RespEar: Earable-Based Robust Respiratory Rate Monitoring

Yang Liu*, Kayla-Jade Butkow*, Jake Stuchbury-Wass*, Adam Pullin*, Dong Ma†, Cecilia Mascolo*

*University of Cambridge, UK, †Singapore Management University, Singapore

{yl868, kjb85, js2372, alp78}@cam.ac.uk, dongma@smu.edu.sg, cm542@cam.ac.uk

Abstract—**Continuous respiratory rate (RR) monitoring is essential for understanding physical and mental health, as well as tracking fitness. However, performing reliable and non-obtrusive RR monitoring across diverse daily routines and activities is still an open research problem. In this work, we present RespEar, a pipeline for robust RR monitoring across various sedentary and active scenarios using earphones. RespEar relies solely on in-ear microphones, repurposing them for continuous RR monitoring purposes. Specifically, leveraging the unique properties of in-ear audio, RespEar enables the use of respiratory sinus arrhythmia (RSA) and locomotor respiratory coupling (LRC), physiological couplings between cardiovascular activity, gait and respiration, to determine the RR. This effectively addresses the challenges posed by the almost imperceptible breathing signals encountered during common daily activities. Additionally, RespEar uniquely identifies and addresses three key practical issues for the RSA and LRC-based solutions and introduces a suite of meticulously crafted signal processing techniques to enhance the accuracy of RR measurements. With data collected from 18 subjects over 8 activities, RespEar measures RR with a mean absolute error (MAE) of 1.48 breaths per minute (BPM) and a mean absolute percent error (MAPE) of 9.12% in sedentary conditions, and a MAE of 2.28 BPM and a MAPE of 11.04% in active conditions, respectively. To the best of our knowledge, RespEar is the first earable-based system capable of accurately determining RR in a variety of realistic settings.**

Index Terms—**earable, breathing rate, in-ear audio, respiratory sinus arrhythmia, locomotor respiratory coupling**

I. INTRODUCTION

Respiratory Rate (RR) is a fundamental vital sign that relays pivotal information about health and fitness conditions of the human body. Clinically, it is critical for diagnosing and managing various pathologies, acting as an early indicator of health deterioration, such as potential cardiac arrest or respiratory illnesses [1]. In daily life, RR indicates the presence of physical and mental stressors including emotional stress, emotional response, and cognitive load [1]. Additionally, RR is a key indicator of exertion levels during physical activities, offering valuable insights for managing and optimizing workout routines and detecting exercise-induced fatigue [2]. By integrating continuous RR monitoring into daily routines, individuals can seamlessly track their health and fitness across diverse settings, thereby enhancing the accessibility and practicality of vital health insights.

Existing RR monitoring solutions designed to facilitate RR monitoring in daily life primarily rely on three principles, *all with limitations and applicable only in specific scenarios.* Namely, 1) methods that detect breathing-induced body movements typically utilize IMUs [3], [4], [5], acoustic [6] or wireless signals [7], [8], [9]. However, these approaches require the user to remain still. Pressure sensors in chest straps [10] function during active conditions but require the use of additional obtrusive wearable devices. 2) Studies monitoring airflow through the nose and mouth typically use intrusive nose-worn sensors [11], or employ microphones [12], [13], which only work in the presence of audible breathing sounds. 3) For indirect measurement, studies [14], [15] use the relationship between RR and physical behaviors to indirectly estimate RR, but are applicable only to specific running scenarios. Studies [16], [17], [18] estimating RR indirectly from physiological signs only work at rest and struggle to provide reliable estimates when RR changes dynamically. Although existing RR monitoring solutions perform well under specific conditions, there is currently no unified system capable of operating seamlessly and reliably across diverse daily activities. Integrating these technologies to function consistently in various scenarios is extremely challenging. *Therefore, a new approach is needed – one that transcends the limitations of existing technologies to provide continuous, non-obtrusive RR monitoring that is genuinely effective across a range of daily activities.*

This work presents RespEar, an earable-based system offering robust RR monitoring across both sedentary (*e.g.,* sitting, standing, working, cooling down after exercise), and active (*e.g.,* walking, running, rowing, and step aerobics) activities. Considering their widespread popularity and the promising sensing location of the human ear (*i.e.,* a stable part of the human body close to the respiratory organs), we chose earbuds, a mainstream consumer wearable device with daily usage (e.g., entertainment, exercise, and work), as our RR sensing device. Moreover, we identified that in-ear microphones (measure sounds inside the ear canal) are uniquely positioned to measure breathing-related signals (*e.g.,* breathing sounds, heartbeats, and footsteps) (Section II-A), thus enabling our solution.

While designing RespEar, we faced these challenges:

1) Almost imperceptible breathing sounds. The intensity of breathing sounds is minimal when the user is sedentary and overwhelmed by other sounds, like footsteps, when the user is active (Section II-A). Thus, directly estimating RR using breathing sounds proves unreliable (Section IV). To address

this, we proposed a unified RR monitoring system through identifying in-ear audio for the following purposes:

- *RSA-based RR monitoring:* When clear heartbeat sounds can be captured using the in-ear microphone (predominantly when the user is sedentary), we can derive heart rate variability (HRV) from in-ear audio. RR is then indirectly estimated using the RSA-based physiological coupling between the cardiovascular activity and respiration, *i.e.,* the association between RR and HRV;

- *LRC-based RR monitoring:* However, when clear heartbeat sounds are not available (*e.g.,* in the presence of footstep sounds), RSA-based solutions are hindered due to unreliable HRV estimation (validated in Section IV). Therefore, when rhythmic footsteps are present (i.e., when the user is active), we rely on the in-ear microphone to capture low-frequency footstep sounds, which are used to derive the stride rhythm. Alongside faint high-frequency breathing sounds, RR is estimated by leveraging the LRC-based physical coupling between gait and respiration, specifically the interaction of RR with stride rhythm.

2) Accurate and reliable estimation. Based on the above system, several technical challenges should be addressed to achieve accurate and reliable RR estimation:

- *RSA-based solutions under varying RRs:* In the literature, deriving RR from HRV using RSA has typically relied on a fixed association between RR and HRV. However, we made a key observation that this association should adapt to changes in RR over time (Section III-B). To this end, we are the first to propose a novel algorithm for dynamically extracting the breathing signal from HRV by searching for the best association through formulating and solving an optimization problem, thereby adapting to the variability in the relationship between RR and HRV.

- *Respiration-related feature extraction for LRC-based RR monitoring:* Although LRC shows synchronization between stride rhythm and RR, the variable and unknown LRC ratio prevents direct RR estimation from stride frequency alone. It necessitates extracting respiration-related features from in-ear audio, combined with stride rhythm, to accurately estimate RR. However, while the user is under active conditions, faint breathing sounds are heavily interfered with by other sounds, such as footsteps. To address this challenge, we propose an advanced processing pipeline that estimates the probability of each in-ear audio frame containing breathing by leveraging our uniquely extracted audio features, compared against a defined breathing template. This generates an effective probability curve, which can then be decomposed to reveal the underlying breathing pattern.

- *Varying LRC ratio:* RR is typically estimated for a window during which the LRC ratio (*i.e.,* the ratio of stride rhythm to the number of breaths) may vary. We are the first to address this variability by applying Singular Spectrum Analysis (SSA) [19] to the generated probability curve to isolate components related to breathing, and propose a method that aggregates these breathing-related components from SSA by considering a range of possible LRC ratios, rather than
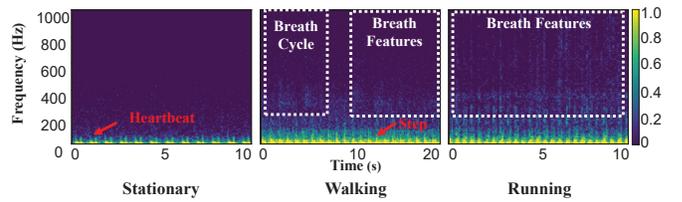

Fig. 1: In-ear audio under different activities.

choosing a fixed one.

We implemented RespEar with an earable prototype and measured the system overhead through deploying it on an iPhone 12 Pro. RespEar was evaluated across 8 different activities involving 18 subjects, achieving an overall MAE of 1.71BPM and MAPE of 9.68%, with errors of 1.48BPM (9.12%) and 2.28BPM (11.04%) under sedentary and active conditions, respectively. We compared RespEar with recent earable-based solutions using IMUs [20], out-ear microphones [21], [14], and in-ear microphones [13], and found that RespEar outperformed in both sedentary and active conditions while additionally being able to cater for both sets of scenarios, unlike these recent works limited to functioning only under specific conditions. Moreover, we tested RespEar under a range of realistic conditions.

In summary, this paper makes the following contributions:
- We propose RespEar, the first wearable system offering continuous and non-obtrusive RR monitoring across diverse daily routines and activities.
- Leveraging solely the in-ear microphone, a sensor naturally present in many earables, we present a holistic and optimized solution for RR estimation which leverages intrinsic relationships of our cardiovascular, gait and respiratory systems and uniquely identifies and addresses three key practical issues.
- We implement RespEar and describes our extensive dataset and evaluation. Our results demonstrate that RespEar outperforms the state-of-art and is uniquely able to generalize beyond what other systems have been able to do in terms of activity intensity while remaining robust under different environmental conditions.

II. PRELIMINARY INVESTIGATION

A. In-ear Microphone

We first investigate the signals captured by in-ear microphones for RR monitoring, which are typically used for active noise cancellation (ANC) in commercial earphones [22].

We collect in-ear audio when a subject was naturally breathing under three conditions: sitting still, walking, and running on a treadmill. The Zephyr BioHarness 3.0 chest strap [10] was worn to collect reference signals. As spectrograms shown in Figure 1, under stationary conditions, the faint breathing sounds are undetectable by in-ear microphones. However, sounds of heartbeats are clearly captured due to the occlusion effect [23]. When walking or running, we observe that 1) footsteps can be clearly detected because of the occlusion effect [24]; 2) due to the variations in breathing intensity, only certain breathing sounds are discernible, while the full breathing cycle cannot be

distinguished; 3) the in-ear microphone is resilient to ambient noise as it resides inside the ear canal.

B. Design Primer

In-ear microphones can capture versatile audio data in different conditions, yet the method to accurately correlate these signals with RR estimation remains unclear. Inspired by the physiological couplings between cardiovascular activity and respiration [16], *e.g.,* RSA, we are initially exploring methods grounded on this physiological principle.

1) RSA-based RR Estimation: RSA is the natural variations in HR that occur due to synchronization with the respiratory cycle [16]. Due to RSA, HR increases during inhalation and decreases during exhalation, resulting in a breathing related modulation of the HRV (*i.e.,* the variation in time interval between successive heart beats) [25]. Rhythms in the low frequency (LF) range of the HRV, spanning from near 0.04 to 0.15 Hz, serve as indicators of sympathetic modulation [26]. Those within the high frequency (HF) range (near 0.15 to 0.4 Hz) encapsulate rhythms governed by parasympathetic activity, which is closely related to respiration [26].

The clearly captured heartbeat sounds from in-ear microphones under sedentary conditions offer the possibility of monitoring RR using RSA. RSA exists under intense full body motions [27] (*i.e.,* in active conditions); however, accurately extracting heartbeat locations from in-ear audio for HRV estimation in such conditions remains an unsolved and challenging issue [23], [28], [29], as validated in Section IV. Thus, it is necessary to identify an alternative method for RR monitoring in active conditions.

2) LRC-based RR Estimation: We observe that audio from in-ear microphones under active conditions, such as walking, running, or other activities with rhythmic footsteps, is dominated by footstep sounds and breathing sounds are not always discernible (Section II-A). Thus, we explore the physiological couplings between gait and respiration here.

LRC is a universal phenomenon in activities that produce and utilize energy rhythmically [30], such as walking, running, swimming, and rowing [31], [32]. It reveals the interconnected dynamics between RR and stride rhythm [33], indicating the synchronization between an individual's stride rhythm and their RR. This implies that there will normally be a certain number of steps for each breath (*i.e.,* inhalation or exhalation). In human locomotion, a number of LRC ratios are observed, *e.g.,* 4:3, 3:2, 2:1, where an LRC of 2:1 means two steps are taken for one breath. Thus, the in-ear audio, containing clear footsteps and partly discernible breathing sounds, offers the possibility for estimating RR based on LRC.

III. SYSTEM DESIGN

Figure 2 shows the system architecture. RespEar uses 60s estimation windows with 30s overlap to produce a RR estimate per window. RespEar uses two paths for RR estimation depending on the presence of rhythmic footsteps or clear heartbeat sounds in the audio. If clear heartbeat sounds are present (*i.e.,* sedentary conditions), RR will be estimated
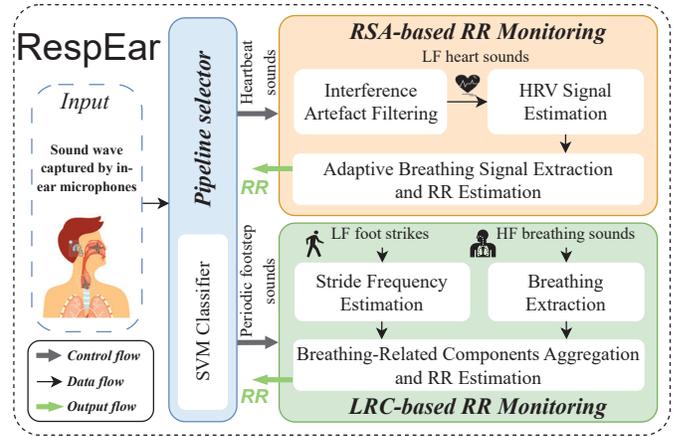

Fig. 2: Illustration of the RespEar architecture.

through the **RSA-based RR monitoring** pipeline. If rhythmic footsteps are present (i.e., active conditions), RR will be estimated using the **LRC-based RR monitoring** pipeline.

A. Pipeline Selector

The pipeline selector determines which processing pipeline should be selected for RR estimation, *i.e.,* RSA or LRC-based, based on the presence of either clear heartbeat sounds or footsteps in the input signal. If the LRC-based pipeline is selected, we further differentiate this into low-intensity and high-intensity rhythmic footstep activities, so that the correct algorithmic parameters can be applied to the pipeline.

We train our pipeline detector using support vector machines (SVM). The in-ear audio is split into 5s segments and Mel-frequency cepstral coefficients (MFCCs) are extracted from each segment and used as the input features to the SVM. We use a two-stage classifier whereby first we classify a segment as sedentary (*i.e.,* strong presence of heart sounds) or active (*i.e.,* strong presence of rhythmic footstep sounds). If active, we further classify it into low-intensity and high-intensity rhythmic footstep activities. Consequently, there are 12 detection results from one model during each 60s estimation window, and we determine the scenario of the whole window through majority voting. Specifically, we empirically determine that only consistent results obtained for more than 75% of segments leads to reliable pipeline selection. Voting aims to handle transition windows between two states that could result in inappropriate pipeline selection. If there is no convergence, the window will be discarded.

B. RSA-based RR Monitoring

1) Design Principle: The high-level process of RSA-based RR monitoring can be summarized in three steps: **1) HRV signal estimation**: Heartbeats are detected, and the HRV signal is computed as the time difference between successive heartbeats. **2) Breathing signal extraction**: A bandpass filter (BPF) isolates respiration-related rhythms in the HRV signal, extracting the high-frequency (HF) range as the extracted breathing signal. **3) RR estimation**: The final RR is determined
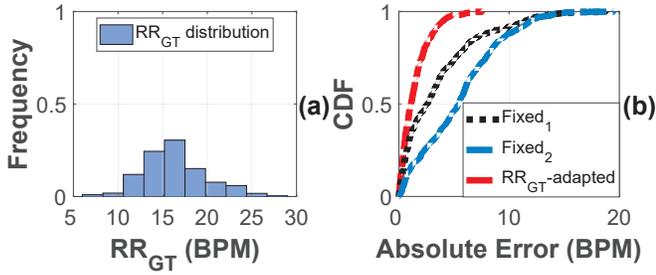
Fig. 3: (a) Distribution of RR_{GT} while sedentary. (b) Comparison of RR estimation performance using different BPFs.

Fig. 4: (a) In-ear audio and heartbeats. (b) Peak detection.

either by counting peaks in the breathing signal or by using Fast Fourier Transform (FFT) to identify the dominant frequency.

Observation. Prior RSA-based methods using photoplethysmogram (PPG) or electrocardiogram (ECG) signals [16], [18], [34], [17] typically use default and fixed cutoff frequencies for the BPF. However, we observe that this leads to suboptimal RR estimation when the true RR shifts. Since the HF range should be centered around the true RR, when the true RR changes over time, the HF range should change accordingly [26].

To validate this, we applied the RSA-based RR monitoring method to in-ear audio data collected while sedentary, following the steps outlined above. The distribution of the Ground Truth (GT) RR (denoted as RR_{GT}) for this data is shown in Figure 3(a), where the minimum RR is 6BPM (0.1Hz) and the maximum is 29BPM (0.5Hz). We compared RR estimation using: (i) a fixed BPF range from prior studies ($Fixed_1 = [0.15, 0.35]Hz$), (ii) a BPF covering the full RR range ($Fixed_2 = [0.1, 0.5]Hz$), and (iii) an adaptive BPF centered on the true RR (RR_{GT}-adapted). There is a large performance gain by using the RR_{GT}-adapted frequency range (MAE = 1.45BPM) compared to the $Fixed_1$ (MAE = 3.54BPM), and the $Fixed_2$ (5.45BPM), as shown in Figure 3. This is because the fixed BPF is effective only if RR_{GT} falls within the BPF range, but even then, a non-centered HF range can degrade the performance.

Our design. We propose a novel approach whereby we formulate and solve an optimization problem to dynamically localize the HF range. To the best of our knowledge, RespEar is the first work to achieve dynamic HF range localization for RSA-based RR estimation. We believe our methodology could also benefit other RSA-based solutions using various sensing modalities, *e.g.,* ECG and PPG. We also propose a series of techniques to enable the full pipeline for RSA-based RR monitoring using in-ear audio.

2) HRV Signal Estimation: **Heartbeats detection.** A low-pass filter with 30Hz cutoff is used to remove high frequency noise from the in-ear audio [23] (Figure 1). To obtain the HRV signal (denoted as S_{HRV}), the heartbeats need to be identified by detecting the peaks in the filtered audio (Figure 4(a)). To accurately detect peaks while accommodating variations in amplitude and morphology changes, RespEar uses peak detection with an adaptive peak detection threshold. As shown
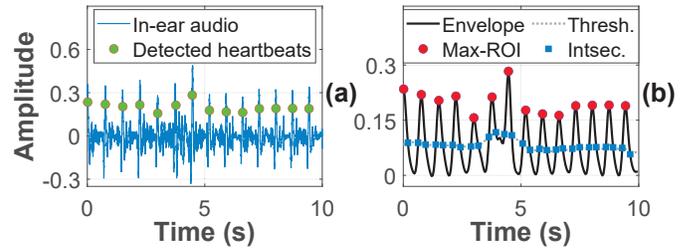
in Figure 4(b), we first compute the smoothed Hilbert envelope of the filtered audio. Next, a moving average is computed for each point in the envelope, serving as an adaptive threshold (Thresh. in Figure 4(b)). Regions of interest (ROIs) are identified between the points where the envelope intersects the threshold. Heartbeat peaks are marked at the maximum point within each ROI (Max-ROI) between two intersection points (Intsec.), provided the amplitude is greater than that of the two surrounding intersection points. Finally, the HRV signal, S_{HRV}, is calculated as the time difference between successive detected peaks (Figure 5(a)).

Automatic channel selection. Leveraging the unique ability of earables to give two in-ear audio channels (*i.e.,* left and right ear), RespEar automatically selects the channel with lower standard deviation (STD) in the estimated HRV signal from each ear. This is because heartbeats are regular signals, so the lower STD implies a less noisy and more robust signal due to more regular heartbeat peaks.

3) Adaptive Breathing Signal Extraction: **Design insight.** To enable adaptive HF range localization on S_{HRV}, we formulate an optimization problem: from a list of RR candidates, the best RR candidate minimizes the difference between its estimate and the true RR where its estimate is the one made using a BPF with a frequency range centered around the candidate.

Our algorithm. To solve this optimization problem, we propose the following algorithm:

i) RR candidate sampling: We filter S_{HRV} using a BPF with cutoffs [0.15,0.35]Hz, and perform a FFT on the filtered signal. The frequency component with the highest amplitude is converted to BPM to determine RR_c, the central RR candidate. The list of candidates (RR_{list}) is generated through incrementing or decrementing RR_c with a step of 0.5BPM until the following conditions are met, where 7.5BPM and 42.5BPM are the smallest and largest human RRs respectively [35], and w is the predefined length of RR_{list}:

$$\min(RR_{\text{list}}) = \max(7.5, RR_{\text{c}} - w/2) \quad (1)$$
$$\max(RR_{\text{list}}) = \min(42.5, RR_{\text{c}} + w/2) \quad (2)$$

ii) Best RR candidate search: For each candidate (RR_{list}^i), filter S_{HRV} using a BPF with low (l^i) and high (h^i) cutoffs defined by that candidate according to:

$$l^i = 0.65 \cdot RR_{list}^i/60; h^i = 1.35 \cdot RR_{list}^i/60 \quad (3)$$

We then perform a FFT on the filtered signal ($Breath^i$) to estimate the respiratory rate (RR_{est}^i) by selecting the frequency
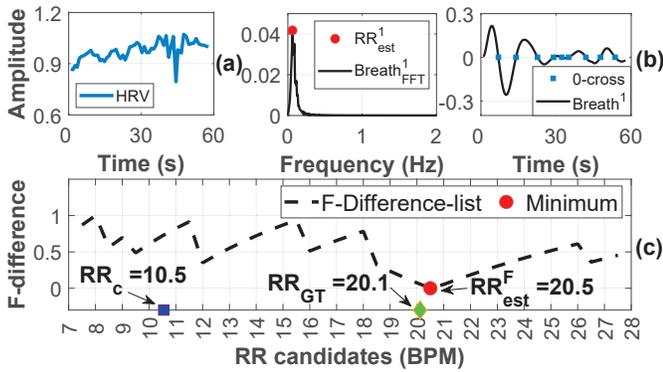
Fig. 5: (a) Extracted HRV signal. (b) FFT determination and zero crossing counting. (c) Best RR candidate searching.

Fig. 6: (a) Best RR candidate searching on the frequency domain. (b) Calibration from the time domain.

component with the highest amplitude and converting the frequency to BPM. Figure 5(a-b) illustrate the calculation of RR_{est}^1. The difference between each candidate and its estimated respiratory rate is calculated by $|RR_{est}^i - RR_{list}^i|$.

This process is repeated for all candidates, and used to create the frequency difference list (F-difference list), which is then min-max normalized as shown in Figure 5(c). Finally, the candidate with the smallest difference is selected as the best RR candidate (RR_{est}^F in Figure 5(c)).

iii) Calibration from time domain: As shown in Figure 6(a), sometimes we find RR_{est}^F, yet is not the optimal RR candidate, likely due to the lower quality of S_{HRV}. To enhance robustness, we further repeat *Step ii)* but we estimate RR of $Breath^i$ in the time domain by counting the zero-crossing points, as shown in Figure 5(b). This generates a time difference list (T-difference list). We observe that the optimal RR candidate typically appears among the three candidates with the top-3 smallest local minima in the F-Difference list (Figure 6(a)). Furthermore, the T-Difference list suggests the trend leading towards the optimal candidate. Therefore, we first smooth the T-Difference list to highlight its underlying trends, and then sum the T-difference value and the F-difference value of these three candidates respectively. We select the candidate with the smallest sum as the estimated RR as depicted in Figure 6(b).

4) Interference Artefact Filtering: While the body is undergoing non-full body motions (*i.e.,* the user is sedentary), in-ear audio is prone to interference artefacts. Examples of these include head motions (drinking, speaking *etc.*), motions of the arms, movement of the trunk, respiratory-related sounds (swallowing, coughing, *etc.*). It must be noted that breathing adapts to speaking (inhalation at syntactic pauses and exhalation during speech [36]), which eliminates the need for respiratory rate estimation while speaking. To ensure accurate sedentary RR estimation even in the presence of interference artefacts, we present an interference filtering approach.

Interference artefact detection. Without interference artefacts, in-ear audio maintains a consistent waveform with clear heart sounds and thus stable statistics over time. Conversely, artefacts, such as one-time head motion, cause significant statistical variations. Hence, we propose a statistics-based
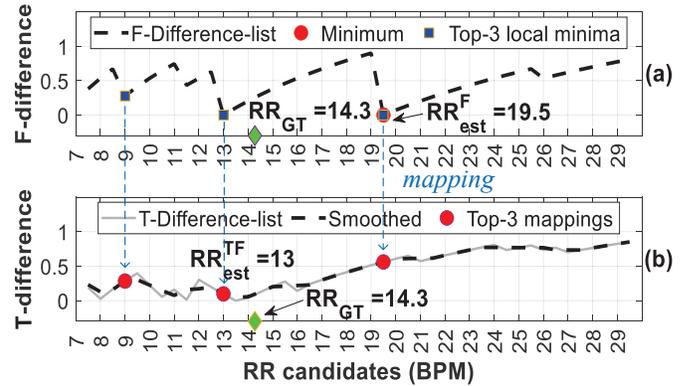
approach to detect the presence of artefacts. For each 60s window, we segment the audio signal into 3s segments. We compute STD of each segment (to measure signal dispersion) and if it is larger than an empirical threshold, this segment is marked as an interfered segment.

Adaptive filter. If an interfered segment is detected, we use an adaptive filter implemented using the recursive least squares (RLS) algorithm [37] to remove the interference artefact. The RLS algorithm recursively finds filter coefficients that minimize a least squares cost function with a reference signal. For the interfered segment, we select the nearest segment without interference as the reference signal. After filtering, the interference artefacts in the segment are mostly removed for reliable RR estimation.

C. LRC-based RR Monitoring

1) Design Challenges.: Although LRC indicates synchronization between stride rhythm and RR, the LRC ratio between stride rhythm and RR is variable and unknown. Inspired by previous studies [14], [15] which linked stride frequency with breathing signals to estimate LRC ratio (but only under running scenarios with assuming a fixed LRC ratio per estimation window), we propose our pipeline. This pipeline addresses two unique challenges identified in RespEar working scenarios:

- Respiration sounds are strongly interfered by other sounds in in-ear audio, especially footstep sounds which are strong and amplified due to the occlusion effect [24] (Figure 1).
- The LRC ratio varies within an estimation window, *e.g.,* for walking and non-regular runners, as demonstrated in Figure 7. We analyze in-ear audio from two participants (User A, a non-regular runner, and User B, a regular runner) by segmenting their audio into 10-second intervals and calculating the mean LRC for each. We then compute the *changing ratio* of mean LRC values between adjacent segments to gauge irregularity. As shown in Figure 7, this variability is especially evident in non-regular runners and during walking. Thus, a constant LRC ratio cannot be assumed, requiring our system to adapt to changing LRC ratios within a single estimation.

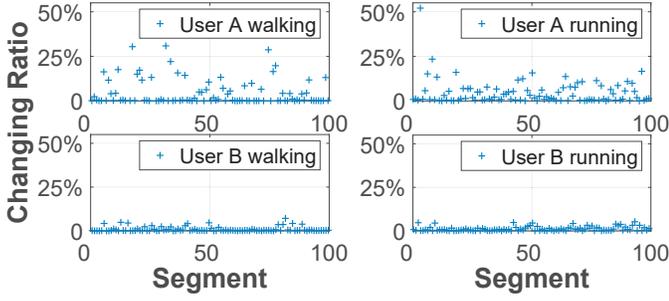We elaborate on our pipeline in the following sections.

Fig. 7: Changing ratio of GT LRC from two users.

Fig. 8: (a) Probability curve. (b) One component related to step. (c) Extracted breathing pattern with peak detection.

2) Stride Frequency Estimation: We first conduct footstep detection, which is done using the same approach as heartbeat detection, as discussed in Section III-B2, while we filter with a 50Hz low pass filter [23]. By counting the number of detected footsteps, the stride frequency can be estimated.

3) Breathing Extraction: The breathing-related features are then extracted from the in-ear audio as follows.

Pre-processing. Human breathing sounds typically fall within the range of 300Hz to 1800Hz [38]. Therefore, a BPF with cutoff frequencies from 300Hz to 1800Hz is used on the input audio during low-intensity rhythmic footstep activities, such as walking. For high-intensity rhythmic footstep activities, where step sounds severely overwhelm the breathing sounds in this frequency range, such as during running, we use a BPF with cutoffs of 2000Hz to 9000Hz to capture harmonics of breathing sounds in these higher frequencies.

Breathing template generation. We generate a breathing template (*i.e.,* signal features of a strong, clear breathing sound) to identify the probability of each frame containing breathing within the estimation window. To generate this template, we collected in-ear audio from a single user while sitting stationary and breathing loudly in a quiet environment. We conduct *Pre-processing* and performed *FFT feature generation* on it to generate the breathing template.

FFT feature generation: We divide the audio window into 40ms frames with a 20ms overlap and calculate the periodogram of each frame. Thereafter, we subdivide the breathing frequency range into 15 bins and sum the signal power in each bin from the periodogram. We therefore generate a feature vector with 15 features for each frame, one corresponding to each frequency bin. The breathing template is finally calculated by averaging the feature vectors of all frames.

Probability curve generation. For each estimation window, we perform the *FFT feature generation* for all frames within it. For each feature vector (*i.e.,* corresponding to each frame), we calculate its similarity (S) with the breathing template using the cosine similarity [14]. Then, the probability of this frame containing breathing ($P(f)$) is computed as:

$$P(f) = \begin{cases} \frac{S-T}{1-T} & \text{if } S > T \\ 0 & \text{if } S \leq T \end{cases} \quad (4)$$

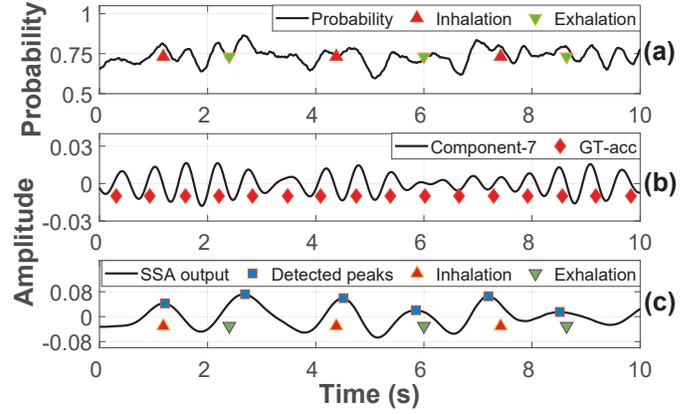where T is a predefined threshold. The probabilities of all

frames within the estimation window generate a breathing probability curve as shown in Figure 8(a).

Probability curve decomposition. Due to the low SNR from strong interference from footstep sounds and light breathing sounds, the breathing pattern is overwhelmed by patterns of interference (Figure 8(a)). To remove the interference patterns, we decompose the probability curve into its constituent components using the SSA algorithm [19]. SSA is able to effectively separate the underlying components of the curve, allowing for the isolation of periodicity occurring at various time scales, in order of significance, even within highly noisy time series data. Figure 8(b) shows component 7 of the decomposed probability curve which corresponds to the steps taken while the user is walking (*i.e.,* GT accelerometer data in Figure 8(b) - "GT-acc"). Once the probability curve is decomposed into periodic components, we exclude the components related to interference and aggregate the breathing-related components for respiratory rate estimation.

4) Breathing-Related Components Aggregation: RespEar leverages a loose constraint which adapts to changes in the LRC ratio to exclude respiration-unrelated components. Specifically, for each decomposed component of the probability curve, we count the number of peaks using peak detection. If the number of peaks falls outside the range of the minimum possible breathing rate (RR_{min}) to the maximum possible breathing rate (RR_{max}), this component is regarded as respiration-unrelated and removed. RR_{min} and RR_{max} can be computed as:

$$RR_{min} = (SF_{est} * (N/fs))/LRC_{max} \quad (5)$$
$$RR_{max} = (SF_{est} * (N/fs))/LRC_{min} \quad (6)$$

where SF_{est}, fs and N are the estimated step frequency, sampling rate, and number of samples in the estimation window, respectively. LRC_{max} and LRC_{min} are the largest and smallest values of the LRC ratios in humans.

We use the LRC range of 1.9 to 4.9 for low-intensity rhythmic footstep activities [39], and 1.8 to 5.6 for high-intensity rhythmic footstep activities [40]. These ranges cover common LRC ratios in humans under each set of scenarios [39],
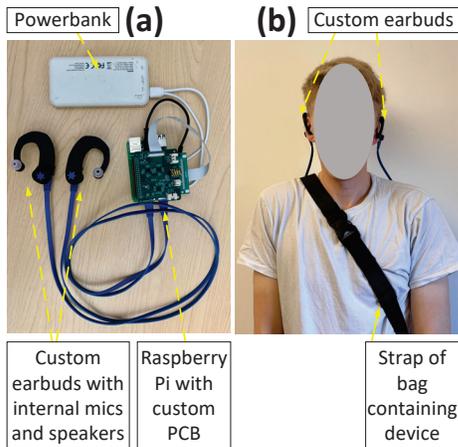
Fig. 9: (a) Custom hardware prototype and (b) one participant wearing the device.

Fig. 10: Bar plot of (a) MAE and MAPE. (b) Bland-Altman plot of RespEar.

[40], and fully cover the LRC ratios appearing in our collected dataset. After excluding all breathing-unrelated components, we sum the remaining components into the extracted breathing pattern. Peak detection is then applied to this signal to estimate the final RR (Figure 8(c)).

IV. IMPLEMENTATION AND EVALUATION

A. Implementation

1) Prototyping: Although in-ear microphones are becoming ubiquitous in ANC earbuds for active noise cancellation, no commercial earbud grants access to the raw data. As such, we developed our own custom earbud prototype to collect data which consists of 3D printed earbuds containing microphones facing inside the ear canal. On the left earbud, we embedded a speaker behind the microphone to enable audio playback. We capture microphone signals using a Raspberry Pi 4 with an audio codec hat and a custom PCB. To make the system portable, we placed the Raspberry Pi into a chest-worn bag and power it with a portable power bank (Figure 9).

2) Data Collection: 18 participants (9 male and 9 female) participated in our data collection which was approved by the Ethics Committee of Department of Computer Sciencec and Technology, University of Cambridge (No. 2029). The participants' ages ranged between 22 and 51. The participants underwent sedentary and active activities, with each activity was performed for 5 minutes. The sedentary activities are: (1) sitting; (2) standing; (3) lying down; (4) listening to music (performed for the duration of one song); (5) working in the wild; (6) uncontrolled cooldown after exercise. The active activities are (7) walking; and (8) running. The activities encompass typical scenarios when a person uses earables and requires RR monitoring. Our data collection involved controlled activities and also in-the-wild scenarios to ensure the applicability of our methods to real-world use. No breathing rates were imposed, and participants were free to breathe as they wished. Breathing after exercise was performed immediately after running to capture natural cool-down breathing. While sitting and standing,
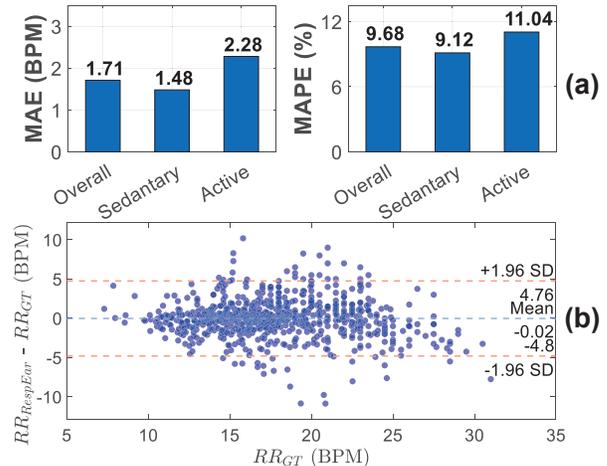
users were asked to move their heads three times to capture head motions to assess the impact of our interference artefact filtering algorithm. Active activities were performed on a treadmill and participants chose their comfortable paces to walk/run. We use the Zephyr BioHarness 3.0 chest strap [10] to collect GT RR with 25 Hz sampling rate and the audio data from in-ear microphone is collected at 22050 Hz.

B. Evaluation

1) Metrics: We evaluate system performance using the Mean Absolute Error (MAE) [13], [20] which is the average absolute error between the GT RR and the calculated RR for each estimation window. We also use the Mean Absolute Percentage Error (MAPE), the average percentage absolute error.

2) RespEar Overall Performance: **Overall performance.** We present the overall performance of RespEar in Figure 10(a). RespEar achieves an overall MAE of 1.71BPM (MAPE of 9.68%), with a MAE of 1.48BPM (9.12%) and 2.28BPM (11.04%) for sedentary and active respectively. The Bland-Altman plot for RespEar is provided in Figure 10(b). We achieve a very low mean error of -0.02BPM with narrow limits of agreement of -4.8 to 4.76. This indicates very good agreement between RespEar and ground truth breathing rate measurements, highlighting the strength of our system.

Performance per activity. Figure 11(a) provides a boxplot of the overall performance of RespEar for each activity. The performance of each sedentary activities are comparable. Slightly higher errors exist while listening to music (MAE=1.98BPM), and working (1.56BPM). This is because working is an uncontrolled activity and thus participants were more active during this task, leading to more interference artefact. The estimation errors while walking and running are satisfactory, *i.e.,* walking (MAE = 1.75BPM; MAPE = 9.17%), and running (MAE = 3.12BPM; MAPE = 14.01%). The slightly higher running errors are due to the higher footstep interference.

Individualised performance. Figure 11(b) reports the overall performance of RespEar for each participant. It is
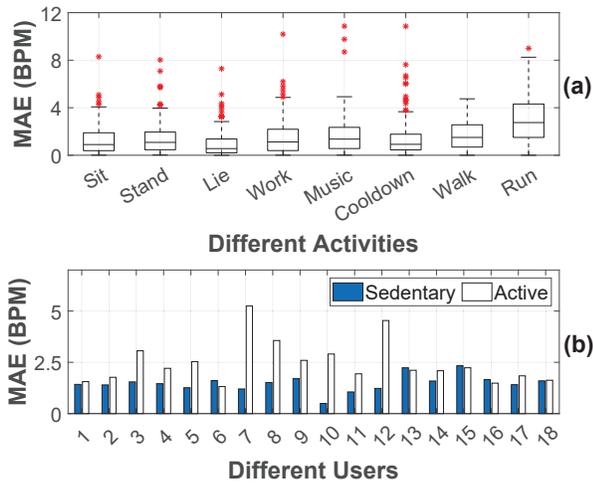
Fig. 11: RR estimation errors for (a) different activities and (b) different users.

Fig. 12: Baseline comparison: (a) IMU-based, and (b-c) audio-based approaches.

evident that the MAE while sedentary is consistent amongst users, with no user exceeding 2.3BPM error. There is more variation amongst estimation errors while active: the smallest MAE is 1.26BPM for user 16, with the largest MAE being 5.25BPM for user 7. The majority of error comes from 2 users (user 7 and 12). User 7 ran at 5KPH, which has slightly worse performance than faster running speeds. This may be attributed to the lower SNR of low-speed running, which causes similar interference from running footsteps but induces weaker breathing sounds compared to higher speeds. User 12's running generated large noise because their feet kept hitting the side of the treadmill, resulting in high-energy noise across all frequencies in the in-ear audio. However, regardless of this, the system still generalizes well for the majority of users.

3) Baseline Comparison: **Earable-based RR estimation.** We first compared RespEar with the state-of-art earable-based RR estimation solutions using IMUs [20], out-ear microphones [14], [15], and in-ear microphones [13], respectively, which are designed to function under specific conditions. We collected data from 11 users in both indoor and outdoor settings. The data was gathered during a range of activities, including sitting still and cooling down after running (sedentary), as well as walking and running (active). We implemented the three IMU based algorithms for RR estimation under sedentary conditions in [20]: an FFT approach (FFT), a peak detection approach (Peak) and a zero-crossing rate (ZCR) approach (Figure 12(a)). For the in-ear and out-ear microphones, we implemented the state-of-art algorithm for sedentary estimation in [13] (Figure 12(b)). For active, we implemented the algorithm employed by [14], [15] (LRC), and expanded upon it to calculate RR (Figure 12(c)). From Figure 12 , it is evident that under both sedentary and active scenarios, our system significantly outperforms these methods, while additionally being able to cater for both sets of conditions, unlike these recent works limited to functioning only under specific conditions.
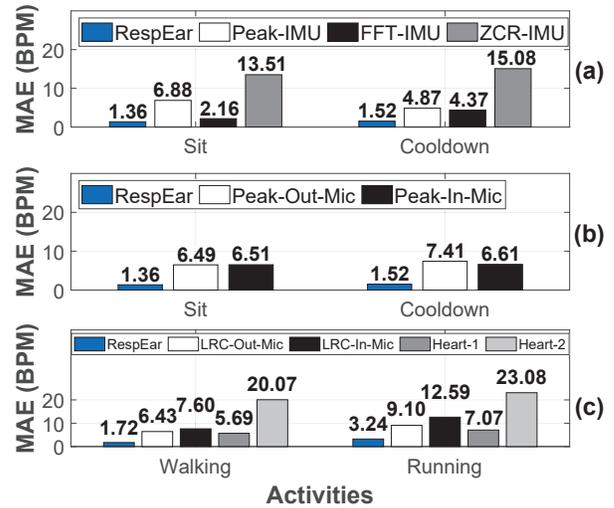
Motion-resilient heart rate (HR) estimation. We then reproduced two motion-resilient HR estimation methods [23], [29] to assess their applicability in RespEar while active. Specifically, we derived HRV from the generated "ECG" signal by [23] during walking and running, integrating it into our RSA-based RR monitoring pipeline (Heart-1). Following [29], we eliminated walking/running frequency components from in-ear audio and then applied a [0.1, 0.8]Hz band-pass filter for RR estimation (Heart-2). Figure 12(c) shows that these methods cannot be applied for RR monitoring while active, necessitating our LRC-based pipeline. This is because, although HR can be estimated in active conditions, accurately estimating HRV by identifying precise heartbeat locations during activity remains an unsolved and challenging issue. Moreover, simply removing motion-related frequency components is not sufficient for RR estimation due to the faintness of the breathing sounds.

4) Benchmark Evaluations: **Other Active Activities.** We assessed RespEar's performance when users engaged in other activities of varying intensities involving rhythmic footsteps, including step aerobics (StepA), climbing up and down stairs (StairUD), and rowing on an indoor rower (Row). We recorded 5 minutes of data per activity per participant, with activities performed in an uncontrolled manner. Figure 13(a) indicates RespEar works properly during different rhythmic activities using the LRC-based pipeline, demonstrating RespEar's effectiveness for activities with rhythmic footsteps.

Outdoor Performance. We also assessed performance outdoors in an uncontrolled environment. The tests were performed on a concrete pavement outside an academic building next to a building site with active construction happening. We assessed the performance under different activities, including sitting, walking, running and cooling down. We recorded 5 minutes of data per activity for each participant. When walking and running, participants were free to select their preferred pace and move around the area. There were thus natural
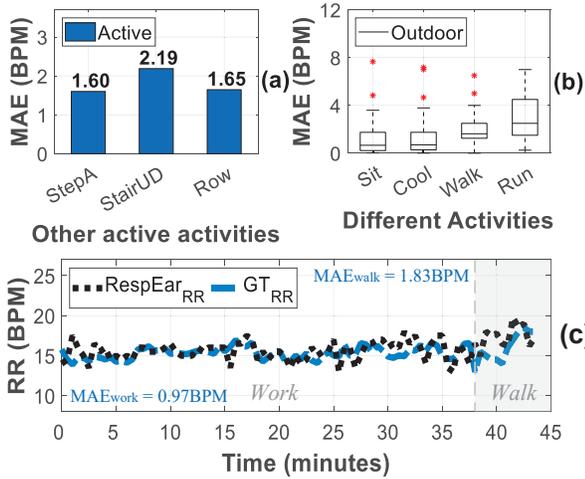
Fig. 13: Errors of (a) other rhythmic activities, (b) outdoors, (c) in-the-wild tracking.

Fig. 14: Performance of system components: (a) Pipeline selector, (b) HRV, (c) stride frequency.

changes in pace throughout the experiment to test whether our system functions under both controlled and uncontrolled speeds. Figure 13(b) shows the results, indicating RespEar achieves robust performance outdoors.

5) In-the-Wild Performance: We asked users to wear the device for near an hour in a busy office while undergoing standard daily activities. The users worked at their desks, listened to music, walked around the office, and performed other activities as they wished, such as sipping coffee or using their cell phones. Figure 13(c) showcases the tracking performance for one user, *i.e.,* continuously estimated RR from RespEar (one RR estimation per 30s) compared with GT_{RR}. This user achieved a MAE of 0.97BPM and 1.83BPM while working and walking respectively. These errors are consistent with the results obtained in the laboratory study for this user, proving that RespEar has excellent performance both in controlled laboratory settings and uncontrolled, real-world settings. It is also clear that RespEar can accurately track RR longitudinally, even in an uncontrolled setting.

6) System Components Evaluation: **Pipeline Selector.** Figure 14(a) provides the results of our pipeline selector using SVM on 5s segments. The SVM is trained on 13 users' data (randomly chosen during training) and tested on the remaining 5 users to ensure user independence of the train and test sets. We implemented 5-fold cross validation and report the average results over 5 folds. Our system is able to select pipelines with excellent performance, achieving 100% accuracy for determining whether a window is active or sedentary (SVM-Stage-I), and 99% accuracy for determining whether an active window is low- or high-intensity footstep activities (SVM-Stage-II). With majority voting of results across an estimation window, the detection accuracy on both tasks is 100%.

Accuracy of HRV Estimation and Stride Detection. To assess the accuracy of our HRV estimation, we compute the MAPE between the GT HRV from ECG and our estimated HRV on beat-to-beat level (Figure 14(b)), where we compare perfor-
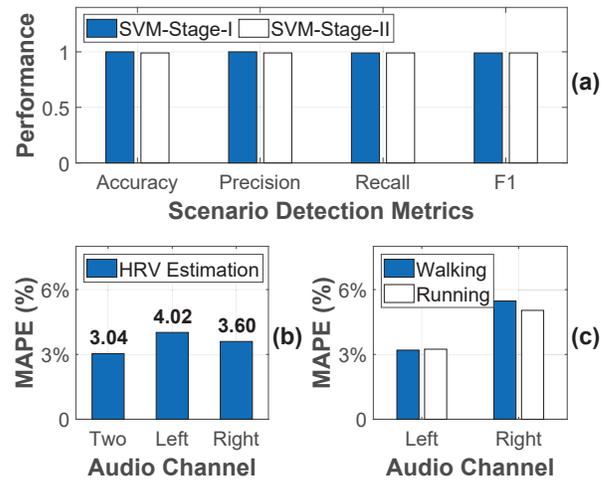
mance on the left channel, right channel and *automatic channel selection*. Using *automatic channel selection*, we achieve the best performance with a MAPE of 3%, which is competitive with reported results for ECG on well-known datasets, *e.g.,* the MIT-BIH Arrhythmia Database [41]. Our results for stride frequency estimation are provided in Figure 14(c). Our system detects strides with a MAPE of less than 3% for both walking and running using channel fusion, again competitive with literature on in-ear step counting [24].

7) System Overhead on Smartphone: To create a portable solution, we deployed RespEar on an iPhone 12 Pro with 8GB of memory, and a 2477 mAH battery capacity (2815 mAH battery with 88% battery health) to measure system overhead. RespEar's latency is 3.11s per window while sedentary and 12.27s per window while active. Since processing occurs in 60s windows, with a 30s overlap, a new RR estimate can be made every 30s under both conditions, implying that our system can run in real time. We have not considered possible data transfer costs via radio frequency (*e.g.,* BLE) which would add additional small delays depending on scenarios. When run continuously for an hour, RespEar consumed 4% and 14% battery for the RSA-based and LRC-based pipelines, respectively. The LRC-based pipeline consumes more power due to the longer latency on account of the more complex algorithm. To contextualise this, if playing music for an hour, the same phone battery decreased by 8%, showing that our application lies within standard levels for battery consumption. However, standard processing optimizations could be applied to reduce the power consumption in the future. Additionally, RespEar consumes a maximum of 49MB per window equating to 0.3% of the available memory on the device. Overall, we see that RespEar can feasibly be run on a smartphone longitudinally to potentially provide real-time RR monitoring.

V. RELATED WORK

A. Non-earable based RR monitoring

Smartwatch. IMUs: [3], [42] use IMUs on smartwatches for RR monitoring. [3] aims to estimate RR only under stationary conditions. [42] demonstrates the possibility of RR estimation under both stationary and walking conditions using learning-based techniques. However, it detects and rejects sensor data that are unsuitable for RR extraction, resulting in poor data retention. Some commercial smartwatches [43], [44] have integrated the RR estimation function, which works effectively only at rest. **PPG:** [45], [46] perform RR estimation on smartwatches using PPG. [45] employs learning-based solutions, working under both sedentary and moving conditions, but it struggles to provide reliable estimations during moving activities (MAE = 3.94 BPM). [46] works only for scenarios where the user is engaged in discontinuous activities while sitting. *Smartwatch-based solutions have shown promise with learning techniques, yet their reliability is compromised by motion artifacts, leading to low data retention rates or high estimation errors, particularly under moving conditions.*

Smartphone. [47], [48], [12], [6] employ sensors on smartphones for RR monitoring. These works require the user holds the smartphone against the chest [47], [48] or abdomen [47], utilizing IMUs to capture breathing movements. [12] records breathing sounds by placing the built-in microphones and headset microphones of the smartphone near the suprasternal notch and nose, respectively. [6] utilizes active acoustic sensing to monitor breathing-induced chest movements, requiring the smartphone to be held or placed in a specific posture. *These solutions can only operate under stationary conditions and preclude continuous and real life monitoring due to the requirement of active user involvement.*

B. Earable-based RR monitoring

PPG: [49] employ in-ear PPG for RR monitoring, but are only functional under stationary conditions. Additionally, [49] requires controlled breathing at specific rates for optimal operation. [50] investigates the use of in-ear PPG for RR estimation in various user activities, *i.e.,* stationary, talking, walking, and running. However, the accuracy is significantly affected by physical activity, with error rates up to approximately 31% under motions, due to the lack of a specific design for motion-resilient estimation. Moreover, PPG is not commonly found in commercial earables, unlike microphones which are commonly integrated. **IMUs:** [4], [5], [20] use IMUs on earphones to estimate RR, but they are effective only under stationary conditions by discarding data from periods with motions. **Microphones:** [13] utilizes in-ear microphones on earphones to determine RR but only works for high-intensity breathing when the user is stationary, as natural breathing is sometimes imperceptible. [51] estimates RR using out-ear microphones on AirPods, employing deep learning techniques. It solely relies on audible breathing sounds, retained through perceptual annotation for model training and testing, which means it only works effectively for heavy breathing. Moreover,

out-ear microphones are inherently vulnerable to environmental noises as breathing sounds are weak and attenuate significantly in air. **Multiple sensors:** [52] employs out-ear microphones and IMUs on earphones for estimating RR, tailored for stationary conditions involving head motion. [14], [15] utilize out-ear microphones on earphones and IMUs on smartphones to estimate a RR-related factor - LRC, when the user is running, which can be used for RR estimation. However, these works assume that the LRC remains constant during one estimation window, which makes the system less robust in daily settings. [53] proposes a learning-based solution for breathing phase detection, which can be used to estimate RR specifically during outdoor running using both in-ear and out-ear microphones. *None of these solutions can reliably and consistently monitor RR across a wide range of daily life conditions, including both sedentary and active scenarios, while maintaining high performance and without discarding any data.*

VI. DISCUSSION

Music Playback. We conducted preliminary experiments to examine the impact of music listening on system performance during sedentary activities. Our findings indicate that listening to music produces error levels similar to other sedentary activities (Figure 11(a)), likely due to the limited overlap between heart sound frequencies and music frequencies. In the future, we plan to explore this in greater depth, including evaluating the effects of different music genres and listening volumes on performance. Additionally, we aim to assess system performance during active states while listening to music.

Environment Noise. We conducted experiments outdoors in a noisy environment (next to an active construction site) to evaluate the impact of ambient noise on system performance. Figure 13(b) demonstrates that RespEar achieves consistent performance across sedentary and active activities. In the future, we plan to perform a comprehensive evaluation of system performance under varying levels of ambient noise.

Clinical Study. In future work, we plan to conduct clinical studies to evaluate system performance in individuals with respiratory conditions and breathing abnormalities to ensure its reliability in diverse populations. These studies will involve collaborations with medical professionals to recruit patients with diverse respiratory profiles, such as asthma, chronic obstructive pulmonary disease (COPD), and sleep apnea.

VII. CONCLUSIONS

This paper presented RespEar, the first earable system for continuous, non-obtrusive, and reliable RR monitoring across both sedentary and active conditions. RespEar employs in-ear microphones and leverages unique relationships of our cardiovascular, gait and respiratory systems to present a holistic and optimized solution for RR estimation, while uniquely identifying and addressing three key practical issues. We implemented RespEar prototype and conducted extensive experiments to evaluate its performance. The results demonstrate that RespEar outperforms the state-of-art, and is robust in a variety of contexts.

REFERENCES

[1] A. Nicolò *et al.*, "The Importance of Respiratory Rate Monitoring: From Healthcare to Sport and Exercise," *Sensors (Basel, Switzerland)*, vol. 20, no. 21, p. 6396, Nov. 2020.

[2] Adão *et al.*, "Fatigue Monitoring Through Wearables: A State-of-the-Art Review," *Frontiers in Physiology*, vol. 12, p. 790292, Dec. 2021.

[3] J. Hernandez *et al.*, "BioWatch: Estimation of Heart and Breathing Rates from Wrist Motions," in *In Proc. of PervasiveHealth 2015*. Istanbul, Turkey: ICST, 2015.

[4] T. Röddiger *et al.*, "Towards Respiration Rate Monitoring Using an In-Ear Headphone Inertial Measurement Unit," in *In Proc. of EarComp 2019*. United Kingdom: ACM, Sep. 2019, pp. 48–53.

[5] M. M. Rahman *et al.*, "Towards Motion-Aware Passive Resting Respiratory Rate Monitoring Using Earbuds," in *In Proc. of BSN 2021*. Athens, Greece: IEEE, Jul. 2021, pp. 1–4.

[6] X. Wang *et al.*, "Smartphone sonar-based contact-free respiration rate monitoring," *ACM Transactions on Computing for Healthcare*, vol. 2, no. 2, pp. 1–26, 2021.

[7] J. Liu *et al.*, "WiPhone: Smartphone-based Respiration Monitoring Using Ambient Reflected WiFi Signals," *In Proc. of IMWUT 2021*, vol. 5, no. 1, pp. 1–19, Mar. 2021.

[8] S. Yue *et al.*, "Extracting Multi-Person Respiration from Entangled RF Signals," *In Proc. of IMWUT 2018*, vol. 2, no. 2, pp. 1–22, Jul. 2018.

[9] F. Adib *et al.*, "Smart Homes that Monitor Breathing and Heart Rate," in *In Proc.of CHI 2015*. Seoul Republic of Korea: ACM, Apr. 2015, pp. 837–846.

[10] "Zephyr bioharness 3.0 chest strap," https://www.zephyranywhere.com/media/download/bioharness3-user-manual.pdf, 2023.

[11] C. Massaroni *et al.*, "Contact-Based Methods for Measuring Respiratory Rate," *Sensors*, vol. 19, no. 4, p. 908, Feb. 2019.

[12] Y. Nam *et al.*, "Estimation of Respiratory Rates Using the Built-in Microphone of a Smartphone or Headset," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 6, pp. 1493–1501, Nov. 2016.

[13] A. Martin *et al.*, "In-Ear Audio Wearable: Measurement of Heart and Breathing Rates for Health and Safety Monitoring," *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, vol. 65, no. 6, p. 8, 2018.

[14] T. Hao *et al.*, "RunBuddy: A smartphone system for running rhythm monitoring," in *In Proc. of UbiComp 2015*. Osaka, Japan: ACM Press, 2015, pp. 133–144.

[15] F. Gu *et al.*, "Detecting breathing frequency and maintaining a proper running rhythm," *Pervasive and Mobile Computing*, vol. 42, pp. 498–512, 2017.

[16] A. Natarajan *et al.*, "Measurement of respiratory rate using wearable devices and applications to covid-19 detection," *NPJ digital medicine*, vol. 4, no. 1, pp. 1–10, 2021.

[17] W. Karlen *et al.*, "Respiratory rate estimation using respiratory sinus arrhythmia from photoplethysmography," in *In Proc. of EMBC 2011*. IEEE, 2011, pp. 1201–1204.

[18] Z. Xu *et al.*, "Toward a robust estimation of respiratory rate using cardiovascular biomarkers: Robustness analysis under pain stimulation," *IEEE Sensors Journal*, vol. 22, no. 10, pp. 9904–9913, 2022.

[19] J. B. Elsner and A. A. Tsonis, *Singular spectrum analysis: a new tool in time series analysis*. Springer Science & Business Media, 1996.

[20] T. Ahmed *et al.*, "RRMonitor: A Resource-Aware End-to-End System for Continuous Monitoring of Respiration Rate Using Earbuds," in *In Proc. of EMBC 2021*. Mexico: IEEE, Nov. 2021, pp. 2463–2467.

[21] T. Hao *et al.*, "Mindfulwatch: A smartwatch-based system for real-time respiration monitoring during meditation," *In Proc. of IMWUT 2017*, vol. 1, no. 3, pp. 1–19, 2017.

[22] X. Fan *et al.*, "APG: Audioplethysmography for Cardiac Monitoring in Hearables," in *In Proc. of MobiCom 2023*. Madrid Spain: ACM, Oct. 2023, pp. 1–15.

[23] K.-J. Butkow *et al.*, "heart: Motion-resilient heart rate monitoring with in-ear microphones," in *In Proc. of PerCom 2023*. IEEE, 2023, pp. 200–209.

[24] D. Ma, A. Ferlini, and C. Mascolo, "OESense: Employing occlusion effect for in-ear human sensing," in *In Proc. of MobiSys 2021*. Virtual Event Wisconsin: ACM, Jun. 2021, pp. 175–187.

[25] F. Yasuma and J.-i. Hayano, "Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm?" *Chest*, vol. 125, no. 2, pp. 683–690, 2004.

[26] B. Aysin and E. Aysin, "Effect of respiration in heart rate variability (hrv) analysis," in *In Proc. of EMBC 2006*. IEEE, 2006, pp. 1776–1779.

[27] G. Blain, O. Meste, and S. Bermon, "Influences of breathing patterns on respiratory sinus arrhythmia in humans during exercise," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 288, no. 2, pp. H887–H895, Feb. 2005.

[28] K.-J. Butkow *et al.*, "An evaluation of heart rate monitoring with in-ear microphones under motion," *Pervasive and Mobile Computing*, p. 101913, 2024.

[29] X. Sun *et al.*, "Earmonitor: In-ear Motion-resilient Acoustic Sensing Using Commodity Earphones," *In Proc. of IMWUT 2022*, vol. 6, no. 4, pp. 1–22, Dec. 2022.

[30] C. P. Hoffmann, G. Torregrosa, and B. G. Bardy, "Sound stabilizes locomotor-respiratory coupling and reduces energy cost," 2012.

[31] J.-F. Gariépy, K. Missaghi, and R. Dubuc, "The interactions between locomotion and respiration," in *Progress in Brain Research*. Elsevier, 2010, vol. 187, pp. 173–188.

[32] D. A. Mahler *et al.*, "Locomotor-respiratory coupling develops in novice female rowers with training," *Medicine and Science in Sports and Exercise*, vol. 23, no. 12, pp. 1362–1366, Dec. 1991.

[33] M. A. Daley *et al.*, "Impact Loading and Locomotor-Respiratory Coordination Significantly Influence Breathing Dynamics in Running Humans," *PLoS ONE*, vol. 8, no. 8, p. e70752, Aug. 2013.

[34] A. Schäfer and K. W. Kratky, "Estimation of breathing rate from respiratory sinus arrhythmia: comparison of various methods," *Annals of Biomedical Engineering*, vol. 36, pp. 476–485, 2008.

[35] "Human respiration rate range," https://www.medicalnewstoday.com/articles/324409, 2023.

[36] S. Fuchs and A. Rochet-Capellan, "The respiratory foundations of spoken language," *Annual Review of Linguistics*, vol. 7, pp. 13–30, 2021.

[37] J. Cioffi and T. Kailath, "Fast, recursive-least-squares transversal filters for adaptive filtering," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 2, pp. 304–337, 1984.

[38] A. Oliveira *et al.*, "Respiratory sounds in healthy people: a systematic review," *Respiratory medicine*, vol. 108, no. 4, pp. 550–570, 2014.

[39] J. O'Halloran *et al.*, "Locomotor-respiratory coupling patterns and oxygen consumption during walking above and below preferred stride frequency," *European Journal of Applied Physiology*, vol. 112, no. 3, pp. 929–940, 2012.

[40] M. A. Daley, D. M. Bramble, and D. R. Carrier, "Impact loading and locomotor-respiratory coordination significantly influence breathing dynamics in running humans," *PloS one*, vol. 8, no. 8, p. e70752, 2013.

[41] A. Arefeen *et al.*, "Inter-beat interval estimation with tiramisu model: A novel approach with reduced error," 2021.

[42] D. Liaqat *et al.*, "WearBreathing: Real World Respiratory Rate Monitoring Using Smartwatches," *In Proc. of IMWUT 2019*, vol. 3, no. 2, pp. 1–22, Jun. 2019.

[43] "Garmin," https://support.garmin.com/en-GB/?faq=2yEgS0Pax53UDqUH7q4WC6, 2023.

[44] "Apple watch," https://support.apple.com/sr-rs/guide/watch/apd830528336/watchos, 2023.

[45] L. Zhao *et al.*, "Robust Respiratory Rate Monitoring Using Smartwatch Photoplethysmography," *IEEE IoT Journal*, vol. 10, no. 6, pp. 4830–4844, Mar. 2023.

[46] R. Dai *et al.*, "Respwatch: Robust measurement of respiratory rate on smartwatches with photoplethysmography," in *In Proc. of IoTDI 2021*, 2021, pp. 208–220.

[47] M. M. Rahman *et al.*, "Instantrr: Instantaneous respiratory rate estimation on context-aware mobile devices," in *In Proc. of BODYNETS 2018*. Springer, 2020, pp. 267–281.

[48] H. Aly and M. Youssef, "Zephyr: Ubiquitous accurate multi-sensor fusion-based respiratory rate estimation using smartphones," in *In Proc. of INFOCOM 2016*. IEEE, 2016, pp. 1–9.

[49] K. Taniguchi and A. Nishikawa, "Earable POCER: Development of a Point-of-Care Ear Sensor for Respiratory Rate Measurement," *Sensors*, vol. 18, no. 9, p. 3020, Sep. 2018.

[50] A. Ferlini *et al.*, "In-Ear PPG for Vital Signs," *IEEE Pervasive Computing*, vol. 21, no. 1, pp. 65–74, Jan. 2022.

[51] A. Kumar *et al.*, "Estimating Respiratory Rate From Breath Audio Obtained Through Wearable Microphones," in *In Proc. of EMBC 2021*. Mexico: IEEE, Nov. 2021, pp. 7310–7315.

[52] T. Ahmed *et al.*, "Remote Breathing Rate Tracking in Stationary Position Using the Motion and Acoustic Sensors of Earables," in *In Proc. of CHI 2023*. Hamburg Germany: ACM, Apr. 2023, pp. 1–22.

[53] C. Hu *et al.*, "Breathpro: Monitoring breathing mode during running with earables," *In Proc. of IMWUT 2024*, vol. 8, no. 2, pp. 1–25, 2024.