Sparsity-Based Multi-Person Non-Contact Vital Signs Monitoring via FMCW Radar

Yonathan Eder, *Graduate Student Member, IEEE*, and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*—Non-contact technology for monitoring the vital signs of multiple individuals, such as respiration and heartbeat, has been investigated in recent years due to the rising cardiopulmonary morbidity, the risk of disease transmission, and the heavy burden on medical staff. Frequency-modulated continuous wave (FMCW) radars have shown great promise in meeting these needs, even using a single-input-single-output (SISO) setup. However, contemporary techniques for non-contact vital signs monitoring (NCVSM) via SISO FMCW radar, are based on simplistic models and present difficulties in coping with noisy environments containing multiple objects. In this work, we first develop an extended model for multi-person NCVSM via SISO FMCW radar. Then, by utilizing the sparse nature of the modeled signals in conjunction with human-typical cardiopulmonary features, we present accurate localization and NCVSM of multiple individuals in a cluttered scenario, even with only a single channel. Specifically, we provide a joint-sparse recovery mechanism to localize people and develop a robust method for NCVSM called Vital Signs-based Dictionary Recovery (VSDR), which uses a dictionary-based approach to search for the rates of respiration and heartbeat over high-resolution grids corresponding to human cardiopulmonary activity. The advantages of our method are illustrated through examples that combine the proposed model with in-vivo data of 30 individuals. We demonstrate accurate human localization in a noisy scenario that includes both static and vibrating objects and show that our VSDR approach outperforms existing NCVSM techniques based on several statistical metrics. The findings support the widespread use of FMCW radars with the proposed algorithms in healthcare.**

***Index Terms*—Frequency modulated continuous wave, joint sparse recovery, multi-person localization, non-contact vital signs monitoring.**

I. Introduction

In the last decade, the rise in chronic health conditions alongside increase in the elderly population, have resulted in a growing need for healthcare approaches that emphasize long-term monitoring in addition to urgent intervention [1], [2], [3], [4], [5]. Monitoring of human vital signs, however, entails numerous difficulties. First, current monitoring devices are typically in physical contact with the measured body, therefore may lead to irritation or general discomfort to the patient, and can be easily detached. Second, usually monitoring devices are connected to patients by medical staff, whether in clinics or hospitals, by a time-consuming interaction that increases the risk of infections and disease transmission, especially during times of pandemics, such as COVID-19. In addition, the manner of connection greatly affects the results, thus it requires considerable skill and experience. Beyond that, many medical teams suffer from high workloads which ultimately lead to an increase in mortality, infections and duration of hospitalization, e.g. in intensive care units [6], [7], [8], [9].

Remote sensing technology such as radar systems can be ideal in these situations since they do not require users to wear, carry, or interact with any additional electronic device [10]. In recent years, several works addressed this issue, attempting to remotely monitor human vital signs such as respiration rate (RR) and heart rate (HR), using radars [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27]. Initially, the family of continuous wave (CW) radars was proposed as simple and reliable devices for remote measurement of cardiac-related chest movements and respiratory activity [11], [12], [13], [14], [15], having the advantages of low transmission power and high sensitivity. However, they do not provide information about patient distance from the radar nor can they separate returns from different objects. To overcome this limitation, many works have turned to use frequency modulated continuous wave (FMCW) radars [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27]. FMCW technology allows spatial separation and potential monitoring of several individuals simultaneously, even by using a single transmitter and a single receiver, i.e. a single-input-single-output (SISO) setup [19], [20], [21], [22], [23], [24], [25], [26], [27], which can reduce processing times and costs of using cumbersome hardware for real-time implementation. However, accurate simultaneous extraction of multiple subjects' cardiopulmonary activity using SISO FMCW radars, is still a challenge in terms of performance, and currently lacks adequate mathematical modelling, leading to sub-optimal solutions.

The conventional algorithmic framework for multi-person non-contact vital signs monitoring (NCVSM) using SISO FMCW radars, is a recurring process in which an estimate of human vital signs is evaluated and recorded at each fixed time interval, as depicted in the lower diagram of Fig. 1. In each

Manuscript received 7 May 2022; revised 25 December 2022 and 9 February 2023; accepted 4 March 2023. Date of publication 10 March 2023; date of current version 6 June 2023. This work was supported by the Israeli Council for Higher Education (CHE) through the Weizmann Data Science Research Center. *(Corresponding author: Yonathan Eder.)*

The authors are with the Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: yoni.eder@weizmann.ac.il; yonina.eldar@weizmann.ac.il).

Digital Object Identifier 10.1109/JBHI.2023.3255740

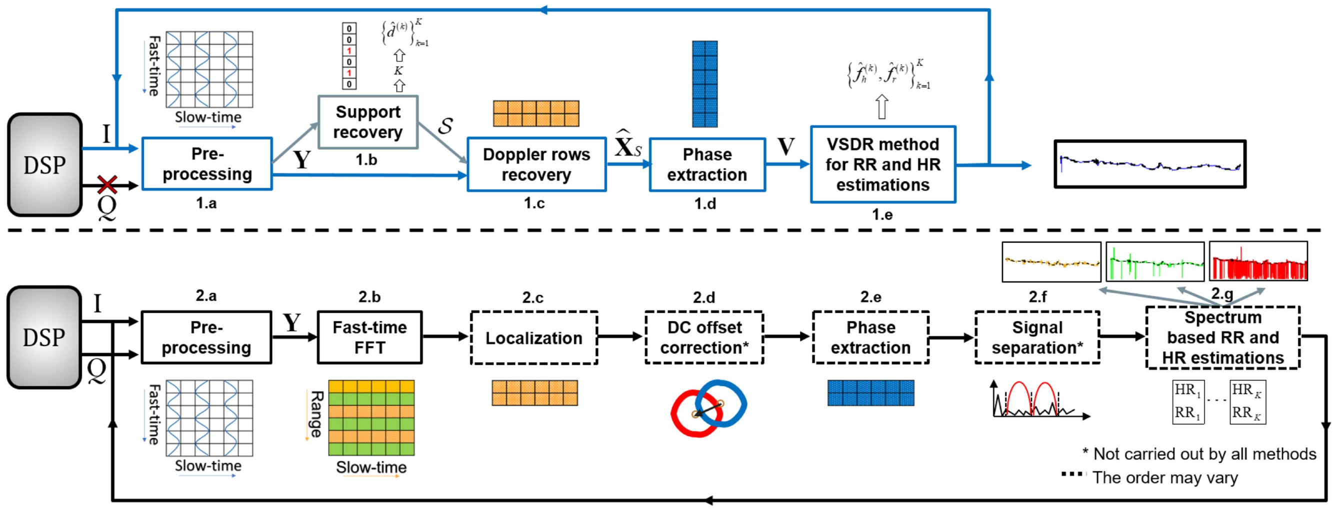


Fig. 1. Block diagram of multi-person NCVSM via SISO FMCW radar. **(Top)** The proposed approach: yields accurate estimates of RR and HR by exploiting the sparse composition of the input data in accordance with human cardiopulmonary properties, using only a single channel. **(Bottom)** The conventional framework: lacks adequate theoretical explanations and presents difficulties in dealing with noisy, cluttered environments.

iteration, a processing matrix is compiled from the raw data obtained by both the *In-phase* (I) and *Quadrature* (Q) channels, through which the samples attributed to human cardiopulmonary activity are located and extracted. Given estimated human thoracic vibrations, various techniques can be used to monitor each individual's RR and HR. In the following, we review existing processing techniques with the help of the lower diagram of Fig. 1.

Traditionally, both the I and Q channels are used to assemble a convenient complex matrix for processing the data of each monitoring iteration [19], [20], [21], [22], [23], [24], [25], [26], [27] (Fig. 1, block 2.a). The main drawbacks of using both channels are the lack of perfect orthogonality, and difference in gain levels in each channel, namely the I/Q imbalance limitation [28], [29], [30], which may corrupt the desired information to be extracted. This imbalance can be compensated for by methods such as the Gram–Schmidt orthogonalization procedure (GSOP) [31], but there is no guarantee for optimal corrections in realistic noisy environments.

The assembled map is then transformed using the fast Fourier transform (FFT) algorithm to contain range bins versus time (Fig. 1, block 2.b), which is then used for human localization (Fig. 1, block 2.c). The authors in [20] performed manual localization based on the intensities of the map, knowing in advance the number of subjects to be monitored, and their true distance from the radar. In [19], the range bin with the maximal average power was selected. In real-world scenarios, information about the number of subjects is not typically available, thus relying solely on spectral magnitudes can produce erroneous decisions due to strong signal reflections obtained from static objects in the radar's field of view (FOV). Several works tried to circumvent the effect of clutters on human localization for NCVSM. The authors in [24] subtracted consecutive time measurements to eliminate reflections off static objects. In [21], the standard deviation was used to distinguish between human and static objects in a room containing furniture. In [22], the authors performed a clipping procedure on a zero-padded FFT map to isolate the target from interfering objects. However, these methods lack adequate theoretical explanations and may be sensitive to non-human vibrating clutters, such as fans or pets.

Once the human-related vectors have been correctly located, the thoracic vibration pattern of each individual is extracted based on a phase extraction process (Fig. 1, block 2.e). Commonly, the extraction is carried out by the arctangent demodulation algorithm [15], followed by an unwrapping procedure. To compensate for the bounds restriction of the arctangent function, [22] and [27] used the extended differentiate and cross-multiply (DACM) algorithm [32] which is based on the derivative of the arctangent function. However, it is highly sensitive to noise and computationally heavy. To improve the extraction quality due to the non-linearity of the operators, [19], [21] performed correction algorithms for DC component offset, prior to this step (Fig. 1, block 2.d). Though, as we discuss in Section III-B, this operation is not required with proper processing.

Given the extracted phase terms, the most commonly used methods to estimate the considered vital signs are based on the discrete Fourier transform (DFT) spectrum (Fig. 1, block 2.g), utilizing the property that in resting state, the frequency bands of heartbeat and respiration do not overlap [19], [20], [21], [22], [24]. Particularly, [19], [22] and [24] used band-pass filters to separate the domains and enhance the relatively weak heartbeat signal (Fig. 1, block 2.f). Aiming to improve frequency resolution and to avoid spectral leakage, [23] performed zero-padding prior to the *slow-time* FFT. Although extremely fine FFT resolution can be obtained this way, two coarsely separated frequencies cannot be resolved. In contrast, [24] performed linear regression on the phase of a filtered complex time-domain signal, to obtain more precise measurements. The latter may improve the accuracy of the estimates, however, the maximal peak is limited by the underlying frequency grid of the DFT. Despite all the well-known benefits of DFT-spectrum analysis, it presents limitations for the considered problem in terms of both resolution and signal representation which ultimately impairs estimation performance.

In this article, we first develop an extended mathematical signal model for the problem of multi-person NCVSM in a cluttered scenario, using SISO FMCW radar. Based on the sparse representation of people via this model, we propose a human localization scheme using a joint-sparse recovery (JSR) mechanism [33], [34], [35]. This approach allows for computationally efficient extraction of Doppler information throughout the monitoring process, even when using a single channel to reduce processing times and avoid dealing with orthogonality issues associated with combining two channels. Then, we demonstrate high-resolution NCVSM of multiple individuals using the developed Vital Signs based Dictionary Recovery (VSDR) technique, which employs a dictionary-based approach to effectively search for the vital signs over high-resolution frequency grids, corresponding to human cardiopulmonary activity.

The performance of the proposed methodology is verified through simulations that incorporate synthetic signals based on the developed model with in vivo data of 30 monitored individuals from [36]. This study demonstrates both precise human localization in a multiple object scenario and superior accuracy results for NCVSM, when compared to contemporary techniques based on [19], [20], [21], [22], [23], [24], using several statistical metrics.

The rest of the article is organized as follows: In Section II, we present the principles of FMCW radar and formulate the problem of single-person NCVSM. In Section III, we present an extended FMCW signal model for multi-person NCVSM and develop the proposed methodology. We evaluate the performance of our algorithms and compare them to existing techniques in Section IV. Finally, Section V summarizes the main points of this work.

Throughout the article, we use the following notation: Scalars are denoted by lowercase letters (a), vectors by boldface lowercase letters $(\mathbf{a})$, sets are given by calligraphic font $(\mathcal{S})$ and matrices are denoted by boldface capital letters $(\mathbf{A})$. The (i,j)'th element of a matrix $\mathbf{A}$ is written as $\mathbf{A}(i,j)$, and $\mathbf{a}_l$ is the l'th column of $\mathbf{A}$. The notations $(\cdot)^T$, and $(\cdot)^H$ indicate the transpose and Hermitian operations, respectively.

II. Standard SISO FMCW Model for Single-Person NCVSM

In this section, we provide the standard 2-D SISO FMCW model for single-subject NCVSM at a given time window, based on previous works. An extended representation for multiple individuals and clutter is presented in Section III, through which the proposed NCVSM approach is developed.

As shown in both Figs. 2 and 3, a typical FMCW radar transmits a series of signals at a given time frame, called chirps [37], whose instantaneous frequency linearly increases over time, forming a saw-tooth waveform. The reflected echo signals split between the I/Q channels through which they are mixed with versions of the transmitted one followed by a low-pass-filter (LPF) to obtain analog base-band signals, known as the intermediate-frequency (IF) signals [27], [38], which are also called beat signals [19], [23]. For each frame, the beat signals in each channel are sequentially sampled by the ADC, resulting in discrete base-band signals of length N, which are sent from the radar to a local computer for further processing.

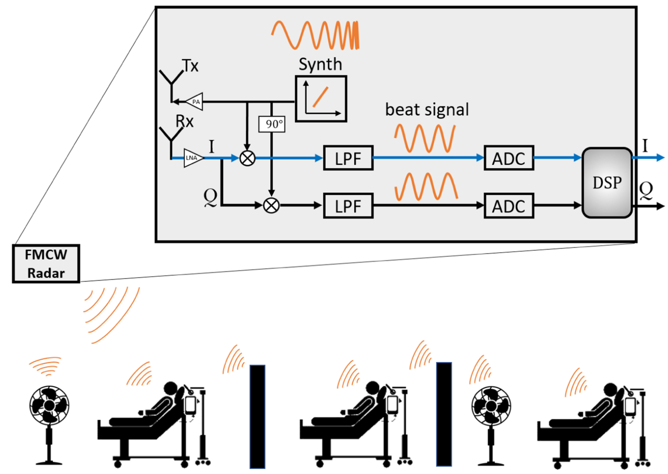


Fig. 2. A schematic illustration of SISO FMCW radar's main components and the examined multiple objects scenario. In this study, we show that we can accurately localize humans in a cluttered environment and monitor their vital signs, using only a single channel (marked in blue).

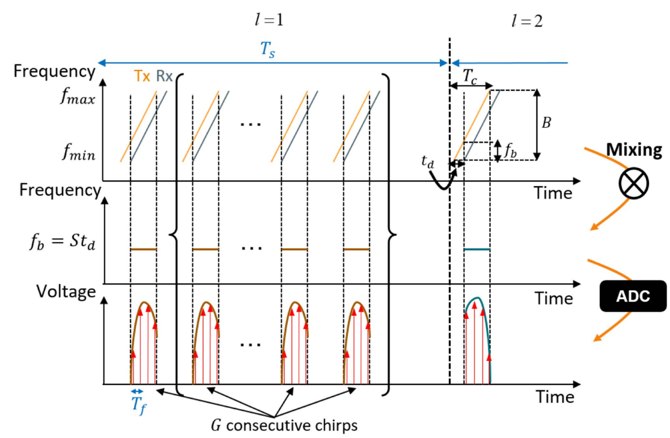


Fig. 3. Transmitted and received chirp sequences at each frame. The brackets illustrate the case of transmitting G consecutive chirps per frame, instead of a single one. **(Top)** Frequency sawtooth waveform. **(Middle)** Analog beat signal after mixing. **(Bottom)** Discrete beat signal after ADC sampling.

Suppose a static human target is located d_0 meters from the radar, with the antenna facing its thorax. The phenomena of respiration and heartbeat produce small time-varying changes in its relative position, so that its actual radial distance from the radar is

$$d(t) = d_0 + v(t), \tag{1}$$

where $v(t) \ll d_0$ denotes a human thoracic vibration due to its cardiopulmonary activity. After hitting a human thorax, a transmitted chirp is reflected back to the radar receiver and appears as an attenuated and shifted version of the transmitted one. As illustrated in Fig. 3, the shifting is expressed by the round-trip delay between the radar and the human object, which is approximated by (1) to $t_d \approx 2d_0/c$ with c denoting the speed of light.

Relying on derivations from [19], [22], [27] for the FMCW signal model, the continuous beat signal of the *In-phase* channel for a single chirp at a given frame, can be described as

$$s_b(t) \triangleq x_b \cos\left(2\pi f_b t + \psi_b(t)\right), \quad t \in [t_d \ \ T_c], \tag{2}$$

with x_b, f_b and $\psi_b(t)$ respectively denoting the amplitude, frequency and phase terms of the beat signal, due to the mixing process between the received and transmitted signals in the overlapping time interval $[t_d \ \ T_c]$, where T_c is the duration of a single chirp, as illustrated in Fig. 3.

The constant beat frequency is defined as

$$f_b \triangleq S t_d = \frac{2S}{c} d_0, \tag{3}$$

where $S \triangleq B/T_c$ corresponds to the rate of the frequency sweep with B being the chirp's total bandwidth. The time-varying beat phase is

$$\psi_b(t) \triangleq \frac{4\pi}{\lambda_{\max}} \left(d_0 + v(t)\right), \quad t \in [t_d \ \ T_c], \tag{4}$$

where $\lambda_{\max}$ denotes the maximal wavelength of the chirp. In practice, thoracic displacement is approximately constant w.r.t. a chirp's duration [19]. In order to extract the temporal variation of a human thorax, consecutive chirps should be transmitted at intervals of $T_s \gg T_c$ seconds, denoted as the *slow-time* sampling interval of $v(t)$.

Hence, based on (4), a discrete phase signal over L given frames can be defined as

$$\psi_b[l] \triangleq \frac{4\pi}{\lambda_{\max}} \left(d_0 + v[l]\right), \quad l = 1, \ldots, L, \tag{5}$$

where $v[l] \triangleq v(lT_s)$. Then, the continuous beat signal at each frame l, denoted by $\tilde{s}_b(t, lT_s) = x_b \cos(2\pi f_b t + \psi_b[l])$, is sampled by the ADC component at the sampling interval T_f, considered as the *fast-time* sampling period of $\tilde{s}_b(t, lT_s)$ (Fig. 3, bottom). This yields the following 2-D discrete beat signal of the *In-phase* channel:

$$y_I[n, l] \triangleq \tilde{s}_b\left(nT_f, lT_s\right) = x_b \cos\left(2\pi f_b n T_f + \psi_b[l]\right), \tag{6}$$

where $n = 1, \ldots, N$ and $l = 1, \ldots, L$. The discrete beat signal obtained by the parallel *Quadrature* channel, which is a 90° shifted version of (6), is used to compose the following complex exponential term

$$y[n, l] \triangleq \tilde{x}_b \exp\left(j\left(2\pi f_b n T_f + \psi_b[l]\right)\right), \quad \begin{cases} n = 1, \ldots, N \\ l = 1, \ldots, L \end{cases}. \tag{7}$$

The samples in (7) form a 2-D measurement matrix $\mathbf{Y} \in \mathbb{C}^{N \times L}$ where $\mathbf{Y}(n, l) = y[n, l]$ (Fig. 1, block 2.a). The samples along the row dimension are referred to as the *fast-time* samples and are related to the distance of the human d_0 for the l'th frame via (3). The samples along the column dimension of (7) are referred to as the *slow-time* samples and are associated with the human thoracic vibration function $v[l]$, which is reflected in varying phase values between successive frames (5). Thus, by estimating f_b and $\{\psi_b[l]\}_{l=1}^{L}$, it is possible to respectively evaluate d_0 and $\{v[l]\}_{l=1}^{L}$ by relations (3) and (5), from which various methods can be used to extract the corresponding RR and HR at the given time window.

III. Sparsity-Based Multi-Person NCVSM: Extended SISO FMCW Model and Solution

In this work, we examine simultaneous vital signs monitoring of several subjects located at different radial distances from the radar, in a realistic environment that contains clutter and noise. To this end, we develop an extension to the model in (7) for FMCW radar signals addressing NCVSM of multiple individuals, using only a single channel and a SISO configuration. We then detail our proposed sparsity-based methodology using this model.

A. Extended SISO FMCW Model for Multi-Person NCVSM

Observing the model in (7), one sees that a monitored individual is characterized by frequency and phase terms of a complex wave, that respectively correspond to its distance from the radar (3) and its thoracic motion pattern (5), for L given frames. In the case of multiple individuals and clutter, each object in the radar's FOV, whether vibrating or stationary, can be modelled based on (7) using an appropriate beat frequency and phase. Consequently, the measured FMCW signal in this case consists of a combination of several signal reflections. To this end, we extend the signal model in (7) to

$$y[n, l] = \sum_{m=1}^{M} x_m \exp\left(j\left(2\pi f_m n T_f + \psi_m[l]\right)\right) + w[n, l], \tag{8}$$

where $n = 1, \ldots, N$, $l = 1, \ldots, L$ and $\{w[n, l]\}$ is a 2-D sequence of i.i.d. zero mean complex Gaussian noise with some variance σ^2. Therefore, the received 2-D beat signal $y[n, l]$ is composed of $M \leq N$ components where the triplets $\{x_m, f_m, \psi_m[l]\}$ denote the amplitude, frequency and phase of each m'th complex wave.

The frequency f_m of each object is proportional to its radial distance from the radar, d_m:

$$f_m \triangleq \frac{2S}{c} d_m, \quad m = 1, \ldots, M. \tag{9}$$

We note that each frequency is distinct which complies with the SISO limitation that allows for the detection of only a single object at any radial distance from the radar. The *slow-time* varying phase $\psi_m[l]$ of each component is given by

$$\psi_m[l] \triangleq \frac{4\pi}{\lambda_{\max}} \left(d_m + v_m[l]\right), \quad l = 1, \ldots, L, \tag{10}$$

where the vibration function $v_m[l]$ is generally modeled for both human and clutter objects by

$$v_m[l] \triangleq \sum_{q=1}^{Q} a_{m,q} \cos\left(2\pi g_{m,q} l T_s\right), \quad l = 1, \ldots, L. \tag{11}$$

The pairs $\{a_{m,q}, g_{m,q}\}_{q=1}^{Q}$ are the corresponding amplitudes and frequencies, with the latter being limited by the *slow-time* frame rate $f_s \triangleq 1/T_s$ according to $\{g_{m,q}\}_{q=1}^{Q} \in [0 \ \ f_s/2)$, for $m = 1, \ldots, M$. We note that while most works did not model the vibration function, [20] and [26] used a sum of 2 sines corresponding to the rates of respiration and heartbeat, at a given time window. Here, the generic vibration model in (11) allows for adequate representation of both static and vibrating objects, using appropriate values for $\{a_{m,q}\}_{q=1}^{Q}$ and $\{g_{m,q}\}_{q=1}^{Q}$.

Particularly, for the case of monitoring K individuals (where K is not known in advance), we denote their unknown radial distances from the radar by $\{d^{(k)}\}_{k=1}^{K}$, and we assume that the corresponding frequency set $\{g_{m,q}\}_{q=1}^{Q}$ (11) includes their unknown HR and RR, denoted by $f_h^{(k)}$ and $f_r^{(k)}$, respectively, for $k = 1, \ldots, K$. As shown in Section III-B2, the objects can be distinguished by utilizing sparse properties of the input signal, within the frequency limits that characterize normal respiration and heartbeat.

Define the *slow-time* varying complex amplitude $\tilde{x}_m[l]$, for $m = 1, \ldots, M$, as

$$\tilde{x}_m [l] \triangleq x_m \exp (j\psi_m [l]), \quad l = 1, \ldots, L. \tag{12}$$

Then, (8) can be represented as

$$y [n, l] = \sum_{m=1}^{M} \tilde{x}_m [l] \exp (j2\pi f_m n T_f) + w [n, l]. \tag{13}$$

For each $l = 1, \ldots, L$, we assemble the *fast-time* samples of $y[n, l]$ into a vector, resulting in

$$\mathbf{y}_l = \mathbf{A}\tilde{\mathbf{x}}_l + \mathbf{w}_l, \quad l = 1, \ldots, L, \tag{14}$$

where $\mathbf{y}_l \triangleq [y[1, l], \ldots, y[N, l]]^T \in \mathbb{C}^N$, $\mathbf{A} \in \mathbb{C}^{N\times M}$ is a Vandermonde matrix, whose entries are given by $\mathbf{A}(n, m) \triangleq \exp(j2\pi f_m n T_f)$, $\tilde{\mathbf{x}}_l \triangleq [\tilde{x}_1[l], \ldots, \tilde{x}_M[l]]^T \in \mathbb{C}^M$ and $\mathbf{w}_l \triangleq [w[1, l], \ldots, w[N, l]]^T \in \mathbb{C}^N$ is the noise vector.

In order to perform continuous NCVSM, the FMCW radar should operate and generate data frames throughout the entire monitoring duration. To this end, at each predefined time interval T_{int}, the sequence $\{\mathbf{y}_l\}_{l=1}^{L}$ (14) is formed by collecting the last L frames up to that time. The number of frames to be processed, L, is determined by a predefined time window T_{win} according to $L = T_{\text{win}} f_s$, where the units of T_{win} and f_s are $[s]$ and $[1/s]$, respectively. For convenience, we reformulate the observations in (14) for each T_{int}, using the following matrix form

$$\mathbf{Y} = \mathbf{A}\tilde{\mathbf{X}} + \mathbf{W}, \tag{15}$$

where $\mathbf{Y} \triangleq [\mathbf{y}_1, \ldots, \mathbf{y}_L] \in \mathbb{C}^{N\times L}$ is the given measurement matrix, $\tilde{\mathbf{X}} \triangleq [\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_L] \in \mathbb{C}^{M\times L}$ and $\mathbf{W} \triangleq [\mathbf{w}_1, \ldots, \mathbf{w}_L] \in \mathbb{C}^{N\times L}$ is the noise matrix. The above data model is assumed to have the following properties:

1) The sequence $\{\mathbf{w}_l\}_{l=1}^{L}$ which forms the noise matrix $\mathbf{W}$, can be viewed as L i.i.d realizations of a zero mean complex Gaussian noise vector with covariance matrix $\sigma^2\mathbf{I}_N$, where $\mathbf{I}_N$ denotes a size-N identity matrix.
2) Only $K \ll M$ subjects are being monitored, where each is stationary except for minor movements caused by breathing or speaking. The latter induces a row-wise sparsity in $\tilde{\mathbf{X}}$, meaning that the vectors $\{\tilde{\mathbf{x}}_l\}_{l=1}^{L}$ share a joint support [34], [35].

Based on the model in (15), the first goal is to recover the row-coordinates of $\tilde{\mathbf{X}}$ associated with humans in the radar's FOV, denoted by the support $\mathcal{S}$, from which we can estimate the number of subjects K and their radial distances from the radar $\{d^{(k)}\}_{k=1}^{K}$. The second goal, using the recovered $\mathcal{S}$, is to continuously evaluate the RR and HR of each detected subject throughout the complete monitoring duration. Mathematically, every T_{int} we seek to estimate the corresponding $\{f_r^{(k)}, f_h^{(k)}\}_{k=1}^{K}$.

B. Sparsity-Based Localization and NCVSM of Multiple Individuals

With the aid of the upper diagram of Fig. 1, below we detail each stage of the proposed sparsity-based approach, based on the model developed in Section III-A.

1) Pre Processing: At the start of each monitoring iteration determined by T_{int}, we perform preliminary processing to assemble $\mathbf{Y}$ according to (15) with an increased SNR (Fig. 1, block 1.a).

Recall from Assumption A-1, that each element of $\mathbf{W}$ is derived from a Gaussian distribution with variance σ^2. Hence, we utilize the slowness of thoracic motion relative to a single chirp [19] to reduce the variance of each element, by averaging several data observations at each frame. To accomplish this, we define a transmission scheme in which $G > 1$ consecutive chirps are transmitted in each frame instead of a single one, with G being limited by the frame duration T_s, as illustrated in Fig. 3.

For every T_{int}, this process generates G beat signal duplicates per frame that differ only in the noise impact, i.e., arranging the input data similarly to (15) results in $\mathbf{Y}_{\text{pre}} = [\{\mathbf{y}_{1,g}\}_{g=1}^{G}, \ldots, \{\mathbf{y}_{L,g}\}_{g=1}^{G}] \in \mathbb{C}^{N\times GL}$. By averaging the *fast-time* row samples every G columns, we get $\mathbf{Y} = [\bar{\mathbf{y}}_1, \ldots, \bar{\mathbf{y}}_L] \in \mathbb{C}^{N\times L}$, where $\bar{\mathbf{y}}_l \triangleq \frac{1}{G}\sum_{g=1}^{G} \mathbf{y}_{l,g}$, $l = 1, \ldots, L$. This procedure yields the model in (15) while reducing the noise variance of each data element by a factor of G, w.r.t. to the use of a single chirp per frame.

2) Support Recovery and Human Localization: Next, we estimate the support $\mathcal{S}$ of $\tilde{\mathbf{X}}$, which allows us to detect the number of subjects and their radial distance from the radar, as well as to efficiently extract the corresponding Doppler samples for the remainder of the monitoring process, as detailed in Section III-B3. Additionally, we show that this localization can be performed only at the first monitoring iteration (Fig. 1, block 1.b).

To this end, first, the *fast-time* frequencies $\{f_m\}_{m=1}^{M}$ (9) are assumed to lie on the Nyquist grid, i.e.,

$$f_m = \frac{f_{\text{ADC}}}{N} i_m, \quad i_m = 0, \ldots, M-1, \tag{16}$$

where $f_{\text{ADC}} \triangleq 1/T_f$ is determined by the ADC component. We note using (9) and (16), that the maximal detectable distance is $d_{\text{max}} = \frac{cf_{\text{ADC}}}{2SN}(M-1)$, and the range resolution is $d_{\text{res}} = \frac{cf_{\text{ADC}}}{2SN} = \frac{c}{2B}$ since $N = f_{\text{ADC}}T_c$ and $ST_c = B$. Interestingly, since N is the number of *fast-time* samples, generally, $M = N$, although as detailed in Section III-B3, by selecting $M = N/2$

we can employ the model in (15) even using data from only a single channel, without jeopardizing estimation performance.

Second, we denote by $B^{(R)}$ and $B^{(H)}$ the frequency bands of normal respiration and heartbeat, respectively, where $B^{(R)}, B^{(H)} \in [0 \;\; f_s/2)$. This prior knowledge of human-typical pulse and breathing frequencies aids in the separation of humans from static or vibrating clutter, such as fans. Hence, based on the *slow-time* frequency modulation structure of the vibration signal $v_m[l]$ (10), (11) in (8), we perform spectral filtering of $\mathbf{Y}$ in the *slow-time* axis, according to

$$\bar{\mathbf{Y}} = \frac{1}{L}\left(\mathbf{F}_L^H\left(\mathbf{\Pi} \odot \mathbf{F}_L \mathbf{Y}^T\right)\right)^T, \tag{17}$$

where $\mathbf{F}_L$ is a full L-size DFT matrix, $\mathbf{\Pi}$ denotes an ideal window corresponding to the vital frequencies in $B^{(H)} \cup B^{(R)}$ and $\odot$ denotes the element-wise product.

Since by Assumption A-2, $\tilde{\mathbf{X}}$ is a row-sparse matrix, inspired by [33], we now recover it from $\bar{\mathbf{Y}}$ given $\mathbf{A}$, using a JSR technique formulated by the following optimization problem:

$$\min_{\tilde{\mathbf{X}} \in \mathbb{C}^{M\times L}} \left\|\bar{\mathbf{Y}} - \mathbf{A}\tilde{\mathbf{X}}\right\|_F^2 + \gamma\left\|\tilde{\mathbf{X}}\right\|_{2,1}. \tag{18}$$

Here, to promote the row sparsity of $\tilde{\mathbf{X}}$, we use the regularization parameter $\gamma \geq 0$ and the mixed $l_{2,1}$ norm defined by $\|\mathbf{X}\|_{2,1} \triangleq \sum_i \|\mathbf{x}^i\|_2$, with $\mathbf{x}^i$ denoting the i'th row of a matrix $\mathbf{X}$. Similarly to [33], we solve (18) and find the support $\mathcal{S}$ using the fast iterative soft-thresholding algorithm (FISTA) [39], [40].

Now, $\mathcal{S}$ is obtained according to the average power [19] over the rows of the recovered $\tilde{\mathbf{X}}$. The support $\mathcal{S}$ is first used to estimate the number of subjects K by the cardinality of $\mathcal{S}$, as well as their radial distances from the radar $\{\hat{d}^{(k)}\}_{k=1}^K$ through (9) and (16), that is $\{d_m\}$, for $m \in \mathcal{S}$. Then, since we are studying the case of monitoring stationary subjects, the coordinates of $\mathcal{S}$ are fixed throughout the monitoring, implying that we can recover $\mathcal{S}$ only once, and use it for all subsequent iterations.

3) Doppler Rows Recovery: The support evaluated in the previous step allows us to efficiently recover only the human-related Doppler samples of $\tilde{\mathbf{X}}$ given $\mathbf{Y}$, throughout the remainder of the monitoring process (Fig. 1, block 1.c).

Using $\mathcal{S}$ and Assumption A-2, the model in (15) can be written as

$$\mathbf{Y} = \mathbf{A}_{\mathcal{S}}\tilde{\mathbf{X}}_{\mathcal{S}} + \mathbf{W}, \tag{19}$$

with $\mathbf{A}_{\mathcal{S}} \in \mathbb{C}^{N\times K}$ and $\tilde{\mathbf{X}}_{\mathcal{S}} \in \mathbb{C}^{K\times L}$ respectively being the atoms of $\mathbf{A}$ and the rows of $\tilde{\mathbf{X}}$ corresponding to $\mathcal{S}$. By knowing the support for each T_{int}, we directly estimate $\tilde{\mathbf{X}}_{\mathcal{S}}$ from $\mathbf{Y}$, using the solution of the following Least-Squares (LS) problem [41]

$$\min_{\tilde{\mathbf{X}}_{\mathcal{S}} \in \mathbb{C}^{K\times L}} \left\|\mathbf{Y} - \mathbf{A}_{\mathcal{S}}\tilde{\mathbf{X}}_{\mathcal{S}}\right\|_F^2, \tag{20}$$

given by

$$\widehat{\mathbf{X}}_{\mathcal{S}} \triangleq \left(\mathbf{A}_{\mathcal{S}}^H\mathbf{A}_{\mathcal{S}}\right)^{-1}\mathbf{A}_{\mathcal{S}}^H\mathbf{Y}. \tag{21}$$

By the Vandermonde structure of $\mathbf{A}$ defined below (14) and since $K \ll M \leq N$, we have that $\mathbf{A}_{\mathcal{S}}^H\mathbf{A}_{\mathcal{S}}$ is invertible. Explicitly, using (16), $\mathbf{A}(n,m) = e^{j2\pi\frac{im}{N}n}$ which results in $\mathbf{A}_{\mathcal{S}} = \mathbf{F}_{\mathcal{S}}^H$ and $\mathbf{A}_{\mathcal{S}}^H\mathbf{A}_{\mathcal{S}} = N\mathbf{I}_K$, meaning that $\widehat{\mathbf{X}}_{\mathcal{S}}$ in (21) can be represented as

$$\widehat{\mathbf{X}}_{\mathcal{S}} = \frac{1}{N}\mathbf{F}_{\mathcal{S}}\mathbf{Y}, \tag{22}$$

where $\mathbf{F}_{\mathcal{S}}$ denotes a partial DFT matrix corresponding to the *fast-time* frequencies $\{f_m\}$ (16), for $m \in \mathcal{S}$.

Note that for $M = N/2$, the frequencies $\{f_m\}_{m=1}^M$ (16) correspond to the positive tones of the sinusoidal combination $\sum_{m=1}^M x_m \cos(2\pi f_m n T_f + \psi_m[l])$, which would have been obtained in (8) when using the *In-Phase* channel solely. Hence, for $M = N/2$ the estimator in (22) when using both the I and Q channels is equivalent to that obtained from the use of only a single channel (I or Q), up to a constant factor. To avoid hardware overload and potential issues of using both channels, we assume here that $M = N/2$ and use only the *In-Phase* channel.

As opposed to existing methods that compute the entire *Range vs. Slow-Time* map (Fig. 1, block 2.b), which is equivalent to replacing $\mathbf{F}_{\mathcal{S}}$ in (22) with $\mathbf{F}_M$ that corresponds to all $M = N/2$ frequencies in (16), at each iteration the proposed estimator recovers only the relevant DFT samples, which is considerably more efficient since $|\mathcal{S}| = K \ll M$.

Finally, since there are no humans within $d_m = 0$ meters of the radar, the estimator in (22) is always filtering the DC component, corresponding to $m = 1$ in (16). As a result, we do not include a mechanism for DC offset correction unlike other techniques (Fig. 1, block 2.d).

4) Phase Extraction: Using relation (12) and the definition of $\tilde{\mathbf{X}}$ below (15), we have that

$$\tilde{\mathbf{X}}_{\mathcal{S}}(m,l) = x_m \exp\left(j\psi_m[l]\right), \quad m \in \mathcal{S}. \tag{23}$$

That is, $\widehat{\mathbf{X}}_{\mathcal{S}}$ in (22) estimates the *slow-time* varying phasor terms associated with humans in the radar's FOV. In order to estimate the appropriate RR and HR from the thoracic vibrations of each individual, i.e., $\{f_r^{(k)}, f_h^{(k)}\}_{k=1}^K$ from $\{v_m[l]\}_{l=1}^L$, $m \in \mathcal{S}$, we must first extract an approximation of the phase terms $\{\psi_m[l]\}_{l=1}^L, m \in \mathcal{S}$ (10) from $\widehat{\mathbf{X}}_{\mathcal{S}}$ (Fig. 1, block 1.d). Hence, we perform the following element-wise angle extraction operation on $\widehat{\mathbf{X}}_{\mathcal{S}}$, which yields the vibration matrix $\mathbf{V} \in \mathbb{R}^{L\times K}$:

$$\mathbf{V}(l,k) \triangleq unwrap\left(\angle\left(\widehat{\mathbf{X}}_{\mathcal{S}}\left(\mathcal{S}\{k\}, l\right)\right)\right)^T, \quad \begin{cases} k = 1,\ldots,K \\ l = 1,\ldots,L \end{cases}. \tag{24}$$

Here, $unwrap(\cdot)$ denotes the unwrapping procedure based on [19], used since the unambiguous phase range is limited by $(-\pi \;\; \pi]$ and the angle extraction operator $\angle(\cdot)$ is based on the four quadrant arctangent function.

The matrix $\mathbf{V}$ can be viewed as a chain of column vectors corresponding to the thoracic vibration pattern of each detected subject, i.e., $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_K]$ where each $\mathbf{v}_k \in \mathbb{R}^L$ contains a scaled approximation of the samples $\{v_m[l]\}_{l=1}^L$ (11), for compatible $m \in \mathcal{S}$.

5) VSDR Method for Estimating RR and HR: In the final stage of each iteration, both the RR and HR of each individual, $\{f_r^{(k)}, f_h^{(k)}\}_{k=1}^K$, are estimated simultaneously given $\mathbf{V}$ and recorded for multi-person NCVSM (Fig. 1, block 1.e). Below, we develop a procedure for selecting high-resolution estimates of vital signs out of two unique dictionaries based on human-typical pulse and breathing frequencies.

First, using (10) and (11), the vibration vectors $\{\mathbf{v}_k\}_{k=1}^K$ that comprise the matrix $\mathbf{V}$ in (24) can be described by

$$\mathbf{v}_k = \mathbf{D}\mathbf{a}_k + \mathbf{n}_k, \quad k = 1, \ldots, K, \tag{25}$$

where each vector $\mathbf{a}_k \in \mathbb{R}^Q$ consists of Q amplitudes $\{\tilde{a}_q^{(k)}\}_{q=1}^Q$ and $\mathbf{D} \in \mathbb{R}^{L\times Q}$ is a cosine-based dictionary matrix with entries $\mathbf{D}(l,q) \triangleq \cos(2\pi\tilde{g}_q lT_s)$ where the frequencies $\{\tilde{g}_q\}_{q=1}^Q \in [0 \quad f_s/2)$. Finally, $\{\mathbf{n}_k\}_{k=1}^K$ denotes a sequence of length-L i.i.d. noise vectors as a result of the non-linear operations in (24).

Ideally, to achieve optimal frequency resolution, one has to use $T_{\text{win}} = 60$ seconds which correspond to the number of heartbeats or breath cycles per minute definition, i.e., bpm. However, this comes at the expense of reduced temporal localization. Therefore, to allow for smaller time windows but with increased resolution we uniformly divide the segment $[0 \quad f_s/2)$ according to a resolution of 1 bpm, i.e., the frequencies $\{\tilde{g}_q\}_{q=1}^Q$ satisfy

$$\tilde{g}_q = h_q\frac{f_s}{Q}, \quad h_q = 0, \ldots, \frac{Q}{2} - 1, \quad Q = 60f_s. \tag{26}$$

We assume that for every T_{win}, the 2 most dominant frequencies of each thoracic vibration $\mathbf{v}_k$, are the rates of heartbeat and respiration. However, the amplitude of the heart signal is much smaller than that of respiration, so in order to facilitate its detection - similarly to [19], [20], [21], [22], [24] - we utilize the phenomenon that at rest, the frequency bands are usually separated from each other. We note that unlike [19], [22] and [24], we do not perform a signal separation procedure prior to this stage (Fig. 1, block 2.f). Here, we exploit this band-separation property to effectively focus the frequency search on limited dictionaries corresponding to each band. Explicitly, we define the vital signs based dictionaries $\mathbf{D}^{(R)} \in \mathbb{R}^{L\times Q_R}$ and $\mathbf{D}^{(H)} \in \mathbb{R}^{L\times Q_H}$ as

$$\mathbf{D}^{(R)}(l,q) \triangleq \cos\left(2\pi\tilde{g}_q^{(R)} lT_s\right) \quad \text{and}$$
$$\mathbf{D}^{(H)}(l,q) \triangleq \cos\left(2\pi\tilde{g}_q^{(H)} lT_s\right), \tag{27}$$

where both the frequencies $\{\tilde{g}_q^{(R)}\}$ and $\{\tilde{g}_q^{(H)}\}$ constitute a subset of (26), satisfying $\{\tilde{g}_q^{(R)}\} \in B^{(R)}$ and $\{\tilde{g}_q^{(H)}\} \in B^{(H)}$.

Using the notation from (27) and the assumption that only the vital frequency bands $B^{(R)}$ and $B^{(H)}$ contribute to the thoracic vibration of each $\mathbf{v}_k$, we represent the model in (25) as

$$\mathbf{v}_k = \mathbf{D}^{(R)}\mathbf{a}_k^{(R)} + \mathbf{D}^{(H)}\mathbf{a}_k^{(H)} + \mathbf{n}_k, \quad k = 1, \ldots, K, \tag{28}$$

where $\mathbf{a}_k^{(R)} \in \mathbb{R}^{Q_R}$ and $\mathbf{a}_k^{(H)} \in \mathbb{R}^{Q_H}$ are both amplitude vectors with only a single non-zero element whose coordinate indicates the corresponding rate (RR/HR), of the k-th human subject.

To recover $\mathbf{a}_k^{(R)}$ and $\mathbf{a}_k^{(H)}$ from each $\mathbf{v}_k$ in (28), we apply the following estimators

$$\hat{\mathbf{a}}_k^{(R)} = \mathbf{D}^{(R)T}\mathbf{v}_k \quad \text{and} \quad \hat{\mathbf{a}}_k^{(H)} = \mathbf{D}^{(H)T}\mathbf{v}_k, \tag{29}$$

for $k = 1, \ldots, K$. The maximum's coordinate of $\hat{\mathbf{a}}_k^{(R)}$ and $\hat{\mathbf{a}}_k^{(H)}$ points to the RR estimation $\hat{f}_r^{(k)}$ and the HR estimation $\hat{f}_h^{(k)}$, from $\{\tilde{g}_q^{(R)}\}$ and $\{\tilde{g}_q^{(H)}\}$, respectively. To enhance estimation stability, we replace the computed $\hat{f}_r^{(k)}$ and $\hat{f}_h^{(k)}$ with the average of the estimates from the last 3 and 1.5 seconds, respectively. This vital signs based dictionary recovery is referred to as the VSDR method.

Algorithm 1 summarizes our approach with L_{lip} and $I_{\max}$ respectively denoting the Lipschitz constant, and the maximal number of iterations in the FISTA algorithm [39].

Algorithm 1: Sparsity-Based Multi-Person NCVSM via SISO FMCW Radar.

Input: T_{int}, $\{y[n,l]\}$, $\mathbf{A}$, γ, L_{lip}, $I_{\max}$, $B^{(R)}$, $B^{(H)}$.
At each T_{int} **do:**
First iteration:
1: Assemble $\mathbf{Y}$ according to Section III-B1
2: Filter spatial interference by (17) to obtain $\bar{\mathbf{Y}}$
3: Perform JSR by (18) using FISTA [39] and save $\mathcal{S}$
4: Estimate $K = |\mathcal{S}|$ and $\{\hat{d}^{(k)}\}_{k=1}^K$ via $\{d_m\}$, $m \in \mathcal{S}$ using (9) and (16)
5: Recover $\widehat{\mathbf{X}}_\mathcal{S}$ given $\mathcal{S}$ and $\mathbf{Y}$ by (22)
6: Extract $\mathbf{V}$ from $\widehat{\mathbf{X}}_\mathcal{S}$ according to (24)
7: Estimate $\{\hat{f}_r^{(k)}, \hat{f}_h^{(k)}\}_{k=1}^K$ given $\mathbf{V}$ by VSDR (29)
Output: $\{\hat{d}^{(k)}, \hat{f}_r^{(k)}, \hat{f}_h^{(k)}\}_{k=1}^K$
In all other iterations:
1: Assemble $\mathbf{Y}$ and skip to steps **5-7** using $\mathcal{S}$
Output: $\{\hat{f}_r^{(k)}, \hat{f}_h^{(k)}\}_{k=1}^K$

TABLE I
SETUP OF MULTIPLE OBJECTS SCENARIO

Object type	x_m	d_m [m]	$\{v_m[l]\}_{l=1}^L$
Vibrating fan #1	0.7	1.5	$0.1\cos(2\pi 40 lT_s)$, $l = 1, \ldots, L$
Human #1	0.5	2	Impedance data from [36]
Static clutter #1	1	2.3	0, $\forall l = 1, \ldots, L$
Human #2	0.45	2.6	Impedance data from [36]
Static clutter #2	0.9	2.9	0, $\forall l = 1, \ldots, L$
Vibrating fan #2	0.6	3.1	$0.1\cos(2\pi 40 lT_s)$, $l = 1, \ldots, L$
Human #3	0.4	3.5	Impedance data from [36]

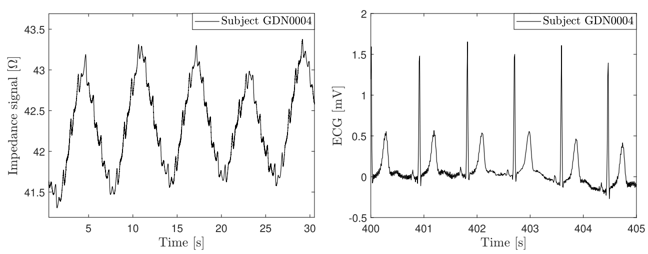
IV. NUMERICAL EXAMPLES

In this section, the performance of the proposed method is evaluated and compared to existing techniques, using a simulation that combines the measurement model in (15) with real Electrocardiography (ECG) and impedance data of 30 individuals from [36]. Note that Algorithm 1 is divided such that during the first monitoring iteration, a localization procedure is performed, after which the vital signs of the detected subjects are estimated throughout the rest of the monitoring period. As a result, we present here two simulation studies. The first investigates the localization of multiple individuals in a cluttered environment while the second examines simultaneous NCVSM, given their extracted thoracic vibrations.

To this end, we used relations (8)–(11) that form the model in (15), to compose 7 different objects in the radar's FOV (of which $K = 3$ human subjects), each characterized by a corresponding x_m (8), d_m (9) and $\{v_m[l]\}_{l=1}^L$ (11), as detailed in Table I and illustrated in Fig. 2. To create a realistic environment, we adjusted

TABLE II
FMCW RADAR PARAMETERS

Parameter	Symbol	Value
Maximal chirp wavelength	$\lambda_{\max}$	3.9 [mm]
Chirp duration	T_c	57 [μs]
ADC sampling rate	f_{ADC}	4 [MHz]
Rate of frequency sweep	S	70 [MHz/μs]
Frame duration	T_s	10 [ms]
# of selected *fast-time* samples	$\bar{N}$	200
# of chirps per frame	G	150



(a) (b)

Fig. 4. Impedance and ECG signals from [36]. **(a)** Impedance [Ω] - used as a reference for comparing the RR estimates and to simulate human thoracic vibration. **(b)** ECG [mV] - used as a reference for comparing the HR estimates.

each x_m so that static objects have the strongest reflections, followed by fans, and finally humans, all of which fade as a function of the distance d_m. Furthermore, rather than generating synthetic thoracic vibration signals for $\{v_m[l]\}_{l=1}^{L}$, here we used real thoracic vibration data of 30 individuals from [36], based on the impedance signal which provides insight into the impedance change of the thorax and can be used to estimate parameters such as RR and HR [42], [43]. In this work, we considered for Table I's vibrations $\{v_m[l]\}_{l=1}^{L}$, the 10-minute long 100 [Hz] impedance cardiography signals (Fig. 4(a)) of subjects 1-30, from [36]'s resting scenario (in which participants were told to breath calm and avoid large movements), normalized to millimeter movements.

Since a variety of respiratory parameters can be extracted from the impedance signal, including RR, [42], [43], the raw signal serves as a reference for comparing the RR estimates. As to the HR reference, we used the gold-standard 2000 [Hz] lead-2 ECG signal (Fig. 4(b)) from [36], and down-sampled it to 100 [Hz] to correspond to $T_s = 10$ [ms]. The main FMCW radar parameters for assembling the model in (15) are based on Texas Instruments IWR1642 76 to 81 [GHz] mmWave sensor [37], and summarized in Table II. We note that for bandwidth $B = 4$ [GHz], $d_{\text{res}} = \frac{c}{2B} = 3.75$ [cm]. In addition, to examine the impact of environmental noise, we used an SNR term that controls the variance of $\{w[n, l]\}$ (8) via SNR $\triangleq 1/\sigma^2 \in [-2 \;\; 2]$ [dB]. Finally, the frequency bands of respiration and heartbeat were set to $B^{(R)} = [0.1 \;\; 0.4]$ [Hz] and $B^{(H)} = [0.78 \;\; 1.67]$ [Hz], respectively, corresponding to a normal resting state.

Using all the above specifications, for each SNR we simulated 10 separate and independent trials, each consisting of localization and 10-minute NCVSM of 3 subjects simultaneously, with RR and HR estimates computed every $T_{\text{int}} = 0.05$ [s], using L data frames from the last $T_{\text{win}} = 30$ [s], starting at T_{win}. We note that each trial included 3 different subjects out of the 30, according to their order in [36]. That is, subjects 1–3 were tested in the first trial, subjects 4–6 in the second, and so on.

A. Multi-Person Localization

This study examined the localization of several subjects in a clutter-rich environment given the pre-processed $\mathbf{Y}$ of the first iteration, as a preliminary step to monitor their vital signs. Particularly, in each of the considered 10 trials, we compared the proposed JSR localization method (17)–(18) to those based on the *Maximum Average Power* [19] and *std* approaches [21].

The parameters of the proposed JSR were set as follows: The vital *slow-time* frequencies of $\mathbf{\Pi}$ in (17) were drawn from the length-L Nyquist grid determined by f_s. The parameters for solving (18) using FISTA [39] were set to $\gamma = 30$, $L_{\text{lip}} = 4.5e6$ and $I_{\max} = 1000$. Due to the consistency of the results in all trials for each SNR, we report here only the localization of the first trial, that is, given impedance data of subjects 1-3 from [36], for SNR $= 0$ [dB].

By replacing $\mathbf{F}_S$ in (22) with $\mathbf{F}_M$, $M = N/2$, the estimator in (22) evaluates the complete *Range vs. Slow-Time* map by $\widehat{\mathbf{X}}_M = \frac{1}{N}\mathbf{F}_M\mathbf{Y}$ and adjusting the *fast-time* axis by (9). Fig. 5(a) depicts the *Range vs. Slow-Time* map via the magnitudes of $\widehat{\mathbf{X}}_M$. The visible row-wise intensities correspond to the DC component as well as the reflections from Table I's objects. We note that although the standard practice is to use this map for localizing humans (Fig. 1, block 2.c, works [19], [20], [21]), the proposed method does not.

One can observe in Fig. 5(b) that the JSR method indicates the correct locations of the subjects compared to the intensity-based *Maximum Average Power* method [19] and the *std* approach [21]. Notice that the former inadvertently selects the strong reflections obtained from statics clutters, whereas the latter seeks highly oscillating objects and thus incorrectly selects the vibrating fans over humans. In contrast to the compared localization techniques, the proposed approach exploits both characteristics of human-typical vital frequencies and prior knowledge of the sparse structure of $\tilde{\mathbf{X}}$.

B. NCVSM vs. SNR

For each examined trial, given a successful human localization from the previous study, we examined the performance of VSDR (29) for simultaneous NCVSM and compared it to other state-of-the-art techniques based on [19], [20], [21], [22], [23], [24].

To compare fairly (regardless of localization performance), we applied the first iteration of Algorithm 1 assuming that all considered subjects were successfully identified. Then, we only compared the last step (7) of the algorithm, in which the vital signs of each subject are estimated given the extracted matrix $\mathbf{V}$ (24). In addition, all methods used the same frequency bands $B^{(R)}$ and $B^{(H)}$ and all other settings that preceded this step were the same.

Our proposed VSDR was first compared to the method detailed in [24] for estimating RR and HR given the phase of an

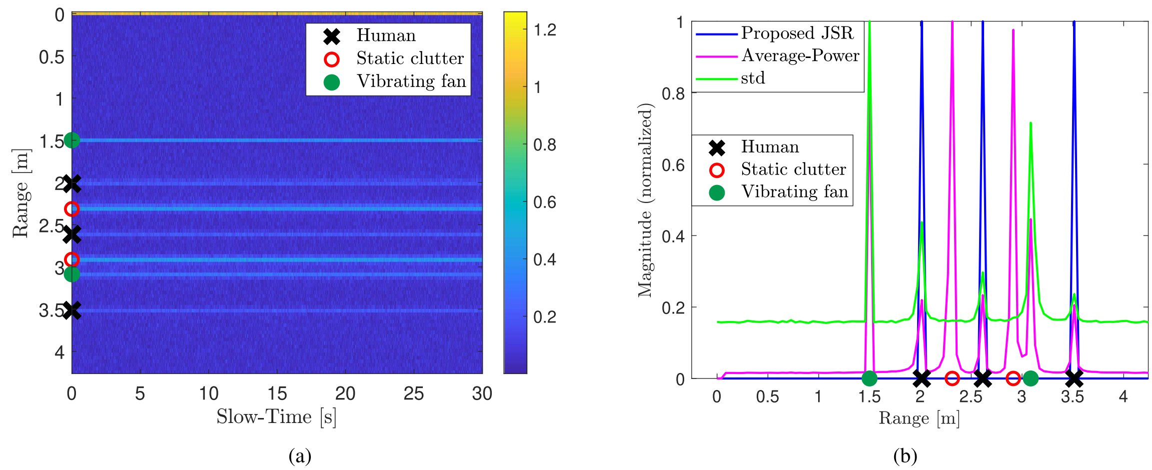


Fig. 5. Multi-person localization by the setup of Table I (subjects 1–3 [36]), for $\mathrm{SNR} = 0$ [dB]. **(a)** The *Range vs. Slow-Time* map via the magnitudes of $\widehat{\mathbf{X}}_M$. The row-wise intensities correspond to the DC component and the reflections of Table I's objects. **(b)** Several localization techniques based on the average power [19], the *std* estimate [21] and the proposed JSR (Section III-B2). Only JSR localization properly detects the subjects in the given scenario.

FMCW signal, called here Phase-Reg. Moreover, VSDR was compared to a FFT-based peak selection in each frequency band, with zero-padding (FFT w/ ZP) [23] and without (FFT w/o ZP) [19], [20], [21], [22], [24]. The padding of FFT w/ ZP was set to fit a 60-second time window corresponding to frequency resolution of 1 [bpm]. All methods were implemented using MATLAB. As to the reference data, we found that it is common to estimate the RR and HR of the reference signals via the DFT spectrum [19], [20], [21], [23], [24]. Assuming that both the ECG and impedance references are noise-free, we padded them similarly to FFT w/ ZP, for optimal results.

To evaluate the monitoring accuracy, we used the following statistical metrics, based on the RR and HR estimates of the compared methods w.r.t. those of the references. 1. Success-Rate (SR) - 2 bpm, defined here as the percentage of times the estimate differed from the reference output by less than 2 [bpm]. 2. Pearson Correlation Coefficient (PCC) [44], showing results in the range [0 1]. 3. Mean-Absolute Error (MAE), and 4. Root-Mean-Square Error (RMSE). The Success-Rate can predict the percentage of time an estimation error greater than 2 [bpm] is expected, the PCC measures the linear correlation between the two data sets, and while in the MAE metric, each error contributes in proportion to its absolute value, the RMSE is a well-known metric that emphasizes the occurrence of coarse errors. In this study, we investigate various SNR cases, each of which involves monitoring data of 30 individuals from [36]. Hence, the performance score produced for each metric, given SNR, is taken as the median across all 30 participants. We note that 10 minutes of monitoring starting at $T_{\text{win}} = 30$ [s], with output obtained at each $T_{\text{int}} = 0.05$ [s] brings to 11400 HR/RR estimates for comparison to the references, for each participant.

Fig. 6 depicts NCVSM of subjects 4-6 (trial #2) given the extracted $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$ w.r.t. the references, by VSDR and the other examined techniques, for $\mathrm{SNR} = 0$ [dB]. It can be seen how both the HR and RR estimates by the VSDR approach show great resemblance to those of the references, compared to the others in which the noisy setup impairs their assessments.

TABLE III
MEDIAN AND MEAN ACCURACY SCORES OF THE PROPOSED VSDR (III-B5) COMPARED TO FFT W/ ZP [23], FFT W/O ZP [19], [20], [21], [22] AND PHASE-REG [24], FOR $\mathrm{SNR} = 0$ [DB]

	Method	SR - 2 bpm	PCC	MAE	RMSE
Median (HR)	FFT w/ ZP	92.28	0.50	0.82	2.98
	FFT w/o ZP	94.03	0.70	0.70	1.28
	Phase-Reg	85.79	0.72	0.84	1.35
	VSDR	**97.01**	**0.89**	**0.54**	**0.78**
Median (RR)	FFT w/ ZP	81.40	0.39	1.20	2.27
	FFT w/o ZP	94.55	0.82	0.62	1.03
	Phase-Reg	80.17	0.74	0.96	1.44
	VSDR	**96.26**	**0.88**	**0.60**	**0.84**
Mean (HR)	FFT w/ ZP	83.30	0.54	2.02	3.92
	FFT w/o ZP	87.80	0.640	1.52	2.85
	Phase-Reg	79.74	0.645	1.71	2.90
	VSDR	**88.64**	**0.77**	**1.18**	**1.87**
Mean (RR)	FFT w/ ZP	76.11	0.47	1.76	2.76
	FFT w/o ZP	93.11	0.78	0.72	1.27
	Phase-Reg	74.29	0.72	1.11	1.66
	VSDR	**94.12**	**0.82**	**0.69**	**1.01**

Fig. 7 shows the median NCVSM performance, in terms of SR - 2 bpm, PCC, MAE and RMSE, for both the HR and RR estimations by all examined methods, as a function of the SNR. One sees that VSDR outperforms the other compared methods in all 4 metrics, for every SNR value. Also, note the considerable difference in performance in the more challenging task of HR monitoring due to the weak heartbeat signature, in favor of the proposed approach. As for the RR monitoring case, the small difference between VSDR and the popular FFT w/o ZP can be explained by the relative dominance of the respiratory signal that facilitates its detection, leading to fine results in both techniques.

At last, in Fig. 8 we show how the HR-RMSE and RR-RMSE of each method are distributed among all 30 tested subjects, for the case of $\mathrm{SNR} = 0$ [dB]. The proposed VSDR achieved the

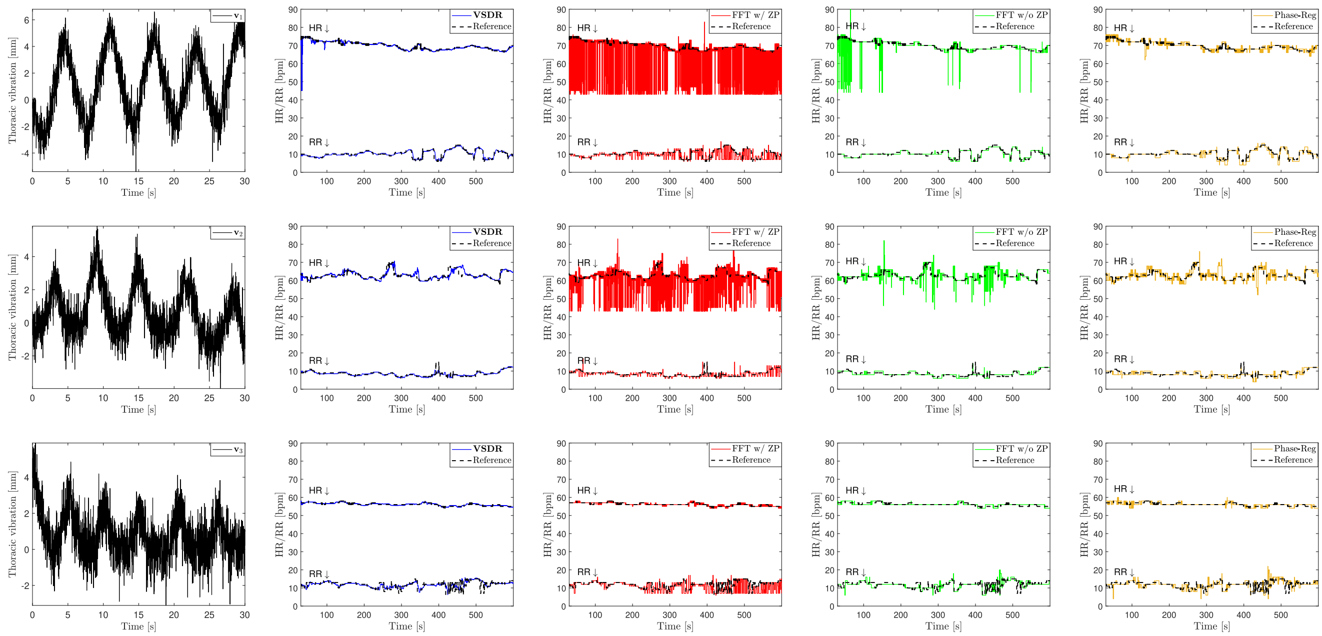


Fig. 6. Simultaneous NCVSM of 3 subjects given $\mathbf{V}$, w.r.t. known references, for $\mathrm{SNR} = 0$ [dB]. **Rows:** Subjects 4-6 [36] (trial #2). **Columns:** Extracted thoracic vibrations $\mathbf{v}_1$-$\mathbf{v}_3$ for some T_{int}, proposed VSDR (III-B5), FFT w/ ZP [23], FFT w/o ZP [19], [20], [21], [22], Phase-Reg [24].

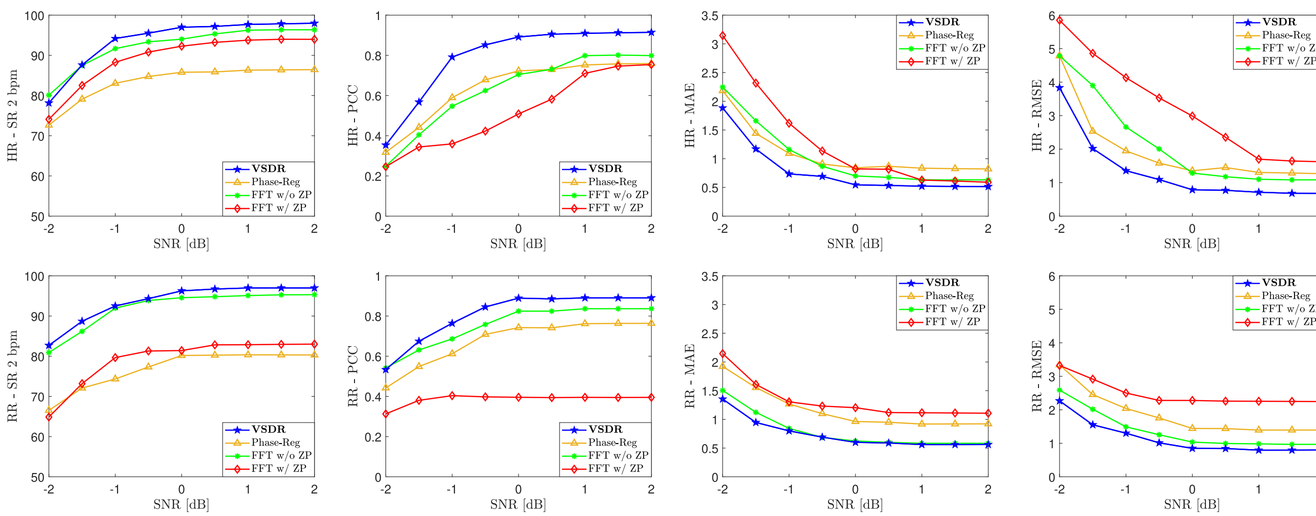


Fig. 7. Median NCVSM performance of the proposed VSDR (Section III-B5) compared to FFT w/ ZP [23], FFT w/o ZP [19], [20], [21], [22] and Phase-Reg [24], vs. SNR. **Rows:** HR, RR. **Columns:** SR - 2 bpm, PCC, MAE, RMSE.

best RMSE scores for the majority of the subjects, regardless of the type of vibration (and as a result of the subject's distance from the radar), while the deviations for the remaining subjects are not substantial, which justifies the VSDR's low median RMSE values, as depicted in Fig. 7. Similar findings are obtained for the remaining accuracy metrics, thus they are not plotted. Numerical median and mean accuracy scores for all 4 metrics, for the case of $\mathrm{SNR} = 0$ [dB] can be found in Table III. The table demonstrates the VSDR's preeminence even when a mean score is used, which is more sensitive to abnormal measurements.

The performance advantage is a consequence of the property that, unlike all other compared spectrum-based methods for NCVSM given a thoracic vibration, the proposed VSDR employs a frequency search over high-resolution grids corresponding to human-typical cardiopulmonary frequencies via a dictionary-based approach.

V. Conclusion

In this article, an extended SISO FMCW signal model for multi-person NCVSM was derived, allowing the interpretation of a realistic noisy environment containing multiple objects. By exploiting the sparse composition of the input data via this model, we presented a JSR approach that accurately localizes

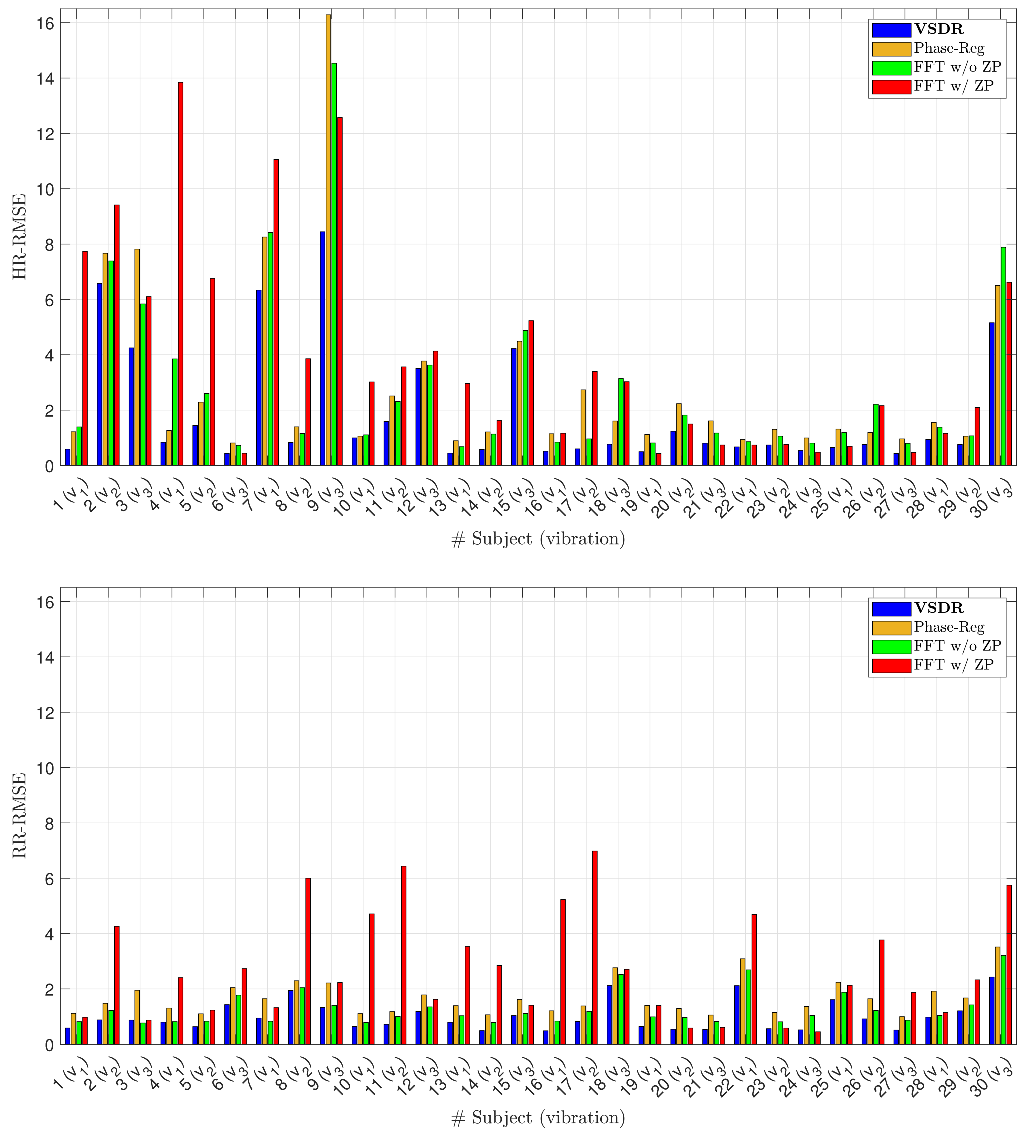


Fig. 8. HR-RMSE and RR-RMSE separated by subjects, of the proposed VSDR (Section III-B5), compared to FFT w/ ZP [23], FFT w/o ZP [19], [20], [21], [22] and Phase-Reg [24], for $\mathrm{SNR} = 0$ [dB].

humans in a clutter-rich scenario, where known localization techniques struggle. We then developed the VSDR method which performs accurate NCVSM given human thoracic vibrations, by leveraging human-typical cardiopulmonary characteristics using a dictionary-based approach. The robustness of the proposed VSDR is reflected in superior performance results using in vivo data of 30 monitored individuals, outperforming state-of-the-art alternatives using multiple statistical criteria.

Acknowledgment

The authors would like to thank Daniel Khodyrker for his assistance with the numerical study.

References

[1] G. Crossley, A. Boyle, H. Vitense, Y. Chang, R. Mead, and C. Investigators, "The CONNECT (clinical evaluation of remote notification to reduce time to clinical decision) trial: The value of wireless remote monitoring with automatic clinician alerts," *J. Amer. College Cardiol.*, vol. 57, no. 10, pp. 1181–1189, 2011.

[2] S. J. Brownsell, G. Williams, D. A. Bradley, R. Bragg, P. Catlin, and J. Carlier, "Future systems for remote health care," *J. Telemed. Telecare*, vol. 5, no. 3, pp. 141–152, 1999.

[3] O. Boric-Lubeke and V. M. Lubecke, "Wireless house calls: Using communications technology for health care and monitoring," *IEEE Microw. Mag.*, vol. 3, no. 3, pp. 43–48, Sep. 2002.

[4] V. Nangalia, D. Prytherch, and G. Smith, "Health technology assessment review: Remote monitoring of vital signs-current status and future challenges," *Crit. Care*, vol. 14, no. 5, pp. 1–8, 2010.

[5] J. C. Lin, "Applying telecommunication technology to health-care delivery," *Eng. Med. Biol. Mag.*, vol. 18, no. 4, pp. 28–31, 1999.

[6] P. Ceballos-Vásquez, G. Rolo-González, E. Hérnandez-Fernaud, D. Díaz-Cabrera, T. Paravic-Klijn, and M. Burgos-Moreno, "Psychosocial factors and mental work load: A reality perceived by nurses in intensive care units1," *Revista Latino-Americana de Enfermagem*, vol. 23, pp. 315–322, 2015.

[7] J. J. Hillhouse and C. M. Adler, "Investigating stress effect patterns in hospital staff nurses: Results of a cluster analysis," *Social Sci. Med.*, vol. 45, no. 12, pp. 1781–1788, 1997.

[8] E. R. Greenglass, R. J. Burke, and L. Fiksenbaum, "Workload and burnout in nurses," *J. Community Appl. Social Psychol.*, vol. 11, no. 3, pp. 211–215, 2001.

[9] A. A. Almenyan, A. Albuduh, and F. Al-Abbas, "Effect of nursing workload in intensive care units," *Cureus*, vol. 13, no. 1, 2021, Art. no. e12674.

[10] F. Fioranelli, J. L. Kernec, and S. A. Shah, "Radar for health care: Recognizing human activities and monitoring vital signs," *IEEE Potentials*, vol. 38, no. 4, pp. 16–23, Jul./Aug. 2019.

[11] Y. Xiao, J. Lin, O. Boric-Lubecke, and V. M. Lubecke, "A ka-band low power Doppler radar system for remote detection of cardiopulmonary motion," in *Proc. 27th Annu. Conf. IEEE Eng. Med. Biol.*, 2006, pp. 7151–7154.

[12] C. Gu et al., "Accurate respiration measurement using DC-coupled continuous-wave radar sensor for motion-adaptive cancer radiotherapy," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 11, pp. 3117–3123, Nov. 2012.

[13] C. Gu, Y. He, and J. Zhu, "Noncontact vital sensing with a miniaturized 2.4 GHz circularly polarized Doppler radar," *IEEE Sens. Lett.*, vol. 3, no. 7, pp. 1–4, Jul. 2019.

[14] H. Zhao et al., "A noncontact breathing disorder recognition system using 2.4-GHz digital-IF Doppler radar," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 1, pp. 208–217, Jan. 2019.

[15] B. K. Park, O. Boric-Lubecke, and V. M. Lubecke, "Arctangent demodulation with DC offset compensation in quadrature Doppler radar receiver systems," *IEEE Trans. Microw. Theory Techn.*, vol. 55, no. 5, pp. 1073–1079, May 2007.

[16] T. K. Vodai et al., "Enhancement of remote vital sign monitoring detection accuracy using multiple-input multiple-output 77 GHz FMCW radar," *IEEE J. Electromagn., RF, Microw. Med. Biol.*, vol. 6, no. 1, pp. 111–122, Mar. 2022.

[17] G. Wang, C. Gu, T. Inoue, and C. Li, "A hybrid FMCW-interferometry radar for indoor precise positioning and versatile life activity monitoring," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 11, pp. 2812–2822, Nov. 2014.

[18] A. Ahmad, J. Roh, D. Wang, and A. Dubey, "Vital signs monitoring of multiple people using a FMCW millimeter-wave sensor," in *Proc. IEEE Radar Conf.*, 2018, pp. 1450–1455.

[19] M. Alizadeh, G. Shaker, J. D. Almeida, P. Morita, and S. Safavi-Naeini, "Remote monitoring of human vital signs using mm-Wave FMCW radar," *IEEE Access*, vol. 7, pp. 54958–54968, 2019.

[20] M. Mercuri, I. Lorato, Y. H. Liu, F. Wieringa, C. V. Hoof, and T. Torfs, "Vital-sign monitoring and spatial tracking of multiple people using a noncontact radar-based sensor," *Nature Electron.*, vol. 2, no. 6, pp. 252–262, 2019.

[21] G. Sacco, E. Piuzzi, E. Pittella, and S. Pisa, "An FMCW radar for localization and vital signs measurement for different chest orientations," *Sensors*, vol. 20, no. 12, 2020, Art. no. 3489.

[22] E. Antolinos, F. García-Rial, C. Hernández, D. Montesano, J. I. Godino-Llorente, and J. Grajal, "Cardiopulmonary activity monitoring using millimeter wave radars," *Remote Sens.*, vol. 12, no. 14, 2020, Art. no. 2265.

[23] E. Turppa, J. Kortelainen, O. Antropov, and T. Kiuru, "Vital sign monitoring using FMCW radar in various sleeping scenarios," *Sensors*, vol. 20, no. 22, 2020, Art. no. 6505.

[24] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, 2015, pp. 837–846.

[25] H. Lee, B. H. Kim, J. K. Park, and J. G. Yook, "A novel vital-sign sensing algorithm for multiple subjects based on 24-GHz FMCW Doppler radar," *Remote Sens.*, vol. 11, no. 10, pp. 12–37, 2019.

[26] S. Kim and K. K. Lee, "Low-complexity joint extrapolation-MUSIC-based 2-D parameter estimator for vital FMCW radar," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2205–2216, Mar. 2019.

[27] Y. Wang, W. Wang, M. Zhou, A. Ren, and Z. Tian, "Remote monitoring of human vital signs based on 77-GHz mm-wave FMCW radar," *Sensors*, vol. 20, no. 10, 2020, Art. no. 2999.

[28] I. Roudas, M. Sauer, J. Hurley, Y. Mauro, and S. Raghavan, "Compensation of coherent DQPSK receiver imperfections," in *Dig. IEEE/LEOS Summer Topical Meetings*, 2007, pp. 19–20.

[29] R. Mahendra, S. K. Mohammed, and R. K. Mallik, "Compensation of receiver IQ imbalance in mm-wave hybrid beamforming systems," in *Proc. IEEE 92nd Veh. Technol. Conf.*, 2020, pp. 1–6.

[30] J. Tubbax, B. Côme, L. Van der Perre, L. Deneire, S. Donnay, and M. Engels, "Compensation of IQ imbalance in OFDM systems," in *Proc. IEEE Int. Conf. Commun.*, 2003, pp. 3403–3407.

[31] I. Fatadin, S. J. Savory, and D. Ives, "Compensation of quadrature imbalance in an optical QPSK coherent receiver," *IEEE Photon. Technol. Lett.*, vol. 20, no. 20, pp. 1733–1735, Oct. 2008.

[32] J. Wang, X. Wang, L. Chen, J. Huangfu, C. Li, and L. Ran, "Noncontact distance and amplitude-independent vibration measurement based on an extended DACM algorithm," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 1, pp. 145–153, Jan. 2014.

[33] U. Rossman et al., "Rapid quantum image scanning microscopy by joint sparse reconstruction," *Optica*, vol. 6, no. 10, pp. 1290–1296, 2019.

[34] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[35] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[36] S. Schellenberger et al., "A dataset of clinically recorded radar vital signs with synchronised reference sensor signals," *Sci. Data*, vol. 7, no. 1, pp. 1–11, 2020.

[37] "IWR1642 single-chip 76- to 81-GHz mmWave sensor (Rev. B)," [Online]. Available: https://www.ti.com/document-viewer/IWR1642/datasheet

[38] J. Liu, Y. Li, and C. Gu, "Radar-based vital signs monitoring," in *Contactless Vital Signs Monitoring*. Cambridge, MA, USA: Academic Press, 2022, ch. 9, pp. 181–203.

[39] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[40] D. P. Palomar and Y. C. Eldar, *Convex Optimization in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[41] D. Ruppert and M. P. Wand, "Multivariate locally weighted least squares regression," *The Ann. Statist.*, vol. 22, pp. 1346–1370, 1994.

[42] J. M. Ernst, D. A. Litvack, D. L. Lozano, J. T. Cacioppo, and G. G. Berntson, "Impedance pneumography: Noise as signal in impedance cardiography," *Psychophysiol.*, vol. 36, no. 3, pp. 333–338, 1999.

[43] A. Sherwood, M. T. Allen, J. Fahrenberg, R. M. Kelsey, W. R. Lovallo, and L. J. V. Doornen, "Methodological guidelines for impedance cardiography," *Psychophysiol.*, vol. 27, no. 1, pp. 1–23, 1990.

[44] R. A. Fisher, "Statistical methods for research workers," in *Breakthroughs in Statistics*. New York, NY, USA: Springer, 1992, pp. 66–70.