SpaceBeat: Identity-aware Multi-person Vital Signs Monitoring Using Commodity WiFi

BOFAN LI, Florida State University, USA
YILI REN, University of South Florida, USA
YICHAO WANG, Florida State University, USA
JIE YANG, Florida State University, USA

Vital signs monitoring has gained increasing attention due to its ability to indicate various human health and well-being conditions. The development of WiFi sensing technologies has made it possible to monitor vital signs using ubiquitous WiFi signals and devices. However, most existing approaches are dedicated to single-person scenarios. A few WiFi sensing approaches can achieve multi-person vital signs monitoring, whereas they are not identity-aware and sensitive to interferences in the environment. In this paper, we propose SpaceBeat, an identity-aware and interference-robust multi-person vital sign monitoring system using commodity WiFi. In particular, our system separates multiple people and locates each person in the spatial domain by leveraging multiple antennas. We analyze the change of signals at the location of each person to achieve identity-aware vital signs monitoring. We also design a contrastive principal component analysis-contrastive learning framework to mitigate interferences caused by other moving people. We evaluate SpaceBeat in various challenging environments, including interference scenarios, non-line-of-sight scenarios, different distances, etc. Our system achieves an average accuracy of 99.1% for breathing monitoring and 97.9% for heartbeat monitoring.

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

Additional Key Words and Phrases: WiFi sensing, Vital signs monitoring, Multi-person, Identity-aware, Interference robust, Deep contrastive learning

ACM Reference Format:
Bofan Li, Yili Ren, Yichao Wang, and Jie Yang. 2024. SpaceBeat: Identity-aware Multi-person Vital Signs Monitoring Using Commodity WiFi. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 113 (September 2024), 23 pages. https://doi.org/10.1145/3678590

1 INTRODUCTION

Vital signs, especially breathing and heartbeat, play a critical role in the indication of various health conditions in humans. They are essential components to detect and track diseases in many areas. For example, breathing signals can be utilized to diagnose shortness of breath [37] and airway obstruction [7]. Heartbeat monitoring can achieve early detection of many heart diseases such as heart arrhythmia [6] and heart failure [14]. Moreover, various emerging well-being applications, including emotion recognition [53], sleep monitoring [47], and identification of patient deterioration [12], also heavily depend on the monitoring of vital signs.

Authors' addresses: Bofan Li, Florida State University, 1017 Academic Way, Tallahassee, Florida, USA, 32306, boli@cs.fsu.edu; Yili Ren, University of South Florida, 4202 E. Fowler Avenue, Tampa, Florida, USA, 33620, yiliren@usf.edu; Yichao Wang, Florida State University, 1017 Academic Way, Tallahassee, Florida, USA, 32306, ywang@cs.fsu.edu; Jie Yang, Florida State University, 1017 Academic Way, Tallahassee, Florida, USA, 32306, jie.yang@cs.fsu.edu.

2474-9567/2024/9-ART113
https://doi.org/10.1145/3678590

Traditional solutions for vital sign monitoring are invasive and relatively cumbersome. They require the person to wear or carry wearable devices [16] (e.g., breathing belt and pulse oximeter) during the whole monitoring procedure and are hard to apply to long-term monitoring [22]. To overcome the drawbacks of traditional solutions, contact-free vital sign monitoring systems, especially radio frequency (RF) sensing-based systems, are developed to realize more convenient, long-term, and ubiquitous vital sign monitoring. However, their reliance on specialized hardware, such as IR-UWB radar [11, 54, 55] and FMCW radar [48], restricts their consumer-oriented applications and impedes their widespread deployment due to their high costs.

Recently, WiFi sensing-based systems have gained numerous research interests as they can reuse commodity WiFi devices and can be easily integrated with smart home and Internet of Things (IoT) environments [1, 10, 22, 40, 41, 51]. Nevertheless, most existing WiFi-based vital sign monitoring systems can only work for a single human subject and are sensitive to interferences in the environment [35]. Although a few WiFi-based systems can extract vital signs from multiple people, they cannot achieve identity-aware monitoring. In other words, they can separate the vital sign signals of multiple people in the signal domain but do not know to whom each vital sign signal belongs. Besides, they are not interference-robust. This is because WiFi-based vital sign monitoring relies on the signal reflections from the human body, whereas the vital sign signals can be disturbed by the signals reflected off other people who are breathing or moving around (i.e., dynamic interferences). These limitations dramatically restrict the applications that require accurate and identity-aware vital sign monitoring.

In this paper, we propose SpaceBeat, an identity-aware and interference-robust multi-person vital sign monitoring system using commodity WiFi. Specifically, in multi-person scenarios, our system can locate each person and identify the vital signs (i.e., breathing and heartbeat) of each specific person and thus achieve identity-aware monitoring. Moreover, our system provides accurate vital signs monitoring which is robust to dynamic interferences introduced by nearby people in motion. Also, our system can reuse commodity WiFi to continuously monitor the vital signs of people and thus it is possible to be widely adopted in the smart home and IoT environments.

In particular, we propose to separate multiple people in the spatial domain and further extract vital signs for each specific person. Current WiFi devices support multiple antennas and the new-generation WiFi devices tend to support more and more antennas (e.g., WiFi 6 or 7 supports up to 8 or 16 antennas). Thus, we exploit the use of multiple spatially distributed antennas on the WiFi receiver to form a two-dimensional (e.g., L-shaped) antenna array. The antenna array enables us to estimate spatial information of the subjects in terms of the two-dimensional angle of arrival (2D AoA) (i.e., azimuth and elevation) of WiFi signal reflections. By obtaining such spatial information, our system could separate multiple people based on their positions in the 2D AoA spatial domain.

Accordingly, we can locate each individual in the 2D AoA spatial domain and extract his/her vital signs by analyzing the changes in WiFi signals originating from his/her location. Therefore, we can achieve identity-aware monitoring which can identify whom each vital sign belongs to. It is worth noting that our identity-aware monitoring is fundamentally different from identification [30]. Our identity-aware vital signs monitoring system intends to find the vital signs that match specific people based on spatial information. The spatial information makes our system identity-aware while identifying refers to the process that the system recognizes and differentiates between different individuals based on their unique signatures [18].

Although the intuition is straightforward, we face some unique challenges that hinder us from accurately extracting vital sign signals from multiple people. The first challenge is the coupling of signals reflected off multiple people. The motion of vital signs or body movements of people can modulate WiFi signal reflections and these signals will be superimposed on each other in 2D AoA spatial domain. Thus, the coupling includes two aspects. On the one hand, the signal reflections caused by the breathing and heartbeat of multiple people can interfere with each other. On the other hand, the movements of nearby people lead to dynamic interferences which severely change the multipath propagation and significantly affect the vital sign extraction of a target person. To address this challenge, we design a contrastive principal component analysis (cPCA)-contrastive

learning (CL) framework to achieve the decoupling of signals reflected off multiple people. In the cPCA-CL framework, we designate the vital sign signals of the target person as the foreground, while considering the vital sign signals of other people or dynamic interferences as the background. Then, our cPCA-CL framework will separate and highlight the vital sign signals of the target person. In particular, cPCA is employed to remove a background from a single foreground in the 2D AoA spatial domain. Furthermore, CL is utilized to remove backgrounds from a sequence of foregrounds in the temporal domain. In addition, we iteratively perform the cPCA-CL framework for further decoupling, which will lead to cleaner and more refined vital sign signals for each individual.

The second challenge is the extraction of the subtle heartbeat. In contrast with the centimeter-level breathing movements of the chest and abdomen, the heartbeat is orders of magnitude smaller. Thus, heartbeat signals can be easily overwhelmed by breathing signals, harmonics of breathing, and other signals such as interferences in the environment. These lead to a low signal-to-noise ratio (SNR) of heartbeat signals and thus make it difficult to detect and extract accurate heart rates. To overcome this challenge, we propose to precisely locate each person and only focus on the person at the target location in the 2D AoA spatial domain to eliminate interferences at other locations. Next, we extract the frequency of breathing from the time-series decoupled signals of each person at the target location. We then develop a harmonic canceller to automatically remove the harmonics of the breathing signal and thus improve the SNR of heartbeat signals.

The third challenge is how to locate each person accurately. 2D AoA can only roughly separate and locate the people as the number of antennas of commodity WiFi devices is still limited, resulting in inadequate resolvability for multiple individuals. Therefore, we further leverage multidimensional information of WiFi signals to improve the resolvability [46]. Specifically, we incorporate time of flight (ToF) and angle of departure (AoD) into the 2D AoA. In addition to the improvement of resolvability, we refine the locations of people by using ToF and AoD to mitigate the interferences of signals that are reflected multiple times in the environment. Our insight is that the signals that are reflected multiple times tend to have larger ToF values (i.e., longer propagation path) as well as different AoD angles compared with the signals reflected off target people.

We experimentally evaluate our system with different numbers of people in various indoor environments including classrooms and laboratories. We conduct experiments in multi-person scenarios, dynamic interference scenarios, different distances, different human orientations, etc. The results demonstrate that our system is highly robust and accurate in monitoring vital signs. We highlight our major contributions in the following:

- We propose an identity-aware and interference-robust vital sign monitoring system for multi-person scenarios using commodity WiFi. Our system extracts vital signs in the spatial domain instead of the signal domain to achieve identity-aware monitoring.
- We develop a cPCA-CL framework to decouple signals reflected off multiple people to achieve interference-robust monitoring. Moreover, we propose to improve the SNR of heartbeat signals using multidimensional information of WiFi signals and harmonic cancellers.
- Experimental results show that our system can achieve accurate identity-aware vital signs monitoring for multiple people. The overall accuracies of breathing rate and heart rate are 99.1% and 97.9%, respectively.

2 RELATED WORK

Vital signs monitoring systems are divided into four categories according to the sensing techniques: wearable devices, dedicated RF devices, commodity WiFi, and other signal modalities.

Wearable Devices. Polysomnography is a typical contact approach to monitoring vital signs, which involves continuously monitoring various physiological variables, such as breathing, heart rate, and muscle activity, making it a viable approach to assess patients displaying moderate to high clinical suspicion of sleep-disordered breathing [9]. Moreover, Khan et al. [16] propose a method to monitor vital signs based on flexible and wearable

Table 1. Summary of commodity WiFi-based vital signs monitoring systems.

Reference	Respiration Rate	Heartbeat Rate	Multi-person	Interference-robust	Identity-aware
Wi-Sleep [23]	✓	×	×	×	×
Liu et al. [22]	✓	✓	×	×	×
PhaseBeat [40]	✓	✓	✓	×	×
Multisense [51]	✓	×	✓	×	×
BreathTrack [52]	✓	×	✓	×	×
TR-Breath [10]	✓	×	✓	×	×
SpaceBeat	✓	✓	✓	✓	✓

medical devices. These devices utilize stretchable and lightweight materials for fabricating the biosensor which is comfortable for people to wear. Wang et al. [39] introduce a system that utilizes a mattress with optical fiber in it to monitor heart rate and breath rate simultaneously.

Dedicated RF Devices. Nowadays, researchers have realized contact-free vital signs monitoring utilizing dedicated RF devices. For example, Adib et al. [5] propose Vital-Radio using dedicated Frequency Modulated Carrier Waves (FMCW) radar to monitor vital signs. The FMCW can sense the chest motion and separate reflectors into different buckets. Therefore, it can extract multi-person vital signs. However, it requires users to be static and relies on specialized hardware. Chen et al. [11] use IR-UWB radar to transmit signals and probe humans. Then they feed the IQ signal into self-designed machine-learning models to achieve fine-grained vital signs monitoring. These systems benefit from the effectiveness of machine learning models, making them robust to various body motions. Zheng and Chen et al. [54] design V^2iFi to monitor in-vehicle multiple people's vital signs by utilizing an impulse radio's temporal resolution. However, these systems all rely on dedicated RF devices and specialized signals, which are unsuitable for consumer-oriented utilization or wide deployment due to their high cost. On the contrary, our system is based on commodity WiFi devices, which are pervasive in daily life and industrial manufacturing.

Commodity WiFi. In recent years, commodity WiFi has enabled various sensing applications [1, 4, 28, 31, 36] including contact-free vital signs monitoring. For example, Wi-Sleep [23] is the first sleep monitoring system based on off-the-shelf WiFi. This system explores the amplitude of Channel State Information (CSI) in WiFi and extracts the single person's respiration to monitor sleep. Liu et al. [22] utilize off-the-shelf WiFi to monitor vital signs during sleep. This system first cleans the WiFi data, then selects subcarriers, and further detects the vital signs periods by examining the peaks in power spectral density (PSD). Wang et al. introduce PhaseBeat [40], which explores the phase difference of CSI data. This system processes signals by discrete wavelet transform (DWT) to obtain the denoised breathing signal and reconstruct the heartbeat signal. It utilizes a root-MUSIC algorithm [27] to extract the breathing rate and FFT to extract the heartbeat rate. Zeng et al. introduce MultiSense [51], which can utilize WiFi signals to reliably and continuously detect respiratory patterns from multiple individuals. They use ICA to achieve blind source separation of WiFi signals in the signal domain. Zhang et al. propose BreathTrack [52] which leverages the dominant path of the AoA-ToF spectrum of WiFi signals to detect breathing rates. TR-Breath [10] processes CSI into the time-reversal resonating strength (TRRS) and applies the root-MUSIC algorithm to extract multiple people's breathing rates. The abovementioned four systems [10, 40, 51, 52] can monitor multi-person vital signs. However, they are not identity-aware and are not robust to dynamic interferences introduced by other people. Our system has the capability to pinpoint each individual within the 2D AoA spatial information and extract their vital signs by analyzing the variations in WiFi signals originating from their respective locations. Consequently, our system can achieve identity-aware vital sign monitoring. We also decouple the signals reflected off multiple people based on the locations to mitigate interferences caused by other people. We summarize the characteristics of commodity WiFi-based vital signs monitoring systems in Table 1.

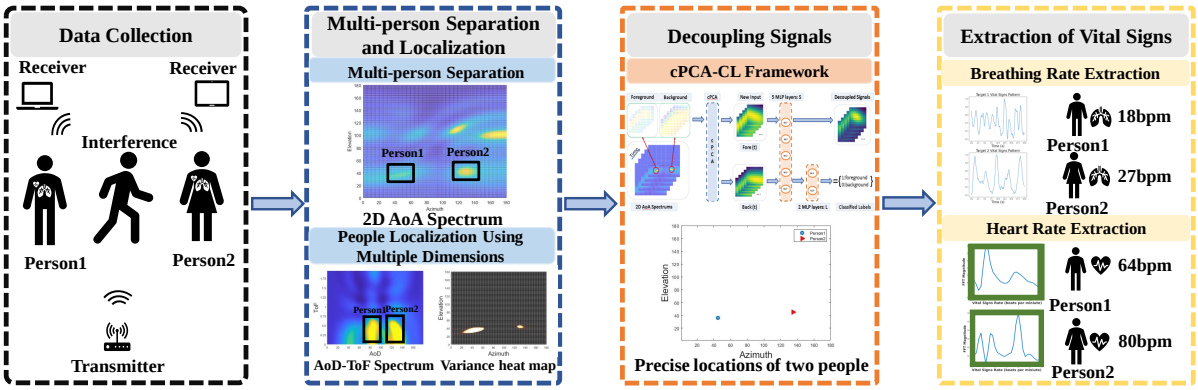


Fig. 1. SpaceBeat system overview.

Other Signal Modalities. Solutions based on other signal modalities could leverage RGB cameras, acoustic signals, etc. to realize vital signs monitoring. For instance, Prannay et al. [17] leverage video frames and image processing techniques to extract vital signs. Zainuddin et al. [49] use a Raspberry Pi camera to collect data and then apply computer vision technology to detect vital signs. Qian et al. [25] adopt acoustic signals and propose an Acousticcardiogram (ACG) using FMCW sonar to collect acoustic signals and monitor Vital signs. Nevertheless, these approaches necessitate line-of-sight conditions and dedicated hardware.

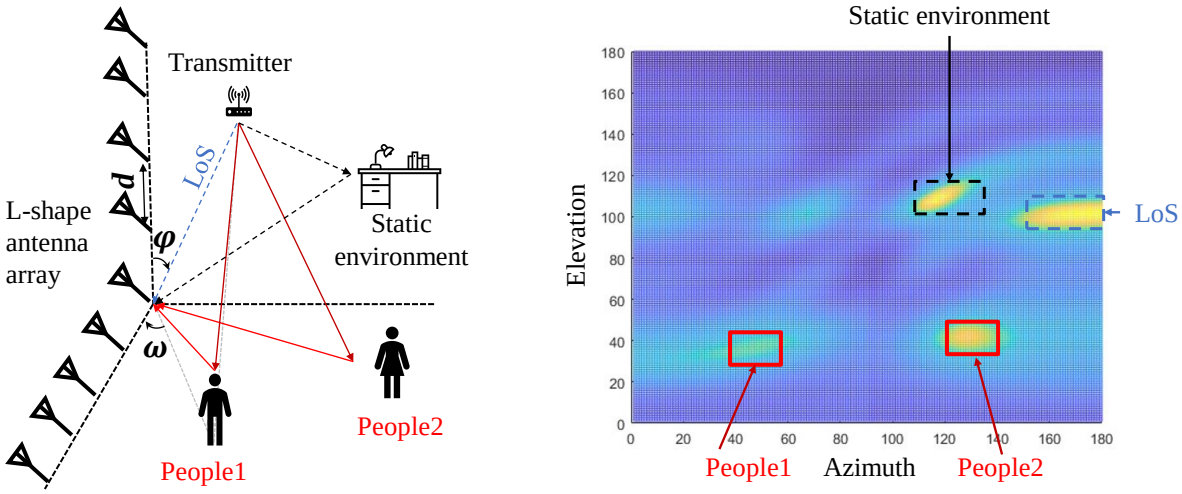
3 SYSTEM DESIGN

3.1 System Overview

The main idea of SpaceBeat is to extract the vital signs of each specific person in the spatial domain. In particular, we utilize multiple antennas on WiFi devices to derive 2D AoA of WiFi signal reflections to separate, locate, and identify the vital signs of each person. As shown in Figure 1, our system can reuse a single existing WiFi transmitter that continuously transmits packets to probe multiple people and acquires time-series channel state information (CSI) measurements from one or more commodity WiFi receivers. The reflected WiFi signals will be modulated by movements of the chest and abdomen caused by respiration, cardiac activities, and other body movements. Moreover, the WiFi signals reflected by different people and objects will arrive at the receivers from different directions. We note that our system aims to achieve identity-aware monitoring, requiring the targeted individual for vital signs extraction to remain stationary while allowing other people to move around.

After the CSI preprocessing, we remove the random phase offsets of signals and the signals reflected off the static background [32, 43]. Then, we can derive the 2D AoA spectrum using multiple antennas to separate and locate multiple people. In order to accurately locate people, we leverage multidimensional information including 2D AoA, angle of departure (AoD), and time of flight (ToF) to improve the resolvability of WiFi signals. We further explore the variance of signals caused by the breathing or body movements of people to refine the location of each person in the 2D AoA spectrums [42, 44]. We select the signals around the location of the person for vital sign extraction. Therefore, our system is aware of whom each vital sign belongs to, which cannot be achieved in previous work that separates people in the signal domain.

Although we can determine the precise location of each person, the signal reflections of different people will affect each other in 2D AoA spectrums and lead to coupling. Hence, we design the cPCA-CL framework to decouple signals reflected off multiple people. Contrastive Principal Component Analysis (cPCA) can identify patterns that are enriched in a target relative to contrastive data that can be used for background elimination [3].



(a) The 2D AoA (azimuth ω and elevation φ) can be estimated by using the L-shaped antenna array.

(b) 2D AoA Spectrum.

Fig. 2. 2D AoA estimation.

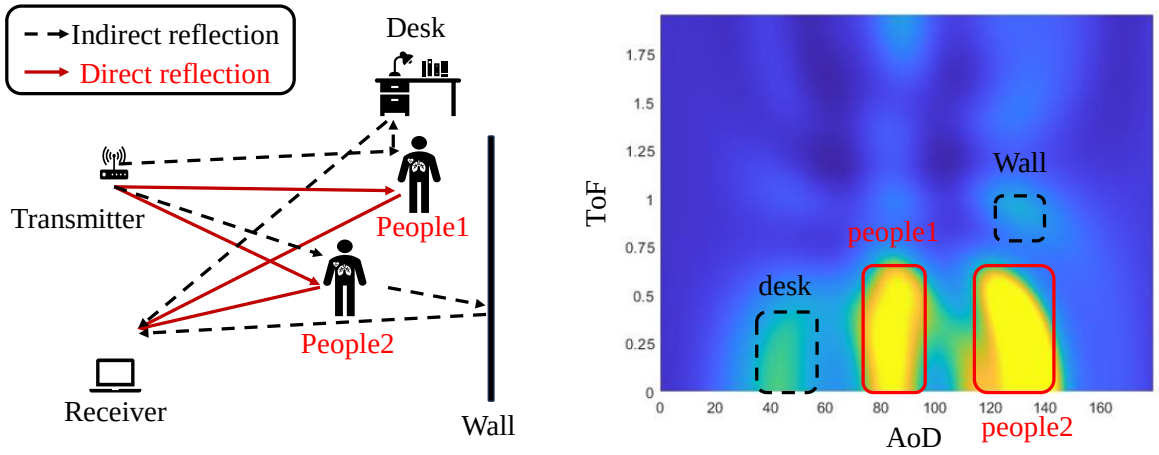
In our work, we employ the cPCA to remove the background from a single foreground 2D AoA spectrum. The signals pertaining to the target individual are identified as the foreground, whereas the signals corresponding to other individuals are regarded as the background. CL represents Contrastive Learning which is employed to eliminate backgrounds from a sequence of foregrounds in the temporal domain. By incorporating both cPCA and CL, the cPCA-CL framework effectively isolates the vital sign signals of the target person from the surrounding background, enabling precise extraction of the target person's vital signs.

Finally, we extract both breathing rate and heart rate from the decoupled signals. We employ the Fast Fourier Transform (FFT) on the signals to extract the breathing rate. However, the breathing signal and even its harmonics are much stronger than the heartbeat, especially if the breathing harmonics frequency is close to that of the heartbeat, which makes it difficult to detect and extract heart rate. We address the issue by developing a cascade of single-delay Moving Target Indicator (MTI) harmonic cancelers. After removing the breathing harmonics, the heartbeat rate can be accurately extracted using the FFT.

3.2 Multi-person Separation and Localization

3.2.1 Multi-person Separation. Our insight is to separate people and extract vital sign signals in the spatial domain. In particular, we derive 2D AoA of signal reflections to represent the spatial information by leveraging a two-dimensional antenna array (i.e., L-shaped antenna array) of the WiFi receiver. Compared to mD-Track [46], our work leverages an L-shape antenna to extract 2D AoA, which provides intuitive 2D spatial information of the physical space that directly illustrates the vital signs of multiple people in the physical environment. By analyzing the phase shift of the received signals at multiple antennas, we are able to calculate the 2D AoA (i.e., azimuth and elevation) of the signals as shown in Figure 2(a). We denote the phase shift of the received signal on the n^{th} antenna as:

$$\Psi_n(\varphi, \omega) = e^{-j2\pi[sin(\varphi)cos(\omega)+cos(\varphi)]/c}, \tag{1}$$



(a) Direct and indirect reflections.

(b) Direct and indirect reflections on AoD-ToF spectrum.

Fig. 3. Illustration of separating direct and indirect reflections on AoD-ToF spectrum.

where φ is the elevation angle, ω is the azimuth angle. Next, we can obtain the phase shifts across the N antennas in the antenna array as follow:

$$\mathbf{s}(\varphi, \omega) = [\Psi_1(\varphi, \omega), \Psi_2(\varphi, \omega), ..., \Psi_N(\varphi, \omega)]^T, \tag{2}$$

where $\mathbf{s}$ is called the steering vector. With the steering vector, the 2D AoA of the signal reflections can be estimated using the MUSIC algorithm [33]. The estimated 2D AoA spectrum is shown in Figure 2(b). We can observe that the locations of the two people are separated by azimuth and elevation.

3.2.2 People Localization Using Multiple Dimensions. Besides relying on 2D AoA, we propose to leverage multidimensional information to improve the resolvability of WiFi signals [26, 34, 46] and thus locate people more accurately. Hence, we estimate ToF and AoD information of WiFi signals in addition to azimuth and elevation.

Specifically, we integrate the phase shifts caused by the spatial diversity exhibited by the transmitting antennas. The phase shift $\Phi(\theta)$ across transmitting antennas as follows:

$$\Phi(\theta) = e^{-j2\pi f d \sin(\theta)/c}, \tag{3}$$

where θ is AoD, d is the distance between two adjacent transmitting antennas, and f is the frequency of the signal.

Subsequently, to estimate the ToF of the n^{th} propagation path, τ_n, we incorporate the phase shifts associated with the frequency diversity of the OFDM subcarriers. In the case of evenly distributed subcarriers, the phase shift between two adjacent subcarriers can be expressed as follows:

$$\Theta(\tau_n) = e^{-j2\pi f_\delta \tau_n/c}. \tag{4}$$

Next, we build the new steering vector $\mathbf{s}(\omega, \varphi, \theta, \tau)$ which can be denoted as:

$$\begin{aligned}\bar{s}(\omega, \varphi, \tau) =&[1, ..., \Theta_\tau^{M-1}, \Psi_{(\omega,\varphi)}, ..., \Theta_\tau^{M-1}\Psi_{(\omega,\varphi)}, ..., \Psi_{(\omega,\varphi)}^{N-1},\\ &..., \Theta_\tau^{M-1}\Psi_{(\omega,\varphi)}^{N-1}]^T,\end{aligned} \tag{5}$$

$$\mathbf{s}(\omega, \varphi, \theta, \tau) = [\bar{s}(\omega, \varphi, \tau), \Phi_\omega \bar{s}(\omega, \varphi, \tau), ..., \Phi_\omega^{K-1}\bar{s}(\omega, \varphi, \tau)]^T, \tag{6}$$

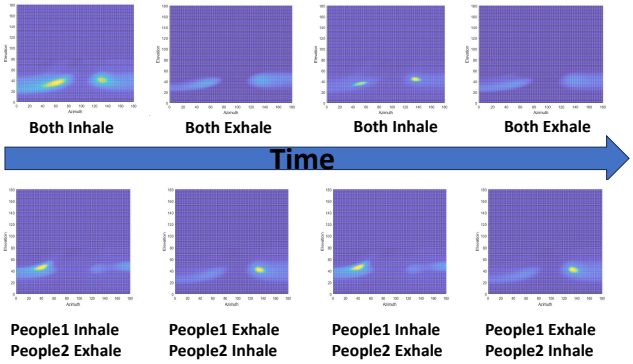


Fig. 4. 2D AoA spectrums of synchronous respiration of two individuals (first row) and asynchronous respiration of two individuals (second row). It is observable that two individuals are breathing in consecutive spectrums.

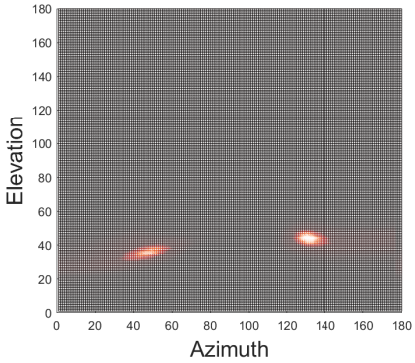


Fig. 5. Variance heat map.

where Θ_τ, Φ_θ and $\Psi_{(\omega,\varphi)}$ are the abbreviations of $\Theta(\tau)$, $\Phi(\theta)$ and $\Psi(\omega,\varphi)$.

Our system jointly utilizes transmitting antennas, receiving antennas, and subcarriers in the MUSIC algorithm, so that the number of signal sources is not limited to the number of receiving antennas [19]. In our work, we empirically determine the number of signal sources in the MUSIC algorithm to be 150. The estimation is presented as follows by maximizing the spatial spectrum function:

$$P(\omega,\varphi,\theta,\tau) = \frac{1}{A^H(\omega,\varphi,\theta,\tau)E_N E_N^H A(\omega,\varphi,\theta,\tau)}. \tag{7}$$

Furthermore, to better utilize multidimensional information, we investigate the spectrum of ToF and AoD information to distinguish between direct reflections and indirect reflections since the ToF and AoD of indirect reflections typically differ from those of direct reflections. We refer to direct reflection as the signals reflected from the target person and directly received by the receiver. While the indirect reflection represents the signals reflected from the target person, then further bounced off the static environment, and finally received by the receiver. We generate an AoD-ToF spectrum based on the derived four-dimensional information in Figure 3. The figure illustrates two people breathing in front of the WiFi devices resulting in the occurrence of direct and indirect reflections. Direct signals are stronger than indirect signals. Therefore, we employ a boundary tracing function extracting direct signals to filter the indirect signals. After filtering, we can obtain more precise AoD and ToF information. Finally, we combine multiple consecutive WiFi packets to enhance the quality of 2D AoA estimation and accumulate the AoA values in the filtered ToF and AoD dimensions to generate the 2D AoA spectrums.

Considering that the single 2D AoA spectrum only captures spatial information [29, 42], we further incorporate temporal information to enhance the location accuracy. As shown in Figure 4, during inhalation, the bright spots in the image are visible. While during the exhalation period, the brightness of the spots in the image is faint. The rationale for this observation is that the strength of the signals fluctuates with individuals' breathing cycles, as the movement of their chests (i.e., contraction and expansion) alters the length of the reflection path, consequently affecting the signal strength. To leverage temporal information, we build a variance heat map on a series of 2D AoA spectrums to detect breathing people, as shown in Figure 5 for detecting static people's location.

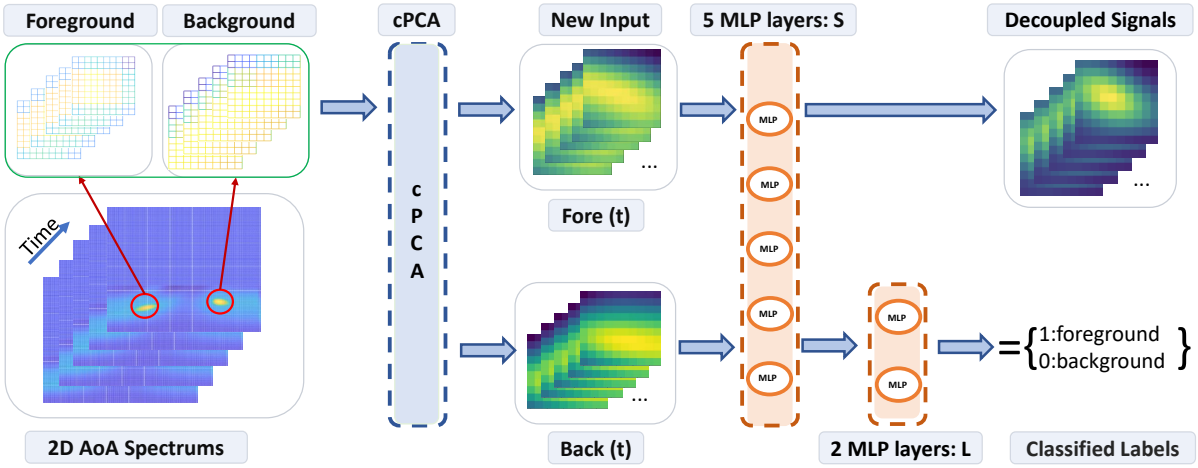


Fig. 6. cPCA-CL framework.

We can obtain the variance map as follow:

$$h(\varphi, \omega) = Var(P(\varphi, \omega, t)), \tag{8}$$

where $h(\varphi, \omega)$ is the variance map of elevation and azimuth spectrums, $P(\varphi, \omega, t)$ is the 2D AoA spectrum, and t is the number of spectrums related to the time. We locate the vital signs of target people by the local highest variance on the 2D AoA spectrum frames instead of the local highest peaks of a single spectrum [19] as the displacement caused by vital signs will lead to a significantly higher variance compared to static objects. It is worth noting that we incorporate the variance map into an iteration in the next section to iteratively refine the locations of people and thus obtain final localization results.

3.3 Decoupling Signals Reflected off Multiple People

Although we can separate and locate each person, the signals reflected off different people still interact with each other in the 2D AoA spectrum, resulting in the signal coupling that strongly distorts the accuracy of vital sign monitoring. The reason is that both the vital signs and physical movements of different people can modulate WiFi signals. Consequently, these signals become intertwined and overlap in the 2D AoA spectrum.

3.3.1 cPCA-CL Framework. To decouple signals, we propose a cPCA-CL framework, as shown in Figure 6, that combines contrastive principal component analysis (cPCA) and contrastive learning (CL) to decouple signals reflected off multiple people and thus enhance the accuracy of vital sign monitoring. The input for cPCA-CL is selected from 2D AoA spectrums where the signals round the highest variance.

The cPCA [2] is a generalization of standard PCA. It employs semi-supervised learning to derive a ratio α from the foreground and background datasets. This ratio can assist in cleaning the foreground signal by subtracting the product of the ratio and the background signal as follows:

$$fore_{new} = fore - \alpha \cdot back, \tag{9}$$

where the $fore$ is the foreground 2D AoA spectrum that is the signals of our target that require decoupling, and $back$ is the background 2D AoA spectrum that is the signals of other individuals or interference as shown in the left part of Figure 6. $fore_{new}$ is the output after applying cPCA which is the unit of $fore(t)$. We also can obtain the new background by switching the foreground and background. By applying cPCA, we can effectively

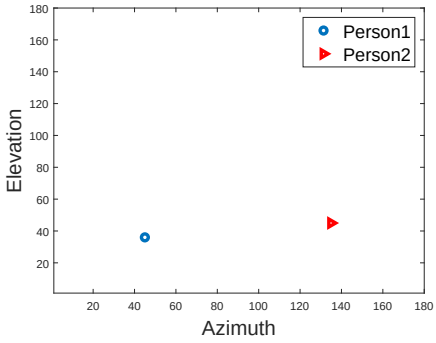


Fig. 7. Final precise locations of two people.

enhance vital signs signals in the spatial domain since the background removal process is applied to each 2D AoA spectrum.

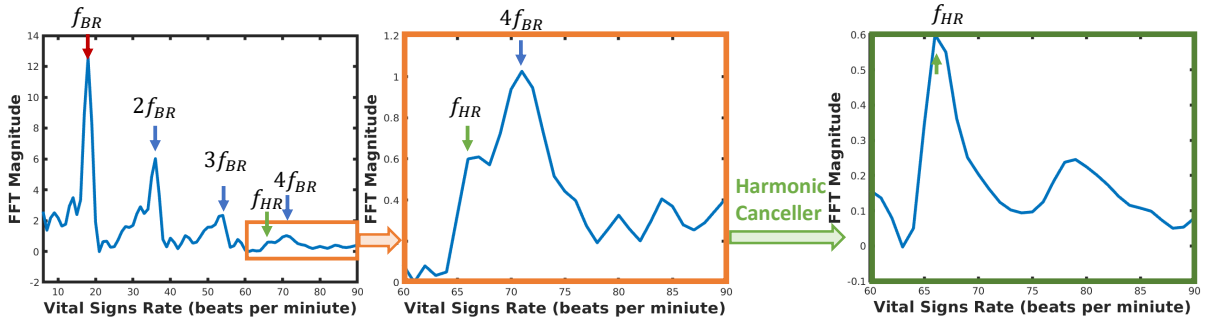
However, utilizing cPCA alone does not take into account sequential information. Therefore, we introduce a contrastive learning process [11] as a subsequent step after applying the cPCA method. Contrastive learning involves constructing contrastive instances. Our contrastive dataset is highly reasonable which is the background dataset. We do not need to generate the contrastive dataset by random signals or by any data augmentation method. The foreground and background datasets derived from the cPCA process are well-suited for instances and contrastive instances. We build a sliding window on the foreground and background datasets to construct these instances $fore(t)$ and contrastive instances $back(t)$ for model input as shown in Figure refCL new input part. The new input instances are 2D AoA frames based on 2D AoA spectrums. We label these instances with labels: foreground and background.

Next, we train a multilayer perceptron (MLP) model $S \times L$ to classify whether the dataset is foreground or background in the right part of Figure6. S is a generator including an encoder-decoder structure [38]. And S is employed for the purpose of extracting features. L contains the other MLP layers performing a binary classification function by minimizing a cross-entropy loss. The rationale is that a successfully trained S has to extract the domain component in $fore(t)$ and $back(t)$. Therefore, we can use S to clean the $fore(t)$ dataset and obtain the decoupled signals. By incorporating both spatial and sequential information, we are able to effectively remove noise and interference from the data, enhancing the overall noise reduction performance of the system.

3.3.2 Iterative Refinement. We further implement iterative steps to refine the location and our decoupling results. The input for each iterative step is the 2D AoA spectrums. Within each step, we perform localization using the variance heat map, as described in Section 3.2. This localization process allows us to obtain the target location based on the variance map. Subsequently, we generate foreground and background datasets incorporating the obtained location information. These datasets are then subjected to the cPCA-CL framework in Section 3.3 for decoupling signals. And we update the 2D AoA frames by incorporating the decoupled signals. We repeat this step iteratively until the final location of the target remains unchanged or the TSER [50] no longer increases. In our paper, the Target-reflected Signal Energy Ratio (TSER) is defined as the ratio of vital signs energy to the overall energy in the frequency spectrum. This ratio can be utilized to determine the quality of the decoupled signals. The resulting decoupled signal after iterative represents our desired output. Finally, following the iterative steps, we are able to precisely determine the locations of the two individuals, as demonstrated in Figure 7. We summarize the iterative algorithm below.

Algorithm 1 Spacebeat: Iterative Steps

Input: 2D AoA frames
Output: Refined decoupled signals
  while TSER increasing OR Location changes **do**
    Location based on variance map
    Decoupling signals based on cPCA-CL
    Update the 2D AoA frames by incorporating the decoupled signals
  end while



(a) Heart rate and harmonic signal.

(b) Heart rate is overwhelmed by 4th harmonic interference.

(c) Heart rate is detectable after harmonic removal.

Fig. 8. Illustration of harmonic removal.

3.4 Extraction of Vital Signs

3.4.1 Breathing Rate Extraction. Once we have obtained the precise locations of the individuals and the decoupled signals, we can extract the breathing rate through the decoupled signals. Given that breathing is a periodic motion, we can utilize a fast Fourier transform (FFT) to find the period of breathing. Considering that the decoupled signals consist of three dimensions, namely azimuth, elevation, and time, we employ the FFT to convert the signals from the time domain to the frequency domain. Subsequently, we utilize Principal Component Analysis (PCA) to aggregate the FFT results. The peak observed in the FFT output corresponds to the dominant frequency, which represents the breathing rate.

Increasing the window size of FFT will result in higher resolution of FFT results. However, when it comes to detecting changes in respiratory patterns caused by breath and heartbeat variations, a larger window size becomes less effective and can lead to inaccurate results since the breathing rate per minute can vary. Empirically, we have set the window size of FFT to 60 seconds. By analyzing the FFT spectrum, we can accurately determine breath rates.

3.4.2 Heart Rate Extraction. Although heartbeat is also a periodic motion, we cannot simply utilize FFT to extract the heart rate. This is because interferences in the environment will disturb the subtle heartbeat signal. Therefore, we only focus on the location of the target person derived in Section 3.2 to eliminate interferences at other locations.

In addition, as shown in Figure 8(a), the harmonics of breathing [21] (i.e., f_{BR}, $2f_{BR}$, $3f_{BR}$, and $4f_{BR}$.) could be close to the frequency of heartbeat (i.e., f_{HR}) and thus significantly reduce the SNR of the heartbeat signal as illustrated in Figure 8(b). To remove the harmonics, we develop a harmonic canceller that could effectively

eliminate the harmonics caused by the breathing signal. Thus, the heartbeat signals can be emphasized and thus improve its SNR as shown in Figure 8(c). Specifically, we develop a cascade of single-delay Moving Target Indicator (MTI) cancellers [21] as follows:

$$Signal_{HR}(t) = Signal_C(t) - Signal_C(t - BT), \tag{10}$$

where $BT = 1/f_{BR}$ is the breathing period, $Signal_{HR}(t)$ is the heartbeat signal, $Signal_C$ is the combination of heartbeat signals and harmonics of breathing singles. To remove all the harmonic periods, we use multiple single-delay filters, which cancel the frequencies that are multiples of K/BT. We can estimate K as follow:

$$K \approx \lceil \frac{110}{BR} \rceil, \tag{11}$$

where BR is breathing rate which is equal to $60/BT$ and 110 is the maximum heart rate [8]. Consequently, the filters effectively cancel out the output at each frequency, eliminating the influence of harmonics so that we can extract the frequency of heartbeat with FFT.

3.5 Data Preprocessing

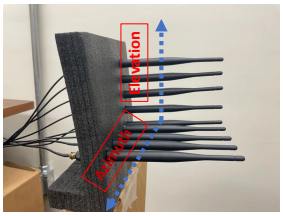
Before we feed the CSI measurements to estimate the 2D AoA, it is necessary to perform CSI de-noising in order to remove phase noises since CSI measurement suffers from a random phase shift since the receiver suffers from a random phase shift caused by the sampling time offset (STO) and packet detection delay (PDD) across packets. To sanitize the random phase offsets, we adopt a linear fit method [19] to unwrapped phase measurements.

It is worth noting that once we have obtained the 2D AoA, a static object removal technique is employed as a preprocessing step before proceeding with people localization. Because the obtained multidimensional information contains comprehensive information about human bodies, movements, and static objects, such as walls and furniture, in the sensing environment. To remove the impact of stationary objects, we apply a static object removal method on the spatial spectrum that leverages the previous consecutive frames and the background information to eliminate reflection from static objects and Line-of-Sight (LoS) signals.

4 PERFORMANCE EVALUATION

4.1 Experimental Setups

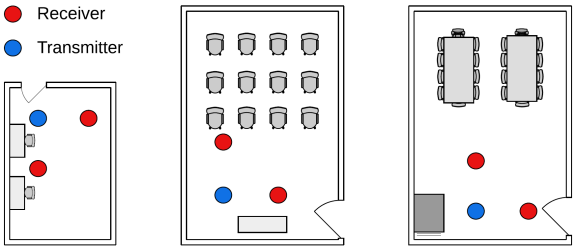
4.1.1 Devices. In our experiments, SpaceBeat is implemented with commodity WiFi devices. The transmitter and receivers are both commodity WiFi devices (i.e., Dell LATITUDE laptops) and both use low-cost Intel 5300 Network Interface Cards (NICs). The transmitter contains three linearly-spaced antennas. We utilize two receivers and each is equipped with nine antennas which form an L-shaped distributed antenna layout (i.e., L-shaped antenna array) as shown in Figure 9(a). This L-shaped antenna array is composed of two uniform linear subarrays in orthogonal directions. Each subarray consists of two NICs interconnected by a signal splitter with a shared antenna. Such a setup could emulate potential antenna configurations of next-generation WiFi devices (e.g., WiFi 7 supports up to 16 antennas). The receivers are operated in monitor mode to capture packets from the transmitter. The antennas on both the receiving and transmitting ends are uniformly arranged, maintaining an equal spacing of half a wavelength (2.8 cm). The WiFi channel frequency is set to 5.24 GHz, with a 40 MHz bandwidth, and the transmitter sends 1000 packets per second. The Linux 802.11 CSI tool [15] is employed for the acquisition of CSI measurements from 30 OFDM subcarriers. Synchronization among all devices is facilitated through the utilization of the network time protocol (NTP). To collect ground-truth respiration data and heartbeat data from multiple individuals, we asked each participant to wear a commercial Neulog respiration belt and a Neulog Heart Rate & Pulse logger sensor on their finger in Figure 9(b) to record the ground truth of vital signs.



(a) The L-shaped antenna array.

(b) Neulog Respiration sensor and Heart Rate sensor.

Fig. 9. Experiment devices.



(a) Lab (b) Classroom1 (c) Classroom2

Fig. 10. Experiment environments and setups.

4.1.2 Data Collection. To evaluate SpaceBeat, we recruit 14 volunteers (8 males and 6 females) aged between 15 and 64 years, with a weight ranging from 45 to 90 kg. All of them are monitored in their natural state. The volunteers' respiration rate ranges from 10 to 30 bpm, and their heartbeat rate varies between 60 to 120 bpm. In this paper, we consider two scenarios. The first scenario is that we ask all people to breathe without significant movements such as walking. Another scenario is that we ask a part of people to breathe without significant movements while other people perform daily activities (i.e., dynamic interferences), including waving hands, jumping, walking, etc. The data collection was approved by the IRB of the authors' institution. We conducted experiments in three different environments: a laboratory and two classrooms as shown in Figure 10. Specifically, environment (a) is a standard laboratory setup with 2 desks, 2 chairs, and computers on the desks. The size of the laboratory is 4m × 6m. Environment (b) is a classroom for lectures with 12 pairs of desks and chairs and a long lectern. The classroom size is 6m × 9.5m. Environment (c) is another classroom with two long tables, 20 chairs, and a projector. The size of this classroom is also 6m × 9.5m. In all three rooms, the desks and tables are constructed from wood, while the chairs are made of a combination of metal and plastic materials. We aim to evaluate our system across diverse real-world environments and showcase that our system can adapt to different environments. Each space varies in layout, size, and furniture such as computers, tables, and chairs. These environmental characteristics can significantly influence the propagation of WiFi signals. In total, we collect around one hundred million WiFi CSI packets to train and test our system for vital signs monitoring over four months.

4.1.3 Baseline. To evaluate our system, we perform a comparative analysis by assessing the performance of our system in comparison to three state-of-the-art WiFi-based vital signs monitoring systems including an FFT-based system [40], an ICA-based system [51], and a 1D AoA-based system [52]. Specifically, the FFT-based system [40] relies on raw CSI measurements which analyze the phase variation on raw CSI measurements. This system utilizes DWT to filter out high-frequency signals, root-MUSIC algorithm to detect multi-person breathing, and FFT to extract heartbeat rate. The ICA-based system [51] utilizes ICA to blindly separate multi-person signals. This system utilizes ICA to separate mixed signals and an optimal algorithm to combine the blindly separated signals for each person. The 1D AoA-based system [52] explores the dominant path utilizing 1D AoA and applies an FFT to the dominant path to estimate breathing rate. All of these systems are evaluated using the same data collected from the same subjects to ensure consistency and fairness in the comparison.

4.1.4 Model Settings and Performance. We transfer the 2D AoA spectrums to the 4D tensors with the size of $600 \times 2 \times 10 \times 10 \times 1$ as the input of our contrastive learning model. The first dimension, 600, is the number of frames. The frames are created by a sliding window based on 2D AoA spectrums, and the second dimension is the number of receivers. The 10×10 is the selected azimuth and elevation degree ranges. Considering the 2D AoA

Table 2. Overall system performance comparison.

	Breathing Rate(%)	Heart Rate(%)
FFT-based (not identity-aware) [40]	91.92	85.22
ICA-based (not identity-aware) [51]	96.63	91.33
1D AoA-based (not identity-aware) [52]	97.36	93.23
SpaceBeat (identity-aware)	**99.11**	**97.98**

spectrum frames, the 4D tensor contains sufficient spatial and temporal information, allowing the deep learning network to decouple the signals effectively. S utilizes an MLP model with seven layers which is an encode-decoder structure. The size of the median layer is the same as the number of input frames. The size of output is the same as our input data size. Therefore, we can readily reconstruct 2D AoA frames. The activation function used after each layer is leaky ReLU, which introduces nonlinearity. L consisting of two other layers serves as the discriminator. During the training phase, we partition the data from 14 subjects into two non-overlay datasets: a training set containing 10 people of the data and a testing set containing the remaining 4 people. This ensures that the testing data includes unseen environments and subjects. Additionally, we perform 5-fold cross-validation to evaluate the robustness of our model. We set a batch size of 16, and the Root Mean Square Propagation optimizer is employed with a learning rate of 0.01. We leverage the standard binary classification for discriminator schemes and the Binary Cross-Entropy (BCE) to backpropagate. Our binary classifier L can achieve an average accuracy of 97% on the test dataset which means the latent space can distinguish the background and foreground effectively.

4.1.5 Evaluation Metrics. We utilize bpm as one of the evaluation metrics for breathing rate and heart rate. Breaths Per Minute (bpm) is a metric that quantifies the number of complete respiratory cycles a person takes within one minute. Beats Per Minute (bpm) is a measure of the heart rate, representing the number of times the heart beats within one minute.

We also utilize the accuracy to evaluate the monitoring of breathing and heartbeat, which is defended as follows:

$$Accuracy = 1 - \frac{|Estimated\,result - Ground\,truth|}{Ground\,truth}, \tag{12}$$

where the estimated result is the breathing or heartbeat rate estimated by our system and the ground truth is the rate recorded by the Neulog respiration belt.

The localization error is calculated by the angle difference between the ground truth and estimated angle. The unit for AoA error is in degree (°) based on the angle difference.

In addition, we use cosine similarity to evaluate the coefficient between reconstructed waveform br and ground truth gt. The cosine similarity is calculated as the cosine of the angle between two vectors (i.e., the reconstructed waveform br and ground truth gt) and then determines to what extent the two vectors point in the same direction in a high-dimensional space. The cosine similarity is defined as follows:

$$CS(br, gt) = \frac{br \cdot gt}{|br||gt|}, \tag{13}$$

whose value lies in the range of [0, 1].

4.2 Overall Performance

Firstly, we study the overall performance of our system by comparing it with three baselines including an FFT-based system [40], an ICA-based system [51], and a 1D AoA-based system [52]. This evaluation includes multi-person and dynamic interference scenarios in which multiple people statically breathe, and at least one

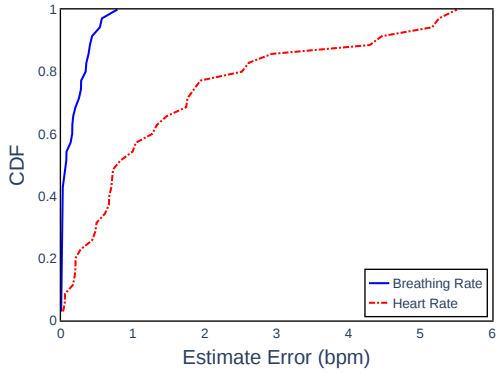


Fig. 11. Overall performance of SpaceBeat system.

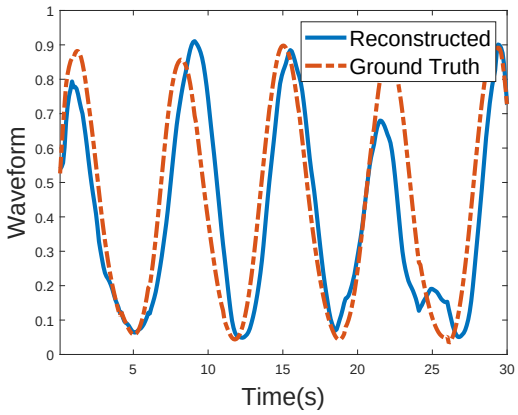


Fig. 12. The reconstructed waveform and ground truth.

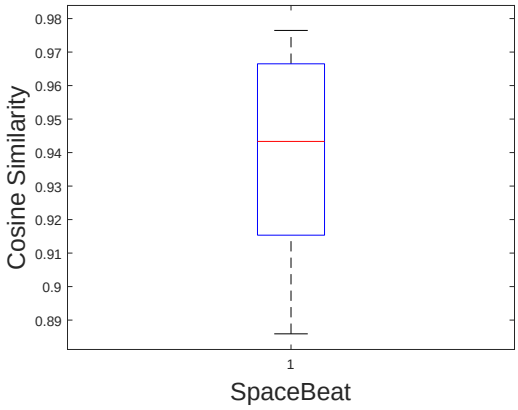


Fig. 13. Cosine similarity between reconstructed waveform and ground truth.

performs random activities. The results of our system and all the baselines are shown in Table 2. It is worth noting that both FFT-based and ICA-based approaches cannot achieve identity-aware monitoring as they separate multiple people in the signal domain and cannot know to whom each vital sign signal belongs [51]. The 1D AoA-based approach [52] investigates the dominant path effect on the 1D AoA spectrum that is vulnerable to interference in scenarios involving multiple people. This approach does not differentiate between individuals and thus it is not an identity-aware monitoring system. Our system is identity-aware as it extracts vital signs from multiple people in the spatial domain and matches vital signs signals to each person based on their location.

Compared with those baselines [40, 51, 52], we can also observe that the SpaceBeat achieves much better performance for both breathing and heartbeat monitoring. Specifically, as shown in Table 2, the overall accuracy of our system is 99.11% and 97.98% for the breathing rate and heart rate, respectively. Moreover, as shown in Figure 11, the median errors for breathing rate and heart rate are 0.07 bpm and 0.82 bpm, respectively. Among all the baselines, the FFT-based approach has the worst performance since it cannot extract the vital sign rates from

Table 3. System performance for movement interference scenarios.

	Breathing Rate(%)	Heart Rate(%)
Walk (FFT-Based)	70.91	58.84
Jump (FFT-Based)	73.81	61.22
Wavehands (FFT-Based)	91.45	85.63
Walk (ICA-Based)	82.37	76.26
Jump (ICA-Based)	85.82	80.49
Wavehands (ICA-Based)	96.25	92.42
Walk (SpaceBeat)	97.42	95.23
Jump (SpaceBeat)	97.82	95.42
Wavehands (SpaceBeat)	98.74	97.66

a variety of frequencies caused by dynamic interferences. Note that the overall evaluation includes a part of data that involves dynamic interferences such as movements caused by nearby people. Besides, it is hard to resolve similar vital sign rates. The ICA-based approach is better than FFT as it can differentiate the vital signs of different people, which are linearly combined. However, it still suffers from non-linear dynamic interferences [11, 55]. The 1D AoA-based approach performs better than the previous two approaches. This is because 1D AoA can provide spatial information in terms of the azimuth angle. The proposed SpaceBeat system achieves the best performance as it can separate multiple people in the 2D AoA spatial domain and decouple signals reflected off different people. The results demonstrate that our system can precisely separate multiple people and identify the vital signs of each person in the spatial domain. We further decouple the signals and thus ensure accurate results. In addition, Figure 12 shows the reconstructed respiration waveforms. We reconstruct the waveform based on the value changes in the 2D AoA spectrums where the people are located and normalize the waveform based on previous work. Figure 13 shows the cosine similarity between the reconstructed waveforms and the ground truth. The performance of our system is comparable to MoRe-Fi [55]. Specifically, the mean cosine similarity between the ground truth and the reconstructed waveform is 94.3%, which indicates our system achieves good performance.

4.3 Impact of Interference

In this section, we investigate the system performance under interference where interferences are consistently present. And we compared the performance of our system with an ICA-based system and an FFT-based system. We conducted experiments for all systems under the same experiment setup where the target people remained stationary and the others performed daily activities to generate interference including walking, jumping, and waving hands. In Table 3, our system demonstrates superior performance compared to the other two systems. Specifically, our system achieves the highest accuracies of 98.74% for breathing monitoring and 97.66% for heartbeat monitoring. This is because our system leverages the spatial information extracted from 2D AoA spectrums to separate vital sign signals from interferences. Additionally, we have implemented a signal decoupling technique to further mitigate interference. Thus, our system is more robust against interferences compared to ICA-based and FFT-based systems. The FFT-based system achieves breathing monitoring accuracies of 70.91%, 73.81%, and 91.45% under walking, jumping, and waving hands scenarios, respectively. The FFT-based system achieves heartbeat monitoring accuracies of 58.84%, 61.22%, and 85.63% under walking, jumping, and waving hands scenarios, respectively. The performance of the FFT-based system is the worst in interference scenarios since the frequencies of interference can overwhelm the frequencies of vital signs signals. The ICA-based system

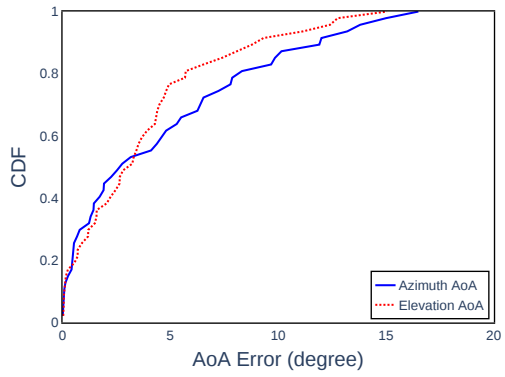


Fig. 14. Multi-person localization error of SpaceBeat.

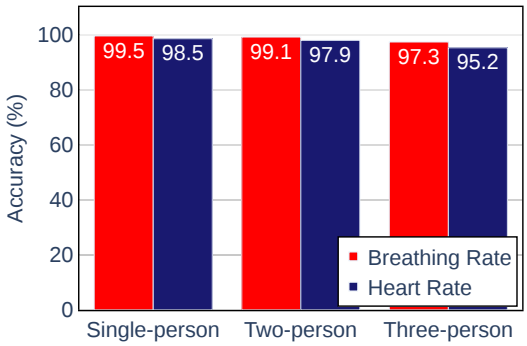


Fig. 15. The mean vital signs accuracy versus the number of persons.

achieves breathing monitoring accuracies of 82.37%, 85.82%, and 96.25% under walking, jumping, and waving hands scenarios, respectively. Besides, it achieves heartbeat monitoring accuracies of 76.26%, 80.49%, and 92.42% under the three interference scenarios, respectively. The performance of the ICA-based system is better than the FFT-based system as it may separate vital sign signals and interferences which are linearly combined. However, it is still sensitive to non-linear interferences.

4.4 Performance of Multi-person Localization

In this section, we evaluate the accuracy of localizing multiple people in the azimuth-elevation coordinate system [24]. This evaluation contains scenarios involving multiple individuals, with a maximum of three people. Some people are in a stationary breathing state and some people engage in random activities. The distance between the transmitter and receiver is set to 2m. The angle difference between two people ranges from 30° to 180° on azimuth and 0° to 60° on elevation. The results of the AoA error are shown in Figure 14. We can observe that our system has a median error of 2.6° in azimuth and 3° in elevation. Moreover, 80% errors are less than 8° in azimuth and 6° in elevation. The results demonstrate that our system accurately extracts azimuth and elevation angles, and efficiently refines the location of each individual through the iterative decoupling process. Consequently, our system achieves precise localization of multiple individuals.

4.5 Impact of Different Numbers of People

In this section, we study the impact of different numbers of people on the performance of our system. We conduct experiments for single-person, two-person, and three-person scenarios. In the single-person scenario, the person is static and breathing normally. And we do not need to decouple signals with the cPCA-CL network. Two-person and three-person scenarios include the case where all people are static and also the case where one person is performing activities. The results in Figure 15 show that even if there are three people in the indoor environment, the accuracy of the breathing rate and heart rate of our system is still relatively high, which are 97.3% and 95.2%, respectively. The accuracy of breathing rate and heart rate increase to 99.5% and 98.5% for single person scenario. This is because signal reflection becomes less complex when there are fewer people. The experimental results demonstrate that our system can effectively separate and decouple multiple people using 2D AoA-based spatial information and then accurately extract vital signs.

Table 4. System performance for different sensing distances.

	Breathing Rate(%)	Heart Rate(%)
50 cm	99.61	99.20
100 cm	99.58	98.21
150 cm	99.14	97.95
200 cm	98.89	97.64

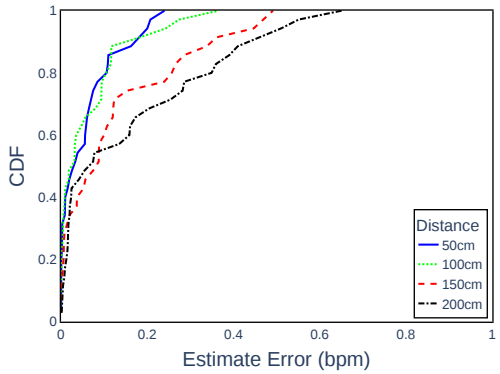


Fig. 16. CDFs of estimation error in breathing rate estimation at different sensing distances.

4.6 Impact of Sensing Distances

As the distance between people and WiFi devices could vary in practical scenes, we evaluate the impact of sensing distance on our proposed system. Specifically, the sensing distance refers to the distance between the people and the WiFi devices. We conduct experiments to change the sensing distance from 50cm to 200cm with a step of 50cm. In the experiment, two people are asked to be static and breathe normally. We show both the accuracy of breathing and heartbeat at different sensing distances in Table 4. The table illustrates that the accuracy of breathing and heartbeat monitoring are 99.6% and 99.2% at a sensing distance of 50cm, respectively. Our system still achieves accuracy around 98.9% and 97.6% for breathing and heartbeat monitoring when the distance increases to 200cm. The reason for the slight degradation is that as the WiFi signals suffer attenuation in long propagation, a longer sensing distance leads to a weaker received signal strength. However, as shown in Figure 16, we can observe that our system has a median error of breathing monitoring of only 0.07 bpm, and 80% errors are less than 0.34BPM even at a sensing distance of 200cm. The results also show that our system can work well in a typical room environment, especially in a smart home environment that contains many IoT and smart devices with WiFi interfaces.

4.7 Impact of Different Orientations

As different orientations of the person will lead to the different amplitude of displacement caused by breath, we also conduct experiments to evaluate our system about the impact of different orientations. In our experiment, two static people are asked to sit facing the receiver, back to the receiver, toward the left and right side of the receiver, separately. As shown in Table 5, the accuracy of breath rate and heart rate is 99.10% and 97.90% for the

Table 5. System performance for two people four body orientation.

	Breathing Rate(%)	Heart Rate(%)
Front	99.10	97.90
Left	98.93	97.28
Right	98.88	97.31
Back	98.65	97.27

Table 6. System performance for two people NLoS scenario.

	FFT Breathing Rate(%)	Heart Rate(%)
Line-of-Sight	92.25	83.21
Non-Line-of-Sight	90.21	81.24
	ICA Breathing Rate(%)	Heart Rate(%)
Line-of-Sight	96.76	92.34
Non-Line-of-Sight	95.13	91.57
	SpaceBeat Breathing Rate(%)	Heart Rate(%)
Line-of-Sight	99.12	97.91
Non-Line-of-Sight	98.74	97.03

front side, 98.93% and 97.28% for the left side, 98.88% and 97.31% for the right side, 98.65% and 97.27% for the backside. The results demonstrate that the orientation has little impact on our system performance and verify our proposed system's robustness.

4.8 Impact of Non-Line-of-Sight Scenario

To study the performance of our proposed system under non-line-of-sight (NLoS) scenarios, we conduct the NLoS experiments using a wood panel as an obstacle between the WiFi devices and the people. The wood panel has a thickness of 2.5 cm, a width of 170 cm, and a height of 110 cm. The results are illustrated in Table 6. Specifically, under NLoS scenarios, the FFT-based system achieves accuracies of 90.21% for breathing rate monitoring and 81.24% for heart rate monitoring. The ICA-based system achieves accuracies of 95.13% and 91.57% under NLoS scenarios for breathing rate and heart rate monitoring, respectively. Our system achieves the best performance under NLoS scenarios. The accuracy for breathing rate monitoring of our system is 98.74% and the accuracy for heartbeat monitoring is 97.03%. In all three systems, the performance under NLoS scenarios decreases compared with the LoS scenarios. This is attributed to the attenuation of the WiFi signal strength when passing through obstacles, leading to a lower Signal-to-Noise Ratio (SNR) of the received signals. Nevertheless, our system can still achieve the best performance since our system can separate vital signs signals of multiple people and mitigate interference even in NLoS scenarios.

4.9 Performance under Complex Scenes

In this section, to evaluate the performance of experiments in complex scenes, we conduct experiments to evaluate the system with more furniture and electrical devices (e.g., an electric fan and computers). Note that common scenes refer to our experiments' setups introduced, while complex scenes denote the experiments with

Table 7. System performance for complex scenes and common scenes.

	Breathing Rate(%)	Heart Rate(%)
Common Scenes	99.11	97.98
Complex Scenes	98.65	97.54

more desks, chairs, and electric devices (e.g., an electric fan and computers). In Table 7, we can observe that the accuracies of our system under common scenes are 99.11% and 97.98% for breathing rate and heart rate monitoring, respectively. The accuracies in complex scenes are 98.65% and 97.54% for breathing rate and heart rate monitoring, respectively. Compared to the common scenes, the accuracy of our system slightly decreases in complex scenes. This is because adding more furniture and electrical devices can lead to more complex signal reflections. However, our system can still separate signals in the spatial domain using 2D AoA spectrums, mitigate the impact of stationary objects and dynamic interferences, and accurately extract vital signs.

5 DISCUSSION

Real-time Implementation. Currently, with limited consumer-grade computational resources, our system is not able to process the 4D MUSIC algorithm in real time. In future work, we would like to further analyze the computational issue with a more powerful server and optimization strategies that could reduce the computational complexity. Specifically, the computational cost of our system is mainly made up of two parts.

The first part of the computational cost involves decoupling signals using our cPCA-CL model. However, once the training of the model is complete, the computational cost for decoupling signals is negligible.

The second part involves the estimation of multidimensional information using the 4D MUSIC algorithm, which incurs a high computational cost and constitutes the primary component of the computational complexity in our system. In particular, we utilize the 4D MUSIC algorithm to jointly estimate multidimensional information, encompassing azimuth (ω), elevation (φ), ToF (τ), and AoD (θ) parameters. This process needs to conduct an exhaustive search across four dimensions. The total number of potential combinations for estimating these parameters in the four-dimensional 4D MUSIC algorithm can be expressed as the product of the step sizes for searching in each dimension: $s_\omega \times s_\varphi \times s_\tau \times s_\theta$, where s_ω, s_φ, s_τ, and s_θ denote the number of steps for searching in azimuth, elevation, ToF, and AoD dimensions, respectively.

Despite the high computational costs associated with the 4D MUSIC algorithm, there are several strategies available to significantly reduce the costs. One approach involves reducing the number of steps for searching in each dimension, which in turn decreases resolution and may impact system performance. Therefore, we need to find a trade-off between performance and computational cost. Another strategy is to explore alternative algorithms that circumvent exhaustive searches in four dimensions. For instance, the Space Alternating Generalized Expectation Maximization (SAGE) algorithm [13], employing a coordinate descent approach [45], can be employed. This method involves iterative processes where three of the four parameters are fixed, and the value of the remaining parameter is optimized to maximize the output. Consequently, the search space is reduced to $s_\omega + s_\varphi + s_\tau + s_\theta$, leading to a significant reduction in computational complexity. Furthermore, another strategy involves employing dimension reduction-based MUSIC algorithms to simplify the 4D estimation challenge into two separate 2D problems [20]. This approach enhances computational efficiency by breaking down the multidimensional problem into more manageable components.

6 CONCLUSION

In this work, we propose SpaceBeat, a commodity WiFi-based identity-aware and interference-robust vital signs monitoring system working for multi-person scenarios. Our system can reuse WiFi devices that already exist at home for long-term monitoring. SpaceBeat separates multiple people and locates each person in the spatial domain. It then extracts each person's vital signs by analyzing the signal change at each person's location. We develop a cPCA-CL framework to decouple the signals reflected off other people and thus eliminate interferences caused by movements of nearby people. We also improve the SNR of subtle vital signs, especially the heartbeat, using the accurate location of people and a harmonic canceller. Extensive experiments show that SpaceBeat achieves high accuracy in vital signs monitoring under various challenging environments, including NLoS, different distances, and different human orientations.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful feedback. This work was partially supported by the NSF Grants CNS-2131143.

REFERENCES

[1] Heba Abdelnasser, Khaled A Harras, and Moustafa Youssef. 2015. UbiBreathe: A ubiquitous non-invasive WiFi-based breathing estimator. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. 277–286.

[2] Abubakar Abid, Martin J Zhang, Vivek K Bagaria, and James Zou. 2017. Contrastive principal component analysis. *arXiv preprint arXiv:1709.06716* (2017).

[3] Abubakar Abid, Martin J Zhang, Vivek K Bagaria, and James Zou. 2018. Exploring patterns enriched in a dataset with contrastive principal component analysis. *Nature communications* 9, 1 (2018), 2134.

[4] Fadel Adib and Dina Katabi. 2013. See through walls with WiFi!. In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*. 75–86.

[5] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. 2015. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 837–846.

[6] Emina Alickovic and Abdulhamit Subasi. 2016. Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier. *Journal of medical systems* 40, 4 (2016), 108.

[7] Ulla Anttalainen, Mirja Tenhunen, Ville Rimpilä, Olli Polo, Esa Rauhala, Sari-Leena Himanen, and Tarja Saaresranta. 2016. Prolonged partial upper airway obstruction during sleep–an underdiagnosed phenotype of sleep-disordered breathing. *European clinical respiratory journal* 3, 1 (2016), 31806.

[8] Robert Avram, Geoffrey H Tison, Kirstin Aschbacher, Peter Kuhar, Eric Vittinghoff, Michael Butzner, Ryan Runge, Nancy Wu, Mark J Pletcher, Gregory M Marcus, et al. 2019. Real-world heart rate norms in the Health eHeart study. *NPJ digital medicine* 2, 1 (2019), 58.

[9] Suzanne Boyer and Vishesh Kapur. 2003. Role of portable sleep studies for diagnosis of obstructive sleep apnea. *Current opinion in pulmonary medicine* 9, 6 (2003), 465–470.

[10] Chen Chen, Yi Han, Yan Chen, Hung-Quoc Lai, Feng Zhang, Beibei Wang, and KJ Ray Liu. 2017. TR-BREATH: Time-reversal breathing rate estimation and detection. *IEEE Transactions on Biomedical Engineering* 65, 3 (2017), 489–501.

[11] Zhe Chen, Tianyue Zheng, Chao Cai, and Jun Luo. 2021. MoVi-Fi: motion-robust vital signs waveform recovery via deep interpreted RF sensing. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 392–405.

[12] Lei Clifton, David A Clifton, Peter J Watkinson, and Lionel Tarassenko. 2011. Identification of patient deterioration in vital-sign data using one-class support vector machines. In *2011 federated conference on computer science and information systems (FedCSIS)*. IEEE, 125–131.

[13] Jeffrey A Fessler and Alfred O Hero. 1994. Space-alternating generalized expectation-maximization algorithm. *IEEE Transactions on signal processing* 42, 10 (1994), 2664–2677.

[14] Amy Groenewegen, Frans H Rutten, Arend Mosterd, and Arno W Hoes. 2020. Epidemiology of heart failure. *European journal of heart failure* 22, 8 (2020), 1342–1356.

[15] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM computer communication review* 41, 1 (2011), 53–53.

[16] Yasser Khan, Aminy E Ostfeld, Claire M Lochner, Adrien Pierre, and Ana C Arias. 2016. Monitoring of vital signs with flexible and wearable medical devices. *Advanced materials* 28, 22 (2016), 4373–4395.

[17] Fatema-Tuz-Zohra Khanam, Ali Al-Naji, and Javaan Chahl. 2019. Remote monitoring of vital signs in diverse non-clinical and clinical scenarios using computer vision systems: A review. *Applied Sciences* 9, 20 (2019), 4474.

[18] Belal Korany, Chitra R Karanam, Hong Cai, and Yasamin Mostofi. 2019. XModal-ID: Using WiFi for through-wall person identification from candidate video footage. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–15.

[19] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. 2015. Spotfi: Decimeter level localization using wifi. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*. 269–282.

[20] Xiang Lan, Wei Liu, and Henry YT Ngan. 2017. Joint 4-D DOA and polarization estimation based on linear tripole arrays. In *2017 22nd International Conference on Digital Signal Processing (DSP)*. IEEE, 1–5.

[21] Antonio Lazaro, David Girbau, and Ramon Villarino. 2010. Analysis of vital signs monitoring using an IR-UWB radar. *Progress In Electromagnetics Research* 100 (2010), 265–284.

[22] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. 2015. Tracking vital signs during sleep leveraging off-the-shelf wifi. In *Proceedings of the 16th ACM international symposium on mobile ad hoc networking and computing*. 267–276.

[23] Xuefeng Liu, Jiannong Cao, Shaojie Tang, and Jiaqi Wen. 2014. Wi-sleep: Contactless sleep monitoring via wifi signals. In *2014 IEEE Real-Time Systems Symposium*. IEEE, 346–355.

[24] Gregory F Masters and Stuart F Gregson. 2007. Coordinate system plotting for antenna measurements. In *AMTA Annual Meeting & Symposium*, Vol. 32.

[25] Kun Qian, Chenshu Wu, Fu Xiao, Yue Zheng, Yi Zhang, Zheng Yang, and Yunhao Liu. 2018. Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart devices. In *IEEE INFOCOM 2018-IEEE conference on computer communications*. IEEE, 1574–1582.

[26] Kun Qian, Chenshu Wu, Yi Zhang, Guidong Zhang, Zheng Yang, and Yunhao Liu. 2018. Widar2. 0: Passive human tracking with a single Wi-Fi link. In *Proceedings of the 16th annual international conference on mobile systems, applications, and services*. 350–361.

[27] Bhaskar D Rao and KV Sl Hari. 1989. Performance analysis of root-MUSIC. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37, 12 (1989), 1939–1949.

[28] Yili Ren, Sheng Tan, Linghan Zhang, Zi Wang, Zhi Wang, and Jie Yang. 2020. Liquid level sensing using commodity wifi in a smart home environment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–30.

[29] Yili Ren, Yichao Wang, Yingying Chen, and Jie Yang. 2022. A Vision-Based Approach for Commodity WiFi Sensing. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. 800–801.

[30] Yili Ren, Yichao Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2023. Person Re-identification in 3D Space: A {WiFi} Vision-based Approach. In *32nd USENIX Security Symposium (USENIX Security 23)*. 5217–5234.

[31] Yili Ren, Zi Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2021. Winect: 3D human pose tracking for free-form activity using commodity WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–29.

[32] Yili Ren, Zi Wang, Yichao Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2022. Gopose: 3d human pose estimation using wifi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–25.

[33] Ralph Schmidt. 1986. Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation* 34, 3 (1986), 276–280.

[34] Elahe Soltanaghaei, Avinash Kalyanaraman, and Kamin Whitehouse. 2018. Multipath triangulation: Decimeter-level wifi localization and orientation with a single unaided receiver. In *Proceedings of the 16th annual international conference on mobile systems, applications, and services*. 376–388.

[35] Sheng Tan, Yili Ren, Jie Yang, and Yingying Chen. 2022. Commodity WiFi sensing in ten years: Status, challenges, and opportunities. *IEEE Internet of Things Journal* 9, 18 (2022), 17832–17843.

[36] Sheng Tan and Jie Yang. 2016. WiFinger: Leveraging commodity WiFi for fine-grained finger gesture recognition. In *Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing*. 201–210.

[37] Evelien ES van Riet, Arno W Hoes, Alexander Limburg, Marcel AJ Landman, Henk van der Hoeven, and Frans H Rutten. 2014. Prevalence of unrecognized heart failure in older persons with shortness of breath on exertion. *European journal of heart failure* 16, 7 (2014), 772–777.

[38] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[39] Senmao Wang, Xiaoling Ni, Liangye Li, Jingyi Wang, Qi Liu, Zhijun Yan, Lin Zhang, and Qizhen Sun. 2020. Noninvasive monitoring of vital signs based on highly sensitive fiber optic mattress. *IEEE Sensors Journal* 20, 11 (2020), 6182–6190.

[40] Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 1230–1239.

[41] Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. TensorBeat: Tensor decomposition for monitoring multiperson breathing beats with commodity WiFi. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 1 (2017), 1–27.

[42] Yichao Wang, Yili Ren, Yingying Chen, and Jie Yang. 2022. Wi-Mesh: A WiFi Vision-based Approach for 3D Human Mesh Construction. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. 362–376.

[43] Yichao Wang, Yili Ren, Yingying Chen, and Jie Yang. 2022. A wifi vision-based 3D human mesh reconstruction. In *Proceedings of the 28th Annual International Conference on Mobile Computing and Networking*. 814–816.
[44] Yichao Wang, Yili Ren, and Jie Yang. 2024. Multi-Subject 3D Human Mesh Construction Using Commodity WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 1 (2024), 1–25.
[45] Stephen J Wright. 2015. Coordinate descent algorithms. *Mathematical programming* 151, 1 (2015), 3–34.
[46] Yaxiong Xie, Jie Xiong, Mo Li, and Kyle Jamieson. 2019. mD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.
[47] Zhicheng Yang, Parth H Pathak, Yunze Zeng, Xixi Liran, and Prasant Mohapatra. 2017. Vital sign and sleep monitoring using millimeter wave. *ACM Transactions on Sensor Networks (TOSN)* 13, 2 (2017), 1–32.
[48] Shichao Yue, Hao He, Hao Wang, Hariharan Rahul, and Dina Katabi. 2018. Extracting multi-person respiration from entangled RF signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–22.
[49] Ahmad Anwar Zainuddin, Sakthyvell Superamaniam, Andrea Christella Andrew, Ramanand Muraleedharan, John Rakshys, Juana Miriam, M Amirul Shafiq Mohd Bostomi, Anas Mustaqim Ahmad Rais, Zulkeflee Khalidin, Ahmad Fairuzabadi Mansor, et al. 2020. Patient monitoring system using computer vision for emotional recognition and vital signs detection. In *2020 IEEE student conference on research and development (SCOReD)*. IEEE, 22–27.
[50] Youwei Zeng, Jinyi Liu, Jie Xiong, Zhaopeng Liu, Dan Wu, and Daqing Zhang. 2021. Exploring multiple antennas for long-range WiFi sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–30.
[51] Youwei Zeng, Dan Wu, Jie Xiong, Jinyi Liu, Zhaopeng Liu, and Daqing Zhang. 2020. MultiSense: Enabling multi-person respiration sensing with commodity wifi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–29.
[52] Dongheng Zhang, Yang Hu, Yan Chen, and Bing Zeng. 2019. BreathTrack: Tracking indoor human breath status via commodity WiFi. *IEEE Internet of Things Journal* 6, 2 (2019), 3899–3911.
[53] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *Proceedings of the 22nd annual international conference on mobile computing and networking*. 95–108.
[54] Tianyue Zheng, Zhe Chen, Chao Cai, Jun Luo, and Xu Zhang. 2020. V2iFi: In-vehicle vital sign monitoring via compact RF sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–27.
[55] Tianyue Zheng, Zhe Chen, Shujie Zhang, Chao Cai, and Jun Luo. 2021. More-fi: Motion-robust and fine-grained respiration monitoring via deep-learning uwb radar. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 111–124.