RespTracker: Multi-user Room-scale Respiration Tracking with Commercial Acoustic Devices

Haoran Wan, Shuyu Shi, Wenyu Cao, Wei Wang, Guihai Chen
State Key Laboratory for Novel Software Technology, Nanjing University
{wanhr, wenyucao}@smail.nju.edu.cn, {ssy, ww, gchen}@nju.edu.cn

Abstract—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 two-stage algorithm to separate and recombine respiration signals from multiple paths in a short period so that it can track the respiration rate of multiple moving subjects. Our experimental results show that our two-stage algorithm can distinguish the respiration of at least four subjects at a distance of three meters.

I. INTRODUCTION

Background and Motivation: Respiration is one of the vital signs that contain valuable information to diagnose assorted diseases, *e.g.*, pulmonary disease [1], heart failure [2], anxiety [3], and sleep disorders [4]. 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 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 [5] or using special devices, including belt integrated with capacitive sensors [6] or smart cushion with air pressure sensors [7]. However, user studies have shown that people are reluctant to deploy these devices due to privacy concerns [5], [8] or the high cost and long-term physical contact requirements [6], [7]. 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 [9]–[11].

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 [12]. 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
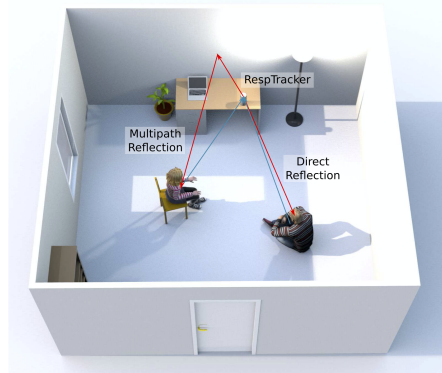
Figure 1. General application scenario of RESPTRACKER.

for a Wi-Fi bandwidth of 40 MHz. Existing works either rely on the differences in the respiration rate [11] or use specialized high bandwidth frequency modulated continuous wave (FMCW) radar and Independent Component Analysis (ICA) [9] 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 [10], *e.g.*, a typical bandwidth of 4kHz 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 [10], [13], [14]. 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, the first continuous, multiple-person respiration tracking system in domestic settings using acoustic-based COTS devices. As shown in Figure 1, the respiration signal of different users may arrive at the receiver through multiple paths. RESPTRACKER proposes a multipath *separation* and *combination* framework for robust respiration signal tracking.

First, RESPTRACKER utilizes inaudible sound signal modulated by the Zadoff-Chu (ZC) sequence 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 measure both the amplitude and the phase of *individual* reflection paths. 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, multiple receiving microphones, and different time-frames, based on the amplitude and phase measurement of the ZC signal. In this way, we can reliably cluster reflection paths to different users even if they have similar respiration rates. With our two-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. And, 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 [9] 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 10cm. In addition, we can measure the features of individual paths in terms of the channel impulse response (CIR). In this way, each path contains less interference of other subjects so that the difficulty 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. Traditional delay-and-sum algorithm for beamforming 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. By clustering these multi-dimensional reflection signals, we can determine whether a given path on a given microphone contains the respiration signal and 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 signal while the subject is moving. 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

(a) CIR amplitude of a single frame.

(b) Time variations of CIR amplitudes.

(c) Filtered CIR amplitude at different distances.

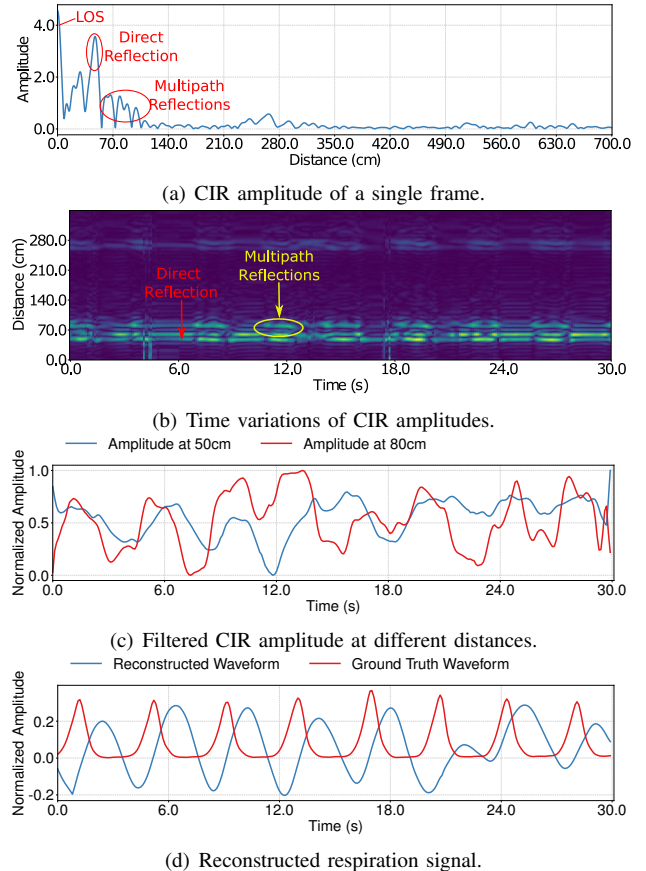

(d) Reconstructed respiration signal.

Figure 2. CIR waveform of a single subject

determine whether there are users' 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.

Summary of Experimental results: 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 signal of more than four users in the same room and achieve an error of less than 1 BPM for each user.

II. 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.

A. Design Motivations

To understand the design challenges for long-range respiration signal detection and separation, we provide a typical respiration signal illustration in Figure 2. Figure 2(a) shows the amplitude of multipath signals at different distances, where
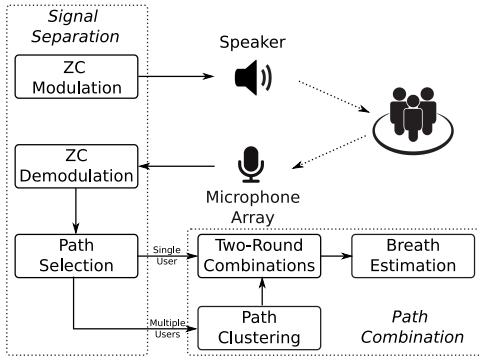
Figure 3. System Overview of RESPTRACKER

each peak corresponds to one signal path. From Figure 2(a), we have two observations. First, due to the high resolution of the sound signal, the width of each peak is less than 10 cm so that theoretically we can separate two users even if they are just 10 cm apart. Second, the sound signal attenuates quickly and it is hard to reliably detect peaks at a distance of 4 meters.

Figure 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 half a minute, *e.g.*, the LOS path and reflections of walls. We observe from Figure 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 before reaching the chest of the user and may reflect from different parts of the chest. While these reflections are weak, they provide important respiration information of the same user. This is because it is well known that the signal quality of a single path largely depends on the posture and angle of the user [11]. 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.

Figure 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.

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?

B. System Design

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

The first stage is signal separation. 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

cross-correlation between the received and the transmitted signal to derive the CIR. We detect each path in random sampled frames and calculate the respiration SNR in the frequency domain to select paths that are candidates of respiration related reflections that will be used in the second stage.

The second stage is path combination. To expand the sensing range, 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. 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. Finally, we use the reconstructed breath signal to perform breath rate estimation for each user.

III. SIGNAL SEPARATION

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

A. ZC Modulation

The transmitting signal used in RESPTRACKER is the ZC sequence modulated by a sinusoid carrier [15]. The ZC sequence with a length of N_{zc} is given by:

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

where the u and q are the parameters of the sequence. We set q to 0, u to 1, and N_{zc} to 199 representing a 2 kHz bandwidth in the modulated signal. Once we get the baseband signal, we use frequency domain interpolation to expand the sequence to a length of L, which is the frame length of our OFDM symbol and is set to 4800 samples in our scheme. We then modulate the signal with a carrier sinusoid at a frequency of f_c 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. Algorithm 1 shows the detailed process, where f_s is the sampling frequency. After we generate one frame of the time-domain real signal $zc_T[n]$, we transmit it repeatedly so that the transmitted signals are cyclical OFDM symbols.

Algorithm 1: Transmitting signal generation

Result: The modulated sequence $zc_T[n]$ 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 $\widehat{ZC}[L - \frac{f_c L}{f_s} - \frac{N_{zc}-1}{2} : L - \frac{f_c L}{f_s} + \frac{N_{zc}-1}{2}] \Leftarrow ZC^*[n]$.
7 Perform IFFT on $\widehat{ZC}$ to the time-domain $zc_T[n]$.

B. ZC Demodulation

After the signal is transmitted from the speaker, the microphone array at receiver side records the signals that comes from both the LOS path and the reflections of subjects and the environment. 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 [15]. Instead of using the time domain down-conversion and correlation as in [15], 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:

$$ zc_R[n] = \sum_{i=1}^{P} A_i e^{-j\phi_i(t)} zc_T \left[n - \frac{\tau_i}{f_s} \right], \tag{2} $$

Where $zc_R[n]$ 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 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 [16], 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 versions of transmitted signal from different paths, as shown in Figure 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 $zc_R[n]$ 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/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 Figure 2(b). The time-domain resolution of 0.1s in the CIR map gives a sampling rate of 10Hz, which is adequate for monitoring respiration signals that have typical frequency of 0.1~0.5 Hz.

C. 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:

$$ zc_{R_b}[t] = A e^{-j\left(\frac{2\pi f d(t)}{c} + p\right)} zc_T \left[n - \frac{d_{body} + d(t)}{c \times f_s} \right] \tag{3} $$

Where d_{body} is the path length of user's body reflection, $d(t)$ is the chest movement during the exhaling and inhaling, which is a periodic signal, and p is the phase shift cause by the software delay and reflection phase inversion. Under this model, the corresponding CIR is:

$$ cir_{R_b}[t] = A e^{-j\left(\frac{2\pi f d(t)}{c} + p\right)} sinc \left[n - \frac{d_{body} + d(t)}{c \times f_s} \right] \tag{4} $$

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 $\frac{1}{f_b}$, 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 for monitoring, we first determine the location of the LOS path by voting for the maximum peak location of the first L_v frames which is set to 20 in experiments and is corresponding to 2 seconds. The LOS localization is an one-time calibration because the distance between speaker and microphone is fixed during the monitoring.

Static Signal Removal and Random Sampling: After localized 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 of each observation slot from each CIR frame. In this way, the remaining non-zero peaks corresponds to dynamical paths. We then randomly sample R frames 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 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. This effectively removes the fluctuation caused by the side-lobes of the sinc function and we set the threshold $\beta = 0.2$. Second, we remove paths that are within T_b sample points to avoid repetition.

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 0.1~0.5 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∼0.5Hz as E_{max}. The breath SNR is defined as $w_1 \frac{E_{max}}{(\sum_{f \in [0.1,0.5]} E_f) - E_{max}} + w_2 \frac{E_{max}}{\sum_{f \in [0.5,5]} E_f}$, which 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 detect the candidate paths that corresponds to breath movements for further path combinations in the next section.

IV. PATH COMBINATION

In this section, we reconstruct the respiration signal through two-round path combinations on the candidate paths detected in the previous section. We also illustrate how to separate the respiration signal of multiple users and how to track users if they move during the monitoring.

A. Two-Round Combinations

Traditional delay-and-sum combinations for beamforming does not work well for respiration signal reconstruction as shown by our experimental results in Section V. Therefore, we propose a two-round combinations scheme to enhance the respiration signal.

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: In this subsection, we first consider path combination for a single user where all candidate paths are from the same respiration movements. 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 [17] 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 pattern, they have different phases and signal details, see Figure 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. Figure
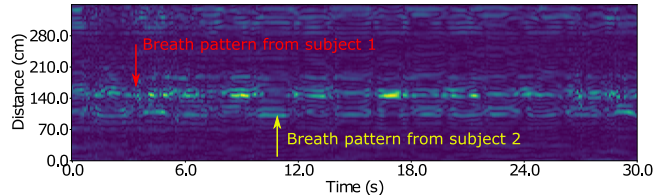

Figure 4. CIR map with two users in the environment.

2(d) compares the reconstructed respiration waveform and the ground truth waveform captured by the respiration belt.

B. Path Clustering for Multiple Users

In real-world scenarios, there might be more than one user in the room. Therefore, we need to distinguish the paths belonging to each user before performing the combination.

Separation of Different Users: As our ZC sequence has a range resolution of around 10 cm, we can separate users by their different distances to the receiver. Figure 4 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 treat the user separation problem as an unsupervised classification problem and use the K-means algorithm to perform clustering of paths. As different users may have similar breath rates and phases, we use the distance as the feature of the clustering algorithm. 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 of each user.

In the multiple users scenario, each user may have different breath SNR. So, we reduce the SNR threshold to cover more paths to include more paths for multipath clustering.

C. 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. 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 pattern 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, we should entirely discard the given slot. In this case, the reconstructed waveform has an abrupt shape with no periodicity

which can be used for motion detection during monitoring and it's beyond the scope of our work. We should also perform a re-synchronization in the next observation slot, especially when there are more than one users in the sensing range. 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 moving, static users will be first matched to their historical clusters and moving users will be relocated to new positions.

D. Breath Rate 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 $0.1Hz$ to $0.5Hz$, we cannot use a fixed empirical window size for the moving average filter. To adapt the length of moving average filter, we 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.

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.

V. IMPLEMENTATION AND EVALUATION

We implement RESPTRACKER on Raspberry Pi 3B+ [18] and desktop computers using Python. The Raspberry Pi is equipped with a speaker and a 6-mic circular microphone array [19] to transmit and receive acoustic signal at a sample rate of 48 kHz. 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 [20] 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 Figure 5.

We use two key metrics to evaluate the performance of RESPTRACKER. The first metric is the 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. For each experiment set, we repeat the processing for 10 times and use the average errors as the experimental error to reduce the impact of random sampling process in our algorithm.

A. Experiments in the Single User Scenario

Effective range: To evaluate the breath detection range of RESPTRACKER, we conduct experiments in the hallway shown in Figure 6, 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 degree to find a coarse-grained elevation and azimuth and then fix the elevation and search with a stride of 1 degree surrounding found azimuth to find the final parameters. For each combination of these parameters,we calculate the breath SNR for all the possible paths to find the combination to maximize 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.

Figure 7 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 Figure 7(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.

Figure 8 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.

Robustness: To evaluate the robustness of our system in different environments, we conduct experiments at different locations in typical indoor environments, including hallway, office room, conference room, and student apartment. Figure 6 shows the sample experimental environments. 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
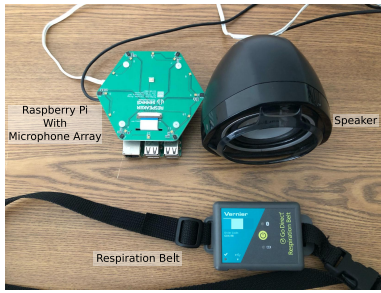
Figure 5. Devices used in the experiments, including Raspberry Pi, speaker, and respiration belt.
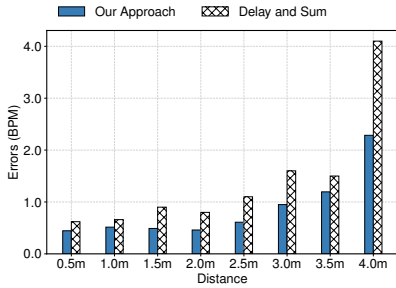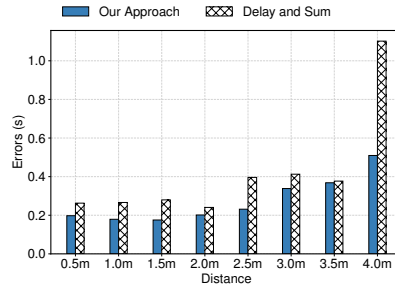

(a) Hallway.

(b) Office room.

Figure 6. Samples of experimental environments.

(a) Errors in BPM.

(b) Errors in breath interval time.

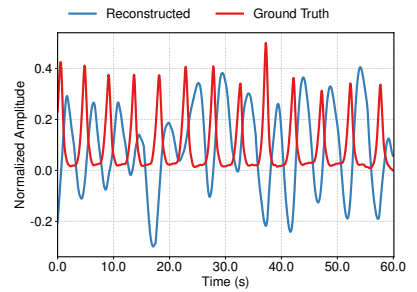Figure 7. Experimental results at different distances.

Figure 8. Reconstructed waveform in the hallway at position 1.

to be one meter. We ask each subject to breath for two minutes for five repetitions while sitting on a chair.

Figure 9 shows that our algorithm is robust to environmental changes. Different environments only slightly affect the performance of our system. The means absolute error at the hallway, office room, conference room and student apartment shown in Figure 9(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 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 Figure 9(c).

Tracking Performance: To evaluate the tracking performance, we requested the subjects to move during the two minutes breathing period, but 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.

Figure 10 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 second for these movement types.

B. 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 0.5m to change the distance between two subjects. At each separation distance, we collect 2 minutes breathing signal for 5 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.

Figure 11 shows that RESPTRACKER can reliably separate the breath signals of two users. We have two observations from Figure 11. First, even if the two subjects sit shoulder to shoulder at the 1.0 m distance, we can still separate their breath signal because the propagation distance of their reflections are still different. However, if their positions lie on the same ellipse with speaker and microphone on the focal point, our system may fail. Second, as subject 2 moves away from subject 1, the error for subject 1 decreases and then stays stable and the error for subject 2 first decreases and then increases. This is because when the spacing between them becomes larger we can fully exploit the multipath reflections to reconstruct a better waveform for both users. But, as
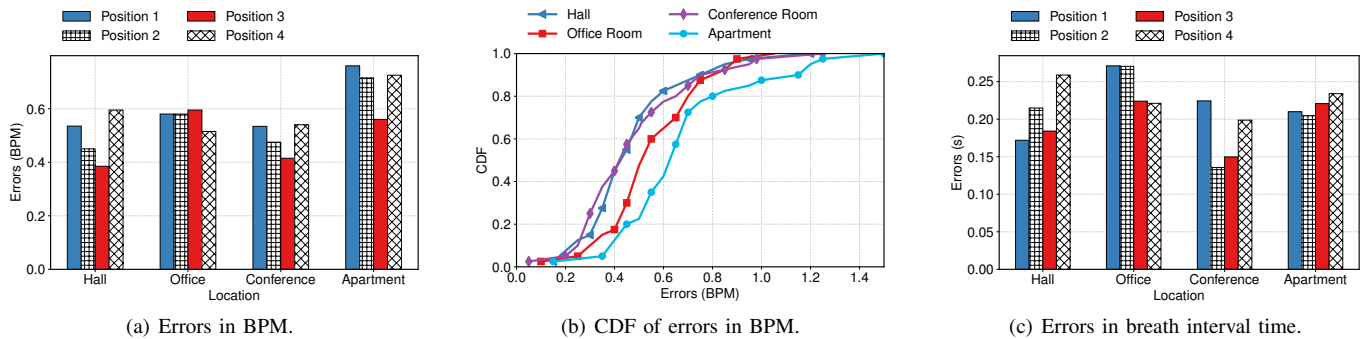
(a) Errors in BPM.

(b) CDF of errors in BPM.

(c) Errors in breath interval time.

Figure 9. Experimental results in different environments.

(a) Error in BPM.

(b) Errors in breath interval time.

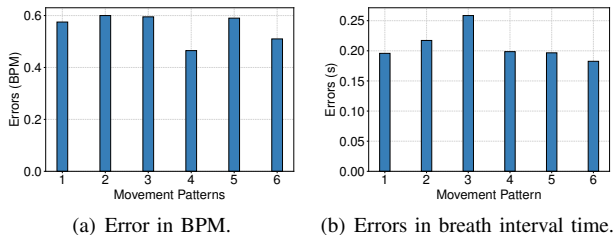Figure 10. Experimental results for tracking.
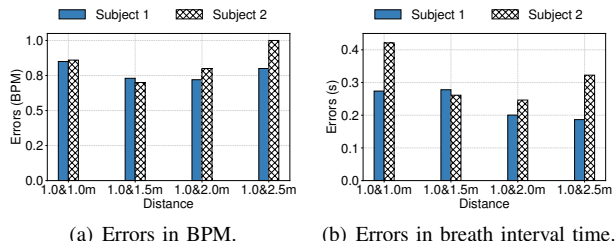
(a) Errors in BPM.

(b) Errors in breath interval time.

Figure 11. Experimental results for spatial resolution at different distances.
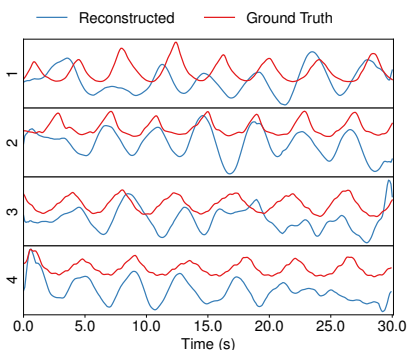
Figure 12. Reconstructed waveform for four users.

(a) Error in BPM.

(b) Errors in breath interval time.
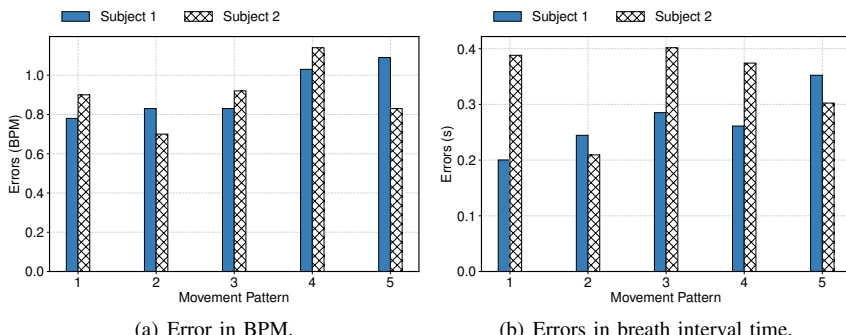
Figure 13. Experimental results for different multi-user movement patterns.

the distance increases the acoustic signal's becomes weaker and the error for subject 2 increases. Figure 12 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 0.91 BPM and 0.285 seconds.

Tracking Performance: In the multiple users scenario, we consider the movement pattern 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 six moving patterns which are moving from 1.0 m to 0.5 m, moving from 1.5 m to 2.0 m, moving simultaneously from 1.0 m to 0.5 m and 1.5 m to 2.0 m, moving from 1.0 m to 2.0 m and moving from 1.5 m to 0.5 m. We denote these movement patterns as pattern 1 to 5 in the figure.

Figure 13 shows that RESPTRACKER can track the movement of multiple users under different moving patterns. In

the first three patterns, two subjects doesn't meet during the movement and the error for these experiments are similar to the static scenario. In the last two patterns, the subjects meet during the movement which may affect the matching process of the paths before and after the movement, so the error is slightly higher for these two cases. The error in breath interval time in Figure 13(b) also shows a similar trend.

C. 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 I shows the computational time for RESPTRACKER to process audio data of one observation slot (20 seconds) on

TABLE I
PROCESSING TIME ON DIFFERENT PLATFORMS

	Single User	Multiple Users
PC (RESPTRACKER)	0.1338s	0.3030s
PC (Delay-and-Sum)	14.5805s	-
Raspberry Pi (RESPTRACKER)	2.0830s	3.7712s

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 ten times. 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 and 3.7712 s for single user and multiple users. The reason why the delay-and-sum method so slow is that it has to search exhaustively in all azimuth and elevation to find the valid delay and combine the signal.

VI. 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 [9], [21]–[23]. BreathTaking [21] leverages the received signal strength between different pairs of network devices to conduct the contactless breath monitoring for single person on the bed. DeepBreath [9] uses multiple FMCW transceivers and the ICA algorithm to separate different users' respiration signal. Liu *et al.* [22] extract breath and heart beats from the CSI gathered by commodity Wi-Fi devices. Wang *et al.* [11] propose the Fresnel Zone model of Wi-Fi sensing in which the subject's respiration can be hardly recognized by the CSI amplitude. To tackle this challenge, Zeng *et al.* [23] exploit the complement-ary 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 [24]. Yang *et al.* [25] leverage the high distance resolution of UWB radars to separate different subjects' respiration and use image processing techniques to detect sleep apnea. ViMo [26] leverages the high spatial resolution (distance, azimuth and elevation) of 60GHz millimeter wave antenna array to extract both the respiration rate and the heart rate of multiple subjects. Although wireless devices can provide strong signal with good quality, they are expensive and the monitoring process might interfere with normal data transmissions.

Respiration Monitoring with Acoustic Signals: Acoustic signal travels much slower than wireless signals. The sampling rate of 48 kHz from COTS microphones provides a fine range resolution of 0.7 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 [10] transmits 18~20 kHz FMCW sound signals to estimate breathing frequency and detect sleep apnea passively. Wang *et al.* [13] 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 [14] uses cross-correlation to increase the accuracy of the acoustic-based breath sensing. Xu *et al.* [27] leverage Energy Spectrum Density (ESD) of a single-frequency acoustic signal, Ensemble Empirical Mode Decomposition, and Generative Adversarial Network to reconstruct the breathing signal in a driving scenario. The vital problem of 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.

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 [28] enables phase array signal processing on COTS device which increases spatial resolution, decreases phase error and suppressed the multipath interference. Wang *et al.* [29] use a blind distributed beamforming on both uplink and downlink to increase the backscatter sensing distance to 64 meters. For wireless sensing and tracking, mdTrack [30] gives a multi-dimensional Wi-Fi localization model and drastically increases passive localization divisibility. Vasisht *et al.* [31] analyze time of flight in different frequency band and between different TX/RX pairs to reach decimeter-level localization accuracy. WiDar 2.0 [32] and FreeSense [33] calculate the AoA and ToF to match and localize the moving subject. In sound signal processing, Roy *et al.* [34] set up a speaker array to achieve long-range ultrasound attacks on voice assistants. Moutinho *et al.* [35] address the inverse problem of localizing microphones with speaker arrays that are playing predefined sounds. Shen *et al.* [36] leverage microphone array and reflections from the wall to localize sound source. Most of the existing beamforming techniques ignore the possibility of utilizing the multipath effect to enhance the received signal.

VII. 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 mul-tipath 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.

VIII. ACKNOWLEDGEMENT

REFERENCES

[1] N. Fens, A. H. Zwinderman, M. P. van der Schee, S. B. de Nijs, E. Dijkers, A. C. Roldaan, D. Cheung, E. H. Bel, and P. J. Sterk, "Exhaled breath profiling enables discrimination of chronic obstructive pulmonary disease and asthma," *American journal of respiratory and critical care medicine*, vol. 180, no. 11, pp. 1076–1082, 2009.

[2] S. Javaheri, T. Parker, J. Liming, W. Corbett, H. Nishiyama, L. Wexler, and G. Roselle, "Sleep apnea in 81 ambulatory male patients with stable heart failure," *Circulation*, vol. 97, no. 21, pp. 2154–2159, 1998.

[3] F. H. Wilhelm, W. T. Roth, and M. A. Sackner, "The lifeshirt: an advanced system for ambulatory measurement of respiratory and cardiac function," *Behavior Modification*, vol. 27, no. 5, pp. 671–691, 2003.

[4] C. Guilleminault, R. Pelayo, D. Leger, A. Clerk, and R. C. Bocian, "Recognition of sleep-disordered breathing in children," *Pediatrics*, vol. 98, no. 5, pp. 871–882, 1996.

[5] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM transactions on graphics*, vol. 31, no. 4, pp. 1–8, 2012.

[6] C. R. Merritt, H. T. Nagle, and E. Grant, "Textile-based capacitive sensors for respiration monitoring," *IEEE Sensors Journal*, vol. 9, no. 1, pp. 71–78, 2009.

[7] M. B. Norman, S. Middleton, O. Erskine, P. G. Middleton, J. R. Wheatley, and C. E. Sullivan, "Validation of the sonomat: a contactless monitoring system used for the diagnosis of sleep disordered breathing," *Sleep*, vol. 37, no. 9, pp. 1477–1487, 2014.

[8] C.-W. Lin and Z.-H. Ling, "Automatic fall incident detection in compressed video for intelligent homecare," in *Proceedings of IEEE ICCCN*, 2007.

[9] S. Yue, H. He, H. Wang, H. Rahul, and D. Katabi, "Extracting multi-person respiration from entangled RF signals," *ACM IMWUT*, vol. 2, no. 2, 2018.

[10] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *Proceedings of ACM MobiSys*, 2015.

[11] H. Wang, D. Zhang, J. Ma, Y. Wang, Y. Wang, D. Wu, T. Gu, and B. Xie, "Human respiration detection with commodity WiFi devices: do user location and body orientation matter?," in *Proceedings of ACM UbiComp*, 2016.

[12] S. Shi, Y. Xie, M. Li, A. X. Liu, and J. Zhao, "Synthesizing wider wifi bandwidth for respiration rate monitoring in dynamic environments," in *Proceedings of IEEE INFOCOM*, pp. 181–189, 2019.

[13] A. Wang, J. E. Sunshine, and S. Gollakota, "Contactless infant monitoring using white noise," in *Proceedings of ACM MobiCom*, 2019.

[14] T. Wang, D. Zhang, Y. Zheng, T. Gu, X. Zhou, and B. Dorizzi, "C-FMCW based contactless respiration detection using acoustic signal," in *Proceedings of ACM UbiComp*, 2018.

[15] K. Sun, T. Zhao, W. Wang, and L. Xie, "VSkin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," in *Proceedings of ACM MobiCom*, 2018.

[16] B. M. Popovic, "Generalized chirp-like polyphase sequences with optimum correlation properties," *IEEE Transactions on Information Theory*, vol. 38, no. 4, pp. 1406–1409, 1992.

[17] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, MobiCom '16, (New York, NY, USA), p. 82–94, Association for Computing Machinery, 2016.

[18] https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/.

[19] https://wiki.seeedstudio.com/ReSpeaker_6-Mic_Circular_Array_kit_for_Raspberry_Pi/.

[20] https://www.vernier.com/product/go-direct-respiration-belt/.

[21] N. Patwari, J. Wilson, S. Ananthanarayanan, S. K. Kasera, and D. R. Westenskow, "Monitoring breathing via signal strength in wireless networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1774–1786, 2014.

[22] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf WiFi," in *Proceedings of ACM MobiHoc*, 2015.

[23] Y. Zeng, D. Wu, R. Gao, T. Gu, and D. Zhang, "Fullbreathe: Full human respiration detection exploiting complementary of CSI phase and amplitude of WiFi signals," *ACM IMWUT*, vol. 2, no. 3, pp. 1–19, 2018.

[24] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "Farsense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, Sept. 2019.

[25] Y. Yang, J. Cao, X. Liu, and X. Liu, "Multi-breath: Separate respiration monitoring for multiple persons with UWB radar," in *Proceedings of IEEE COMPSAC*, 2019.

[26] F. Wang, F. Zhang, C. Wu, B. Wang, and K. R. Liu, "ViMo: Multi-person vital sign monitoring using commodity millimeter wave radio," *IEEE Internet of Things Journal*, 2020.

[27] X. Xu, J. Yu, Y. Chen, Y. Zhu, L. Kong, and M. Li, "Breathlistener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '19, (New York, NY, USA), p. 54–66, Association for Computing Machinery, 2019.

[28] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson, "Phaser: Enabling phased array signal processing on commodity WiFi access points," in *Proceedings of ACM MobiCom*, 2014.

[29] J. Wang, J. Zhang, R. Saha, H. Jin, and S. Kumar, "Pushing the range limits of commercial passive RFIDs," in *Proceedings of Usenix NSDI*, 2019.

[30] Y. Xie, J. Xiong, M. Li, and K. Jamieson, "mD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking," in *Proceedings of ACM MobiCom*, 2019.

[31] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point," in *Proceedings of Usenix NSDI*, 2016.

[32] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2.0: Passive human tracking with a single Wi-Fi link," in *Proceedings of ACM MobiSys*, 2018.

[33] T. Xin, B. Guo, Z. Wang, P. Wang, J. C. K. Lam, V. Li, and Z. Yu, "Freesense: a robust approach for indoor human detection using Wi-Fi signals," *ACM IMWUT*, vol. 2, no. 3, pp. 1–23, 2018.

[34] N. Roy, S. Shen, H. Hassanieh, and R. R. Choudhury, "Inaudible voice commands: The long-range attack and defense," in *Proceedings of Usenix NSDI*, 2018.

[35] J. Moutinho, R. Araújo, and D. Freitas, "Indoor localization with audible sound — towards practical implementation," *Pervasive and Mobile Computing*, vol. 29, pp. 1 – 16, 2016.

[36] S. Shen, D. Chen, Y.-L. Wei, Z. Yang, and R. R. Choudhury, "Voice localization using nearby wall reflections," in *Proceedings of ACM MobiCom*, 2020.