

Extraction of Multi-person Vital Signs using Single-antenna Doppler Radar

De-Ming Chian, Chao-Kai Wen, Fu-Kang Wang, and Kai-Kit Wong

Abstract

Noninvasive monitoring is an important internet-of-things (IoT) application, which is made possible by the advances in radio-frequency (RF) based detection technologies. Existing techniques however rely on the use of antenna array, and their size greatly limits the applicability of in-home monitoring. Another deficiency is that the technology so far is not applicable to multi-person scenarios. In this paper, we propose our system termed ‘DeepMining’ which is a *single-antenna* Doppler radar system that can simultaneously track the breathing rates and heartbeats of multiple persons with high accuracy. DeepMining uses a number of signal observations over a period of time as input and returns the trajectory of the breathing and heartbeat rates of each person. The extraction is based on frequency separation algorithms using successive signal cancellation. The proposed system is implemented using the self-injection locking (SIL) radar architecture and tested in a series of experiments, showing accuracies of 95% and 90% for two and three subjects, respectively, even for closely located persons.

Index Terms

Multi-person vital sign, wireless sensing, blind source separation, frequency estimation.

I. INTRODUCTION

Vital signs such as breathing and heartbeat rates are useful to characterize a person’s basic physiological functions. They are often used by doctors to evaluate an individual’s physical health, identify clues of underlying disease, and monitor the progress of recovery. For example, sleep-disordered breathing can be diagnosed through overnight polysomnography, which records

D.-M. Chian and C.-K. Wen are with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan (e-mail: icefreeman123@gmail.com, chaokai.wen@mail.nsysu.edu.tw).

F.-K. Wang is with the Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan (e-mail: fkw@mail.ee.nsysu.edu.tw).

K.-K. Wong is with Department of Electronic and Electrical Engineering, University College London, UK (e-mail: kai-kit.wong@ucl.ac.uk).

physiological changes during sleep by detecting chest and abdominal movements, respiratory airflow, and electrocardiogram. However, traditional methods for monitoring vital signs during sleep require patients to visit hospitals and wear specialized sensors that are invasive and costly. A hospital environment can be stressful, and the distress may bias measurement results. Furthermore, long-term sleep monitoring in a clinical setting is difficult. Therefore, an alternative solution that is noninvasive, has low-cost, and can perform long-term vital signs monitoring at home is much preferred.

Noninvasive methods for vital signs monitoring are mostly based on radio frequency (RF) technologies, as they do not require measuring devices to be attached to the users, e.g., [1–13]. The methods operate by using RF signals to monitor the small movements from inhalations, exhalations, and heartbeats. Fundamentally, such a system transmits an RF signal, captures its reflections, and measures the time (or strength) the RF signal takes to be reflected back to a measuring device. A person’s chest expands (or contracts) when the person inhales (or exhales). Thus, the reflection time (or strength) of the RF signal changes, which can be interpreted for medical purposes.

Devices used to perform RF-based vital signs detection can be briefly classified as specialized radars [2–8] or WiFi devices [9–12]. For example, [2–5] implemented a Doppler radar at 2.4 GHz to capture a person’s breathing and heartbeat movements with high accuracy in complex indoor environments. Studies in [6–8] utilized frequency-modulated continuous-wave (FMCW) radars to monitor the breathing and heartbeat rates of multiple persons by detecting chest fluctuations from breathing and heartbeats. These radars were designed specifically for target localization and displacement tracking and as a consequence provided highly accurate and reliable cardiopulmonary monitoring. Recently, commodity WiFi devices have also been used for contactless monitoring of respiration and heartbeats [10–12]. Although they have demonstrated promising performance for respiration monitoring, they require directional antennas to achieve accurate heart rate monitoring. The use of directional antennas can help substantially reduce multipath effects and hence increase the vital signs detection sensitivity. However, current solutions are bulky and not accurate for multi-person respiration and heartbeat monitoring.

In this paper, we employ a specialized radar for its higher sensitivity than WiFi for vital signs detection for multi-person monitoring. In this context, prior work [6] dealt with multi-person monitoring by requiring the participants to stay away from each other. An antenna array and/or an FMCW radio can be used for relatively distant people to position a particular individual and

can receive the signals reflected from their body without interference. As the participants become closer in proximity, nevertheless, the RF signals they reflect become superimposed on all the received signals. As a consequence, even systems that combine antenna arrays and FMCW still require a minimum distance of 1.5 m to 2 m. Recently, [7] proposed a system called DeepBreath to recover the breathing signals of multiple individuals even when they are co-located. DeepBreath uses an FMCW radio equipped with an antenna array to separate the RF signals from different locations. Then it further disentangles the mixed signals using independent component analysis. Unfortunately, the antenna array used in [7] was large, which is undesirable for deployment. Furthermore, heart rate monitoring has yet to be achieved.

In contrast to FMCW radars, Doppler radars are another option which have the implementation advantages of a simple architecture, narrowband operation, and low cost but the latter cannot measure the absolute distance of a target. Recently, a single-antenna self-injection locking (SIL) radar was proposed in [3] to achieve a high signal-to-noise ratio (SNR) at the demodulator output that could increase the sensing distance by a factor of four by doubling the operating frequency. The SIL radar uses a single antenna to transmit a continuous-wave (CW) signal and then receives the echo signal without the need to isolate both signals. Therefore, the architecture avoids the problem of setting up two antennas and is particularly useful for high-gain antennas. In addition, given that no mixing occurs between RF and local oscillator signals in the reception process, the SIL radar is inherently immune to stationary clutters, such as those produced by background reflection and antenna coupling. For this reason, the SIL radar has also been used in the livestock industry to monitor animal health in harsh backscattering environments [14].

Figure 1 illustrates a block diagram of a SIL radar system that was built by the industry. The SIL radar is inexpensive, in which the voltage-controlled oscillator (VCO) and low noise amplifier (LNA) are the only active component, and others are passive coaxial components. Extensive experiments in [3–5] have demonstrated its promises in enabling high-sensitivity single-object vital signs detection even for tiny heartbeats. Therefore, a relevant question would be to answer whether the low-cost SIL radar can be applied to provide full-night vital signs monitoring. More specifically, *can the vital signs of multiple persons be monitored through the use of only a single-antenna system?* This paper aims to provide a positive answer to this question by designing, implementing, and validating a system comprising an SIL radar that tracks the instantaneous breathing and heartbeat rates of multiple persons with high accuracy.

Our key observation is that the mixed signals from multiple individuals can still be charac-

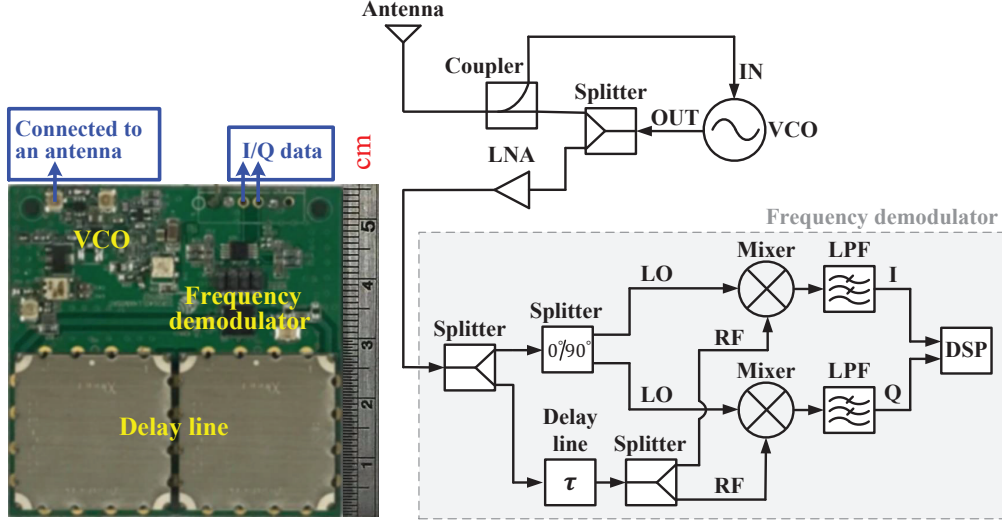


Fig. 1. A block diagram of the SIL radar system [3].

terized by different amplitudes with varied frequencies even without an antenna array. However, the challenge is that the vital signs of multiple people cannot be distinguished from the spectra because of spectral leakage. Also, the noise in an SIL radar system is not white. Thus, the multiple signal classification (MUSIC) algorithm, which is a classical super-resolution technique, cannot be used to perform frequency separation. To tackle this, we propose to estimate the frequencies of multiple people through successive detection and cancel the significant periodic signals from the mixed signals. Specifically, we identify the strongest signal first, subtract it from the combined signal, and then determine the difference as the weak signal. A long period needs to be segmented into multiple short periods for continuous monitoring. Although we can obtain several frequencies for each stable period, identity matching between estimated frequencies and multiple persons remains an issue. This issue is then resolved by frequency tracking.

In summary, we propose our system, referred to as DeepMining, which regards a period of continuous observations as input and returns the trajectory of breathing and heartbeat rates of each person. To the best of our knowledge, DeepMining is the first system that achieves this goal without relying on a bulky antenna array. We conduct a series of experiments to show the high accuracy of DeepMining. Particularly, we show that DeepMining can achieve a mean accuracy that surpasses 90% even with three subjects sitting shoulder-to-shoulder.

The rest of this paper is organized as follows. In Section II, we provide the context of our work. The principle of the SIL radar system is introduced in Section III, where we formulate the

problem as blind source separation. Section IV discusses the challenges and our observations in the multi-person vital signs estimation problem. We also present DeepMining in that section. Our experiments are presented in Section V. Finally, we conclude the paper in Section VI.

II. RELATED WORK

Our work is related to RF-based vital signs detection. The work in [6] used a *T-shaped antenna array* and an *ultra-wide band* FMCW radar to monitor breathing and heart rates. However, the participants need to be far away from each other to allow the signal to reflect off their bodies with limited to no interference. EQ-Radio in [8] then extended [6] to obtain heart rate variability and was used for emotion recognition. Following the architecture of [6], DeepBreath in [7] further disentangles the mixed signals using independent component analysis to recover the breathing signals of multiple individuals. DeepBreath can monitor the respiration of multiple people even when they have zero distance between them. However, multi-person heart rate monitoring has not been addressed.

Some studies have attempted to use WiFi signals to address similar problems. For example, [10] leveraged the amplitude information of the channel state information (CSI) to monitor breathing and heart rates during sleep. PhaseBeat in [11] also achieved the same goal by leveraging CSI phase differences. Studies in [10, 11] used directional antennas or antenna arrays for the heartbeat monitoring scenario to boost the radio signal quality. Overall, the WiFi-based systems cannot directly separate mixed signals from different people. Thus, frequency-domain analysis and signal decomposition algorithms are required in order to extract the respiration rates of multiple people. The power spectrum density method [10], Root-MUSIC [15], and tensor decomposition [16] have been tried to separate respiration patterns. However, they are unable to match identity between multiple persons' respiration rates.

The main distinction between the proposed DeepMining and existing RF-based monitoring systems is the accurate estimation of *both* breathing and heartbeat rates without relying on a bulky antenna array or a beamforming scheme. In contrast to FMCW radars, DeepMining has a simple low-cost architecture based on narrowband operation. Compared with the WiFi-based systems [10, 11], DeepMining exhibits a high sensitivity in heartbeat detection without an antenna array. For source separation, contrary to [10, 15, 16], DeepMining uses frequency tracking to address the identity matching problem. DeepMining also leverages the literature on line spectral estimation (LSE) [17–20], which is a fundamental problem in statistical signal processing and has been used

in numerous domains. In this work, we present an approximation that allows us to formulate the concerned problem as blind source separation and solve it using a properly designed LSE approach.

III. SIL RADAR AND SIGNAL MODEL

A. Principle of SIL Radar

We first review how the SIL radar system [3, 14] can be used to extract vital signs from a single reflector. As shown in Figure 1, the SIL radar system includes an antenna, a coupler, a VCO with an injection terminal, a frequency demodulator composed of a splitter, a passive mixer, a delay line, and a low-pass filter, and a digital signal processing (DSP) unit.

The sensing procedure is described as follows. For transmission, the VCO emits a CW signal $s_{\text{osc}}(t)$ with instantaneous frequency $\omega_{\text{osc}}(t)$ towards a human object. The CW signal is then reflected back from the human object to serve as an injection signal $s_{\text{inj}}(t)$ with instantaneous frequency $\omega_{\text{inj}}(t)$ into the VOC via the injection port. Afterwards, the VCO output signal $s_{\text{out}}(t)$ is frequency demodulated and then low-pass filtered to subsequently detect vital signs in the DSP unit.

We consider a measurement setup with a distance d between the radar and a person. When the person inhales, their chest expands and moves forward, thereby reducing d . By contrast, when the person exhales, their chest contracts and moves away from the device, thereby increasing d . We use $x(t)$ to denote the time-varying distance of the chest wall from inhalations, exhalation, and heartbeats. The distance traveled from the transmission antenna to the stationary person (reflector) and back to the receiving antenna can be expressed as

$$d(t) = 2d + x(t). \quad (1)$$

The periodic displacement of the chest wall causes a Doppler shift. Therefore, the received signal's instantaneous frequency is a delayed version of the VCO output instantaneous frequency plus the reflected Doppler shift, which gives

$$\omega_{\text{inj}}(t) = \omega_{\text{osc}}(t - \tau) + \frac{2\omega_{\text{osc}}}{c} \frac{dx(t)}{dt}, \quad (2)$$

where c is the speed of light, and τ is the round-trip propagation delay (given by $2d/c$). Given $\omega_{\text{osc}}(t - \tau) \approx \omega_{\text{osc}}$, the injection frequency can be viewed as the oscillation frequency modulated

with the displacement rate $dx(t)/dt$. Unlike in the conventional CW radar, the injection signal is fed into the VCO and brought into an SIL state. Afterwards, the VCO's output frequency can be derived as [3]

$$\omega_{\text{out}}(t) = \omega_{\text{osc}} - \omega_{\text{LR}} \sin\left(\frac{d(t)}{c}\right), \quad (3)$$

where the locking range of VCO, ω_{LR} , is given by

$$\omega_{\text{LR}} = \frac{\omega_{\text{osc}}}{2Q} \frac{V_{\text{inj}}}{V_{\text{osc}}}. \quad (4)$$

In (4), Q is a quality factor of the VCO's tank circuit, and V_{inj} and V_{osc} are the injection and the free-running amplitudes, respectively.

Under the SIL condition, the instantaneous phase modulation from the Doppler effect is given by

$$\phi_{\text{out}}(t) = \omega_{\text{out}} \frac{2x(t)}{c}, \quad (5)$$

where ω_{out} is a steady-state value of $\omega_{\text{out}}(t)$. A frequency demodulator is used to process the VCO output signal and to extract the phase modulation signal. Figure 1 shows that the mixer is multiplied by its delay version and outputs the signal with their phase difference to the baseband. As such, variations from breaths and heartbeats can be identified by measuring the resulting variations in the phase of the reflected signal.

B. Signal Model

Although breathing and heartbeat signals are complicated and cannot be characterized easily, they can still be modeled as signals with strong periodicity. To simplify the modeling, we assume that the displacement of the chest wall is a sinusoidal function of time, that is,

$$x(t) = \Delta d_{\text{R}} \sin(2\pi f_{\text{R}} t + \phi_{\text{R}}), \quad (6)$$

where f_{R} and ϕ_{R} represent the respiratory frequency (rate) and phase, respectively, and Δd_{R} is the maximum displacement of the chest wall. Breaths and heartbeats are typically within the ranges of $0.2 - 0.8 \text{ Hz}$ and $0.8 - 2 \text{ Hz}$, respectively. Therefore, the model in (5) remains reasonable after the phase-modulated signal (6) is filtered with a $0.2 - 0.8 \text{ Hz}$ bandpass filter.

If multiple reflectors exist at distances d_1, d_2, \dots, d_K with the maximum displacement of the chest wall being $\Delta d_{\text{R}_1}, \Delta d_{\text{R}_2}, \dots, \Delta d_{\text{R}_K}$, then the total reflected signal for the K reflectors

results in the phase modulation of

$$\phi_{\text{out}}(t) = \frac{2\omega_{\text{out}}}{c} \left(\sum_{k=1}^K \Delta d_{R_k} \sin(2\pi f_{R_k} t + \phi_{R_k}) \right). \quad (7)$$

In an actual situation, thermal and phase noises should be considered. All the imperfect terms are considered, including the modeling error and the noises, and the resulting baseband signal can be modeled as

$$y(t) = \sum_{k=1}^K g_k \sin(2\pi f_k t + \phi_k) + \epsilon(t), \quad (8)$$

where g_k represents the corresponding amplitude of the vital signs at frequency f_k , and $\epsilon(t)$ is the noise, including all the undesirable effects. In the above model for the samples $y_n = y(nT_s)$ at instants nT_s , $n = 1, 2, \dots, N$, we have

$$y_n = \sum_{k=1}^K g_k \sin(\omega_k n + \phi_k) + \epsilon_n, \quad (9)$$

where T_s denotes the sampling period, $\omega_k = 2\pi f_k T_s$, and $\epsilon_n = \epsilon(nT_s)$.

If we filter the phase-modulated signal (5) with a 0.8 – 2 Hz bandpass filter, then (9) can also be used to represent the resulting baseband signal for heartbeats. Rather than using ω_{R_i} and ω_{H_i} to denote the frequencies for the breaths and heartbeats, respectively, we simply use notation ω_i in (9) to denote the frequency of the vital signs of interest. Consequently, vital signs estimation is now recast into frequency estimation ω_i based on the baseband signal y_n .

IV. DEEPMINING

Frequency estimation in the SIL radar system is non-trivial. We plot the FFT output of the baseband signal y_n in Figure 2 to gain insights into the challenges of frequency estimation. In the example, three people are present in the measurement area. First, their breathing cannot be distinguished from the spectra due to the spectral leakage from the short measurement window (e.g., 20 s). A wide measurement window (e.g., 60 s) provides superior frequency resolution but is poorly configured to track changes in the breathing rate. Second, except for certain peak voltages associated with breathing, several sub-peak voltages are associated with unknown interference signals. Therefore, the noise ϵ_n in (9) is not white. The interference signals will lead to poor resolution in the FFT outputs. In this example, the local peaks occur at 0.1 Hz, 0.15 Hz, 0.2

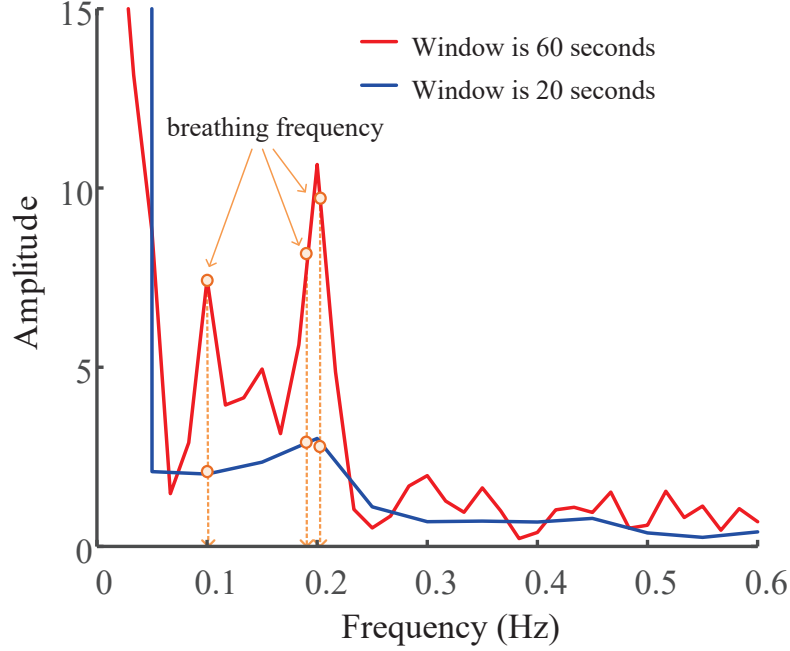


Fig. 2. An example output of the fast Fourier transform (FFT) of the baseband signal y_n after a rectangular time window is applied and filtering between $[20 - 60]$ beats/min.

Hz, whereas the ground truth breathing rates are approximately 0.1 Hz, 0.19 Hz, 0.21 Hz, respectively.

Obviously, we cannot simply detect multiple people's vital signs by selecting the absolute and local maximums of the FFT. This approach is referred to as the power spectral density (PSD) method. The MUSIC algorithm, which is a classical *super-resolution* technique that estimates the frequency of periodic signals, is expected to perform more accurately than the PSD method. The MUSIC algorithm works by separating the signal and noise subspaces. However, the noise ϵ_n in the SIL radar system is not white.¹ Thus, determining clean-noise subspaces is difficult and the classical MUSIC algorithm is not applicable to the SIL radar system.

Our key insight into frequency estimation in the SIL radar is that the SIL mechanism restores the phase modulation signal without distortion and pushes the phase noise out of the modulation bandwidth to higher frequencies. Although background noise exists in several unknown interference signals, they tend to be small. Therefore, we can still separate the vital signs of multiple people through successive detection by cancelling the significant periodic signals from

¹Since the SIL radar works as a first-order delta-sigma modulator equipped with noise shaping capability [3], the noise is not white.

y_n sequentially. Our frequency estimation mechanism is inspired by the successive interference cancellation technique in wireless data detection. In particular, we identify the strongest signal first, subtract it from the combined signal, and then determine the difference as the weak signal. The procedure is repeated iteratively to improve the signal quality from coarse to fine because the accuracy of the other estimated sinusoid will affect the one currently being estimated. We call the entire procedure for extracting multi-person vital signs as DeepMining. It regards a period of observations as input and returns the trajectory of the breathing and heartbeat rates of each person. DeepMining works by following two procedures, namely, frequency estimation and frequency tracking (Figures 3 and 4). The following subsections describe these procedures in greater detail.

A. Frequency Estimation

For ease of notation, we express the signal model (9) in a complex exponential form as

$$y_n = \sum_{k=1}^K g_k e^{j\omega_k n} + \epsilon_n, \quad (10)$$

where g_k is a complex coefficient involving the phase ϕ_k . By stacking (10) for all n in a column, we have

$$\mathbf{y} = \sum_{k=1}^K g_k \mathbf{s}(\omega_k) + \boldsymbol{\epsilon}, \quad (11)$$

where $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$, $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_N]^T$, and

$$\mathbf{s}(\omega) = [1, e^{j\omega}, \dots, e^{j\omega(N-1)}]^T. \quad (12)$$

Let $\mathbf{g} = [g_1, \dots, g_K]^T$ and $\boldsymbol{\omega} = [\omega_1, \dots, \omega_K]^T$. The least-squares (LS) estimator of $(\mathbf{g}, \boldsymbol{\omega})$ is given by

$$(\hat{\mathbf{g}}, \hat{\boldsymbol{\omega}}) = \underset{\mathbf{g}, \boldsymbol{\omega}}{\operatorname{argmin}} \left\| \mathbf{y} - \sum_{k=1}^K g_k \mathbf{s}(\omega_k) \right\|^2. \quad (13)$$

Solving this problem is difficult because the frequencies $\boldsymbol{\omega}$ and amplitudes \mathbf{g} are continuous variables. However, the problem becomes reasonable when $K = 1$ is considered.

For $K = 1$, we write (13) as

$$\underset{g, \omega}{\operatorname{argmin}} \| \mathbf{y} - g \mathbf{s}(\omega) \|^2, \quad (14)$$

which is equivalent to minimizing the cost function

$$J(g, \omega) = -2\text{Re}\{g\mathbf{y}^H\mathbf{s}(\omega)\} + |g|^2\mathbf{s}(\omega)^H\mathbf{s}(\omega). \quad (15)$$

In this case, given ω , the LS estimate of g is calculated by

$$\hat{g} = \frac{\mathbf{y}^H\mathbf{s}(\omega)}{\mathbf{s}(\omega)^H\mathbf{s}(\omega)}, \quad (16)$$

which can be substituted into (15), and the LS estimate of the frequencies solves the following problem:

$$\hat{\omega} = \underset{\omega}{\text{argmin}} J(\omega) \quad (17)$$

with

$$J(\omega) = -\frac{|\mathbf{y}^H\mathbf{s}(\omega)|^2}{\mathbf{s}(\omega)^H\mathbf{s}(\omega)}. \quad (18)$$

Although ω can assume any value in the interval $[0, 2\pi)$, $J(\omega)$ is differentiable. Thus, several efficient search methods, such as gradient and Newton's methods, can be employed to determine the solution of a function of a single variable [21]. Also, ω can be determined by Newton's method through

$$\hat{\omega}^{(t+1)} = \hat{\omega}^{(t)} - \frac{J'(\hat{\omega}^{(t)})}{J''(\hat{\omega}^{(t)})}, \quad (19)$$

where $\hat{\omega}^{(t)}$ represents the t -th iteration, and J' and J'' are the first-order and second-order derivatives of $J(\omega)$, respectively. We use $\hat{\omega}^{(0)}$ and $\hat{\omega}$ to denote the initial and final points as the Newton method converges (or stops), respectively.

For $K > 1$, we identify the strongest signal first, subtract it from the combined signal, and then determine the difference as the weak signal. We use a detection-estimation method to identify each sinusoidal signal, such as those in [17–20]. The method involves two steps, namely, coarse frequency detection and fine frequency estimation. Initialization is performed by the FFT approach to determine the strongest frequency $\hat{\omega}_1^{(0)}$ when no prior frequency information exists about the strongest signal (i.e., in the first round). Fine frequency estimation is performed via Newton's method (19) after the coarse frequency is detected, and $\hat{\omega}_1$ is obtained. Then we remove this sinusoid component from \mathbf{y} and obtain $\mathbf{y}_r = \mathbf{y} - \hat{g}_1\mathbf{s}(\hat{\omega}_1)$. These two steps are repeated until we determine all K frequencies. Before removing each sinusoid component, we

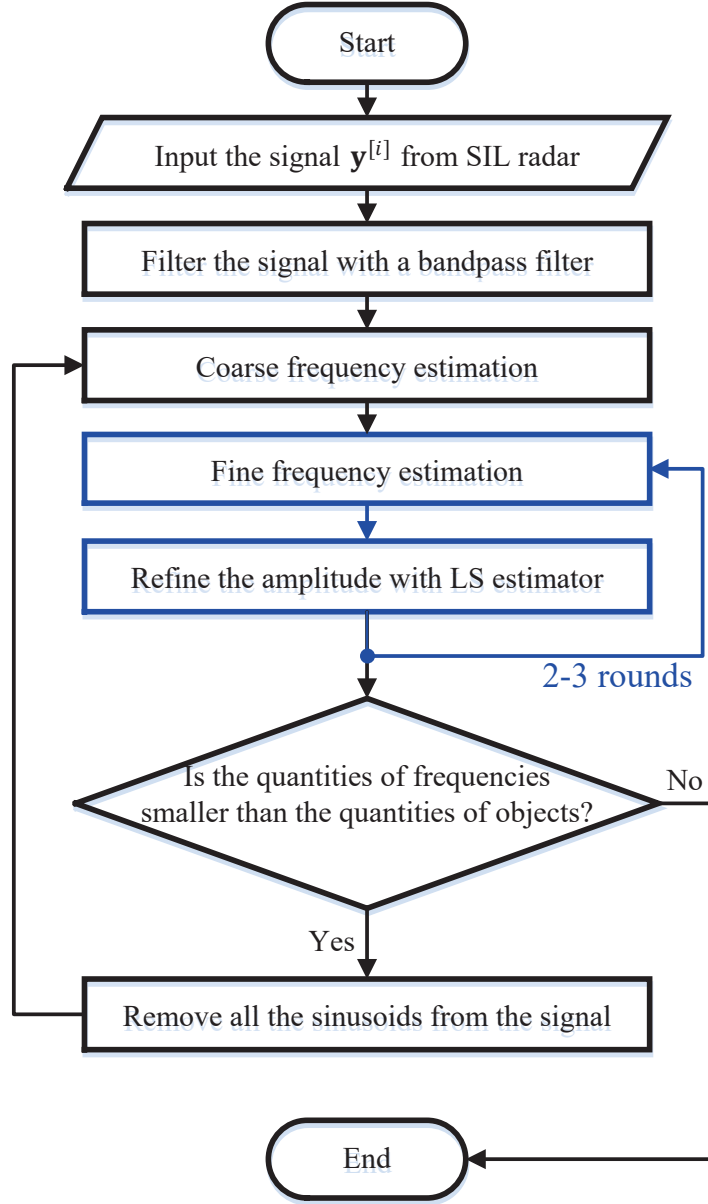


Fig. 3. A flowchart for frequency estimation.

refine the coefficients \mathbf{g} using the LS estimator

$$\hat{\mathbf{g}} = [\mathbf{S}(\hat{\boldsymbol{\omega}})^H \mathbf{S}(\hat{\boldsymbol{\omega}})]^{-1} \mathbf{S}(\hat{\boldsymbol{\omega}})^H \mathbf{y}, \quad (20)$$

where $\mathbf{S}(\hat{\boldsymbol{\omega}}) = [\mathbf{s}(\hat{\omega}_1) \cdots \mathbf{s}(\hat{\omega}_{K'})]$ and $K' \leq K$. This refinement ensures that the energy can be concentrated in \mathbf{g} on the current estimated frequencies such that the residual energy is the minimum for the present set of estimated frequencies. Figure 3 is a flowchart of the frequency

estimation.

Though $\{\hat{\omega}_1, \dots, \hat{\omega}_K\}$ can be obtained sequentially through the above detection-estimation steps, the accuracy of the other estimated sinusoids will affect the one currently being estimated. That is, if the estimation is inadequately accurate, then the current strongest component will not be properly removed, thereby possibly degrading the detection performance of the other sinusoids. Therefore, we repeat the fine frequency estimation step similar to [19] before detecting the next sinusoidal signal. In particular, if frequency $\hat{\omega}_k$ is currently our refined target, then we first remove all the sinusoids from the considered signal except the k -th one. Next, we refine the frequency estimation of the k -th sinusoid using Newton's method and then update $\hat{\omega}_k$ with the obtained value. The above refinement procedure is performed across the entire estimated frequencies and repeated for two to three rounds.

We assume that the number of objects K is available. Our method is not sensitive to inaccurate K values, and a rough guess will suffice. The number only tells the algorithm how many frequencies to extract during the measurements. For example, we might set $K = 2$ in a bedroom to extract the two strongest signals to represent the vital signs of two people. In an actual situation, we set the number K^+ larger than the true object number to estimate the frequency. For example, we set $K^+ = K + 2$ in all our experiments. The setting is explained by the fact that ϵ_n exists in unknown interference signals, which can also degrade the performance of the frequency estimate. Therefore, removing significant interference signals is essential to improve the estimation accuracy. DeepMining can identify the number of objects even with a completely undefined object number by observing the time variation of $\{\hat{\omega}_1, \dots, \hat{\omega}_{K^+}\}$. This identification procedure is explained in Section IV-C.

B. Frequency Tracking

We can apply the frequency estimation algorithm to each stable period and obtain several frequencies for each period. For example, with people #1 and #2, in each stable period, we obtain frequencies $\hat{\omega}_1$ and $\hat{\omega}_2$. Two choices are available to match these frequencies: $\hat{\omega}_1$ is #1's vital signs, and $\hat{\omega}_2$ is #2's vital signs, or vice versa. Such selection should be made for each stable period to obtain the full duration of person's vital signs. Identity matching can become considerably complex when the number of people or periods is larger than two.

Nonetheless, the proposed frequency estimation method allows us to easily address such a task through frequency tracking. Let $\mathbf{y}^{[i]}$ be the observation in the i -th stable period. Given the

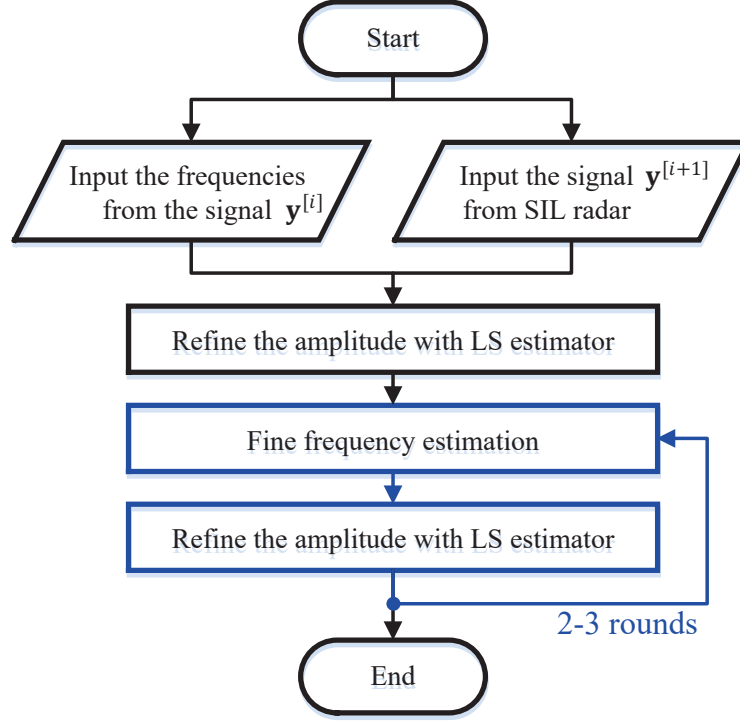


Fig. 4. A flowchart for frequency tracking.

same example above, we obtain frequencies $\hat{\omega}_1$ and $\hat{\omega}_2$ in the stable period i . In the next stable period $i + 1$, we skip the coarse frequency estimation stage and directly use $\hat{\omega}_1$ and $\hat{\omega}_2$ as the estimated result. Next, we implement the frequency refinement stages. The refined frequency of $\hat{\omega}_k$ naturally serves as the vital signs of the same person. We call this procedure frequency tracking, and its flowchart is shown in Figure 4.

Notably, the coefficients \mathbf{g} differ in various stable periods because of the phase offset. Therefore, we must re-estimate these coefficients from the considered signal before proceeding to frequency tracking. Specifically, we re-estimate \mathbf{g} using the LS estimator (20), which is based on $\mathbf{y}^{[i+1]}$. The computational complexity of frequency tracking is much lower than that of frequency estimation. Additionally, our subsequent experiments illustrate that frequency tracking also improves frequency estimation accuracy.

C. Object Identification

Setting a large number K^+ is useful in object identification. We can plot the estimated frequencies and corresponding amplitudes across continuous periods and apply a fit clustering algorithm. The intuition is that the frequency of unknown interference signals varies more rapidly

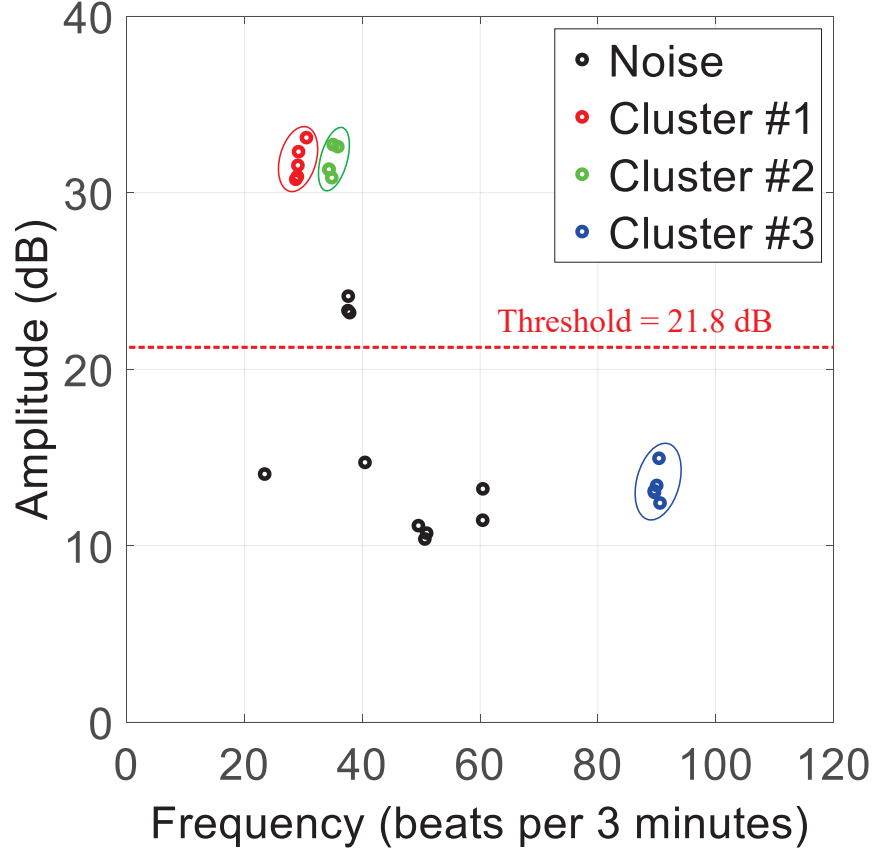


Fig. 5. Sample of clustered points based on five periods in a scenario with two persons. The breathing frequencies of the persons are approximately 10 and 12 beats/min.

than the frequency of the vital signs signals because they are unstable and occasionally seem to disappear. Therefore, interference signals can be identified through long-term observation.

In particular, we apply the well-known density-based spatial clustering of applications with noise (DBSCAN) [22]. Given a set of points in a certain space, DBSCAN groups points that are closely packed together, and the points that lie alone in low-density regions are marked as outliers. A sample of the clustering process is given in Figure 5. Three candidate clusters are clearly seen. Each cluster should have five points. The third cluster can be removed due to its small amplitude. The first and second clusters are identified as the objects. The third cluster belongs to the third harmonic of a person.

In a full-night measurement task, DeepMining begins with object identification to determine K if no information regarding the object number is available. Next, DeepMining, successively performs frequency estimation and tracking. We restart the frequency estimation every few minutes to prevent the frequency tracking results from consistently being dropped into bad

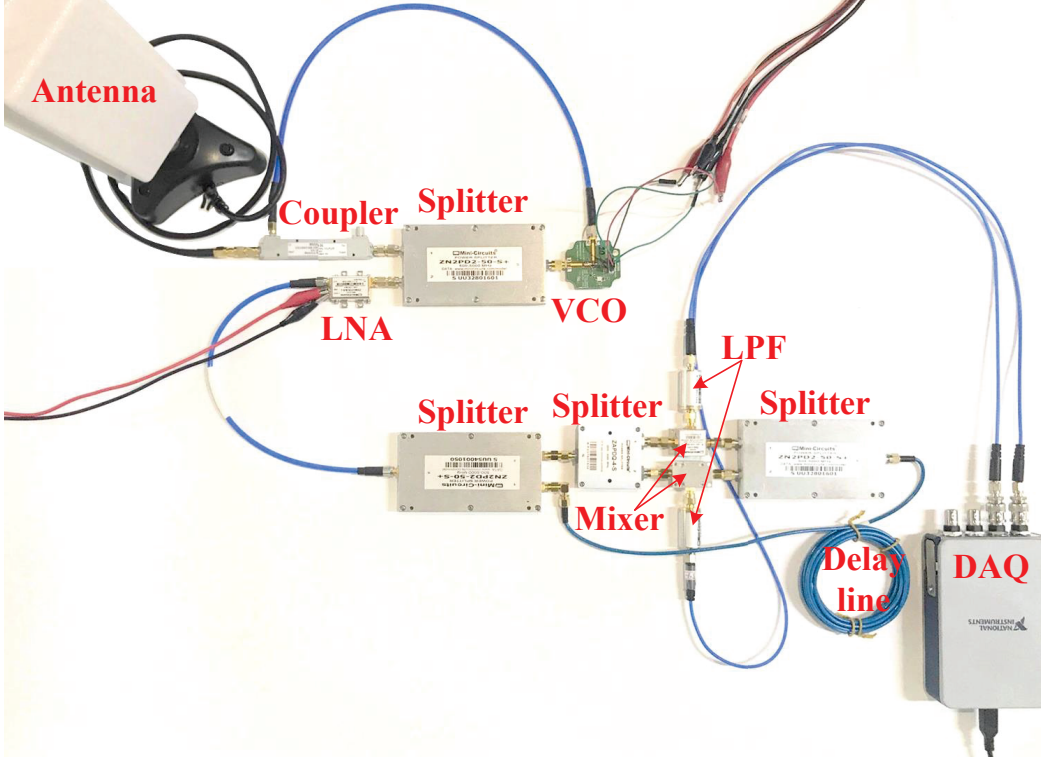


Fig. 6. The hardware configuration of the single-antenna SIL radar system. The architecture follows that in Figure 1.

solutions due to unexpected perturbations.

V. PERFORMANCE EVALUATION

DeepMining is designed to separate and track each vital sign using only one SIL radar that receives mixed signals from multiple human bodies. We conduct a series of experiments to evaluate the performance of DeepMining in the following subsections. We use the PSD method, which examines the peaks in PSD, as a benchmark.

A. Device and Implementation

Hardware—Following the architecture as shown in Figure 1, we reproduce a state-of-the-art single-antenna SIL radar system in Figure 6. The SIL radar operates at a 2.4 GHz ISM band. This VCO, fabricated on an FR4 substrate, uses a Clapp configuration along with an injection port connected to the gate of the transistor. The VCO has a tuning range of 2.3 – 2.65 GHz. The SIL radar uses the same one-panel antenna with 65° in horizontal beamwidth to transmit a CW output signal with a power level of 0 dBm towards the subject and receive the reflected signal.

The SIL radar connects to a computer over the USB. The received signal is sampled (digitized) and transmitted over the USB to the computer for real-time processing. The sampling rate is 100 Hz.

Software—We implement DeepMining using MATLAB. The code runs in real time. For instance, the code running with an Intel i7 CPU at 3.4 GHz takes approximately 1 ms to generate the estimate frequencies from 2,000 sampling points (i.e., the 20 s measurement window). The code operates on shifted overlapping measurement windows and generates new estimates every second.

B. Performance Metrics

Mean Accuracy—To compare the performance of the algorithms, we define the mean accuracy as follows:

$$\text{Mean Accuracy} = \frac{1}{N} \sum_{i=1}^N \left(100\% - \frac{|\omega^{[i]} - \hat{\omega}^{[i]}|}{\omega^{[i]}} \right), \quad (21)$$

where $\hat{\omega}^{[i]}$ is the estimate from the concerned algorithm, and $\omega^{[i]}$ is the ground truth frequency, which we know in advance or measure from the respiration and heart belts. By comparing $\hat{\omega}^{[i]}$ with $\omega^{[i]}$, we can obtain an accuracy of one measurement. After we obtain the average of the accuracy in time-continuous N measurements, we obtain the mean accuracy.

SNR—We define the SNR to understand the quality of the received signal and its effect on frequency estimation. We set the frequency band of respiration from 6 beats/min to 36 beats/min and the frequency band of heartbeat from 60 beats/min to 120 beats/min in this paper. Unless specified otherwise, we take an FFT to obtain a frequency spectrum with 20 s data. We then find the amplitude of the ground truth frequency (or the estimated frequency) as the signal energy. The calculated result is affected by the frequency resolution. Hence, the number of frequency samples is increased with the interpolation method. Next, we calculate the noise energy by subtracting the signal energy from the sum of all amplitudes of the frequency sampling points in the concerned frequency band. We consequently obtain the SNR from a finite-length portion of the sequence by calculating the ratio between the signal and noise energies.

The accuracy and the SNR are updated in every second. Finally, the mean accuracy and the average SNR can be obtained from the continuous testing period. For example, if the testing period is 15 min, then we obtain the average of 900 data samples of accuracy and SNR.

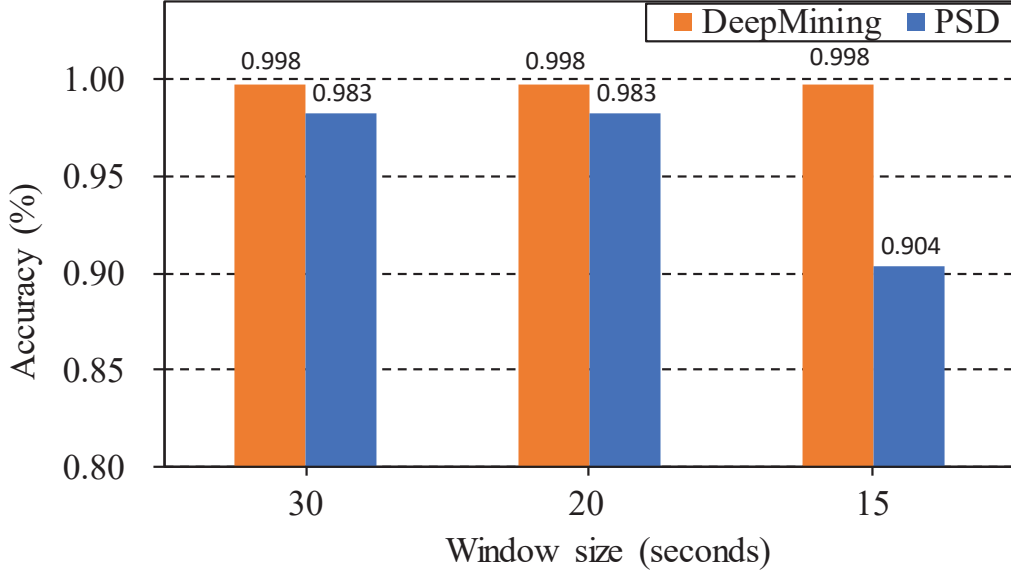


Fig. 7. The accuracy results against the measurement window size. The figure compares the mean accuracy of DeepMining with the PSD method under different measurement window sizes.

C. Accuracy in Clean Room

Different measuring devices for vital sign (e.g., respiration belt and heartbeat belt) have different parameter calculating methods, which result in different outcomes even in the exact same environment. In addition, the results from respiration and heartbeat belts may be biased due to the unconscious body pulling in the testing process. To exclude these uncertainties, we initially design a clean room setup. In the setup, a metal plate ($20\text{ cm} \times 25\text{ cm}$) is simulated as the area of human chest to reflect the electromagnetic wave from the SIL radar. Using an actuator-controlled moving metal plate with 5 mm vibration amplitude and constant velocity along y-axis, we generate the vibration frequency data of 17.7 beats/min. The distance between the panel antenna and the metal plate is 1 m.

Figure 7 compares the mean accuracy of DeepMining with the PSD method under different measurement windows. The mean accuracy is calculated on the basis of a 15-min testing period. For the PSD method, a wide measurement window provides good frequency resolution but poor capability of tracking changes. If the window size is 30 or 20 s, the mean accuracy of the PSD method will be 98.3%. The accuracy of the PSD method degenerates significantly as the window size decreases because the frequency resolution is proportional to the window size. For instance, if the window size is 15 s, then the frequency resolution will be only 4 beats/min. In contrast

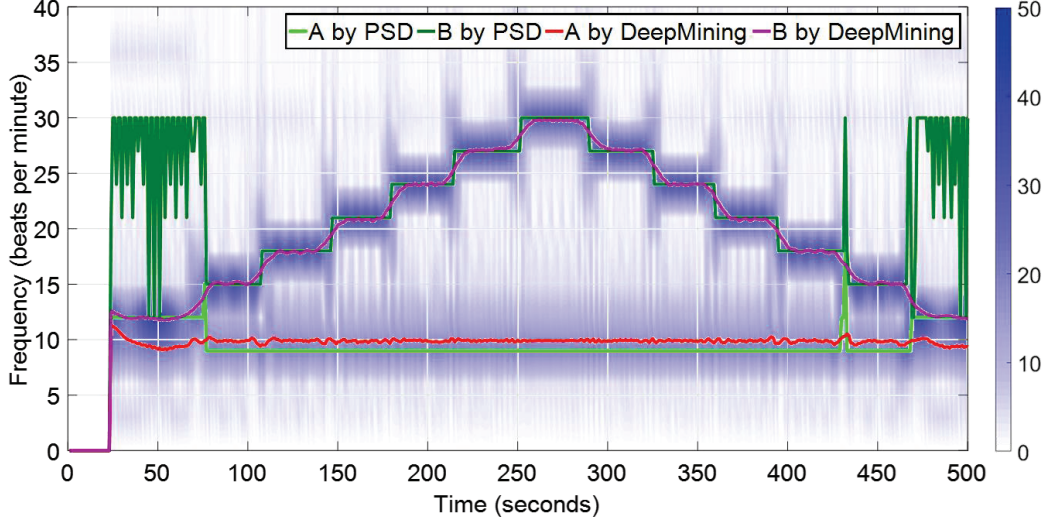


Fig. 8. The spectrogram of the mixed signals from two metal plates. The A metal plate keeps vibrating at 10 beats/min, while the B metal plate increases the vibration frequency from 12 beats/min to 30 beats/min and then decreases back to 12 beats/min. The results calculated by the PSD method and DeepMining are also plotted.

to FFT, the mean accuracy of DeepMining is above 99.8% in all the considered cases and it is unaffected by the frequency resolution. The results demonstrate that DeepMining has the immediacy of acquiring vital sign.

The vibration frequency of the metal plate is in the frequency band of respiration. The average SNR is calculated to be 8.35 dB. We consequently know that DeepMining can obtain 99.8% accuracy under the SNR of 8.35 dB.

D. Multi-frequency Extraction Capability

We design the second experiment similar to the first one to analyze the multi-person vital sign extraction ability of DeepMining. However, in the second experiment, we use two actuators to control two metal plates. The A metal plate keeps vibrating at 10 beats/min. Meanwhile, the B metal plate increases the vibration frequency from 12 beats/min to 30 beats/min and then decreases back to 12 beats/min. It changes 3 beats/min at a 30 s interval. The total testing time is 500 s.

The vibration frequencies of metal plates calculated by the PSD method are determined by selecting the absolute and second maximum values in the respiration frequency band with a 20 s measurement window. Given the two frequencies, we need to handle the identity matching problem. For example, the results from the first period of time comes out as 10 beats/min and

14 beats/min, whereas those from the second period of time comes out as 12 beats/min and 16 beats/min. For the PSD method, we cannot confirm whether the result of 10 beats/min in the first period corresponds to 12 beats/min or 16 beats/min in the second period; as well as the result of 14 beats/min in the first period. Therefore, we sort the vibration frequency of the PSD method in every second and connect them in accordance with high and low frequencies. For DeepMining, the identity matching problem can be easily resolved by the frequency tracking.

Figure 8 shows the result calculated by the PSD method and DeepMining. To confirm the correctness of our calculated result, we plot the spectrogram as the background of Figure 8. The horizontal axis is the time index, and the vertical axis is the frequency variable. At each time index, every frequency variable is corresponding to the magnitude calculated by FFT with the finite length of 20 s, and the different magnitude is plotted with a coloured bar. We compare the spectrogram with the result calculated by the PSD method and DeepMining. When the two metal plates are vibrated at 10 beats/min and 12 beats/min, the PSD method cannot correctly separate the two frequencies and can only roughly obtain the 9 beats/min corresponding to the A metal plate as the frequency resolution is insufficient.² By contrast, DeepMining not only separates two frequencies but results in small difference to the ground truth (lower than 1 beats/min). The result indicates that DeepMining can dissect mixed signal of multiple vital signs.

E. Single-user Accuracy

We use an SIL radar to monitor the vital signs of a person in a real measurement. The experimental setup simulates the situation of a person taking a nap, as shown in Figure 9(a). The person wears the respiration and heartrate belts during a testing period of 15 mins. The respiration belt is calculated once every 10 s, whereas the heartbeat belt is calculated once every 3 s. We have 20 participants in the test.

The measurement result from the respiration belt may be distorted when the subject may move slightly because a human cannot stay still for a long time even asleep. In Figure 9(b), the red line is the original data calculated from the respiration belt, whereas the green line corresponds to the tension data measured from the respiration belt. When the amplitude of tension varies irregularly, the frequency value also changes drastically for a while. This phenomenon may introduce errors to the ground truth frequency. We fit the original data with the fifth order regression line (the

²For a window size of 20 seconds, the frequency resolution is 3 beats/min.

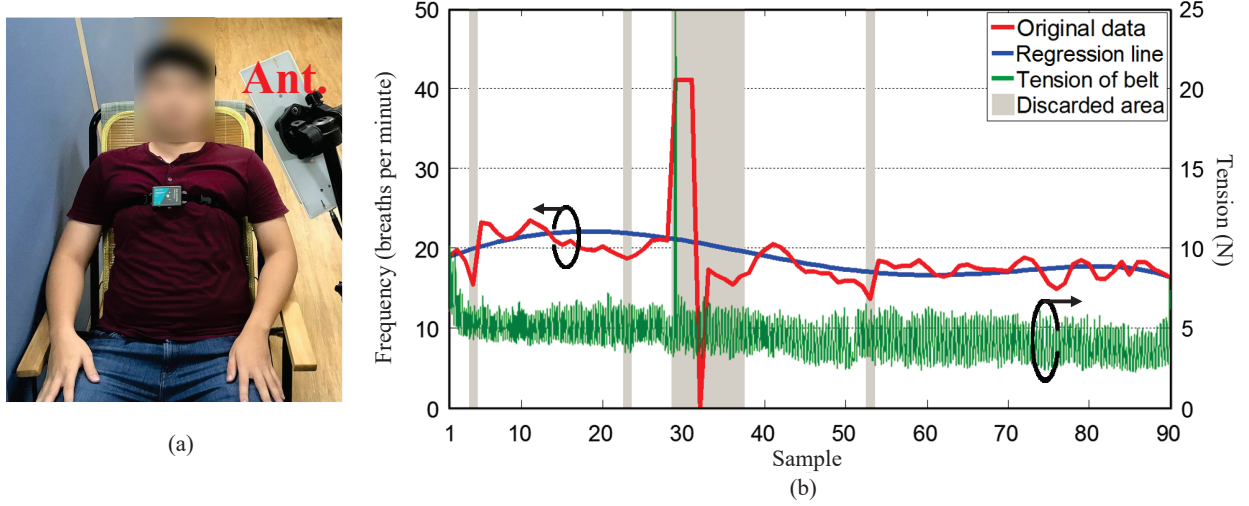


Fig. 9. (a) The experimental setup for a single user scenario. (b) Measurement result from the respiration belt. The red line is the original data calculated from the respiration belt. The green line is the tension measured from the respiration belt. The blue line is the fifth-order regression line of the original data.

blue line in Figure 9(b)) because human vital signs change smoothly. If the original data and the regression line have a difference of more than 3 beats/min as the gray area in Figure 9(b), then we discarded the data.

From our discussion in Section IV-C, the mean accuracy of the PSD method decreases drastically when the window size is 15. We set the window size as 20 s to maintain the performance of the PSD method. We collect 20 data samples from the testers and calculate the mean accuracy of respiration and heartbeat rates. The results are shown in Figure 10(a). We average the 20 data samples of the mean accuracy in the last column. On average, DeepMining can achieve 95.8% and 96.3% of respiration and heartbeat mean accuracy, respectively. The results of using DeepMining are better than those by using the PSD method (93.0% and 95.3%, respectively). We find that in the worst case of the 20 data samples, the mean accuracy of the PSD method is 88.3%. However, the minimum mean accuracy by using DeepMining is 92.4%. Therefore, DeepMining can acquire vital signs immediately and monitor human body stably in practical environments.

Figure 10(b) provides the SNRs of the 20 data samples. The average SNRs for the respiration and heartbeat are 2.31 dB and 2.77 dB, respectively. In comparison with the first experiment by using actuator-controlled moving metal plate, the SNR of the experiment drastically decreases approximately 6 dB from 8.35 dB. However, the mean accuracy only decreases approximately

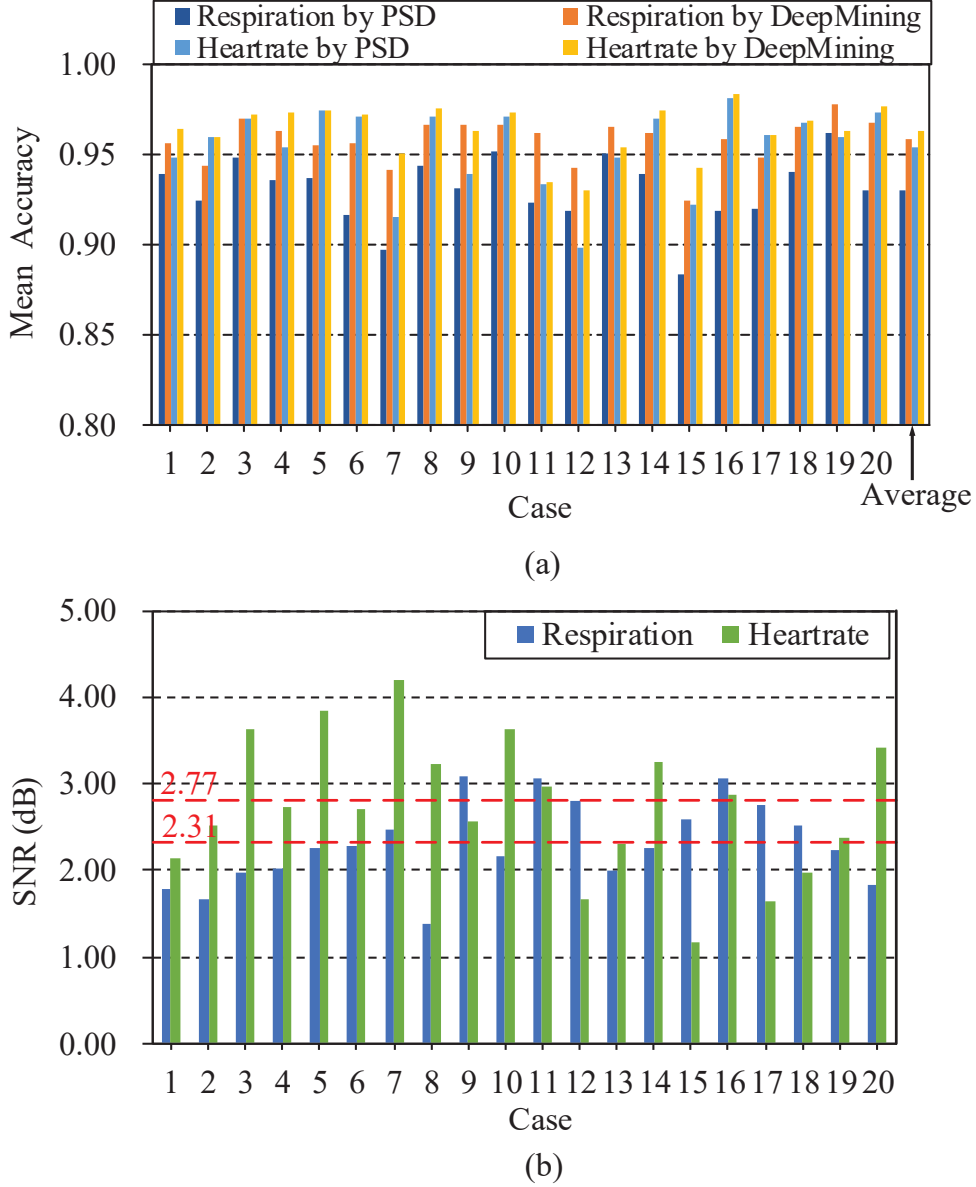


Fig. 10. (a) Mean accuracy of breathing and heartbeat rates. (b) Mean SNR of breathing and heartbeat rates.

4% from 99.8%. Therefore, using the frequency tracking method, DeepMining is more robust to noise.

F. Multi-user Accuracy

In this subsection, we analyze the performance of DeepMining for monitoring multiple people. The PSD method not only suffers from the problem of frequency resolution in the multiple people case but the identity matching problem. Nonetheless, the problems are solved with DeepMining.

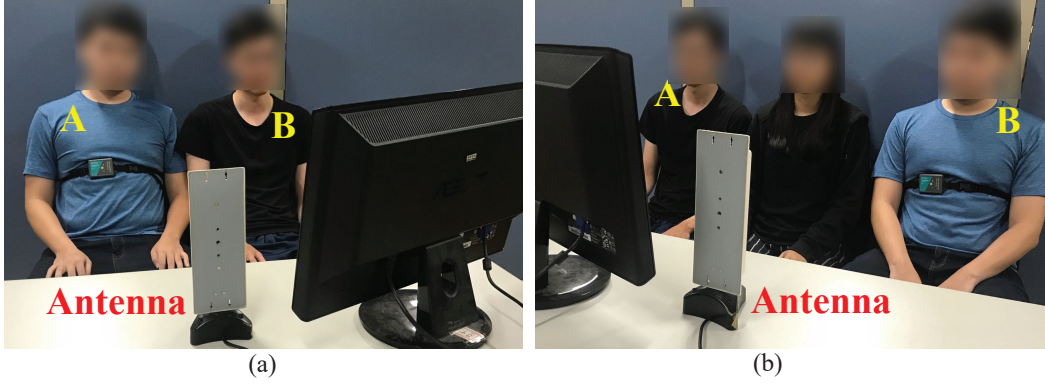


Fig. 11. The experimental setup for multiuser scenarios: (a) two subjects; (b) three subjects.

Hence, we only analyze the performance boundary of DeepMining in the following experiments of monitoring multiple people.

Figure 11 shows the experimental setup of multiple people. Two or three people sit side by side in front of the antenna. The method of collecting data in the multiple-people experiment is to monitor the person sitting on location A or B in 1 min because staying still for a long time is difficult for humans. Then, we collect 10 data samples on every location.

Table I summarizes the mean accuracy and the SNR of the different cases from one person to three people. Results show that DeepMining can achieve a mean accuracy that surpasses 90% in every considered location. The results illustrate that DeepMining can efficiently separate the respiration rate and the heartbeat rate of multiple people with a mixed signal. For the calculation of the SNR in the multiple people case, in addition to the subject chosen for vital sign, others are classified as the noise in the frequent band. Therefore, the noises may interfere the signal from the subject and further affect the calculated result. In addition, in the testing process, people's slight movements are difficult to avoid. Slight movements may cause the signal quality to decrease. In the case of monitoring three people, the SNR of the subjects who stay far from the antenna are approximately 0 dB. Thus, the signal energy is approximately equal to the noise energy. By constraining to the signal quality, the mean accuracy and the number of people analyzed from the mixed signal push the limit. If we increase the transmitted power or use a antenna array system, then we may separate additional vital signs from many more people to obtain a high mean accuracy.

TABLE I
COMPARISON OF THE DIFFERENT CASES.

Case	Respiration		Hearttrate	
	Accuracy	SNR (dB)	Accuracy	SNR (dB)
One person	95.8%	2.31	96.3%	2.77
Two people (A)	94.4%	1.06	95.9%	1.45
Two people (B)	94.5%	1.73	95.8%	0.44
Three people (A)	90.5%	-0.01	90.7%	-0.02
Three people (B)	91.0%	0.9	90.8%	-0.08

V. CONCLUSION

Tracking vital signs of multiple closely located people is challenging. In this study, we presented DeepMining, which is the first complete system comprising a single-antenna Doppler radar that can track the instantaneous breathing and heartbeat rates of multiple persons with high accuracy. Compared with existing solutions, DeepMining has the advantages of a simple architecture, narrowband operation, and low cost. Therefore, it has the potential to be widely deployed. A series of experimental results showed that DeepMining can achieve high accuracy in breathing and heartbeat rate estimation even when the individuals have zero distance between them.

REFERENCES

- [1] N. Patwari, L. Brewer, Q. Tate, O. Kaltiokallio, and M. Bocca, "Breathfinding: A wireless network that monitors and locates breathing in a home," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 1, pp. 30–42, Feb. 2014.
- [2] F.-K. Wang, C.-J. Li, C.-H. Hsiao, T.-S. Horng, J. Lin, K.-C. Peng, J.-K. Jau, J.-Y. Li, and C.-C. Chen, "A novel vital-sign sensor based on a self-injection-locked oscillator," *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 12, pp. 4112–4120, Dec. 2010.
- [3] F.-K. Wang, T.-S. Horng, K.-C. Peng, J.-K. Jau, J.-Y. Li, and C.-C. Chen, "Single-antenna Doppler radars using self and mutual injection locking for vital sign detection with random body movement cancellation," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 12, pp. 3577–3587, Dec. 2011.
- [4] F.-K. Wang, C.-H. Fang, T.-S. Horng, K.-C. Peng, J.-Y. Li, and C.-C. Chen, "Concurrent vital sign and position sensing of multiple individuals using self-injection-locked tags and injection-locked I/Q receivers with arctangent demodulation," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 12, pp. 4689–4699, Dec. 2013.
- [5] F.-K. Wang, P.-H. Juan, S.-C. Su, M.-C. Tang, and T.-S. Horng, "Monitoring displacement by a quadrature self-injection-locked radar with measurement- and differential-based offset calibration methods," *IEEE Sensors J.*, vol. 19, no. 5, pp. 1905–1916, Mar. 2019.

- [6] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, “Smart homes that monitor breathing and heart rate,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Republic of Korea, April 18-23 2015, pp. 837–846.
- [7] S. Yue, H. He, H. Wang, H. Rahul, and D. Katabi, “Extracting multiperson respiration from entangled RF signals,” in *Proc. ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (Ubicomp)*, vol. 2, 2018.
- [8] M. Zhao, F. Adib, and D. Katabi, “Emotion recognition using wireless signals,” *Communications of the ACM*, vol. 61, no. 9, pp. 91–100, Sep. 2018.
- [9] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, “Tracking vital signs during sleep leveraging off-the-shelf WiFi,” in *Proc. 16th ACM Int. Symp. Mobile Ad Hoc Networking and Computing*, New York, NY, USA, 2015, pp. 267–276.
- [10] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang, “Monitoring vital signs and postures during sleep using WiFi signals,” *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2071–2084, Jun. 2018.
- [11] X. Wang, C. Yang, and S. Mao, “PhaseBeat: exploiting CSI phase data for vital sign monitoring with commodity WiFi devices,” in *2017 IEEE 37th Int. Conf. Distributed Computing Systems (ICDCS)*, Atlanta, GA, USA, June 2017, pp. 1230–1239.
- [12] A. Khamis, C. T. Chou, B. Kusy, and W. Hu, “CardioFi: enabling heart rate monitoring on unmodified COTS WiFi devices,” in *Proc. 15th EAI Int. Conf. Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous’18)*, New York, NY, USA, 2018, pp. 97–106.
- [13] Y. Yang, J. Cao, X. Liu, and X. Liu, “Multi-Breath: separate respiration monitoring for multiple persons with UWB radar,” in *2019 IEEE 43rd Annu. Computer Software and Applications Conf. (COMPSAC)*, Milwaukee, WI, USA, USA, Jul 2019.
- [14] C. Li et al., “A review on recent progress of portable short-range noncontact microwave radar systems,” *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 5, pp. 1692–1706, May 2017.
- [15] C. Chen, Y. Han, and Y. Chen, “Multi-person breathing rate estimation using time-reversal on WiFi platforms,” in *Proc. IEEE Global Conf. Signal and Information Processing (GlobalSIP)*, Washington, DC, USA, Dec. 2016, pp. 1059–1063.
- [16] X. Wang, C. Yang, and S. Mao, “Tensorbeat: Tensor decomposition for monitoring multiperson breathing beats with commodity WiFi,” *ACM Trans. Intelligent Systems and Technology (TIST)*, vol. 9, no. 1, p. 8, 2017.
- [17] I. Ziskind and M. Wax, “Maximum likelihood localization of multiple sources by alternating projection,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.
- [18] J. Selva, “ML estimation and detection of multiple frequencies through periodogram estimate refinement,” *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 249–253, Mar. 2017.
- [19] B. Mamandipoor, D. Ramasamy, and U. Madhow, “Newtonized orthogonal matching pursuit: frequency estimation over the continuum,” *IEEE Trans Signal Process*, vol. 64, no. 19, pp. 5066–5081, Oct 2016.
- [20] S. Djukanović and V. Popović-Bugarin, “Efficient and accurate detection and frequency estimation of multiple sinusoids,” *IEEE Access*, vol. 7, pp. 1118–1125, 2019.
- [21] E. K. P. Chong and S. H. Žak, *An Introduction to Optimization*. John Wiley & Sons, Inc., 2013.
- [22] M. Ester, H.-P. Kriegel, J. Xu, and W. Sander, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining (KDD-96)*, 1996, p. 226–231.