

Multi-user Room-scale Respiration Tracking using COTS Acoustic Devices

HAORAN WAN, State Key Laboratory for Novel Software Technology, Nanjing University, China

SHUYU SHI, State Key Laboratory for Novel Software Technology, Nanjing University, China

WENYU CAO, State Key Laboratory for Novel Software Technology, Nanjing University, China

WEI WANG, State Key Laboratory for Novel Software Technology, Nanjing University, China

GUIHAI CHEN, State Key Laboratory for Novel Software Technology, Nanjing University, China

Continuous domestic respiration monitoring provides vital information for diagnosing assorted diseases. In this paper, we introduce *RESPTRACKER*, the first continuous, multiple-person respiration tracking system in domestic settings using acoustic-based COTS devices. *RESPTRACKER* uses a multi-stage algorithm to separate and recombine respiration signals from multiple paths so that it can track the respiration rate of multiple moving subjects. And it leverages features from multiple dimensions to separate different users in the same area. Our experimental results show that our two-stage algorithm can distinguish the respiration of at least four subjects and cover a distance of three meters.

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

Additional Key Words and Phrases: Ultrasound Signal, Wireless Sensing, Respiration Monitoring.

ACM Reference Format:

Haoran Wan, Shuyu Shi, Wenyu Cao, Wei Wang, and Guihai Chen. 2023. Multi-user Room-scale Respiration Tracking using COTS Acoustic Devices. *ACM Trans. Sensor Netw.* 1, 1, Article 1 (January 2023), 28 pages. <https://doi.org/10.1145/3594220>

1 INTRODUCTION

Respiration is one of the vital signs that contain valuable information for diagnosing assorted diseases, *e.g.*, pulmonary disease [2], heart failure [3], anxiety [4], and sleep disorders [5]. Clinical instruments, such as capnography or plethysmography, provide reliable respiration measurements. However, they need professional operators and cannot be deployed in the domestic scenario to perform long-term monitoring, which is vital to early diagnoses of chronic diseases, such as

¹This is an extended version of a prior conference paper that appeared in IEEE INFOCOM 2021 [1].

²Shuyu Shi is the corresponding author.

³This work is partially supported by National Natural Science Foundation of China under Numbers 61872173, 61902177, 61972254 and 61832005, Natural Science Foundation of Jiangsu Province of China under number BK20190298 and Collaborative Innovation Center of Novel Software Technology.

Authors' addresses: Haoran Wan, wanhr@smail.nju.edu.cn, State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, Jiangsu, China, 210093; Shuyu Shi, ssy@nju.edu.cn, State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, Jiangsu, China, 210093; Wenyu Cao, wenyucao@smail.nju.edu.cn, State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, Jiangsu, China, 210093; Wei Wang, ww@nju.edu.cn, State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, Jiangsu, China, 210093; Guihai Chen, gchen@nju.edu.cn, State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, Jiangsu, China, 210093.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

1550-4859/2023/1-ART1 \$15.00
<https://doi.org/10.1145/3594220>

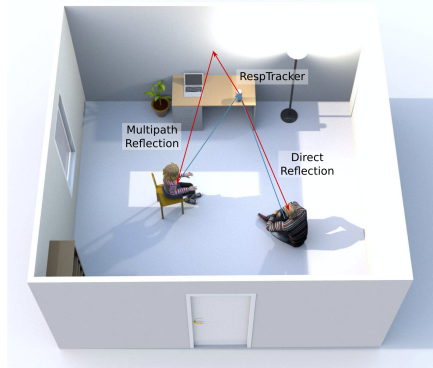


Fig. 1. General application scenario of RESPTRACKER.

obstructive sleep apnea syndrome (OSAS) and chronic obstructive pulmonary disease (COPD). As a result, the development of domestic continuous respiratory monitoring systems has attracted increasing research interest in recent years.

There are domestic respiratory monitoring systems based on cameras [6] or using special devices, including belt integrated with capacitive sensors [7] or smart cushion with air pressure sensors [8]. However, user studies have shown that people are reluctant to deploy these devices due to privacy concerns [6, 9] or the high cost and long-term physical contact requirements [7, 8]. A more promising solution is enabling device-free respiratory monitoring with ubiquitously available wireless signals emitted by commercial off-the-shelf (COTS) devices in domestic settings [10–12].

Limitations of Prior Art: Existing device-free respiratory monitoring systems leverage two types of signals emitted by COTS devices: radio frequency (RF) signals and ultrasound signals. One popular solution for RF-based systems is collecting Wi-Fi channel state information (CSI) for further respiration measurements [13]. However, due to the narrow bandwidth of Wi-Fi signals, the range resolution of CSI is too low to separate two nearby respiration signals. For example, the aliasing range between two non-resolvable paths is 7.5 m for a Wi-Fi bandwidth of 40 MHz. Existing works either rely on the differences in the respiration rate [12] or use specialized high bandwidth frequency modulated continuous wave (FMCW) radar and Independent Component Analysis (ICA) [10] to separate multiple users. These solutions impose extra assumptions on respiration patterns or need specialized devices that increase the domestic deployment cost. Acoustic-based systems turn the speaker-microphone pair integrated with COTS devices, such as mobile phones and smart speakers, into an active sonar to perform the respiration monitoring task. The advantage of acoustic-based systems is the higher range resolution [11], *e.g.*, a typical bandwidth of 4 kHz leads to a range resolution of 8.5 cm for ultrasound signals. However, due to the fast attenuation of sound signals, most acoustic-based systems have a limited range of 0.7~1.1 m [11, 14–16]. Therefore, their applications are limited to sleep monitoring instead of room-scale domestic deployment for continuous respiratory monitoring and tracking.

Proposed Approach: In this paper, we introduce RESPTRACKER, a continuous, multiple-person respiration tracking system in domestic settings using acoustic-based COTS devices. As shown in Fig. 1, the respiration signal of different users may arrive at the receiver through multiple paths. RESPTRACKER proposes a multipath *separation*, user *separation* and path *combination* framework for robust respiration signal tracking.

First, RESPTRACKER utilizes inaudible sound signal modulated by the Zadoff-Chu (ZC) sequence [17] to separate sound reflections from different users. Compared to traditional FMCW-based systems, the key advantage of our separation scheme is that we can precisely differentiate *individual*

reflection paths with channel impulse response (CIR) through correlation. Then, RESPTRACKER turns the indoor multipath effect into our friends by recombining the multipath signals belonging to the same user. Our signal combination algorithm performs a multi-dimensional search and analysis among different distances, different angles, multiple receiving microphones, and different time-frames, based on the CIR measurement from the ZC signal. In this way, we can reliably cluster reflection paths to different users even if they have similar respiration rates. With our multi-stage scheme, RESPTRACKER can detect reliable single person respiration signal at a distance of 3 meters and track the movement of each user within 20 seconds after movements. We can also separate multiple subjects' respiration and track each of them in domestic settings.

Technical Challenges and Solutions: The first challenge is to reliably separate multiple breath signals. Existing work for multi-user breath detection [10] leverages the ICA algorithm to extract different subjects' respiration. As multiple reflections of wireless signal are mixed at the receiver due to the limited range resolution, they need a reliable decomposition algorithm to separate them. To address this challenge, we use the ZC sequence to distinguish different sound reflection paths with a high resolution of less than 10 cm (for round trip). On top of distance metrics, we use a microphone array to conduct beamforming and cancellation to separate users in different angles. In addition, we also measure the statistical features and features related to different users' breath to increase the separability. In this way, each path contains less interference of other subjects so that the complexity of signal decomposition is greatly reduced.

The second challenge is to expand the monitor range to the room-scale. Since the ultrasonic signal attenuates quickly in indoor environments, the measurement of a single path could be noisy and inaccurate. Microphone arrays of COTS acoustic devices are often sparsely arranged with an interval around 5 cm to enhance speech signals with a frequency under 3 kHz and a wavelength longer than 11 cm. In this case, traditional delay-and-sum beamforming algorithm blindly combines signals from the same distance and angle where the weak respiration signal may be destroyed by the out-of-phase combination. To resolve this issue, we use a multi-dimensional signal combination scheme to select and recombine the respiration signals from the same user. We first leverage multiple microphones that are common on COTS devices, such as Amazon Echo and Google Home, to collect multiple copies of the sound reflections. Based on the multipath phenomenon, we collect sound reflections on paths at different distances that arrive at the same microphone. We first determine whether these reflections are related to respiration patterns and then extract a group of features to differentiate different users. By clustering these multi-dimensional reflection signals, we can determine which user the respiration signal belongs to. In this way, we are able to combine a large number of weak paths from the same user, thereby reconstructing the respiration signal reliably and achieving long-distance monitoring.

The third challenge is to track the respiration signals when users change their position. As users may not keep static in their daily routine, our monitoring system should be able to keep tracking while users change their position or orientation. To achieve respiration tracking under dynamic position and orientation, we divide the signal into short observation slots with a duration of twenty seconds. Within each slot, we first determine whether there are movements and then track the distance change of each reflection path. Therefore, we can quickly use the historical data to regain synchronization within twenty seconds after the movement through cluster center match and correlation.

Summary of Experiments and Contributions: In the single user scenario, our system can robustly estimate the respiration rate with an error under 0.6 Beats per Minute (BPM) for different environments, such as in the hallway, offices, and conference rooms. RESPTRACKER can also achieve an error of less than 1 BPM within a distance of three meters and maintain an error of less than 0.8 BPM while the user is moving. In the multi-user scenario, RESPTRACKER can separate the respiration

signals of more than four users in the same room and achieve an error of less than 1 BPM for each user.

In summary, our contribution has the following three aspects:

- We design a signal separation technique to separate multiple users' respiration signal from the each other and reflection from environment.
- We propose to combine multiple echoes related to respiration from other object in the environments and expand the sensing range up to 3 meters.
- We combine features of different users in multiple dimensions to reconstruct the respiration signal for different users in the same environment.

Paper Organization: The rest of this paper is organized as follow. Section 2 summarizes related works of this paper. Section 3 provides motivation and overview of our system. Section 4 introduces the signal model used to track the respiration. Section 5 introduces how to select the paths related to respiration from the environmental reflections. Section 6 separate respiration signal from different users in features from multiple dimensions, and Section 7 introduce the models of signal combination. Section 8 showcases the implementation and evaluation of our system. Section 9 discusses the limitation of the system and Section 10 concludes the paper.

2 RELATED WORK

We summarize recent works related to respiration tracking according to the following three categories.

Respiration Monitoring with Wireless Signals: Wireless signals are widely used for non-invasive vital sign monitoring [10, 18–20]. BreathTaking [18] leverages the received signal strength between different pairs of network devices to conduct the contactless breath monitoring for single person on the bed. DeepBreath [10] uses multiple FMCW transceivers and the ICA algorithm to separate different users' respiration signal. Liu *et al.* [19] extract breath and heart beats from the CSI gathered by commodity Wi-Fi devices. Wang *et al.* [12] propose the Fresnel Zone model of Wi-Fi sensing and detect locations where the subject's respiration can be hardly recognized by the CSI amplitude. To tackle this challenge, Zeng *et al.* [20] exploit the complementary between amplitude and phase of complex CSI data to cover the blind point of Fresnel Zone and further use CSI ratio of two antennas to calculate the accurate phase of reflections [21]. Yang *et al.* [22] leverage the high resolution of UWB radars to separate different subjects' respiration and use image processing techniques to detect sleep apnea. ViMo [23] leverages the high spatial resolution (distance, azimuth and elevation) of 60 GHz millimeter wave antenna array to extract both the respiration rate and the heart rate of multiple subjects. Hou *et al.* [24] place an RFID tag on users' chest and leverage the back scatter signal to conduct respiration sensing. While the tags are light and small, the requirement of putting them on body is still a great limitation. Liu *et al.* [25] use smartphone to receive WiFi signal from reflections both from chest and ambient environment to improve the reliability of NLoS condition. Zhang *et al.* [26] use antenna array to steer the LoRa signal on different users' body to spatially separate users and detect their respiration activities. MoReFi [27] conducts respiration waveform reconstruction with a UWB MIMO radar and deep neural network. Although wireless devices can provide strong signal with good quality, they are expensive and the monitoring process might interfere with normal data transmissions, which make them hard to be deployed in domestic environments.

Respiration Monitoring with Acoustic Signals: Acoustic signal travels much slower than wireless signals. The bandwidth of 4 kHz from COTS acoustic devices provides a fine range resolution of 8.5 cm, while similar resolution on RF-based systems requires Gigahertz of bandwidth. So, recent works exploit acoustic signal to perform device-free breath sensing. Apneapp [11]

transmits 18 ~ 20 kHz FMCW sound signals to estimate breathing frequency and detect sleep apnea passively. Wang *et al.* [14] expand the frequency band of acoustic signal by transforming audible white noises into FMCW signals and enhance the signal SNR with receiving-end beamforming to conduct infant respiration monitoring. Unlike traditional processing methods for FMCW signal, CFMCW [15] uses cross-correlation to increase the accuracy of the acoustic-based breath sensing. Xu *et al.* [28] leverage Energy Spectrum Density (ESD) of a single-frequency acoustic signal, Ensemble Empirical Mode Decomposition, and Generative Adversarial Network (GAN) to reconstruct the breathing signal in a driving scenario. BreathPrint [29] uses the acoustic feature of users' breathing to perform breath event detection passively and uses them as biometric signature to conduct user authentication. One of the key challenges for acoustic based approach is the attenuation of sound in the air, which makes it hard to expand the sensing range to more than two meters. Furthermore, the passive methods are not suitable for daily monitoring because the breath sound in audible frequency is very weak and can be affected by many kinds of noise.

Beamforming of Wireless Signals: Beamforming is an important technique in wireless communication, as it can enhance signal strength from/to different direction with receiving/transmitting array. In the wireless transmission scenario, Phaser [30] enables phase array signal processing on COTS device which increases spatial resolution, decreases phase error and suppressed the multipath interference. Wang *et al.* [31] use a blind distributed beamforming on both uplink and downlink to increase the backscatter sensing distance to 64 meters. For wireless sensing and tracking, md-Track [32] gives a multi-dimensional Wi-Fi localization model and drastically increases passive localization divisibility. Vasisht *et al.* [33] analyze time of flight in different frequency band and between different TX/RX pairs to reach decimeter-level localization accuracy. WiDar 2.0 [34] and FreeSense [35] calculate the AoA and ToF to match and localize the moving subject. RFocus [36] synchronizes thousands of passive antennas to form a wallpaper and improves the median signal strength by 9.5 \times and channel capacity by 2.0 \times . In sound signal processing, Roy *et al.* [37] set up a speaker array to achieve long-range ultrasound attacks on voice assistants. Moutinho *et al.* [38] address the inverse problem of localizing microphones with speaker arrays that are playing predefined sounds. Shen *et al.* [39] leverage microphone array and reflections from the wall to localize sound source. X-Array [40] co-phases multiple millimeter wave transmitters to enhance the indoor communication coverage. Most of the existing beamforming techniques ignore the possibility of utilizing the multipath effect to enhance the sensing signal.

3 SYSTEM OVERVIEW

RESPTRACKER aims at multiple-person room-scale respiration tracking. Therefore, the system is supposed to detect and separate the weak reflection signals at a long range reliably. To differentiate different people, the wireless signal should have ability to estimate the absolute distance of individual reflections. Passive sensing by listening to the respiration sound or ambient sound is hard to fulfill the task due to their range limitation and limited separation ability [41].

Comparing to other active sensing medium such as RF signal or millimeter wave signal, acoustic signal has three important advantages. First, the acoustic devices are already ubiquitous in the domestic environments, *e.g.*, TV sets, laptops, voice assistants and *etc.*, and reusing them to conduct breath sensing or human activity recognition (HAR) incurs no extra cost. Moreover, RF and millimeter wave require extra hardware to support the signal acquisition processes including ADC and down-sampling. Second, acoustic signal travels much slower than the wireless signal. Based on the fact that range resolution of broadband signal with bandwidth B and propagation speed of c is proportional to c/B , to achieve the same range resolution with acoustic signal, RF and millimeter wave signal require Gigahertz of bandwidth which requires expensive devices and may violate the ISM band regulation. Third, acoustic signals are more flexible in signal design which can

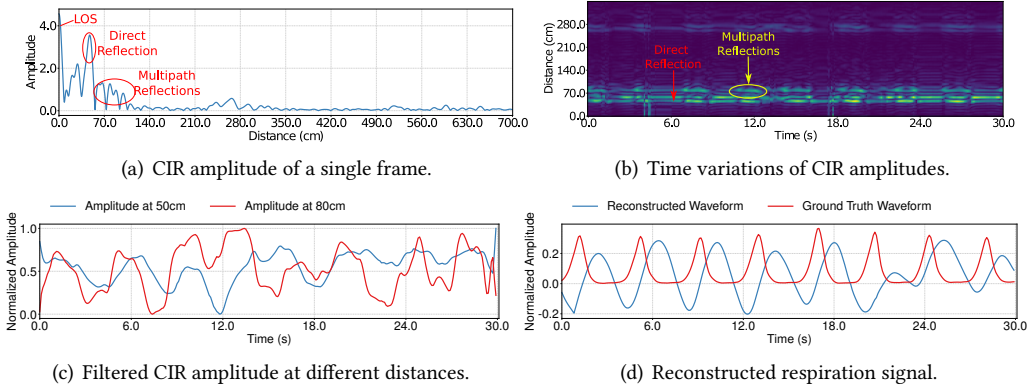


Fig. 2. CIR waveform of a single subject

be adjusted for different conditions. For example, we can transmit shorter supplementary signal in different frequency band simultaneously to locate the user fast and provide timely feedback for the breath sensing signal, while most COTS RF radars can hardly change the signal design.

3.1 Design Motivations

To understand the design challenges for long-range respiration signal detection and separation, we provide a typical respiration signal illustration in Fig. 2. The experiment is performed with the same acoustic devices as used in the evaluation section, shown in Fig. 7, and we conduct this experiment in a hallway shown in Fig. 8(a). The subject is facing the acoustic devices shown in Fig. 8(a) against the wall and is 50 *cm* away from them. Fig. 2(a) shows the amplitude of multipath signals at different distances, where each peak corresponds to one path. From Fig. 2(a), we have two observations. First, due to the high resolution of the sound signal, the width of each peak is less than 20 *cm* so that theoretically we can separate two users even if they are just 10 *cm* apart due to round trip reflections. Second, the sound signal attenuates quickly and it is hard to reliably detect peaks at a distance of 4 meters.

Fig. 2(b) further illustrates the time variations of the paths, where we removed the static components by subtracting the paths that are not changing within a period of 30 seconds. These paths are mainly the LOS path and reflections of walls. The subtraction can be performed in the complex-valued baseband signal because that the speaker and microphones are driven by the same clock on the same board without frequency offset and clock skew between them. We observe from Fig. 2(b) that the respiration of a user causes regular fluctuations in the corresponding path. More interestingly, a single user may incur correlated changes in multiple paths, as the signal may be reflected by the wall or other objects before/after reaching the chest of the user. Also, signals may reflect from different parts of the chest. While these reflections are weak, they provide important respiration information of the same user. It is well known that the signal quality of a single path largely depends on the posture and angle of the user [12]. The fluctuations of a single path may be undetectable for certain user orientations, which lead to interruptions in continuous monitoring. Therefore, it is vital to combine the information of different paths to perform reliable continuous monitoring.

Fig. 2(c) shows the waveform of the respiration signal of the same user at reflection paths at different distances. While the patterns of these signals are similar, they have different phases and signal details. Therefore, directly adding these paths may not be an effective way to enhance the signal.

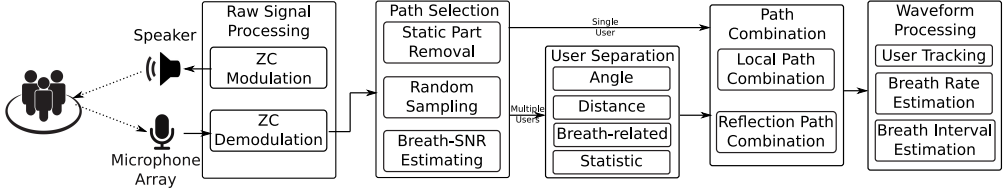


Fig. 3. System Overview of RESPTRACKER

Based on the above observations, we find that RESPTRACKER needs to address two important challenges. First, how to efficiently separate and identify the multipaths of different users? Second, how to reliably combine and reconstruct the breath signals from different paths belonging to a single user?

3.2 System Design

To address the above challenges, RESPTRACKER proposes a multi-stage design as shown in Fig. 3.

The first stage is raw signal processing. We use COTS speakers to transmit ZC modulated sound signals. The reflected signals are received by a microphone array that collects multiple copies of the reflection signal. We perform frequency domain dot products between the received and the transmitted signal to derive the CIR.

The CIRs are fed into the path selection stage. We detect each path in randomly sampled frames and calculate the respiration SNR in the frequency domain to select paths that are candidates of respiration related reflections, which will be used in the second stage.

For multiple users, we use multiple dimensions of features, including distance (CIR), angle (beamforming) and other features, to separate them and perform the path clustering before the following steps.

In the path combination stage, we first perform cross-correlation between the detected paths and their surrounding samples to calculate delay and conduct delay-and-sum in the local paths to expand the sensing range. We then use a Principal Component Analysis (PCA) algorithm to optimally combine the time-domain waveform of the detected paths.

Based on the combined respiration signal, we perform the room-scale tracking by calculating the waveform of each observation slot independently and then match them to form a complete waveform. Finally, we use the reconstructed breath signal to perform breath rate and breath interval time estimations for each user.

4 RAW SIGNAL PROCESSING

We use ZC sequences that have ideal auto-correlation property to separate paths of different users and at different distances.

4.1 ZC Modulation

The transmitting signal used in RESPTRACKER is the ZC sequence modulated by a sinusoid carrier [17, 42]. The ZC sequence with a length of N_{zc} is given by

$$zc[n] = e^{-j \frac{\pi u n(n+2q)}{N_{zc}}}, n = 0, \dots, N_{zc} - 1, \quad (1)$$

where $0 \leq n < N_{zc}$, q is a constant integer and often set to zero which make it a Chu sequence, and N_{zc} is the length of sequence and must be a prime number. The parameter u determines the correlation property and is an integer with $0 < u < N_{zc}$ and $\gcd(N_{zc}, u) = 1$. In this project, we only use a single ZC sequence and set $u = 1$ for convenience. If we need multiple ZC sequences to operate simultaneously [43, 44], we could choose different values that are co-prime with each

other for each sequence, such as $u_1 = 5, u_2 = 17$. Sequences with different u will not exhibit good cross-correlation property. The selection of N_{zc} determines the bandwidth in the final modulated signal. For example, if we set N_{zc} to 163 and the frame length to 1920 sample points, the bandwidth of the modulated signal would be $163/1920 \times 48 = 4.075$ kHz under 48 kHz sampling rate. With this signal setting, we use the acoustic device (shown in Fig. 7) and two plastic plates to demonstrate the cross-correlation property. We place the two plastic plates at 70 cm and 50 cm away from the transceiver respectively, as in Fig. 4(a). The distance between the reflectors and the sound devices are much larger than the distance between microphone and speaker, so the signals' propagation distance is approximate twice of the direct distance (50 cm and 70 cm). We calculate the CIR at the receiving end and we can clearly observe the reflection peaks corresponding to these two objects in Fig. 4(b). If we use different u to calculate the CIR, the result is a noisy sequence and we can't extract any reflection patterns from it. This property may not help to increase the performance but can be used for other purpose, e.g., simultaneously transmitting in the same frequency band or sensing with privacy. The wider the bandwidth is, the higher range resolution we can achieve, which means we can separate different object closer to each other. However, wider bandwidth also means that average signal strength on each frequency will be smaller and may reduces the effective range. The frame length determines the *slow time* frame rate. With longer frame length, we have lower frame rate and longer detection range and vice versa. In our system, we set q to 0, u to 1, N_{zc} to 199 and frame length L to 4,800 sampling points.

Once we get the baseband signal, we first perform FFT and FFT shift on it to get the frequency domain baseband signal. For computational efficiency, the result of FFT has the frequency indexes from 0Hz to $f_s/2$ and then $-f_s/2$ to $-1Hz$ [45]. The FFT shift operation is to reorder the frequency indexes from $-f_s/2$ to $f_s/2$ so that 0Hz is at the center of the FFT result to facilitate the following modulation process. We generate a sequence with length L and then modulate the signal with a carrier sinusoid at a frequency of f_c , which is set as 19 kHz to avoid noise interference to human, by moving the baseband sequence to the higher frequency part. Before performing Inverse Fast Fourier transform (IFFT) for OFDM modulation, we set the negative frequency part to the conjugate counterpart of the signal on the positive frequency to make sure the generated signal is real number in time domain. Algorithm 1 shows the detailed process, where f_s is the sampling frequency. After we generate one frame of the time-domain real signal $z_{cT}[n]$, we transmit it repeatedly so that the transmitted signals are cyclical OFDM symbols.

Algorithm 1: Transmitting signal generation

Result: The modulated sequence $z_{cT}[n]$ with a length of L and a carrier frequency of f_c .

- 1 Generate $zc[n]$ from Eq.1 with a length of N_{zc} .
 - 2 Perform FFT on $zc[n]$ to get $ZC[n]$.
 - 3 Perform FFT shift on $ZC[n]$ to get $ZC_s[n]$.
 - 4 Generate a all zero sequence $\widehat{ZC}[n]$ with a length of L .
 - 5 $\widehat{ZC}[\frac{f_c L}{f_s} - \frac{(N_{zc}-1)}{2} : \frac{f_c L}{f_s} + \frac{(N_{zc}-1)}{2}] \Leftarrow ZC_s[n]$.
 - 6 $\widehat{ZC}[L - \frac{f_c L}{f_s} - \frac{N_{zc}-1}{2} : L - \frac{f_c L}{f_s} + \frac{N_{zc}-1}{2}] \Leftarrow ZC_s^*[n : 1]$.
 - 7 Perform IFFT on \widehat{ZC} to the time domain $z_{cT}[n]$.
-

4.2 ZC Demodulation

After the signal is transmitted from the speaker, the microphone array at the receiver side records the signals that come from both the LOS path and the reflections of subjects and the environment.

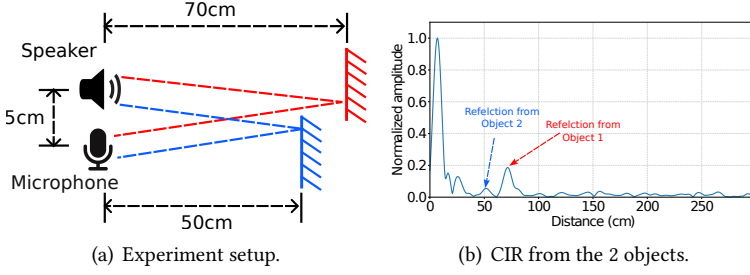


Fig. 4. Demonstration of ZC's range separation ability.

On one pair of speaker/microphone, we can extract one set of CIR per OFDM frame by performing cross-correlation between the received signal and the known transmitted signal [42]. Instead of using the time domain down-conversion and correlation as in [42], we leverage the frequency domain multiplication to perform the frequency-domain correlation which will greatly reduce the computational complexity of correlation.

The received signal is modeled as

$$z_{cR}[n] = \sum_{i=1}^P A_i e^{j\phi_i} z_{cT}[n - \tau_i f_s], \quad (2)$$

where $z_{cR}[n]$ is received signal, P is the number of paths, A_i is attenuation coefficient of path i , ϕ_i is the phase shift caused by the propagation/reflection of path i after modulating 2π and τ_i is the time of flight (ToF) of path i . We first segment the received signal into frames with the same length of L . We then perform FFT on each frame and extract OFDM passband frequency components $ZC_R[n]$ corresponding to the transmitted $ZC_S[n]$. We multiply $ZC_R[n]$ by $ZC_S^*[n]$ to perform cross-correlation in the frequency domain. According to the ideal auto-correlation property of ZC sequence [46], the auto-correlation of $ZC_S[n] \times ZC_S^*[n]$ is all 1 in the frequency domain. Therefore, the cross-correlation gives an ideal CIR under the bandwidth limitation. We use zero-padding to expand the frequency domain baseband length to L then perform an IFFT to get an interpolated time-domain CIR. The peaks in the resulting CIR denote different delayed copies of transmitted signal from different paths, as shown in Fig. 2(a). Algorithm 2 shows the detailed demodulation process.

Algorithm 2: Received signal demodulation

Result: The interpolated time-domain $cir[n]$.

- 1 Perform FFT on $z_{cR}[n]$ to get $ZC_R[n]$.
 - 2 $CIR_{baseband}[n] \leftarrow ZC_R[n] \times ZC_S^*[n]$.
 - 3 Generate an all-zero sequence $CIR[n]$.
 - 4 $CIR[0 : \frac{N_{zc}-1}{2}] \leftarrow CIR_{baseband}[0 : \frac{N_{zc}-1}{2}]$
 - 5 $CIR[L - \frac{N_{zc}+1}{2} : L] \leftarrow CIR_{baseband}[\frac{N_{zc}+1}{2} : N_{zc}]$
 - 6 Perform IFFT on $CIR[n]$ to the time domain $cir[n]$.
-

On each pair of speaker and microphone, we obtain one measurement of $cir[n]$ for an OFDM frame, which has a duration of 0.1 second. We assemble the measurement of CIR in multiple OFDM frames within an *Observation Slot* to form a 2D CIR map as shown in Fig. 2(b). The time-domain resolution of 0.1s in the CIR map gives a sampling rate of 10 Hz, which is adequate for monitoring respiration signals that have typical frequency of 0.1~0.5 Hz.

5 PATH SELECTION

Before we reconstruct the respiration signals from multiple paths, we need to first select correct paths to that contains breath related signal patterns. As modeled in Eq. (2), we can denote breath related reflections as

$$z_{cR_b}[t] = Ae^{-j(\frac{2\pi f d(t)}{c} + p)} z_{cT} \left[n - \frac{d_{body} + d(t)}{c} f_s \right], \quad (3)$$

where d_{body} is the path length of user's body reflection, $d(t)$ is the path length changes caused by chest movement during the exhaling and inhaling, which is a periodic signal, and p is the unknown initial phase shift caused by transmitting and receiving delay. There is no need to estimate p since we don't require the absolute distance of users. And most existing ranging works ask users to put devices on a known location to calibrate and cancel the initial phase [47–49]. Under this model, the corresponding CIR is:

$$cir_{R_b}[t] = Ae^{-j(\frac{2\pi f d(t)}{c} + p)} \text{sinc} \left[n - \frac{d_{body} + d(t)}{c} f_s \right]. \quad (4)$$

As the OFDM signal is band-limited with a rectangular frequency gate function, the corresponding time domain CIR is a convolution of the sinc function with the impulsive response. For a breath movement with a period of $1/f_b$, the corresponding CIR peak will move back-and-forth with an amplitude of d_r around d_{body} . As the LOS and reflection from static environment or static body parts remain almost the same along with time, we can separate the static paths and the breath related paths by their periodicity. After the system starts monitoring, we first determine the location of the LOS path by voting for the maximum peak location of the first L_v frames. The voting period is set to 20 frames based on our experiments and is corresponding to 2 seconds. The LOS localization is an one-time calibration because the distance between the speaker and the microphone is fixed during the monitoring.

5.1 Static Signal Removal and Random Sampling:

After localizing the LOS path, we remove both the LOS path and static reflection. As the LOS and static reflections corresponding to peaks with quasi-static amplitude and phase, we can remove them by subtracting the average complex-valued CIR along the slow time axis of each observation slot from each CIR frame and get cir_{ra} (CIR after removing the average complex values in slow time axis):

$$cir_{ra}[n, t] = cir[n, t] - \frac{1}{T} \sum_{t=0}^{T-1} cir[n, t], \quad (5)$$

where n is the fast time index in frames indicating the propagation distance, t is the frame index in the observation slot, $cir[n, t]$ is the time-stacked result from Algorithm 2, T is the frame number in one *observation slot*. In this way, the remaining non-zero peaks correspond to dynamical paths and the process is shown in Fig. 5. We then randomly sample R frames with uniform distribution in the observation slot to detect the dynamical paths. The random sampling scheme is robust for respiration detection as the paths corresponding to respiration may periodically disappear due to chest movements. We found that using around 10% of the frames in one *observation slot* ($R = 10$) can cover most of dynamical paths.

We use two extra constraints to remove the interference of noisy paths. First, we remove peaks that have an amplitude smaller than a threshold β of the maximum dynamical path. We perform the min-max normalization on each CIR frame. Since the LOS peak normally has the maximum and stable amplitude, we would use the average LOS peak's amplitude of first 10 frames and use it

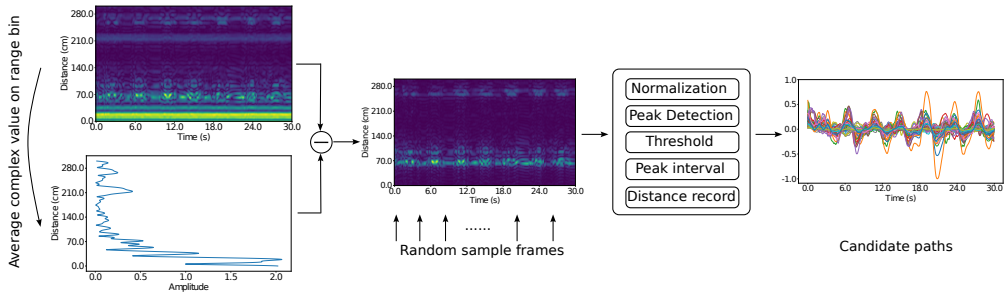


Fig. 5. Process of static signal removal and random sampling (we use the magnitude of the complex signal for illustration).

and 0 to normalize the rest of data. After the normalization, we directly apply a threshold between $0 \sim 1$. This can remove the fluctuation caused by the sidelobes of the sinc function and other noise, the selection of β is a trade-off between the tolerance of noise and retaining weak reflections. With a higher threshold β , it tends to pick stronger reflections and ignore weak reflections. However, weak reflections may be helpful for breath signal reconstruction at long distance. We will further discuss the selection of β in Section 8.3. Second, we remove paths that are within T_b sample points to avoid repetition. Finally, we record the remaining peaks' distance index and take the range bins in that distance across the *observation slot* as candidate paths.

5.2 Breath SNR Calculation:

After detecting the dynamical paths, we use the breath SNR to determine whether the path contains respiration signal or other interfering movements. The breath SNR is based on the observation that the respiration signal will have a strong frequency component within the breath frequency range of $0.1 \sim 0.5$ Hz as indicated in Eq. (4). Therefore, for a specific dynamical path, we first perform an FFT along the time-axis to get the spectrum of the path. We then measure the maximum energy in the FFT bins within the breath frequency range of $0.1 \sim 0.5$ Hz as E_{max} . We modify the s-BNR from [10] and the breath SNR is calculated as $w_1 \frac{E_{max}}{(\sum_{f \in [0.1, 0.5]} E_f) - E_{max}} + w_2 \frac{E_{max}}{\sum_{f \in [0.5, 5]} E_f}$, where E_f is the energy of FFT bin that represents frequency f . This is a weighted sum of the uniqueness of the peak within the breath frequency range and the strength of the peak comparing to other movements. In this way, we can further filter the candidate paths that correspond to breath movements for further path combination in the following sections.

6 USER SEPARATION

In previous sections, we have discussed how to separate paths related to respiration from other unrelated reflections. In real-world scenarios, there might be more than one user in the room and their respiration related paths are all selected in the candidate paths. In this section, we discuss how to separate different users in the same area through features in multiple dimensions.

6.1 Distance Separation

As our ZC sequence has a range resolution of around 8.5 cm and the signal model is shown in Equation 4, we can separate users by their different distances to the receiver. Fig. 6(a) shows the CIR map when there are two users at distance of 1 meter and 1.5 meters. We can clearly observe two traces related to the respiration signal from these two users at corresponding distances. We record each candidates path's distance from LOS as a feature for clustering.

6.2 Angle Separation

Beamforming is a technique to enhance the signal strength from/to one or many directions and suppress the signal from other directions using Rx/Tx arrays [50]. We can use receiver-side beamforming to separate different users from different angles. Different users might appear at the same distance from the microphone and speaker set, so we can't fully rely on the distance metrics derived from CIR [32]. In addition, we use the angular difference to help to resolve multi-user signals. Subspace AoA estimation algorithms, such as multiple signal classification (MUSIC)[51], are not suitable for this situation because reflections of ZC sequences are correlated to each other which violates the assumption of MUSIC. Therefore, we design an iterative cancellation algorithm by iteratively canceling the strongest reflection and then focusing on the weaker reflections.

Fractional Shift. Using the one element of the array as the reference point, to enhance the signal coming from azimuth θ and elevation ϕ , beamforming in normal scenario with unit circle array (UCA) will operate on the original time-domain signal:

$$R[n] = \sum_{i=0}^{M-1} R_i[n - \Delta_i f_s], \quad (6)$$

where R_i is the received signal from i th microphone in total M and Δ_i is the time difference caused by signal from the azimuth ϕ and elevation θ , which is calculated as $\Delta_i = \frac{r \sin(\phi) \cos(\theta - \Theta_i)}{c}$ and r is the radius of the circular microphone array, Θ_i is the relative angle of the microphone with respect to the reference direction. $\rho = 0^\circ$ is usually selected as the reference direction in (r, ρ) polar coordinate system (along x-axis in Cartesian coordinate system) and $\Theta_i = 2\pi i/N$ in this case [50]. Our goal is to enhance paths with breath signals and most of our operation are performed on the CIR, where DC is removed by subtracting the average value along slow time axis of each distance bin and get the cir_{ra} as described in Section 5.1:

Since the time-domain shifting is not computationally efficient when performing sub-sample shifts, we use frequency-domain shift instead. We use i to represent the index of microphones and take traditional delay-and-sum beamforming as an example here. We first perform FFT on the $cir_{ra,i}$ to get the frequency domain signal ($ZC_{ra,i}$) on i th microphone of total M microphones and then apply a phase shift to recalculate the estimation of the average and the beamformed CIR:

$$CIR_{ra,\theta}[n] = \sum_{i=0}^{M-1} (ZC_{ra,i}[n] \times e^{-j2\pi \frac{2Lf_c - (N_{zc} - 1 - 2n)f_s}{2f_s} \frac{\Delta_i f_s}{L}}). \quad (7)$$

After that, we perform zero-padding, IFFT shift, and IFFT on $CIR_{ra,\theta}$ to get the beamformed $cir_{ra,\theta}$. Instead of using all M microphones to estimate several azimuths and elevations as conventional approaches, where the side-lobes are severe, we propose an iterative algorithm to extract the beamformed breath patterns.

Iterative Algorithm. To get the breath patterns from different directions with different signal strength, we iteratively beamform the signal from two microphones, select the breath-related paths and then cancel the strongest incoming signal that related to breath from certain direction to focus on weaker reflections. For each iteration, we use two microphones to perform delay and sum in twelve preset azimuth angles, with the estimated angle, we perform path selection on the beamformed patterns and then cancel the strongest reflection. After we calculate the beamformed CIR from the preset angles and we calculate the Breath-SNR and average signal strength with a distance interval of 10 range bins in each direction. We can get a matrix with each entry's index indicating the angle and distance and the value is multiplication of Breath-SNR and average signal strength which means the bigger value it has, the higher probability it's related to respiration. We use the *RAB* (range, angle and breath) matrix to refer to it. An example of generated Breath-SNR

matrix is show in Fig. 6(b). After we collect the highest paths in this angle, we then focus on weaker reflections from other angle. We use the angle that contains the greatest Breath-SNR and signal strength and data from one microphone (microphone 0 as an example) to cancel other microphone's reflection from this angle:

$$ZC_{ra,i,0}[n] = ZC_{ra,i}[n] - ZC_{ra,0}[n] \times e^{-j2\pi \frac{2Lfc - (N_{zc} - 1 - 2n)fs}{2fs} \frac{(\Delta_0 - \Delta_i)fs}{L}}. \quad (8)$$

Δ_0 and Δ_i are the shifted time for microphone 0 and microphone i caused by the incoming azimuth with max signal strength mentioned above. Using Equation (8), we cancel the signal from the estimated angle on microphone 1 ~ 5. At the end of one iteration, we get CIR of five microphones with the strongest reflection canceled for the next iteration. We repeat this process for five iterations and collect the features, the detailed process is introduced in Algorithm 3.

We use COTS sound devices to implement our system, where the distance between microphones are normally around 4 cm for speech signal enhancement. Our circular microphone array (Fig. 7) has a radius of 4.6 cm and microphones are separated by 60 degrees. We use the microphones that next to each other to mitigate the impact of side-lobe in which condition the number of side-lobes are no more than 2 for each user. Even in the worst condition when the first 3 iterations all focus on the same user (main-lobe and side-lobes), we can still use the next 2 iterations to focus on other users. Fig. 6(b) and Fig. 6(c) show the scenario where two users have nearly the same signal reflection distance but different angles. We can observe that although the side-lobes of the first user are strong after two rounds of cancellation, the main lobe of the second users become the strongest and the position of the second user can be determined in the third round. In each iteration of beamforming and cancellation, we select the paths in that angle and record the angle as one of the features of the selected paths.

Algorithm 3: Process of Angle Estimation

Result: Path candidates after beamforming and feature list.

- 1 Generate an empty array *Candidates* for candidate paths and *Angles* for angle feature.
 - 2 Perform Static Signal Removal on every microphone's CIR according to Equation 5 and get $cir_{ra,i}$, where $i \in [0, M - 1]$.
 - 3 Perform FFT and then FFT shift on $cir_{ra,i}$ to get $ZC_{ra,i}$.
 - 4 for $mic_1 = 0 : M - 2$
 - 5 $mic_2 = mic_1 + 1$
 - 6 for $\theta = \theta_0 : \theta_{11}$
 - 7 Calculate $CIR_{ra,\theta}$ with ZC_{ra,mic_1} and ZC_{ra,mic_2} using Equation 7.
 - 8 Perform IFFT shift and then IFFT on $CIR_{ra,\theta}$ and get $cir_{ra,\theta}$.
 - 9 Calculate *RAB* matrix with $cir_{ra,\theta}$ using Breath-SNR and reflection strength.
 - 10 Maximum value on *RAB* matrix's θ value θ_m is selected as the first echo angle.
 - 11 Perform random sampling on cir_{ra,θ_m} to get the candidate paths with angle feature θ_m .
 - 12 Append the selected paths on *Candidates* and θ_m on *Angles*.
 - 13 for $index = mic_2 : M - 1$
 - 14 Update $ZC_{ra,index}$ with θ_m and ZC_{ra,mic_1} according to Equation 8.
 - 15 Return *Candidates* and *Angles*.
-

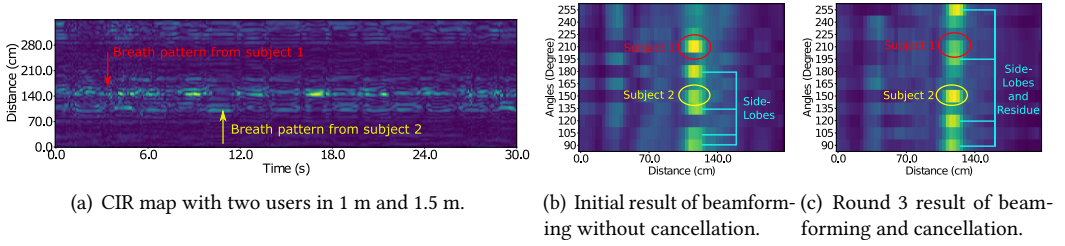


Fig. 6. Separate different users in distance and angle dimension.

Table 1. Feature List for Path Clustering

Types	Features
Distance	Distance from LOS
Angle	Angle used for beamforming
Breathe-related	Times of crossing zeros, Breath SNR, Biggest frequency bin within the BFR, 2nd biggest frequency bin within the BFR, Biggest frequency bin beyond the BFR, 2nd biggest frequency bin beyond the BFR.
Statistic	Standard deviation in TD, Kurtosis in TD and FD, Skewness in TD and FD.

6.3 Breath Related Features

Besides location information, different people may have different breath patterns which may result in different features in time-domain and frequency domain of the signal. For example, different people may have different breath rate and other related patterns. In time-domain, we use the times of crossing zeros and in frequency-domain, we consider Breath SNR in Section 5 and the maximum frequency bin's index within and out of breath frequency range (BFR for short) as the breath related features.

6.4 Statistical Features

We also consider some statistical features in time-domain and frequency-domain of these candidate paths for separation. We calculate Standard Deviation in time-domain and we estimate Kurtosis [52] and Skewness [53] in both time-domain and frequency-domain.

6.5 Path Clustering

The complete feature information is listed in Table 1. Before we apply the clustering algorithm on these features, we normalize the features within each feature's dimension with min-max normalizer. We want to enhance the influence of some features like distance and angle, because they are related to users' position which is unique for each user and other features like breath-related features may be similar since different users may have the same breath patterns. Clustering algorithms mostly rely on the distance between samples under different distance metrics. To enhance the influence, we only need to multiply the normalized features with a weight factor and we set the weights for distance and angle features to 5 and others to 1. Because the distance and angle features are unique for each users, they either have different distances from the devices or have different angles. Other features could increase the separability of different users, but the breath patterns and statistic features could be similar when they happen to have similar breath patterns. We apply Kmeans clustering algorithm on these features. After the clustering, the paths of the same user are more likely to be placed in the same class since the effective multipath reflections are mostly around

the direct reflection. We then perform the two-round combinations algorithm to reconstruct the respiration signal for each user.

7 PATH COMBINATION AND WAVEFORM PROCESSING

We gathered the candidate paths in the previous sections and we introduce how to combine these paths to reconstruct respiration waveform and the processing on the final waveform in this section.

7.1 Two-Round Combinations

In this subsection, we reconstruct the respiration signal through two-round path combinations on the clustered candidate paths selected in the previous section. After the path selection and user separation, we can get breath related paths from different users. To combine signal from multiple paths and reconstruct the respiration signal, we propose a two-round combinations scheme.

Local Path Combination: According to our signal model, the CIR samples surrounding each peak share the same pattern of the path at the peak so that we can combine them to enhance the common features caused by breathing. Specifically, we calculate the cross-correlation between the candidate paths and N_{local} path samples around them to get the weight parameter. We then delay the surrounding paths and use a weighted-sum to add them to the candidate path to reduce the noise of the single sample at the candidate peaks.

Path Combination from Different Distances: We have collected the clustered paths from different users and the underlying respiration patterns for the same user is considered highly related. After the local combination, we gather the candidate paths from different distances and microphones together to form a matrix X with a size of $n \times T_p$, where n is the total number of candidate paths and T_p is the number of frames in the observation slot. We only use the amplitude of the candidate paths to avoid the phase noises in paths. We then remove the static part of each row through the LEVD algorithm [54] and apply a moving average filter with a length of nine samples to smooth the waveform.

Although these data are all from single user and share the same breath patterns, they have different phases and signal details, see Fig. 2(c), caused by the propagation delay and environment reflections. A straightforward method is to use the breath SNR as an indicator and exhaustively search for all possible phase delay parameter to maximize the SNR of generated signal, which is time-consuming. Instead of using this method, we use the PCA algorithm to extract the principal components which are strongly correlated to the respiration signal. The first principal component of the signal matrix gives a low-noise reconstruction of the respiration signal. Fig. 2(d) compares the reconstructed respiration waveform and the ground truth waveform captured by the respiration belt.

7.2 Waveform Processing

Tracking: Users may move during the respiration monitoring period. Therefore, we need to relocate the users and regain synchronization after each movement. To achieve this, we divide the continuous monitoring period into shorter *observation slots* and perform user tracking within each slot.

To balance between the accuracy of movement detection and delay of respiration rate estimation, we choose to set the observation slot length to 20 seconds, which lasts 200 OFDM frames. The reason of using 20 seconds as the observation window is a trade-off between detection accuracy and feedback frequency. We leverage the frequency domain features to distinguish breath patterns from static reflection and other unrelated movements. Given the fact that normal respiration periods are around 3 ~ 5 seconds, if the window size is too short (e.g., around 10 seconds), the number of

periods is small and short sequence makes the frequency resolution coarse-grained, which could affect the path selection process. Using longer window can make the detection of breath patterns more accurate, yet, this will bring more computational burden for resources constraint devices and cause a longer feedback interval. Observation slot with length of 20 seconds is suitable in our scheme to provide acceptable performance in both sides. Within each observation slot, we perform movement detection on the path index change and combination result. When a movement occurs, the peaks found for breath will move largely and the periodic patterns of the result will be devastated. This is because we sample the frames randomly within the observation slot, and the possibility of the small portions of movements being sampled is quite small and the selected paths' major component are still breath related. However, when there are movements across the whole slot, the breath patterns are totally mixed with the movement patterns and our scheme still struggles in this situation. Although we use breath-SNR to select the candidate paths, complex user movements may also have a large power in breath related frequency range. The movement affected paths are mixed with the subtle breath affected paths because they are caused by the same person and close in distance. When we apply PCA on these paths, the majority components are movement related patterns. As a result, the reconstructed waveform has an abrupt shape with no periodicity which can be used for motion detection during monitoring. We further discuss this limitation in Section 9.

When multiple users moved in the environment, we should perform a re-synchronization between observation slot. In the re-synchronization process, we should match the candidate clusters in the current observation slot with those in the previous slots. We use the mean-square of the difference between cluster centers to perform the matching so that when users are moved, static users might be first matched to their historical clusters indicated by distance and angle and moved users will be relocated to new positions.

Breath Rate/Interval Estimation: After we reconstructed the waveform of each observation slot, we estimate the number of breath periods for each user. Within each observation slot, we use a moving average filter to eliminate noise and false respiration peaks. Since the breath rate may vary from 0.1 Hz to 0.5 Hz and the noise level may also vary, we cannot use a fixed empirical window size for the moving average filter. The window length should not be too long to affect the breath peaks or too short to be unable to suppress the noise peaks. We treat the filter length selection as a classification problem and extract the FFT energy within the range of human respiration frequency as the feature. We then use a Support Vector Machine (SVM) to select the filter length based on the FFT energy features. The filter length candidates are odd numbers in [9, 31] (in the sampling frequency of 10 Hz), which makes it a 12 classes classification problem. We collected another set of breath data to train the model. We generate the labels through running the whole estimation process with different filter length and the filter length with minimum errors are selected as the label. If different filter length achieves the same minimum errors, we select the shorter one for less computational cost. In the operation phase, we extract the same features and feed them into the classifier to select the filter length.

After the smoothing process, we normalize the waveform through min-max normalization and perform peak detection. The detected peaks must satisfy two constraints. First, the interval of two adjacent peaks must bigger than 20 sampling points because the breath frequency range in our system is from 0.1 Hz to 0.5 Hz and the smallest possible interval is 20 sampling points. Second, the prominence of the peaks must be bigger than an empirical threshold Thr . Since the waveform is already smoothed and normalized, the Thr is set to 0.05 to avoid false alarms. We then further estimate the BPM and the breath interval time of users based on the detected peaks.

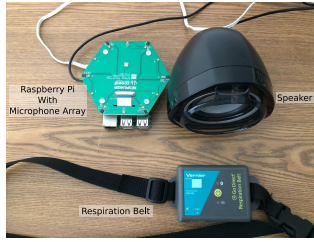


Fig. 7. Devices used in the experiments, including Raspberry Pi, speakers, and the respiration belt.



(a) Hallway.



(b) Office room.

Fig. 8. Samples of experimental environments.

8 IMPLEMENTATION AND EVALUATION

We implement **RESPTRACKER** on Raspberry Pi 3B+ [55] and desktop computers using Python. The Raspberry Pi is equipped with a speaker and a 6-mic circular microphone array [56] to transmit and receive acoustic signal at a sample rate of 48 kHz. The speaker and microphones are driven by the same clock on the board and they can perform playback and recording simultaneously with no frequency offsets. The captured sound signal is sent to PC through Wi-Fi in real-time for further processing. The ground truth of the respiration signal is collected through a Vernier respiration belt [57] that measures the pressure of the chest. As the subjects inhale and exhale, the sensor will record the pressure change of the belt caused by the chest. The devices used in our experiments are shown in Fig. 7.

We use two key metrics to evaluate the performance of **RESPTRACKER**. The first metric is Beats per Minute (BPM) that indicates the average frequency of the breath. The second metric is the breath interval that gives more detailed information about each inhale and exhale, which is vital for diagnosing of chronic diseases. We recruit five volunteers in our evaluation, who are healthy graduate students from 21 to 24 years. During the evaluation, all subjects are asked to breath normally and we find that the resulting BPM are in the range of 10 to 20 which is consistent with the breath rates for healthy people. And the subjects are allowed to occasionally have some body movements like touching faces. For each experiment set, we repeat the processing for 10 times and use the average errors as the experimental result to reduce the impact of random sampling process in our algorithm.

8.1 Experiments in the Single User Scenario

Effective range: To evaluate the breath detection range of **RESPTRACKER**, we conduct experiments in the hallway shown in Fig. 8, at different distances from 0.5 m to 4.0 m. At each distance, we collect two minutes breathing data for five repetitions for each subject. To compare with existing beamforming schemes, we implement a delay-and-sum algorithm to process and combine the same data. In the delay-and-sum scheme, we conduct the delay process in the frequency domain and reuse the intermediate data from the demodulation process to reduce the computational cost. To find the proper elevation and azimuth, we first search with a stride of 10 degrees to find a coarse-grained elevation and azimuth and then fix the elevation and search with a stride of 1 degree in the neighborhood of the azimuth detected by the first step to find the final parameters. For each combination of these parameters, we calculate the breath SNR for all possible paths to find the combination that maximizes the SNR. After that, we get the combined path of six microphones and then calculate BPM and breath interval time in a similar way as in **RESPTRACKER**.

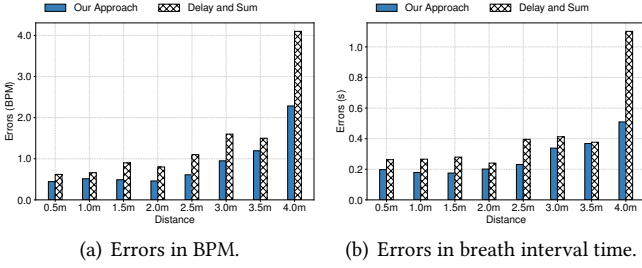


Fig. 9. Experimental results at different distances.

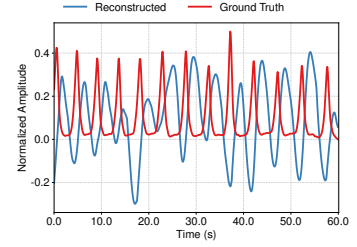


Fig. 10. Reconstructed waveform in the hallway at position 1.

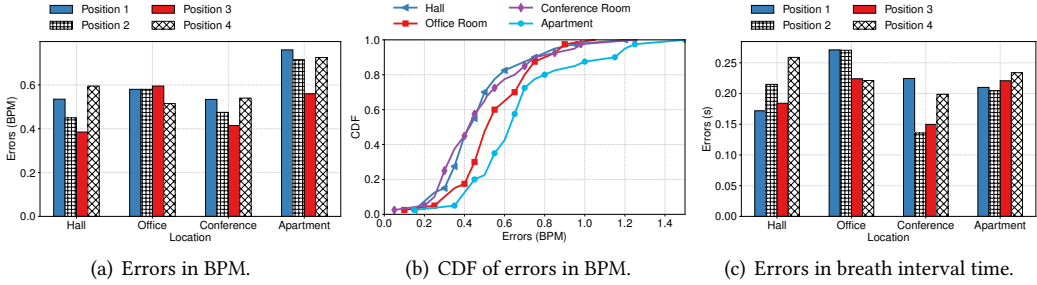


Fig. 11. Experimental results in different environments.

Fig. 9 shows that RESPTRACKER achieves an acceptable measurement error of less than 1 BPM at a distance of 3.0 m. RESPTRACKER outperforms the traditional delay-and-sum method in most cases. This is because the delay-and-sum scheme can only combine the received data at one distance, elevation, and azimuth. When reflected signal is quite weak, the single path set used in delay-and-sum is unstable. In comparison, RESPTRACKER combines multiple path sets thus it can enhance the reflected signal multiple times and the mean absolute errors within 3.0 m is less than 0.95 BPM, while the error of delay-and-sum at 3.0 m is 1.60 BPM. The measurement errors for the breath interval time are shown in Fig. 9(b). The measurement error increases rapidly for the long distances because the ultrasound attenuates quickly in the indoor environment. Although, RESPTRACKER can still reliably work at a distance of three meters.

Fig. 10 further shows the details of the reconstructed waveform. Note that the respiration belt can only detect the inhale, due to the measured chest pressure should always be non-negative, and the acoustic signal can detect both the inhale and the exhale movements. And there is a delay between the two waveform caused by the signal propagation.

Impact of Different Environments: To evaluate impact of different indoor environment, we conduct experiments at different locations in typical indoor environments, including hallway, office room, conference room, and student apartment. Fig. 8 shows the experimental environments in the hallway and the office room. In each environment, we choose four different locations based on the environments' condition, including facing wall, parallel to wall, facing corner, and in the middle of room to cover different reflection conditions. During these experiments, the distance between the user and the microphone/speaker is fixed to be one meter. We ask each subject to breath for two minutes for five repetitions while sitting on a chair.

Fig. 11 shows that our algorithm is robust to environmental changes. Different environments only slightly affect the performance of our system. The mean absolute error at the hallway, office room, conference room and student apartment shown in Fig. 11(a) are: 0.49, 0.57, 0.49, and 0.69 BPM, respectively. In case when there are complex multipath conditions, the breath related paths

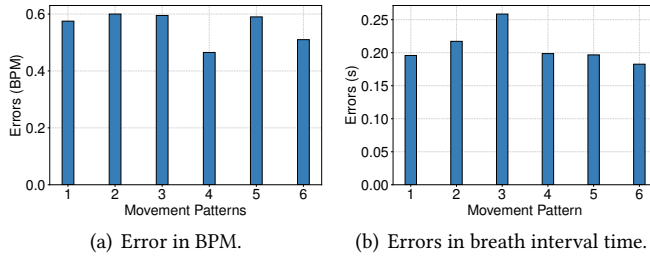


Fig. 12. Experimental results for tracking.

could mix at the same distance, causing ambiguity in the reconstructed respiration signal that leads to higher errors. However, as our algorithm considers paths at different distances, the impact of these multipath signals are mitigated. We also estimate the error of breath interval time between breaths, the average error for these four environments are 0.207, 0.247, 0.209, and 0.217 seconds, respectively, as shown in Fig. 11(c).

Tracking Performance: To evaluate the tracking performance, we requested the subjects to move during the two minutes breathing period, while stay static before and after the movement. We use six different moving patterns which are from 1.0 m to 1.5 m, from 1.5 m to 2.0 m, from 1.0 m to 2.0 m, from 1.5 m to 1.0 m, from 2.0 m to 1.5 m, and from 2.0 m to 1.0 m respectively, during which the subjects are always facing the speaker and microphone array. We denote these movement patterns as pattern 1 to 6 on the result figure. We conduct this experiment in the hallway.

Fig. 12 shows that RESPTRACKER can track different movement patterns. Since our system only measure the respiration signal when the subject is static and the subjects only move once during the measurement, RESPTRACKER achieves similar performance compared to the static experiments. The average absolute error of these six type of movement are 0.58, 0.60, 0.59, 0.47, 0.59, and 0.51 BPM respectively. The breath interval time errors are 0.196, 0.217, 0.258, 0.199, 0.196, and 0.183 seconds for these movement types.

Comparison with Baseline Method. Most previous acoustic signal based respiration monitoring systems [11, 14, 15] monitor the respiration of a single person in different scenarios. To make a fair comparison, we select C-FMCW [15] as the baseline method in this experiment. C-FMCW leverages the correlation function of FMCW frames to obtain the fine-grained distance estimation, which has poorer correlation property than the CIR used in our system [42]. We replace the correlation function with CIR and compare C-FMCW and we can get the results of both systems in a single session of data collection. We compare RESPTRACKER's performance under two different ranges, 1.0 m and 2.5 m in BPM, which is the only metric discussed in [15]. Fig. 13 shows the result of this experiment. In a close distance, the two systems have similar results. However, in a farther distance, RESPTRACKER outperforms the baseline method much, which demonstrates the effectiveness of our algorithm on expanding the sensing range.

Diversities of Users' Age. To evaluate RESPTRACKER's robustness in different age groups, we recruit 6 people in three different age groups, which are 20 ~ 30, 30 ~ 45 and 45 ~ 55 years old. In each age group, there are one healthy male and one healthy female subject. We keep the distance between the subjects and the acoustic devices as 1 m in the typical setting. Fig. 14 shows the result of this evaluation, where RESPTRACKER's performance exhibits little variations among different age groups. This is because the general breath patterns are similar and can be correctly selected by our algorithm.

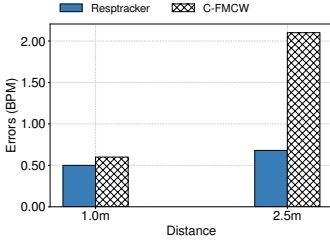
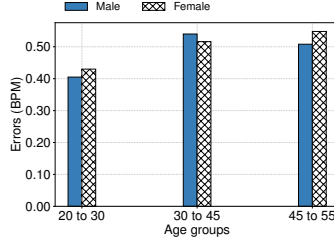
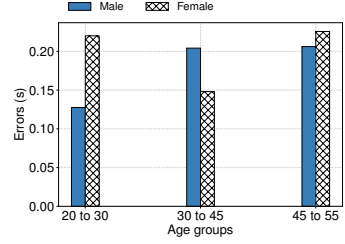


Fig. 13. BPM performance compared with baseline.

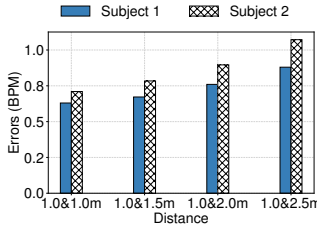


(a) Error in BPM.

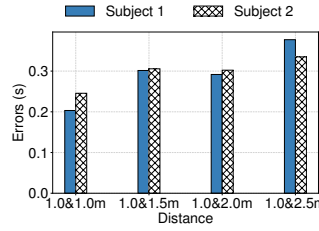


(b) Errors in breath interval time.

Fig. 14. Experimental results in different age groups.



(a) Errors in BPM.



(b) Errors in breath interval time.

Fig. 15. Experimental results for spatial resolution at different distances.

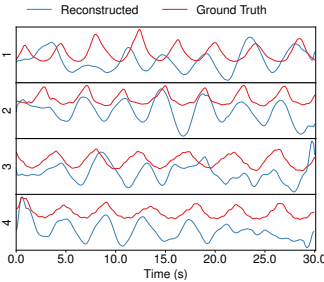
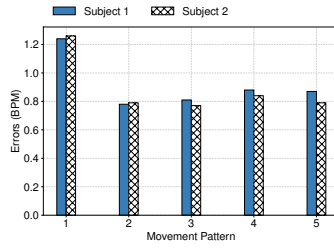
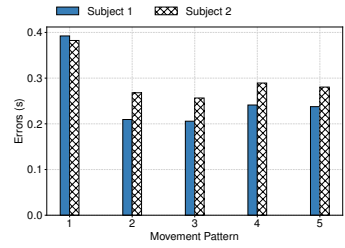


Fig. 16. Reconstruction for four users.



(a) Error in BPM.



(b) Errors in breath interval time.

Fig. 17. Experimental results for different multi-user movement patterns.

8.2 Experiments in the Multiple Users Scenario

Distance Resolution: To evaluate the spatial resolution of RESPTRACKER, we invite two subjects to sit close to each other in the hallway. We fix the distance to microphone/speaker of one user to 1 m and adjust the distance of the other user from 1.0 m to 2.5 m with an interval of 0.5 m to change the distance between two subjects. They sit in the same direction from the microphone array and when they are both 1.0 m from the microphone, they sit shoulder-to-shoulder with each other. At each separation distance, we collect 2 minutes breathing signal for five times. On top of that, we also invite four users to sit together while the distances between them and the sound devices are around 0.5 m, 1.0 m, 1.5 m, and 2.0 m, and try to reconstruct their breath signals.

Fig. 15 shows that RESPTRACKER can reliably separate the breath signals of two users. Even if the two subjects sit shoulder-to-shoulder at the 1.0 m distance, we can still separate their breath signal by using both distance, angle and other features. As subject 2 moves away from subject 1, the error

for subject 1 stays stable and increases slightly and the error for subject 2 increases faster. This is because during the experiments, subject 2 move backward from the microphone in a straight line although the distance between them increases, the angle between them is becoming narrower ($\leq 15^\circ$). And the distance increases the acoustic signal's becomes weaker so the error for subject 2 increases faster. Fig. 16 shows the reconstructed breath signals for four users at four different distances. While the respiration signals are more noisy than in the single user scenario due to the interference between users, we can still reliably measure the breath rate and breath interval time with an average error of 0.80 BPM and 0.295 seconds.

Tracking Performance: In the multiple users scenario, we consider the movement patterns when there is only one user moving or two users are moving away from each other. In the initial state, the distances between two users and sound devices are 1.0 m and 1.5 m. We then request them to conduct five moving patterns which are moving from 1.0 m to 0.5 m, moving from 1.5 m to 2.0 m, moving from 1.5 m to 0.5 m moving, moving from 1.0 m to 2.0m and simultaneously from 1.0 m to 0.5 m and 1.5 m to 2.0 m. We denote these movement patterns as pattern 1 to 5 in the figure.

Fig. 17 shows that RESPTRACKER can track the movement of multiple users under different moving patterns. It appears that except for movement 1, our system can get the same level of BPM and breath interval time error as the static experiments. The reason why the result so bad in movement type may be that subject moved so close to the microphone and blocked the paths of the other subjects. And we can achieve an average BPM error of 0.82, and breath interval time error of 0.286 s.

Angle Separation Performance: In this experiment, we place two users exactly on the ellipse formed by microphone and speaker as focus points which means the reflection distance of two users are exactly the same. We fix the flexible ruler's two ends with length 2.5 meters in the position of speaker and microphone array's center. We do our best to make the reflection distance the same, yet, we can't make the subjects completely static and they occasionally have some body movements like touching face. We adjust users' position to change the angle interval between users as 15° , 30° , 45° , 60° and 90° with the ruler fasten to fix the reflection distance. To estimate the effectiveness of beamforming and cancellation process, we compare the result with or without beamforming process. In the control group, we use six channels' microphone data to directly compute the CIR and select paths like we do in single person condition. And we extract the distance features, statistic features and breath-related features on those paths to conduct path clustering and the process after it.

Fig. 18 shows the result of this experiment. As we can see, beamforming and cancellation process can increase the performance in both BPM and breath interval time in this situation. However, the promotion might not look very remarkable because in real situation, the reflection distances from two users can rarely be exactly the same, let alone with presence of other body movements. Angle features provide a extra dimension to increase the separability.

8.3 Micro-Benchmark

Effect of the Number of Microphones Used: In single user condition, compared to traditional beamforming algorithm, our algorithm can work in any microphone number settings. We change the number of microphones to evaluate the impact of microphone number to our solution. We use the same distance settings as the single user range experiments and at each distance we vary the number of microphones from 1 to 6. At each distance, we repeat the experiment for 100 times and we randomly select different combinations of microphones for each experiment. Intuitively, with more microphones providing more information about the breath patterns, we can get a more accurate estimation on the BPM and breath interval time. Fig. 19 shows the result of this experiment,

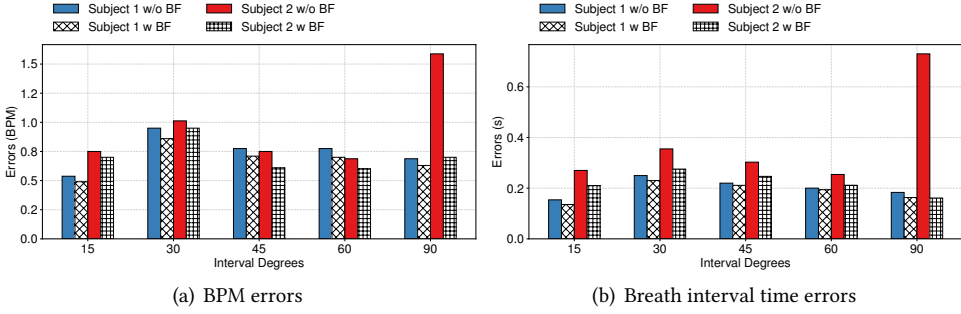


Fig. 18. Effect of beamforming process.

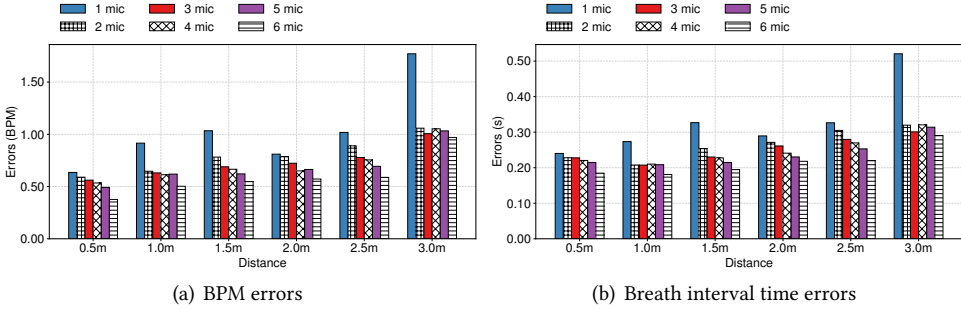


Fig. 19. Effect of the number of microphones used.

which is consistent with our intuition. At the same distance, the results with more microphone combined provide a more accurate estimation. Furthermore, our solution can achieve an acceptable accuracy when using only two microphones.

Effect of Multipaths: In Section 5, we chose the reflection paths with an amplitude threshold $\beta = 0.2$ of the maximum dynamical path. In this experiment, we evaluate the effect of multipaths by changing the threshold of β . With bigger β , we consider less reflection paths to reconstruct the waveform and with smaller β , we take more paths into consideration. We evaluate β in a descending manner with its value $\beta = [0.01, 0.02, 0.1, 0.2, 0.5, 0.7, 0.9]$ at distance 1.0m and 2.5m. Fig. 20 shows the result. Our results show that in closer range, *e.g.*, 1.0m, the number of multipaths hardly affect the result because the reflection signal is strong and we can use smaller number of reflection paths without severe performance drop. For longer range, *e.g.*, 2.5m, the errors grow larger with higher threshold β because the direct reflection is weak in this situation and we need to take more multipaths into account to get a better respiration waveform. Therefore, it is necessary to combine a proper number of multipaths by setting the threshold β around 0.2.

Effect of Users' Orientation and Blockage: We evaluate our system in different user orientations and under blockage. Fig. 21(a) exhibits the experiment setting, where users turn different sides toward the microphones and speaker. Furthermore, we evaluate the performance under blockage, using a three-folding screen made with cardboard placed around 5 cm away from the microphones and speaker. This experiment is conducted in a typical bedroom environment, and the distance between the user and the sound devices is 1.0 m. Fig. 21(b) and 21(c) show the result of this experiment, and we can observe that the orientation of users does not affect the performance too much. However, the cardboard blockage will influence the result, which may be caused by two reasons. First, the cardboard would absorb some sound passing through and make the overall

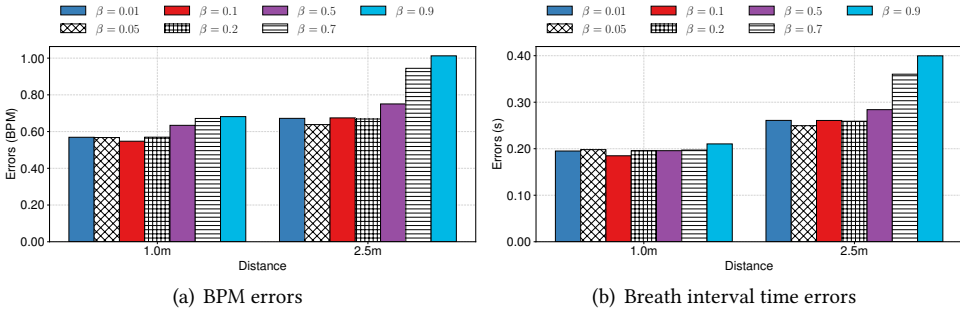


Fig. 20. Effect of multipaths.

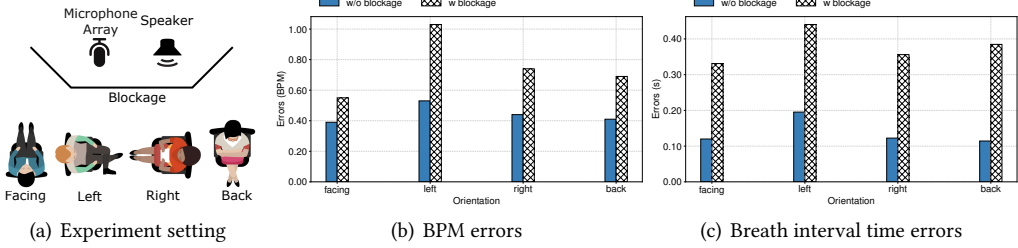


Fig. 21. Effect of users orientation.

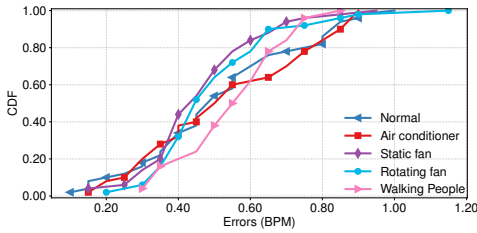


Fig. 22. Effects of different interference.

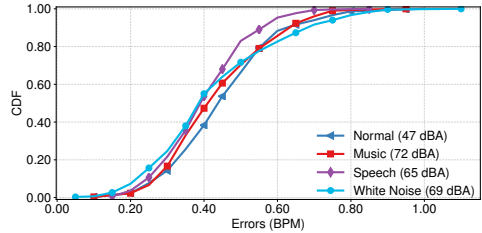


Fig. 23. Effects of Noises.

signal strength weaker. Second, the distance between the screen and the speaker is too close, which blocks some paths reflected from nearby reflectors.

Effect of Interference and Users Activities: To evaluate the performance under interference, we expose our system under different interference during data collection, including air conditioner, static electric fan, rotating electric fan, and another person walking around. The fan is a normal rotating electric fan. The person walks around 0.5 to 1.5 meters away from the subject during data collection, and he is not the target of respiration monitoring. We conduct this experiment in a typical bedroom environment, and the distance between the sound devices and target is 1.0 m. Fig. 22 shows the result of this experiment. We observe that interference from other objects would not degrade the performance of RESPTRACKER too much. This is because external interference has patterns different to respiration, e.g., different in frequency and cross wider range bins, and the breath-SNR calculation step can help to avoid selecting these paths. However, our system can not tackle the situation when the monitoring targets move continuously during the session, and we discuss this situation in Section 9.

Impact of Noise: During the data collection, we expose our system under different types of environmental noise, including music, speech and white noise to test the impact of noise on RESPTRACKER. We play those noise with a mobile phone at a distance of 30 cm from the microphones

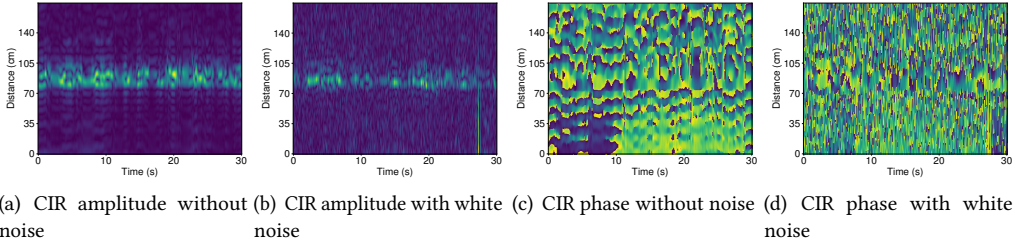


Fig. 24. Impact of white noises on CIR.

Table 2. Processing Time on Different Platforms

	Single User	Multiple Users
PC (RESPTRACKER)	0.1338 s	0.3061 s (w/o BF) 0.5963 s (w BF)
PC (Delay-and-Sum)	14.5805 s	-
Raspi (RESPTRACKER)	2.0830 s	4.4681 s (w/o BF) 12.2938 s (w BF)

and speaker and set the volume to the maximum of the phone. We conduct this experiment in a rest room and fix the distance between the user and microphone to 1.0 m. Since the music and speech has different frequency components from RESPTRACKER in general cases, which will not affect our system's performance. And white noise covers the whole frequency band including our working range, so it may affect the modulated ZC sequence's CIR estimation. Fig. 23 shows the CDF of errors in BPM with different types of noise presence. As we can observe, the BPM estimation error are only slightly degraded with presence of these noises. To understand the reason behind this observation, we plot the CIR map and the phase of CIR map after removing the mean complex CIR value. Fig. 24(a) to 24(d) show the CIR amplitude and phase with and without presence of white noise. As we can observe, the amplitude of the CIR (Fig. 24(a) and 24(b)) is more robust against the white noise's influence and phase is more fragile to them (Fig. 24(c) and 24(d)). Signal with a small amplitude added on the same frequency band might greatly influence the phase of the complex valued CIR. In the extreme condition when the environment are too noisy, the loud noise might cause clipping distortion in the receiver side which will devastate both the amplitude and phase. We can decrease the microphone's gain to handle this situation.

Performance Experiments: Our algorithm are designed as a light-weight algorithm so that it can be deployed on resource-limited mobile devices. To evaluate the computational cost of our algorithm, we run the system on different types of devices, including a Raspberry Pi 3B+ and a desktop computer with an i7-9700 CPU and 16GB memory. On each device, we process the data of both the single user and multiple users experiments and report the average processing time.

Table 2 shows the computational time for RESPTRACKER to process audio data of one observation slot (20 seconds) on different devices using different methods. The source data is six-channel recorded sound signals at a sample rate of 48 kHz with 32 bit float precision. We observe that our algorithm out performs the delay and sum scheme by more than one hundred times on the same platform. Even on resource constrained platforms like Raspberry Pi, our system can handle the incoming data efficiently where the average processing time for a 20 s observation slot is 2.0830 s, 4.4681 s and 12.2938 s for single user and multiple users (with or without beamforming process). It's worth noting that the Raspberry Pi 3B+ has limited memory and we have to use the swap space which is slower. The reason why the delay-and-sum method is so slow is that it has to search exhaustively in all azimuth and elevation in all possible distance bins to find the valid delay and combine the signal from all 6 microphones. And our method perform beamforming

with 2 microphones each time and the times of combination is smaller. There are some possible directions to further improve the performance. First, we could use the high-performance library with multi-thread processing functions like FFTW. Second, we could replace the python with C/C++ to improve the efficiency. Third, we may adjust the loop nest or the data layout to find a way that suits the hardware platform, Raspberry Pi.

Power Consumption. To evaluate the power consumption of our algorithm on Raspberry Pi 3B+, we plug a current and voltage meter between the device and the power source to evaluate the overall power consumption of signal processing and audio playback and recording. We measure the absolute power consumption and the power consumption difference between running the processing and idle standby. The average power consumption for signal processing and audio playback and recording are 349.3 *mW* and 175.4 *mW* when RESPTRACKER is operating continuously.

9 DISCUSSIONS AND LIMITATIONS

While developing RESPTRACKER, we have the following discoveries in the limitation of our system.

Interference from Environments: Our system uses distance and angle estimation to separate different target with a resolution of 8.5 cm and 15 degree and can be increased higher easily. To increase the ranging resolution, we can expand the ZC working bandwidth. To increase the angle resolution, we can increase the number of preset azimuth to compute and physically use an array with more microphones and smaller interval. Distance resolution are very easy to increase with few changes on the parameters and the interference beyond the distance resolution can be filtered out. And in most of condition, distance metrics itself can work well.

Measure while Moving: Our system ignores the signal segment when targets are moving and only performs respiration estimation when all users are static. When the users are moving, the breath related paths are very likely affected by the movements from the limbs that are close to the chest. Those irregular movements are unlikely to be totally removed from the respiration related signal with signal processing techniques. As the signals are not reliable for breath estimation, we choose to ignore these parts of signal. It can be considered to use deep learning based method to recover respiration signals from movements. However, this is beyond the scope of our work.

Using Amplitude instead of Phase: After we conduct the signal processing parts that use the original complex signal, RESPTRACKER reconstructs the respiration waveform with the amplitude of the candidate paths. The reason is that users' chest/belly movements caused reflection distance changes are usually larger than distance between sample points (3.5 mm for round trip), so the breath related patterns on single distance bin may appear and disappear alternately. Therefore, the phase of a single bin is not reliable and we need to combine the phase in different distance bin and different time to reconstruct the breath signal. Such phased based solutions could be more accurate but have a higher complexity. We also observe that amplitude is more robust compare to phase with the presence of white-noise because some minor phase turbulence will greatly affect the original phase shape, and we showcase this observation in Section 8.3. Phase information can be used when the distance bin resolution is poor and the chest caused movements lie within each distance bin, e.g. Wi-Fi/RF signal [20] or acoustic devices working under low sampling rate of 16 kHz [58].

Detailed Breath Patterns: In this work, we focus on the respiration rate and interval between each breath instead of the detailed breath patterns. Because the breath strength measured by the respiration belt could be affected by the tightness of the belt, which is not accurate. Furthermore, in the long distance scenario, the signal's attenuation could be severe and the detailed information of the breath reflection could be lost. We leave this part as the future work.

Reflection from Long Distance: As we previously mentioned, we assume the multipath are around the direct reflection path. However, if there are strong reflectors, e.g., a wall or ceiling

that has a flat surface, the multipath may appear very far from the direct reflection path. Multiple reflection with the nearby wall can also cause this phenomenon. The reflection at around 280 cm in Fig. 2(b) exhibit similar patterns as the breath. Extra breath related features may be extracted from them with knowledge of the surrounding environment although their signal strength is weak. However, we do not explore this part of information in this work.

10 CONCLUSION

In this paper, we present new insights on how to tackle the design challenges for long-range, multiple users domestic respiration tracking systems. We propose to exploit the multipath effect to recombine the reflections in order to improve system sensitivity and robustness. In this way, we expand the sensing range of acoustic respiration patterns from the 0.7 to 1.0 meters in previous works to a room-scale of 3.0 to 4.0 meters. We believe our new insights could bring new opportunity for domestic sensing application.

REFERENCES

- [1] Wan Haoran, Shi Shuyu, Cao Wenyu, Wang Wei, and Chen Guihai. Resptracker: Multi-user room-scale respiration tracking with commercial acoustic devices. In *Proceedings of IEEE INFOCOM*, pages 1–10, 2021.
- [2] Niki Fens, Aeilko H Zwinderman, Marc P van der Schee, Selma B de Nijs, Erica Dijkers, Albert C Roldaan, David Cheung, Elisabeth H Bel, and Peter J Sterk. Exhaled breath profiling enables discrimination of chronic obstructive pulmonary disease and asthma. *American journal of respiratory and critical care medicine*, 180(11):1076–1082, 2009.
- [3] S Javaheri, TJ Parker, JD Liming, WS Corbett, H Nishiyama, L Wexler, and GAI Roselle. Sleep apnea in 81 ambulatory male patients with stable heart failure. *Circulation*, 97(21):2154–2159, 1998.
- [4] Frank H Wilhelm, Walton T Roth, and Marvin A Sackner. The lifeshirt: an advanced system for ambulatory measurement of respiratory and cardiac function. *Behavior Modification*, 27(5):671–691, 2003.
- [5] Christian Guilleminault, Rafael Pelayo, Damien Leger, Alex Clerk, and Robert CZ Bocian. Recognition of sleep-disordered breathing in children. *Pediatrics*, 98(5):871–882, 1996.
- [6] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM transactions on graphics*, 31(4):1–8, 2012.
- [7] Carey R Merritt, H Troy Nagle, and Edward Grant. Textile-based capacitive sensors for respiration monitoring. *IEEE Sensors Journal*, 9(1):71–78, 2009.
- [8] Mark B Norman, Sally Middleton, Odette Erskine, Peter G Middleton, John R Wheatley, and Colin E Sullivan. Validation of the sonomat: a contactless monitoring system used for the diagnosis of sleep disordered breathing. *Sleep*, 37(9):1477–1487, 2014.
- [9] Chia-Wen Lin and Zhi-Hong Ling. Automatic fall incident detection in compressed video for intelligent homecare. In *Proceedings of IEEE ICCCN*, pages 1172–1177, 2007.
- [10] Shichao Yue, Hao He, Hao Wang, Hariharan Rahul, and Dina Katabi. Extracting multi-person respiration from entangled RF signals. *ACM IMWUT*, 2(2):1–22, 2018.
- [11] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. Contactless sleep apnea detection on smartphones. In *Proceedings of ACM MobiSys*, page 45–57, 2015.
- [12] Hao Wang, Daqing Zhang, Junyi Ma, Yasha Wang, Yuxiang Wang, Dan Wu, Tao Gu, and Bing Xie. Human respiration detection with commodity WiFi devices: do user location and body orientation matter? In *Proceedings of ACM UbiComp*, page 25–36, 2016.
- [13] Shuyu Shi, Yaxiong Xie, Mo Li, Alex X Liu, and Jun Zhao. Synthesizing wider wifi bandwidth for respiration rate monitoring in dynamic environments. In *Proceedings of IEEE INFOCOM*, pages 181–189, 2019.
- [14] Anran Wang, Jacob E. Sunshine, and Shyamnath Gollakota. Contactless infant monitoring using white noise. In *Proceedings of ACM MobiCom*, pages 1–16, 2019.
- [15] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. C-FMCW based contactless respiration detection using acoustic signal. In *Proceedings of ACM UbiComp*, pages 1–20, 2018.
- [16] Nacer Khalil, Omprakash Gnawali, Driss Benhaddou, and Jaspal Subhlok. Sonicdoor: A person identification system based on modeling of shape, behavior, and walking patterns. *ACM Trans. Sen. Netw.*, 14(3–4), dec 2018.
- [17] B. M. Popovic. Generalized chirp-like polyphase sequences with optimum correlation properties. *IEEE Trans. Inf. Theor.*, 38(4):1406–1409, 2006.
- [18] Neal Patwari, Joey Wilson, Sai Ananthanarayanan, Sneha K Kasera, and Dwayne R Westenskow. Monitoring breathing via signal strength in wireless networks. *IEEE TMC*, 13(8):1774–1786, 2014.

- [19] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. Tracking vital signs during sleep leveraging off-the-shelf WiFi. In *Proceedings of ACM MobiHoc*, page 267–276, 2015.
- [20] Youwei Zeng, Dan Wu, Ruiyang Gao, Tao Gu, and Daqing Zhang. Fullbreathe: Full human respiration detection exploiting complementarity of CSI phase and amplitude of WiFi signals. *ACM IMWUT*, 2(3):1–19, 2018.
- [21] Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang. Farsense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas. *ACM IMWUT*, 3(3), 2019.
- [22] Yanni Yang, Jiannong Cao, Xiulong Liu, and Xuefeng Liu. Multi-breath: Separate respiration monitoring for multiple persons with UWB radar. In *Proceedings of IEEE COMPSAC*, pages 840–849, 2019.
- [23] Fengyu Wang, Feng Zhang, Chenshu Wu, Beibei Wang, and KJ Ray Liu. Vimo: Multiperson vital sign monitoring using commodity millimeter-wave radio. *IEEE Internet of Things Journal*, 8(3):1294–1307, 2021.
- [24] Yuxiao Hou, Yanwen Wang, and Yuanqing Zheng. Tagbreathe: Monitor breathing with commodity rfid systems. In *Proceedings of ICDCS*, pages 404–413, 2017.
- [25] Jinyi Liu, Youwei Zeng, Tao Gu, Leye Wang, and Daqing Zhang. Wiphone: Smartphone-based respiration monitoring using ambient reflected wifi signals. *ACM IMWUT*, 5(1), 2021.
- [26] Fusang Zhang, Zhaoxin Chang, Jie Xiong, Rong Zheng, Junqi Ma, Kai Niu, Beihong Jin, and Daqing Zhang. Unlocking the beamforming potential of lora for long-range multi-target respiration sensing. *ACM IMWUT*, 5(2), 2021.
- [27] Tianyue Zheng, Zhe Chen, Shujie Zhang, Chao Cai, and Jun Luo. More-fi: Motion-robust and fine-grained respiration monitoring via deep-learning uwb radar. In *Proceedings of ACM SenSys*, SenSys '21, page 111–124, 2021.
- [28] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. Breathlistener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals. In *Proceedings of ACM MobiSys*, page 54–66, 2019.
- [29] Jagmohan Chauhan, Yining Hu, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. Breathprint: Breathing acoustics-based user authentication. In *Proceedings of ACM MobiSys*, page 278–291, 2017.
- [30] Jon Gjengset, Jie Xiong, Graeme McPhillips, and Kyle Jamieson. Phaser: Enabling phased array signal processing on commodity WiFi access points. In *Proceedings of ACM MobiCom*, page 153–164, 2014.
- [31] Jingxian Wang, Junbo Zhang, Rajarshi Saha, Haojian Jin, and Swarun Kumar. Pushing the range limits of commercial passive RFIDs. In *Proceedings of Usenix NSDI*, pages 301–316, 2019.
- [32] Yaxiong Xie, Jie Xiong, Mo Li, and Kyle Jamieson. mD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking. In *Proceedings of ACM MobiCom*, pages 1–16, 2019.
- [33] Deepak Vasisht, Swarun Kumar, and Dina Katabi. Decimeter-level localization with a single WiFi access point. In *Proceedings of Usenix NSDI*, 2016.
- [34] Kun Qian, Chenshu Wu, Yi Zhang, Guidong Zhang, Zheng Yang, and Yunhao Liu. Wider2. 0: Passive human tracking with a single Wi-Fi link. In *Proceedings of ACM MobiSys*, page 350–361, 2018.
- [35] Tong Xin, Bin Guo, Zhu Wang, Pei Wang, Jacqueline Chi Kei Lam, Victor Li, and Zhiwen Yu. Freesense: a robust approach for indoor human detection using Wi-Fi signals. *ACM IMWUT*, 2(3):1–23, 2018.
- [36] Venkat Arun and Hari Balakrishnan. Rfocus: Beamforming using thousands of passive antennas. In *Proceedings of Usenix NSDI*, pages 1047–1061, 2020.
- [37] Nirupam Roy, Sheng Shen, Haitham Hassanieh, and Romit Roy Choudhury. Inaudible voice commands: The long-range attack and defense. In *Proceedings of Usenix NSDI*, page 547–560, 2018.
- [38] J.N. Moutinho, R.E. Araújo, and D. Freitas. Indoor localization with audible sound – towards practical implementation. *Pervasive and Mobile Computing*, 29(25-33):1–16, 2016.
- [39] Sheng Shen, Dagan Chen, Yu-Lin Wei, Zhijian Yang, and Romit Roy Choudhury. Voice localization using nearby wall reflections. In *Proceedings of ACM MobiCom*, page 1–14, 2020.
- [40] Song Wang, Jingqi Huang, Xinyu Zhang, Hyoil Kim, and Sujit Dey. X-array: Approximating omnidirectional millimeter-wave coverage using an array of phased arrays. In *Proceedings of ACM MobiCom*, pages 1–14, 2020.
- [41] Xiangmao Chang, Cheng Peng, Guoliang Xing, Tian Hao, and Gang Zhou. Isleep: A smartphone system for unobtrusive sleep quality monitoring. *ACM Trans. Sen. Netw.*, 16(3), 2020.
- [42] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie. VSkin: Sensing touch gestures on surfaces of mobile devices using acoustic signals. In *Proceedings of ACM MobiCom*, page 591–605, 2018.
- [43] Haoran Wan, Lei Wang, Ting Zhao, Ke Sun, Shuyu Shi, Haipeng Dai, Guihai Chen, Haodong Liu, and Wei Wang. Vector: Velocity based temperature-field monitoring with distributed acoustic devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(3):1–28, 2022.
- [44] Lei Wang, Haoran Wan, Ting Zhao, Ke Sun, Shuyu Shi, Haipeng Dai, Guihai Chen, Haodong Liu, and Wei Wang. Scalar: Self-calibrated acoustic ranging for distributed mobile devices. *IEEE Transactions on Mobile Computing*, pages 1–15, 2023.
- [45] Ronald Schafer Alan Oppenheim. *Discrete-Time Signal Processing*. Pearsom, 2009.
- [46] B. M. Popovic. Generalized chirp-like polyphase sequences with optimum correlation properties. *IEEE TIT*, 38(4):1406–1409, 1992.

- [47] Wenguang Mao, Jian He, and Lili Qiu. CAT: High-precision acoustic motion tracking. In *Proceedings of ACM MobiCom*, page 69–81, 2016.
- [48] Anran Wang and Shyamnath Gollakota. MilliSonic: Pushing the limits of acoustic motion tracking. In *Proceedings of ACM CHI*, page 1–11, 2019.
- [49] Yuzhou Zhuang, Yuntao Wang, Yukang Yan, Xuhai Xu, and Yuanchun Shi. Reflectrack: Enabling 3d acoustic position tracking using commodity dual-microphone smartphones. In *Proceedings of ACM UIST*, page 1050–1062, 2021.
- [50] Wulf-Dieter Wirth. *Radar Techniques Using Array Antennas*. Radar, Sonar & Navigation. Institution of Engineering and Technology, 2013.
- [51] Marvin H.J. Gruber. Statistical digital signal processing and modeling. *Technometrics*, 39(3):335–336, 1997.
- [52] R. Dwyer. Detection of non-gaussian signals by frequency domain kurtosis estimation. In *Proceedings of IEEE ICASSP*, volume 8, pages 607–610, 1983.
- [53] K. V. Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530, 1970.
- [54] Wei Wang, Alex X. Liu, and Ke Sun. Device-free gesture tracking using acoustic signals. In *Proceedings of ACM MobiCom*, page 82–94, 2016.
- [55] Raspberry Pi Foundation. Raspberry pi 3 model b+. <https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/>, 2023.
- [56] Seeed Studio. Respeaker 6-mic circular array kit for raspberry pi. https://wiki.seeedstudio.com/ReSpeaker_6-Mic_Circular_Array_kit_for_Raspberry_Pi/, 2023.
- [57] Vernier Science Education. Go direct respiration belt. <https://www.vernier.com/product/go-direct-respiration-belt/>, 2023.
- [58] Seeed Studio. Respeaker mic array. https://wiki.seeedstudio.com/ReSpeaker_Mic_Array/, 2023.