signal data of a user;

obtaining an out-ear signal data;

determining from the out-ear signal data if an ambient noise level is above a threshold level;

if the ambient noise level is at or above the threshold level, estimating a residual noise level from the out-ear signal data and subtracting the residual noise level from the in-ear signal data to denoise the in-ear signal data; and

if the ambient noise level is below the threshold level, determine that the in-ear signal data need not be denoised.

[Claim 27] The method according to claim 26, wherein estimating the residual noise level from the out-ear signal data comprises

obtaining a plurality of amplitude and phase values of the out-ear signal data at different frequency bands;

obtaining a corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands;

and

computing a transformed plurality of amplitude and phase values of the out-ear signal data at different frequency bands based on a linear correlation between the

corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands and the plurality of amplitude and phase values of the out-ear signal data at different frequency bands; and

subtracting the residual noise level from the in-ear signal data comprises subtracting the transformed plurality of amplitude and phase values of the out-ear signal data

at different frequency bands from the corresponding plurality of amplitude and phase values of the in-ear signal data at different frequency bands; and converting

the subtracted plurality of amplitude and phase values to the in-ear signal data to denoise the in-ear signal data.

5 [Claim 28] A data processing apparatus comprising means for carrying out the method according to any one of claims 1 to 27.

[Claim 29] A system comprising a pair of earbuds; and the data processing apparatus according to claim 28.

10 [Claim 30] A computer program comprising instructions which, when the program is executed by a computer, cause the computer to carry out the method according to any one of claims 1 to 27.

15 [Claim 31] A computer-readable medium comprising instructions which, when executed by a computer, cause the computer to carry out the method according to any one of claims 1 to 27.