MoVi-Fi: Motion-robust Vital Signs Waveform Recovery via Deep Interpreted RF Sensing

Zhe Chen[1,2*], Tianyue Zheng[1*], Chao Cai[1], Jun Luo[1]
[1] School of Computer Science and Engineering, Nanyang Technological University, Singapore
[2] China-Singapore International Joint Research Institute, Guangzhou, China
Email: chenz@ssijri.com, {tianyue002, junluo}@ntu.edu.sg, chriscai@hust.edu.cn

ABSTRACT

Vital signs are crucial indicators for human health, and researchers are studying contact-free alternatives to existing wearable vital signs sensors. Unfortunately, most of these designs demand a subject human body to be relatively static, rendering them very inconvenient to adopt in practice where body movements occur frequently. In particular, *radio-frequency* (RF) based contact-free sensing can be severely affected by body movements that overwhelm vital signs. To this end, we introduce MoVi-Fi as a *motion-robust* vital signs monitoring system, capable of recovering fine-grained vital signs waveform in a contact-free manner. Being a pure software system, MoVi-Fi can be built on top of virtually any commercial-grade radars. What inspires our design is that RF reflections caused by vital signs, albeit weak, do not totally disappear but are composited with other motion-incurred reflections in a *nonlinear* manner. As nonlinear blind source separation is inherently hard, MoVi-Fi innovatively employs *deep contrastive learning* to tackle the problem; this self-supervised method requires no ground truth in training, and it exploits contrastive signal features to distinguish vital signs from body movements. Our experiments with 12 subjects and 80-hour data demonstrate that MoVi-Fi accurately recovers vital signs waveform under severe body movements.

CCS CONCEPTS

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

KEYWORDS

Motion-robust vital signs monitoring, commercial-grade radars, contact-free RF-sensing, deep contrastive learning.

ACM Reference Format:
Z. Chen, T. Zheng, C. Cai, and J. Luo. 2022. MoVi-Fi: Motion-robust Vital Signs Waveform Recovery via Deep Interpreted RF Sensing. In *The 27th Annual International Conference on Mobile Computing and Networking (ACM MobiCom'21), January 31-February 4, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3447993.3483251

* Both authors contributed equally to this research.

ACM MobiCom'21, January 31-February 4, 2022, New Orleans, LA, USA

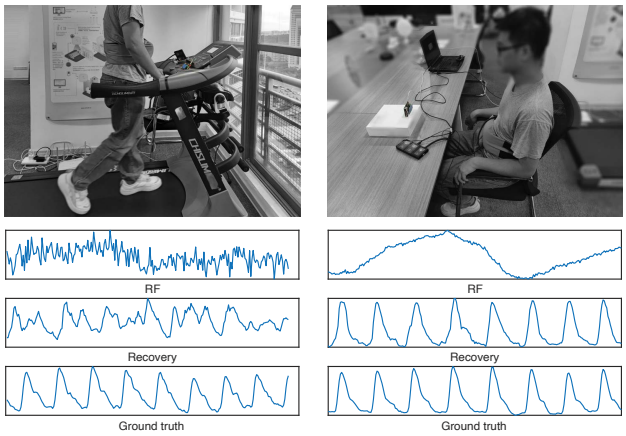
ACM ISBN 978-1-4503-8342-4/22/01…$15.00
https://doi.org/10.1145/3447993.3483251

1 INTRODUCTION

Vital signs (particularly heartbeat and breath) are representative indicators on human physical and mental status [6, 12], so novel applications such as sleep monitoring [53], fatigue detection [61], and emotion recognition [60] all rely on vital signs awareness. Approaches for monitoring vital signs can be roughly categorized into *contact* and *contact-free* sensing; the former often relies on wearable sensors (ranging from smartwear to medical devices) to detect micro-activities (either mechanical or electrical) of human bodies [57]. Unfortunately, the contact nature makes people uncomfortable and may even affect vital signs. Therefore, contact-free sensing has attracted increasing attention from both academia and industry [3, 26, 29, 36, 46, 48, 51–53, 55], in which *radio-frequency* (RF) sensing leveraging various commercial-grade radars has demonstrated a promising future [3, 17, 26, 29, 53, 61].

Although exploiting different media (e.g., light [10, 37, 55], RF [3, 26, 29, 36, 53, 61], and sound [46, 48, 52]), contact-free vital signs sensing shares a common technical basis: analyzing the reflected signals excited by the micro-activities of human bodies. Therefore, these approaches are subject to the same "curse" of *body movements*, simply because these high-energy motions may overwhelm the micro-activities caused by vital signs in signal space. Although RF is known to be more tolerable to background noise than other media, prior RF-sensor designs rely on customized yet sophisticated hardware [26, 32, 33, 44, 54], making it very hard for system developers to reproduce their results. Consequently, existing RF-based system developments mostly focus on refining the granularity of sensing outcome [3, 17, 29, 36, 53, 61], while discarding those contaminated by body movements and thus *leaving motion-robust monitoring as an open problem.*

To better understand the damaging effect of body movements on extracting vital signs, we compare two cases in Figure 1. In both cases, an IR-UWB radar [5] is placed in front of the subject, where the subject walks with a speed of 1 m/s on a treadmill in case (a) but sits statically in case (b). Even with the most up-to-date approach [17] to recover the heartbeat waveform, the motion-corrupted RF signal in case (a) cannot yield correct results compared with case (b). It is also worth noting that, whereas the breath waveform for case (b) is conspicuous even in raw RF signal, it is also corrupted by body movements in case (a). In reality, it is impractical to force a subject to remain static during monitoring for two major reasons. On one hand, we do need to monitor vital signs of a moving subject (e.g., walking-on-treadmill or typewriting); though sophisticated medical equipment can be used for this purpose [15, 43], we may not have access to such equipment in our daily life. On the other hand, a subject may move unconsciously (e.g., turning-over during sleep); existing approaches suspend vital signs monitoring



(a) Walking on a treadmill. (b) Sitting statically.

Figure 1: Heartbeat waveform recovery (a) with and (b) without body movements.

when such body movements are detected [3, 56]. Therefore, it is imperative to endow RF-sensing with *motion-robustness*, so as to deliver practical vital signs monitoring solutions.

In this paper, we present MoVi-Fi, a contact-free sensing system for Motion-robust Vital signs monitoring, with Fi indicating the outcome as fined-grained waveforms. We design MoVi-Fi as a pure software system readily deployable onto virtually all types of commercial-grade radars (including both IR-UWB and FMCW) adopted by existing research proposals [3, 17, 61]. Such a cross-technology design makes MoVi-Fi independent of any hardware features such as the center frequency, number of antennas, and sampling rate, as far as the bandwidth is sufficiently wide. The advantage of MoVi-Fi being cross-technology is evident: different RF technologies can be chosen to suit specific applications, for example, IR-UWB (at 7 GHz range) for extended sensing scope (e.g., through walls) and FMCW (at mmWave range) for finer-grained but close-by monitoring. Meanwhile, MoVi-Fi's motion-robustness is confined by the capability of RF technologies: it works only if a subject's front upper body lies within the sensing scope.

The major challenge faced by designing a motion-robust vital signs RF-sensing system is the complex composition among various motions in the reflected signal space. In particular, large-scale body movements and the micro-activities (e.g., neck vibrations) caused by vital signs are not composed in an additive manner. First, for limb movements (e.g., typewriting) not taking place at the spots of micro-activities, the compositions in signal space manifest in both amplitude and phase, which is apparently nonlinear. Second, torso movements (e.g., walking-on-treadmill) may change the positions of the micro-activity spots, causing the composition to be extended beyond the current range dimension. Last but not least, the reflected signals caused by body movements can exhibit various statistical properties (e.g., non-stationary or cyclostationary), which cannot be readily separable by a single type of algorithm.

To solve above challenges in a cross-technology manner, we take the following steps to gradually construct MoVi-Fi into a software framework driven by an end-to-end deep learning pipeline to recover vital signs waveform.

- We first perform a detailed analysis on the two types of mainstream radars, thus deriving a unified vital signs RF-sensing model for capturing the motion-excited reflection signals. In particular, we realize that, though bandwidth is essential to offer a fine sensing resolution, center frequency and antenna number do not seem to be crucial factors.
- In order to better understand the nature of body movements (hence their impacts on the signal space pertinent to vital signs), we conduct a study on several common instances. We first clarify the feasible scope of body movements, then we categorize these movements into three types, namely stationary (e.g., typewriting), cyclostationary (e.g., walking-on-treadmill), and non-stationary (e.g., standing-up/sitting-down), so that any complex movements can be formed as their combinations. We also study each of these types in terms of how they interfere with vital signs in the reflected signal space.
- It is clear from the earlier studies that the body movement type is a key to the whole signal separation process. To this end, we propose two novel *self-supervised contrastive-learning* algorithms to exploit the distinct patterns of various movement types. In particular, these contrastive-learning algorithms leverage the temporal and spatial diversities in the signal space to properly distinguish and thus separate the interference of body movements from the micro-activities excited by vital signs. Finally, we design an encoder-decoder module trained by a discriminator, in order to reproduce the fine-grained waveform of both heartbeat and breath.

We implement MoVi-Fi on top of three typical commercial-grade radars: i) Novelda's XeThru X4 at 7.29 GHz [5], ii) Infineon's Position2Go at 24 GHz [4], and iii) Texas Instruments (TI)'s IWR1443 at 77 GHz [23]. We evaluate MoVi-Fi on 12 subjects under 8 body movements, obtaining over 380,000 heartbeat and breath cycles. All these experiment results clearly demonstrate the accuracy and motion-robustness of MoVi-Fi in every scenario. In summary, we make the following major contributions in this paper:

- To the best of our knowledge, MoVi-Fi is the first RF-sensing system capable of recovering fine-grained vital signs waveform under major body movements.
- MoVi-Fi is designed as a pure software system readily deployable onto virtually any types of commercial-grade radars, so as to suit different application requirements.
- Inspired by a detailed investigation on common body movements, MoVi-Fi is equipped with a carefully engineered end-to-end deep learning pipeline; it contains novel contrastive-learning models to distill vital signs, and it also employs an encoder-decoder model to refine vital signs waveform.
- We implement MoVi-Fi as a software prototype and test it upon 3 popular radar platforms; the extensive evaluation results evidently demonstrate MoVi-Fi's motion-robustness in recovering fine-grained vital signs waveforms.

To avoid losing the focus on motion-robustness, we do not aim to extract details (e.g., cardiac cycle events) out of the recovered waveform. As MoVi-Fi has exhibit promising ability in recovering vital signs waveform under body movements, we leave the more detailed event extractions to an extended development that leverages existing proposals (e.g., [17]). In the following, we first survey

the literature in Section 2, then we provide a detailed exposition on constructing MoVi-Fi from scratch in Section 3. We explain implementation details in Section 4 and report performance evaluations in Section 5, before finally conclude our paper in Section 6.

2 RELATED WORK

As contact sensing either avoids motion interference by resorting to signals largely immune to such interference [15, 43] or applies conventional filtering to remove interference [27, 59], we focus on discussing contact-free vital signs monitoring, which has witnessed a substantial amount of developments in the past decade. Whereas RF is the mainstream sensing medium, light (hence computer vision) [10, 37, 55] and sound [46, 48, 52] are also adopted. Among all RF-sensing approaches, we differentiate between *sensor design* (e.g., [32, 33, 54]) and *system development* (e.g., [3, 17, 29, 36, 53, 61]): whereas the former focuses on developing radars and equip them with proper signal processing algorithms, the latter aims to build working prototypes based on commercial-grade RF platforms.

Sensor design community has a rather long history on studying motion-robust vital signs monitoring. Early proposals rely on tricky placements of two radars to handle the interference from body movements [28, 33, 44, 47]. However, these proposals are far from practical as a strict synchronization between the two radars are needed (yet hard to achieve), while those tricky placements can be very unrealistic to achieve in our daily environments. Later proposals have shifted their focus to signal processing techniques, in order to avoid hardware complications [32, 45, 54]. Tu *et al.* [45] extract only breath rate out of unrealistic 1-D body movements. Lv *et al.* [32] propose a matched filter to cope with body movements, but their method relies on a strong assumption on the existence of quasi-static periods during movements. Latest proposal [54] applies adaptive noise cancellation to handle body movements, yet their evaluations, without even specifying what body movements are involved, are highly questionable in validity.

Given the unsatisfactory progress from the sensor design community, system developers decide to build their own vital signs monitoring systems based on commercial-grade radars.[1] Most existing systems have been developed to estimate coarse-grained vital signs via RF signals [3, 25, 36, 53, 61]; they leverage either time or frequency analysis to estimate the breath or heart rate within a sliding time windows. Recent papers [14, 17, 29] have utilized FMCW radars to recover the fine-grained heartbeat waveform, relying on delicate signal processing and deep learning techniques, respectively. Although these systems have made a sound progress in refining the monitoring granularity towards even clinic-level applications, all of them require a subject to remain relatively static, i.e., they all lack motion-robustness.

Other contact-free methods are either vision-based or acoustic-based. Vision-based method for vital signs monitoring is also termed remote photoplethysmography (or rPPG); it always adopts a camera to capture video of a subject and then analyzes the subtle color changes on the facial regions of the subject to derive heart rate [10, 37, 55]. The acoustic-based method is convenient to monitor vital signs since smartphone can be used to readily produce and analyze acoustic signals [46, 48, 52]. While [46, 48] are only able to monitor respiration rate, [52] recovers the fine-grained breath waveform using deep learning. It is interesting to note the complimentary nature between vision and acoustic methods: the former works for heartbeat but the latter excels in breath. Unfortunately, both methods are highly susceptible to background interference and may sometimes incur the privacy concerns. Most importantly, they still cannot offer a full-scale motion-robustness.

[1]We neglect the literature on Wi-Fi based respiration sensing [1, 51, 58], as communication systems may not serve the long-term monitoring purpose.

3 DESIGNING MoVi-Fi

We design MoVi-Fi in three steps. We first unify the RF-sensing model in Section 3.1, so that MoVi-Fi can operate across different radar platforms. Then we study the common body movements in Section 3.2, mostly in terms of how they get composited with micro-activities in signal space. Finally, we elaborate the construction of MoVi-Fi to achieve motion-robust vital signs monitoring in Section 3.3.

3.1 Modeling RF Reflections

Equivalent Sensing Model for Radars. RF reflections are represented by the RF Channel Impulse Responses (CIRs) from transmitter to receiver, and *time delay*s are the main elements in CIRs. To understand how various motions affect CIRs, we first model the signal propagation distance that causes variations in time delays:

$$d(t) = \bar{d} + d^{\mathrm{b}}(t) + d^{\mathrm{r}}(t) + d^{\mathrm{h}}(t), \tag{1}$$

where $\bar{d}$ is the mean distance between a radar and a subject, while $d^{\mathrm{b}}(t)$, $d^{\mathrm{r}}(t)$, and $d^{\mathrm{h}}(t)$ are variations respectively caused by body movement, respiration, and heartbeat. The key idea to detect above distance variations with RF signals is to extract the amplitude and phase changes of CIR. Given a transmitted waveform $s(t)$, the received signal becomes:

$$y(t) = \alpha(t)e^{-j2\pi f_c \frac{2d(t)}{c}} s\left(t - \frac{2d(t)}{c}\right), \tag{2}$$

where the c represents the speed of radio wave, and f_c is the carrier frequency. Though there could be multiple reflection paths, radar-based RF-sensing considers only the direct path provided that no other objects block the subject [13, 29].

Although $s(t)$ differs in the two popular waveforms (i.e., IR-UWB and FMCW) adopted by radars, $y(t)$ has the same 2-D *CIR matrix* representation, with the two dimensions respectively denoted as *fast-time* and *slow-time* due to different time scales of sampling [62]. Basically, IR-UWB and FMCW is a time-frequency dual pair: while the former leverages time-domain pulse positions (in fast-time dimension) to indicate distances, the latter exploits frequency-domain pulses (obtained via FFT) to reach the same goal. As the fast-time samples indicate distances, they are often termed *range bin*s (or *bin*s for short). Consequently, both radars essentially have the same distance resolution determined by $\frac{c}{2B}$ with B denoting the bandwidth (i.e., how narrow a pulse is). For a certain bin, multiple samples are acquired from consecutively received frames $y_i(t), i = 0, 1, \cdots$. As the frames (resp. samples) are transmitted (resp. acquired) at a much lower rate (e.g., 512Hz), they form the slow-time dimension.

Spatial Diversity is Limited. Whereas IR-UWB often has only one antenna pair, FMCW is commonly equipped with an antenna

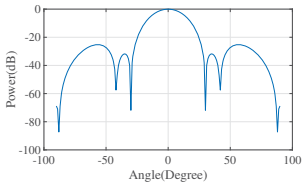


Figure 2: Beamforming spectrum of TI IWR1443.

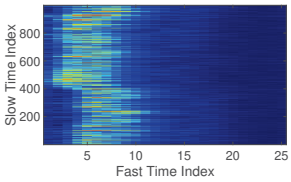


Figure 3: Swaying body from left to right.

array. For example, TI's FMCW radar has 3 tx and 4 rx antennas, forming 12 tx-rx pairs. However, our experiment results in Figure 2 show that the beamforming effect of all these antenna pairs leads to an equivalent 3dB-beamwidth of at least 15°. Considering $\bar{d} = 1\ m$ between the radar and a subject, the field of view (FoV) is already about 30 cm (i.e., half of the body width). Therefore, the diversity provided by these antenna pairs barely offers a sharp focus on only the micro-activities excited by vital signs. In addition, our experiment results in Section 5.2.5 also show that a narrow FoV may not favor motion-robustness. It is true that the beamforming capability can be very useful in monitoring multiple subjects. Nonetheless, as we focus on motion-robust monitoring for a single subject in this paper, we deem the 2-D CIR matrix as the essential input for deriving vital signs, but we make use of the antenna arrays whenever available via the beamforming scheme proposed in [17].

Motion Impact on Signal Space. Since $d^{\mathrm{r}}(t)$ and $d^{\mathrm{h}}(t)$ are both periodic with low frequencies, they cannot be captured by the sequence of bins, but are rather "hidden" in the slow-time dimension of certain bins and represented by signal phase changes [3, 17, 61]. Their impact on the amplitude is often neglected due to their minor scales compared with $\bar{d}$. Different from vital signs incurring $d^{\mathrm{r}}(t)$ and $d^{\mathrm{h}}(t)$, body movements affect both fast- and slow-time samples via $d^{\mathrm{b}}(t)$ with a wider bandwidth, and they also alter the signal amplitude $\alpha(t)$. Figure 3 shows that a subject swaying-body causes the CIR matrix to change in all aforementioned three dimensions. Apparently, the impact of body movements on the CIR matrix is much more complicated than that imposed by vital signs, which significantly handicaps the existing techniques in extracting vital signs waveform. As an example, we monitor vital signs of a subject under both static and dynamic (i.e., playing smartphone games) situations, using an extended version of RF-SCG [17] and two different radars. Figure 4 shows the average *relative errors* (see Section 5.2.1 for the detailed definition of this quantity) of the four cases; the performance is clearly much worse in the dynamic case.

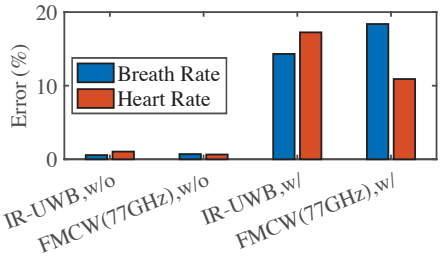


Figure 4: Estimating breath and heart rates with (w/) and without (w/o) body movements.

3.2 Understanding Body Movements

We first use Figure 5 to state the scope of body movements. Essentially, we require that i) the subject does not turn back to the radar, ii) the center of gravity of the subject's body is confined within the FoV of a radar, and iii) the distance $d(t)$ between the radar and the subject lies within a reasonable range (e.g., ±30cm) around its mean $\bar{d}$. While the first two conditions forbid the subject to drastically change his/her posture (e.g., from standing to lying), the last one prevents body movements from significantly altering the subject's position, as otherwise the radar has to keep track of the subject. Under these requirements, we can categorize the common body movements into three types: *stationary*, *cyclostationary*, and *non-stationary*. Note that the micro-activities incurred by vital signs are cyclostationary with very low strength, and an arbitrary body movement can be formed by a certain combination of these types.

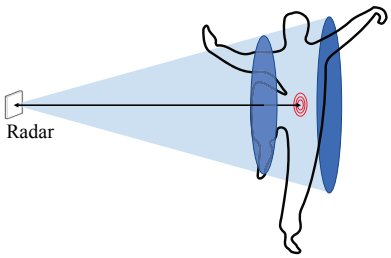
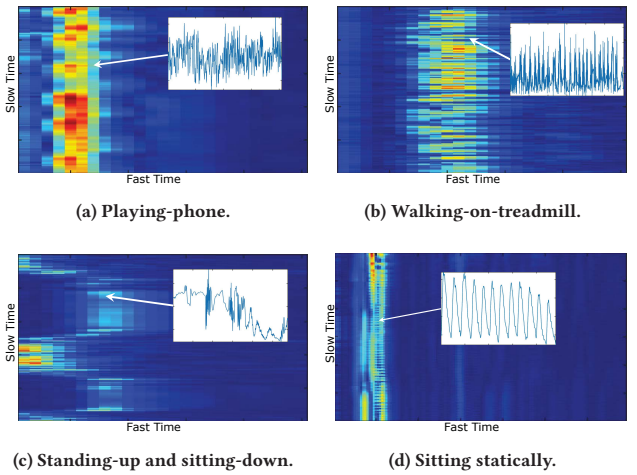


Figure 5: Motion scope in terms of the radar FoV.

Figure 6 uses the CIR matrices (heatmaps) to illustrate the four types of body activities, with the small top-right figures aggregating the slow-time variations over a set of hot-zone bins. We first use playing-phone (games) in Figure 6a as a representative for the stationary type (similar to but not Gaussian noise). This type may also include typewriting and leg shaking. The cyclostationary type



(a) Playing-phone. **(b) Walking-on-treadmill.**

(c) Standing-up and sitting-down. **(d) Sitting statically.**

Figure 6: Four common body activity types: (a) stationary (e.g., playing-phone), (b) cyclostationary (e.g., walking-on-treadmill), (c) non-stationary (e.g., standing-up and sitting-down), and (d) sitting statically with only cyclostationary vital signs waveform.

is exemplified by walking-on-treadmill in Figure 6b, which typically includes various physical exercises (e.g., swaying-body) with a rhythm. Standing-up/sitting-down suddenly, as demonstrated in Figure 6c, are certainly non-stationary due to their burstiness. Another typical non-stationary case is turning-over during sleep, a long-term obstacle for over-night vital signs monitoring [30, 56]. Whereas vital signs are "buried" under above major body movements, they can be clearly visible in Figure 6d when the subject remains relatively static.

We further measure the heart rates under six typical body movements using a template matching method [17]. As shown in Figure 7, all body movements cause large errors in measuring heart rates. While the motion strength seems to contribute to the magnitude of errors, the source of the errors is unclear given that vital signs are largely independent of body movements: if they were superposed linearly, the template matching method should be able to extract vital signals properly. To verify if the *linearity* holds, we take the ground truths of the vital signs waveform recorded simultaneously with CIR matrices under body movements using a wearable sensor [34]. We then leverage Singular Value Decomposition to remove noises and obtain the reference waveform. If the superposition in the CIR matrices between body movements and vital signs were linear, the bin vectors should have high correlation coefficients with the reference waveform. We calculate all correlation coefficients under body movements and normalize them against their respective static correlation coefficients; the results shown in Figure 8, with certain coefficients close to 0.1, unfortunately prove that the composition is far from linear.

Without linear composition, existing solutions (all with related assumptions) are bounded to fail, and we need to look for nonlinear separation schemes for eliminating the impact of body movements. Fortunately, such a separation is intuitively plausible for a few reasons, even for body movements sharing the same cyclostationary

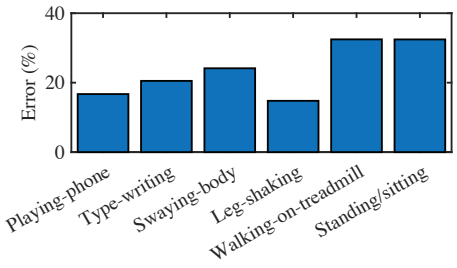


Figure 7: The average relative errors of heart rate under different body movements.

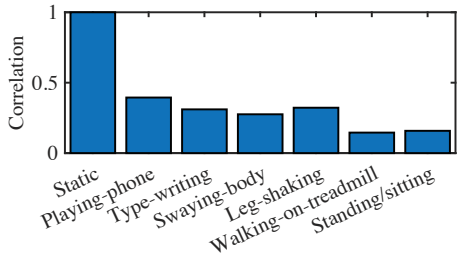


Figure 8: The normalized correlation coefficients of vital signs waveform under different body movements.

nature with vital signs. First of all, they may have different features, such as distinct frequencies, phases, and/or different level of randomness (e.g., heartbeat is definitely more regular than body movements). Secondly, both breath and heartbeat have special time-domain patterns that certainly do not appear in body movements. Last but not least, as different motions take place at slightly different locations, the high range resolution provided by wideband radars may capture this subtlety.

3.3 Constructing MoVi-Fi

According to earlier analysis, this section constructs MoVi-Fi with the following key components.

- **Data Preprocessing**: Given a CIR matrix formed in Section 3.1 as the raw input from an arbitrary radar, MoVi-Fi further adjusts it and also determines the type of the interfering body movements.
- **Separating Stationary Motions**: A *deep contrastive learning* approach is exploited to compare the original time sequences with their randomized versions; it essentially leverages the nonlinear mapping ability of a neural feature extractor to reverse the nonlinear composition between body movements and vital signs.
- **Separating Non-stationary Motions**: For cases dominated by bursty motions, a different contrastive method is adopted to discriminate between distinct time segments of the same sequences; it only removes the bursty motions, but has to rely on the previous module to separate heartbeat from breath.
- **Fine-grained Waveform Recovery**: After largely suppressing the impact of body movements, certain residual noises may still persist. This final component applies an encoder-decoder module to refine and merge the resulting time sequences, in order to recover the vital signs waveform.

3.3.1 Data Preprocessing. According to Section 3.1, RF sensing data can be represented as a CIR matrix $[y_1(t), \cdots y_i(t), \cdots, y_m(t)]^T$ where t indexes the (fast-time) bin and i is also a temporal index but for the slow-time dimension. In order to facilitate further separation, MoVi-Fi reforms this matrix so that it deems a set of bins indexed by $j \in \{1, 2, \cdots\}$ as *observations*, with each observation containing a time sequence $y_j(t)$ where t now becomes the temporal index for slow-time. Essentially, the CIR matrix is transposed to become $\boldsymbol{y}(t) = [y_1(t), \cdots y_j(t), \cdots y_n(t)]^T$, as shown in Figure 9 (left). The width of $\boldsymbol{y}(t)$ is termed *processing window* (taken as 20 seconds in our current implementation). Basically, while the index t keeps increasing during a continuous monitoring session, MoVi-Fi applies a sliding window so that samples are processed in segments. Although all bins are potential observations, MoVi-Fi only takes n (row) of them within a hot-zone for the sake of efficiency; we leave the details on recognizing the hot-zone to Section 4. Consequently, we hereafter denote the hot-zone in a CIR matrix by $\boldsymbol{y}(t)$.

The next step is to determine the nature of the body movements represented by $\boldsymbol{y}(t)$. Essentially, both stationary and cyclostationary types are treated as stationary, i.e., having time dependent (especially periodic) but latent features. MoVi-Fi differentiates stationary and non-stationary cases by employing autocorrelation, so as to handle them differently. The rationale is that the autocorrelation

curves of stationary signals barely decay, but the decay is significant for those from non-stationary signals. We empirically set a decay threshold to construct a hypothesis test for this sake.

3.3.2 Separating Stationary Motions. According to our analysis in Section 3.2, vital signs are very hard to detect under body movements, because their composition in $\boldsymbol{y}(t)$ is highly nonlinear. As a result, commonly used source separation algorithms, in particular *independent component analysis* (ICA), fail to work in this case, mainly due to their assumption of linear composition. Essentially, given a set of source signals (including those excited by vital signs and body movements) $\boldsymbol{x}(t) = [x_1(t), \cdots x_\ell(t), \cdots x_m(t)]^T$, the function $\mathbf{f} : \boldsymbol{y}(t) = \mathbf{f}(\boldsymbol{x}(t))$ can be highly complex and nonlinear. As noted by [20], trying to reverse $\mathbf{f}$ may lead to an infinite number of solutions and is hence an ill-posed problem.

Fortunately, recent progresses in deep learning indicate that, under certain mild conditions, reversing $\mathbf{f}$ is possible with the help of a *deep contrastive learning* [9, 19]. In particular, with conditions i) independent sources and ii) time dependent features (i.e., different manifestations of cyclostationarity in our case), contrastive learning aims to train an approximation of $\mathbf{f}^{-1}$ by maximizing the difference between observations and certain purposefully constructed *contrastive instance*. As illustrated in Figure 9, the workflow of MoVi-Fi starts with a time-dependent extension of the observations: $\boldsymbol{z}(t) = \begin{bmatrix} \boldsymbol{y}(t) \\ \boldsymbol{y}(t-\tau) \end{bmatrix}$, where τ is a time-translation to be specified in Section 4. Then a contrastive instance set is constructed with the original observations in the top half and the bottom half being a random permutation of the top half: $\boldsymbol{z}^\star(t) = \begin{bmatrix} \boldsymbol{y}(t) \\ \boldsymbol{y}(t^*) \end{bmatrix}$, where t^* represents a random time index. This contrastive set has the same marginal distribution as $\boldsymbol{z}(t)$, but its temporal structure is heavily corrupted in the bottom half (the lower-right part of the input in Figure 9). The rationale is that, by drawing sample (column vector) pairs from $\boldsymbol{z}(t)$ and $\boldsymbol{z}^\star(t)$ and contrasting them via a cross-correlation-like metric applied to the two halves of each sample, recognizing the vital signs waveform "hidden" in $\boldsymbol{y}(t)$ becomes possible. Of course, simple cross-correlation fails to work due to the nonlinearity in signal composition (see Figure 8), so we resort to a deep neural network to achieve the separation.

Given $\boldsymbol{z}^\star(t)$ and $\boldsymbol{z}(t)$ as two contrastive datasets with natural labels (i.e., $\star$ or not), we train a multilayer perceptron (MLP) model $g \circ \mathbf{h}(\cdot)$ to discriminate which dataset an arbitrary sample $\tilde{z}_{\hat{t}}$ comes from. Here $\tilde{z}_{\hat{t}}$ is a column vector sampled from either $\boldsymbol{z}^\star(t)$ or $\boldsymbol{z}(t)$ at a random time index $\hat{t}$, and $g(\cdot)$ performs a binary classification by minimizing a cross-entropy loss, relying on the output of the feature extractor $\mathbf{h}(\cdot)$. Intuitively speaking, as a sample from $\boldsymbol{z}(t)$ has its two halves highly correlated but one from $\boldsymbol{z}^\star(t)$ totally loses such a temporal structure, a successfully trained $\mathbf{h}(\cdot)$ has to reproduce the temporal structure for a sample drawn in $\boldsymbol{z}(t)$, in order to effectively distinguish it from samples drawn in $\boldsymbol{z}^\star(t)$. As the most compact characterization of this temporal structure boils down to separating the original signals and recovering their respective temporal features, it is plausible to deduce that the discrimination by g works best when $\mathbf{h}$ separates the source signals. Because two of these reproduced cyclostationary features are introduced by vital signs, $\mathbf{h}$ may well approximate $\mathbf{f}^{-1}$ that maps $\boldsymbol{y}(t)$ back to $\boldsymbol{x}(t)$, or $\boldsymbol{x}'(t)$ as a scaled version of $\boldsymbol{x}(t)$.

Training this contrastive model is totally self-supervised without ground truth vital signs waveform provided by wearable sensors; this is particularly important as otherwise acquiring training data would be much more difficult. Also, whereas contrastive learning for visual interpretation [9] may demand excessive training data, our approach is far more frugal as it handles only 1-D time sequences. After sufficiently training the model $g \circ \mathbf{h}(\cdot)$ (see Section 4 for training details), it can take an observation $\boldsymbol{y}(t)$ and let $\mathbf{h}(\cdot)$ directly output the decomposed $\boldsymbol{x}(t)$. However, several issues remain: i) what about non-stationary body movements? ii) although we are sure that the vector function $\mathbf{h}(\cdot)$ produces the decomposed $\boldsymbol{x}(t)$, we have no idea which neurons output the vital signs waveform (similar situation happens to ICA in linear cases), and iii) as MoVi-Fi is meant for continuous vital signs monitoring, yet the hot-zone $\boldsymbol{y}(t)$ is taken within a processing window, then how to merge the consecutive pieces of decomposed vital signs waveforms? We handle these issues in the following subsections.

3.3.3 Separating Non-stationary Motions. Non-stationary body movements create bursty temporal patterns, so the temporal correlation structure leveraged in Section 3.3.2 is not applicable anymore. Nonetheless, it is indeed this temporal burstiness that allows us to adapt a different deep contrastive scheme [18] to tackle it in a relatively straightforward manner. The essence of contrastive learning is to properly create contrastive training instances somehow characterizing signal features. For stationary signal sources in

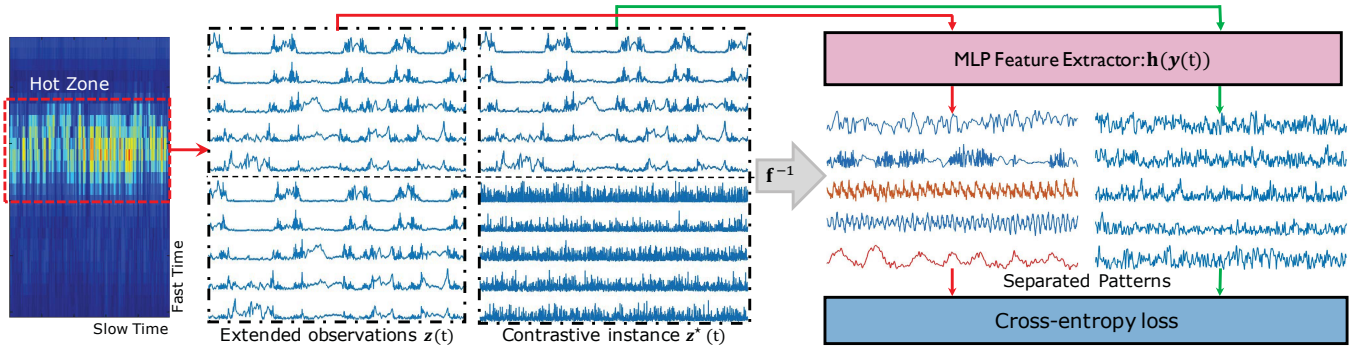


Figure 9: The workflow of separating walking-on-treadmill from vital signs: the resulting heartbeat and breath waveforms marked in distinctive colors are evidently periodic and close to our common-sense perception.

Section 3.3.2, this contrastivity has to be artificially produced. Fortunately, non-stationary sources are by default temporarily varying (hence having time contrastivity). In other words, we can leverage the temporal independent structure created by non-stationary motions to generate natural contrastive datasets without the need for data augmentations. Figure 10 shows the partial workflow of separating non-stationary motions, omitting the procedure of separating heartbeat from breath using $\mathbf{h}$ obtained in Section 3.3.2. Basically, we first divide the observations $\boldsymbol{y}(t) : t = 1, 2, \cdots, T$ into K segments with an equal length; let $\boldsymbol{y}^k$ denotes the k-th segment and k also serves as the natural class label for this segment; i.e., $\boldsymbol{y}^k$ is both the k-th dataset and the k-th class. We now train another neural model $g \circ \mathbf{h}'(\cdot)$ to correctly classify a sample $\tilde{y}_{\hat{t}}$ (a column vector drawn from $\boldsymbol{y}(t)$ at a random index $\hat{t}$) into its own segment. Consequently, the output of the feature extractor $\mathbf{h}'(\cdot)$, given $\boldsymbol{y}(t)$ as the input, should be a linear combination of the source signals $\boldsymbol{x}(t)$, as proven in [18]. In fact, as vital signs are statistically very different from the bursty body movements, the vital signs waveform are still combined in the output of $\mathbf{h}'(\cdot)$ and hence need to be further separated by $\mathbf{h}(\cdot)$ introduced in Section 3.3.2. The structure of $g \circ \mathbf{h}'(\cdot)$ is simpler due to a low demand in capacity.

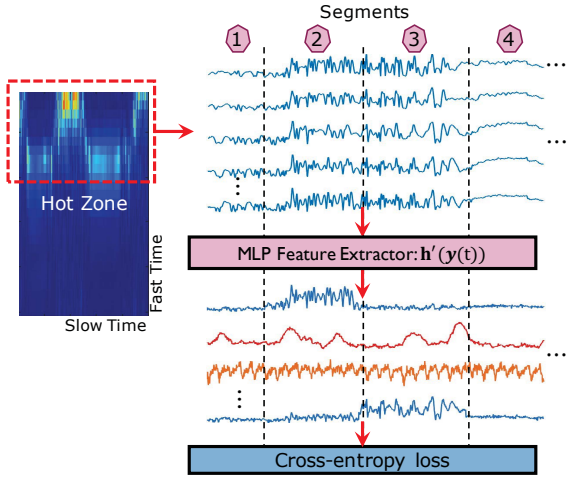


Figure 10: The partial workflow of separating standing-up/sitting-down from vital signs.

3.3.4 Fine-grained Vital Sign Waveform Estimation. MoVi-Fi aims to recover fine-grained vital signs waveform applicable to clinic-level applications, and it should do so continuously under body movements. Although the above algorithms separate body movements and vital signs in each processing window, two problems still require further attention, i.e., select correct waveforms from the output of $\mathbf{h}(\cdot)$, and merge consecutive pieces of waveforms.

Since the reconstructed waveforms of breath and heartbeat exhibit a much higher periodicity than others, we apply FFT on each waveform output by $\mathbf{h}(\cdot)$ and calculate the ratio between the peak power and the residue. Then a hypothesis test is performed on the ratio using an empirically set threshold to select vital signs waveform. In practice, though the selected waveforms contain sufficient features of vital signs, their shapes may still need to be refined to assimilate vital signs waveform under static situation. To tackle this

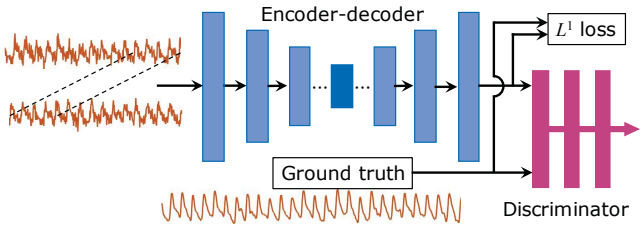


Figure 11: VS-Net for merging and refining vital signs waveform, with an encoder-decoder architecture.

problem while merging waveforms from consecutive processing windows, we adopt an encoder-decoder model shown in Figure 11 to regenerate fine-grained vital signs waveform.

The waveform recovery is conducted for heartbeat and breath separately, but using the same neural model. Each time two consecutive (yet partially overlapped, see Section 3.3.5) waveforms are taken as input, and the encoder-decoder model reproduces a continuous waveform at the output. We have compared the conventional encoder-decoder architecture against another one with skip connections between mirrored layers (i.e., U-Net [40]); it appears that the conventional one is already sufficient. Therefore, our VS-Net adopts the basic encoder-decoder model. However, commonly used loss functions for comparing the model output with ground truth are often based on L^1 or L^2 norms; they generally lead to smoothed waveforms that may lose details. Therefore, we learn from the patchGAN discriminator [24] to run a sliding-window convolutionally across the two waveforms and take the aggregation of all responses to produce the discrimination.

3.3.5 Summary. We hereby summarize the entire pipeline of MoVi-Fi. Initially, the (raw) input data from a radar are regulated to the uniform format discussed in Section 3.1. To strike a balance between resolution and latency, a 20-second sliding (processing) window is taken with 25% overlap between consecutive ones to sample the continuous data stream into a sequence of $\boldsymbol{y}(t)$. MoVi-Fi then determines the nature of the body movements, and it sends $\boldsymbol{y}(t)$ to respective separation procedures (i.e., $\mathbf{h}$ or $\mathbf{h} \circ \mathbf{h}'$) explained in Sections 3.3.2 and 3.3.3. The distilled waveforms are in turn fed to VS-Net (Section 3.3.4) to perform merging and refining, in order to finally produce fine-grained vital signs waveform.

4 IMPLEMENTATION

Hardware Implementations. Our MoVi-Fi prototype is built on three typical radar platforms. First, Novelda's IR-UWB radar X4M05 [5] operates at 7.3 or 8.7 GHz with 1.5 GHz bandwidth; it has a pair of tx-rx antennas with an FoV of 65° in both azimuth and elevation angles. Second, Infineon's FMCW radar Position2Go [4] works at 24 GHz with 200 MHz bandwidth; it has 1 tx antenna and 2 rx antennas; each has a 76° azimuth and 19° elevation FoV. Third, TI's FMCW radar IWR1443BOOST [23] works at 77 GHz with at most 4 GHz bandwidth; it has 3 tx antennas and 4 rx antennas, each with a 56° azimuth and 28° elevation FoV. For unifying the data format of $\boldsymbol{y}(t)$, all radar outputs are sampled at 512 Hz along the slow-time dimension for each range bin. Whereas the TI radar requires a special DCA1000 module [22] to capture real-time data, we develop

a driver on Raspberry Pi to interface the other two radars using C/C++; it receives data and feeds them to a PC with an i9 CPU, 16GB DDR4 RAM, and a GeForce RTX 2070 graphics card.

Software Implementations. We implement MoVi-Fi based on C/C++ and Python 3.7, with the neural network components built upon TensorFlow 2.0 [16]. To align the starting time of ground truth and $\boldsymbol{y}(t)$ to a μs level, we use Ethernet to synchronize the clocks between hardware components (i.e., wearable devices and radars) based on Precision Time Protocol [21]. To achieve cross-technology data pre-processing (given diversified sensing packets delivered from different radars), we implement an abstraction signal data interface before the deep learning pipeline to mask this diversity. The following are more implementation details promised earlier.

- The hot-zone $\boldsymbol{y}(t)$ (i.e., the number n of observations) is recognized by *constant false alarm rate* (CFAR) [2], a common algorithm for radars to detect subjects against noise and interference.
- The time-translation τ for separating stationary motions is set to be between one sample interval and 1.5 seconds, so as to retain time-dependent features (e.g., within one breath cycle or a few heartbeat cycles).
- For separating non-stationary motions, the segment length T/K is taken as 5 seconds, because it likely contains one breath cycles (and hence several heartbeat cycles).
- Both $\mathbf{h}$ and $\mathbf{h}'$ adopt an MLP model (6 layers for $\mathbf{h}$ and 5 layers for $\mathbf{h}'$), with leaky ReLU as the activation function after every layer to add nonlinearity. Another two-layers MLP serves as the classifier g. We set the batch size to 512 for training, and use Stochastic Gradient Descent optimizer with learning rate, momentum, decay step, and decay factor set to 0.001, 0.9, 5e5, and 0.999, respectively.
- Although both $\mathbf{h}$ and $\mathbf{h}'$ are, in principle, sufficiently trained (offline) with several representative body movements, re-training them for adapting to new movement types can be efficiently conducted given their self-supervised nature.
- An encoder layer of VS-Net adopts three CNN kernels of sizes 3×3, 7×7, and 11×11 in a parallel manner to deliver a multi-resolution ability. The layer properties are: stride 1, padding 0, and dilation 1. All outputs with different kernel sizes are concatenated to fed to the subsequent maxpooling layer with a kernel size 2. A decoder layer uses the same kernels as the encoder but connects them sequentially. The discriminator is composed of three convolutional layers with an input size matches the waveform length. We adopt an L^1 loss for waveform matching and a logistic loss for binary classification during training. The batch size is set to 64, and an Adam optimizer with a learning rate of 0.001 is used.

5 PERFORMANCE EVALUATION

In this section, we report a thorough evaluation on MoVi-Fi in several scenarios and under various parameter settings.

5.1 Experiment Setup

To evaluate MoVi-Fi, we recruit 12 subjects (6 women and 6 men), with ages between 15 and 64 and weights in the range of 50 to 80kg. All subjects are healthy, and they are monitored under their natural states without any forceful influence; our experiments have essentially followed the IRB protocol of our institute. We ask the subjects to perform 8 common human body movements: playing-phone, typewriting, swaying-body, leg-shaking, walking-on-treadmill, exercising, standing-up/sitting-down, and turning-over (during sleep), as well as 1 quasi-static sitting posture, all in our daily life environments such as gym, meeting room, and bedroom; see Figures 1 and 12 for four examples on the test sites.

(a) Exercising. **(b) Sleeping.**

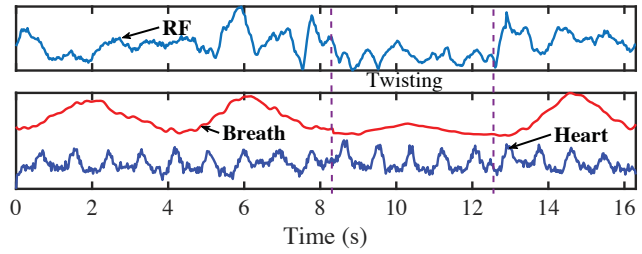
Figure 12: Another two experiment scenarios.

We use wearable devices NeuLog [34, 35] to collect ground truth in all scenarios, with breath sensed by a chest strap, and heartbeat explained later. The RF sensing radars are placed within a range of 0.5 to 2m from a subject (exact range may vary for individual trials). We conduct data collections with different time spans, but guarantee roughly the same total time for each subject. These include minute-long tests (e.g., walking-on-treadmill), hour-long observations (e.g., typewriting), and over-night monitoring (e.g., sleeping with turning-overs). All these amount to a 80-hour dataset of RF and ground truth recordings, including about 330k heartbeat cycles and 68k breath cycles. We first collect 30% of these data involving only 2 women, 2 men, and 3 body movements (typewriting, swaying-body, and standing-up/sitting-down) for training the deep learning modules, then the remaining 70% are collected (involving all subjects and movements) with MoVi-Fi operating in parallel to recover vital signs waveform online (leveraging the trained modules).

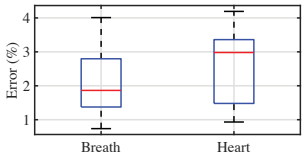
Given radars' ability in sensing 1-D motions/vibrations, we believe that the heartbeat waveform captured by the slow-time dimension of CIR matrix $\boldsymbol{y}(t)$ is defined by the clinical term *blood volume pulse* (BVP): it represents the volume changes in blood passing through a certain blood vessel, apparently driven by heartbeat [31, 38]. Therefore, the ground truth sensor [34] is chosen to measure *photoplethysmography* (PPG) via earlobe or finger tip, the most commonly used wearable sensing method to obtain BVP. Though we are the first to study motion-robust RF vital signs monitoring, we still establish a **baseline** for comparison; it applies the RF-SCG method [17] to recover BVP waveform. We refrain from considering acoustic sensing methods [42] as baselines, because they cannot handle heartbeat monitoring.

5.2 MoVi-Fi Performance

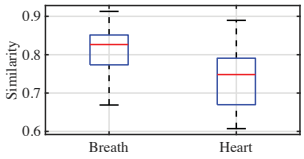
5.2.1 Micro-benchmarking. We hereby evaluate the signal separation performance of MoVi-Fi with its contrastive learning modules, taking IR-UWB (X4M05) as the radar sensor. We first use Figure 13a to illustrate the outcome of signal separation given exercising as the body movements and **RF** referring to a typical bin slice in $\boldsymbol{y}(t)$, then we report the overall quality of extracted vital signs waveforms in the other two subfigures of Figure 13. For the latter, we use *relative*



(a) Signal separation outcome in terms of breath and heartbeat waveforms.



(b) Rates.



(c) Similarity.

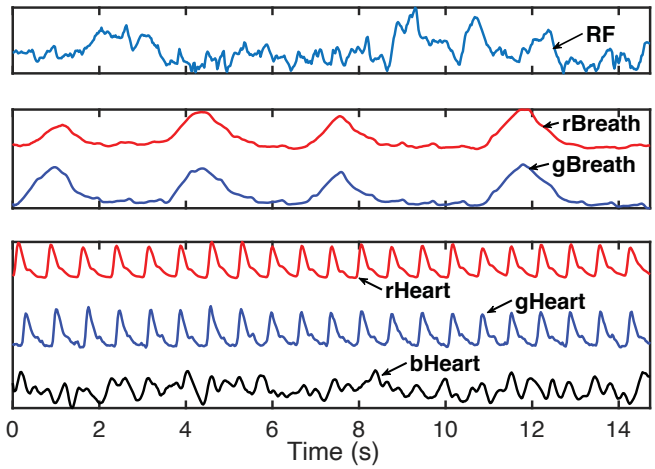
Figure 13: Waveform examples (a), and statistics on rate errors (b) and waveform similarities (c) for both vital signs.

error, defined as the ratio between absolute error and the ground truth rates, to indicate the accuracy of estimating breath and heart rates. We compute the respective rates by using an FFT and identify the peak frequency within the known range around 0.2Hz for breath and 1.2Hz for heart. While this metric only characterizes accuracy in event-level (i.e., average peak interval), we further use *cosine similarity* to measure the similarity between recovered waveforms and their corresponding ground truth waveforms. These metrics shall be adopted throughout the performance evaluation.

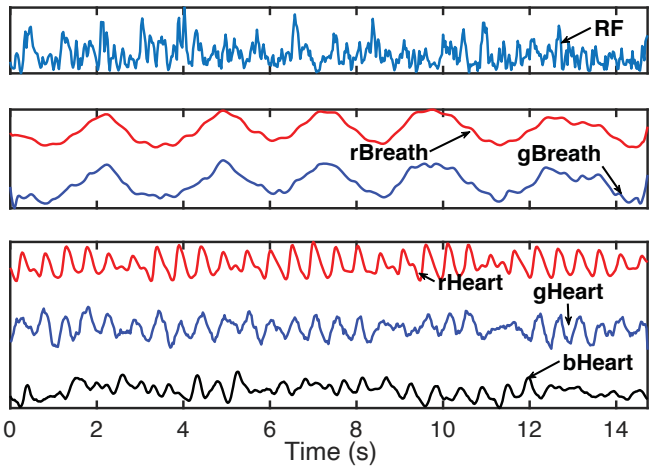
The example shown in Figure 13a is relevant as it demonstrates that breath can become less "visible" to Doppler-based RF-sensing: the exercise practiced by our subjects [8] involves twisting upper-body around and hence causes the radar to face body sides sometimes; we specifically let the subject slow down to give a clear view of this phenomenon. Our experience indicates that, while breath is sensed via chest vibrations, it is highly probable that common carotid arteries (through human neck) cause the most conspicuous BVP to be detected by radars. Therefore, breath signs may sometimes disappear but heartbeat signs always remain unless a subject turns back to the radar. The overall performance of the extracted waveforms can be deemed as satisfactory, because they almost perfectly capture the correct rates (Figure 13b), but the shape of these waveforms can still be rough due to the residual noise (Figure 13c). Therefore, further polishing via VS-Net is needed.

5.2.2 Waveform Recovery. We then evaluate MoVi-Fi's performance in recovering vital signs waveform under all 8 body movements, again using the IR-UWB radar. Similarly, we first provide two examples, before reporting overall statistics, in Figure 14, where the prefix **r**, **g**, and **b** denote MoVi-Fi's recovery results, ground truth, and baseline, respectively. All curves are re-scaled to fit their frames for clear exposition. The results evidently demonstrate the excellent robustness of MoVi-Fi against body movements, while the baseline generally fail to obtain meaningful results; it makes sense as RF-SCG was not designed for motion-robustness. Although the ground truth is supposed to be motion-robust by relying on PPG, the results in Figure 14b reveal its minor limitation: body movements can still interfere PPG, if the contact sensor is not applied properly, to the extent of erasing a few heartbeat cycles, but MoVi-Fi manages to maintain its performance even during these "difficult" periods. Breath waveforms from both MoVi-Fi and ground truth match each other and exhibit perfect motion-robustness.

The two body movements shown in Figure 14, albeit both interfering vital signs monitoring, yield different levels of hardness for MoVi-Fi to handle. Typewriting barely causes neck movements and hence the interference takes place solely in signal propagation, whereas walking may change the neck position, causing far complicated interference. As an SCG ground truth sensor [17] requires on-body accelerometers, its motion-robustness is far inferior to that of PPG, hence SCG ground truth are not valid (for both training and verification purposes) under body movements. Therefore, instead of recovering SCG-related waveforms [17, 39, 49], we adopt BVP that is equally useful to clinical studies as SCG, as they are both strongly correlated with *electrocardiogram* (ECG) [38].



(a) Waveform recovery from typewriting.



(b) Waveform recovery from walking-on-treadmill.

Figure 14: Examples of vital signs waveform recovery under two stationary body movements.

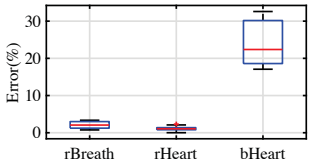


Figure 15: Relative errors against ground truth.

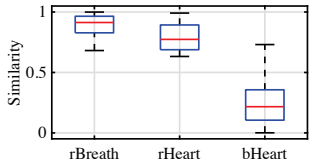


Figure 16: Cosine similarity to ground truth.

The overall performance evaluations on both vital signs under all 8 types of body movements are reported in Figures 15 and 16. Apparently, MoVi-Fi performs significantly better than the baseline, evidently proving its motion-robustness. In addition, the slight discrepancy between MoVi-Fi and the ground truth (under both metrics) should be attributed to the minor defects of both; the ground truth sometimes may miss cycles if the sensor clipper is not tightly applied, as shown in Figure 14b.

5.2.3 Impact of Body Movements. In order to study how individual body movement types affect the performance of MoVi-Fi, we specifically look into the MoVi-Fi's heartbeat monitoring performance against each body movement in Figures 17 and 18. For brevity, we hereafter use the following abbreviated names for the 8 movements: PP (playing-phone), TW (typewriting), SW (swaying-body), LS (leg-shaking), WT (waking-on-treadmill), EX (exercising), SS (standing-up/sitting down), and TO (turning-over). It is rather intuitive to observe that PP and LS lead to the least impact on the performance, as the body parts involved in the movements are far from the neck. Therefore, it is equally reasonable to expect that WT, EX, and SS cause the worst performance (in relative sense), as they both lead to the back and forth motion of necks.

To better evaluate the quality of heartbeat waveform recovery, we specifically verify if *heart rate variability* (HRV) can be captured by the waveforms, because HRV is a crucial feature of cardiac cycles: a high HRV often indicates greater cardiovascular fitness [41]. As HRV checks the natural variation among *interbeat intervals* (IBIs), we first report the absolute errors of IBI against respective body movements in Figure 19, where intervals of a recovered waveform are individually compared against their corresponding ground truth

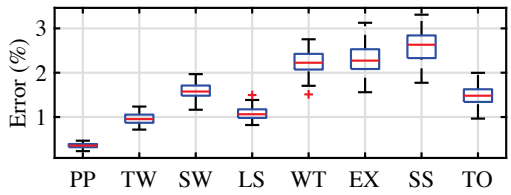


Figure 17: Relative errors of heart rates under different body movements.

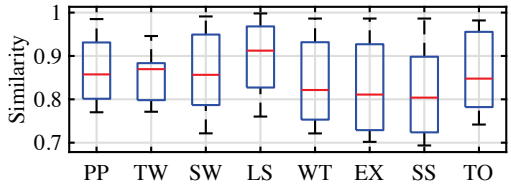


Figure 18: Cosine similarity of heartbeat waveform to ground truth under different body movements.

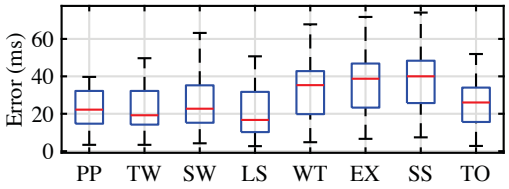


Figure 19: Absolute errors of heartbeat IBI under different body movements.

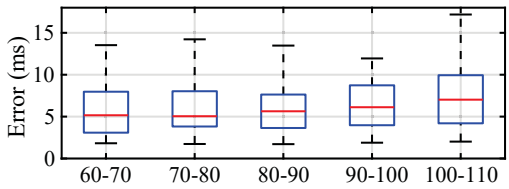


Figure 20: Absolute errors of heartbeat SDNN under different heart rate ranges.

intervals. In reality, what really matters is not the value of individual IBIs, but rather the statistics on HRV, such as SDNN (the standard deviation of IBI) [41]. Therefore, we further evaluate SDNN in Figure 20, where statistics are grouped according to heart rate ranges. Although the IBI errors in Figure 19 have media values up to 40ms (5% of a 800ms nominal IBI), the resulting SDNN errors are negligibly small with values largely below 10ms. Because a subject with a SDNN value beyond 100ms or below 50ms is respectively deemed as healthy and unhealthy, the minor SDNN errors reported in Figure 20 would barely alter any clinical judgement.

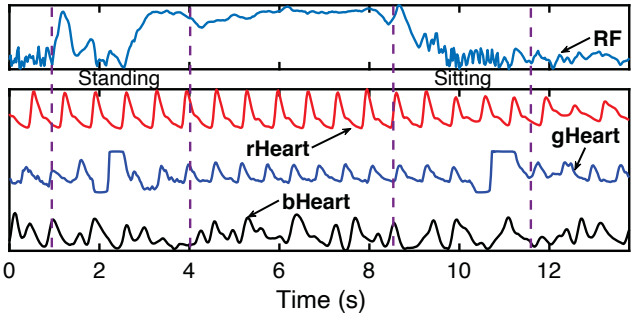
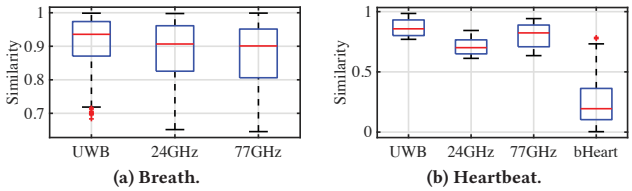


Figure 21: Heartbeat waveform recovery under non-stationary body movements (i.e., SS).

In order to further illustrate the performance of MoVi-Fi under non-stationary body movements, we choose SS as an example and report the various waveforms in Figure 21, with motion periods marked by dash boxes. These results also show that, whereas ground truth can be affected by sudden movements if not properly collected, MoVi-Fi still survives them thanks to the contact-free sensing mode and time contrastive separation.

5.2.4 Cross-Technology Transfer. As claimed earlier, MoVi-Fi is a pure software system readily deployed onto any commercial-grade radars. In this section, we demonstrate the cross-technology transferability of MoVi-Fi in two steps. We first show that this transferability can be achieved by retaining not only the same software architecture but also the trained deep learning modules. Remember



(a) Breath. (b) Heartbeat.

Figure 22: Cosine similarity of (a) breath and (b) heartbeat after a rough cross-technology transfer.

that we produce the earlier results using the IR-UWB radar, so now we directly migrate the whole suit of codes onto another two radars; the evaluation results are shown in Figure 22. Although some degradation can be expected, the performances of a roughly transferred MoVi-Fi on other radars are still rather satisfactory, and the heartbeat waveform delivered by the TI's FMCW radar is still far better than that achieved by the baseline (specifically designed for this radar but without motion-robust consideration).

We further improve the performance on the TI's FMCW radar by re-training the deep learning modules using the data collected by the new radar. The outcome for a specific body movement (SW) is illustrated in Figure 23 and the performance under all types of body movements are summarized in Figure 24. As expected, a well-trained MoVi-Fi performs better on the 77 GHz radar (than the IR-UWB radar in Figure 18) thanks to a higher carrier frequency, yet the performance improvement is minor because the higher sensitivity equally takes in more motion interference.

5.2.5 Time/Spatial Diversity Gains. We first take the IR-UWB radar to verify the time diversity gain in terms of involving different numbers of range bins in the hot-zone. Essentially, we fine-tune the parameter of the CFAR algorithm (see Section 4) to select different rows (bins) in the hot-zone, representing signals reflected at slightly

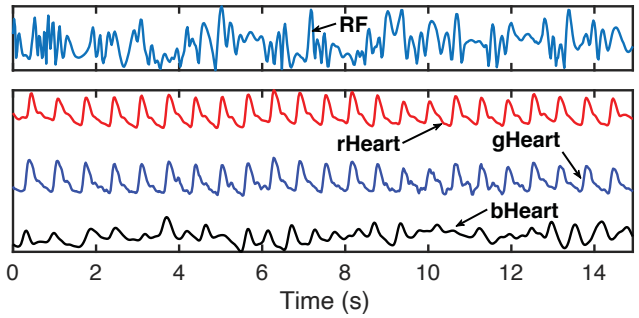


Figure 23: Heartbeat waveform recovery under body swaying (SW) using the 77GHz mmWave radar.

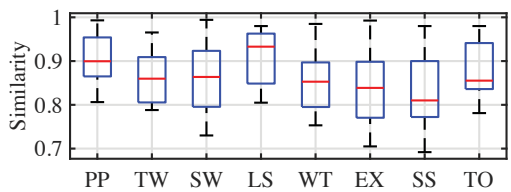
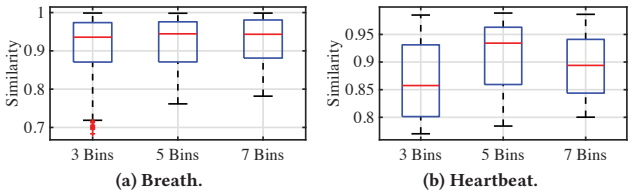
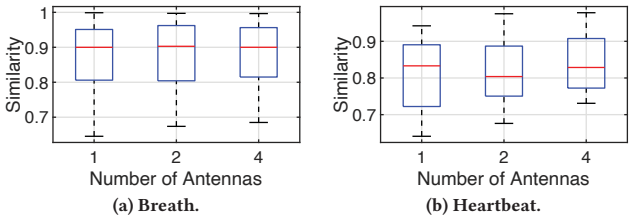


Figure 24: Heartbeat monitoring performance on 77 GHz radar under different body movements.



(a) Breath. (b) Heartbeat.

Figure 25: Cosine similarity of (a) breath and (b) heartbeat under different number of involved range bins.



(a) Breath. (b) Heartbeat.

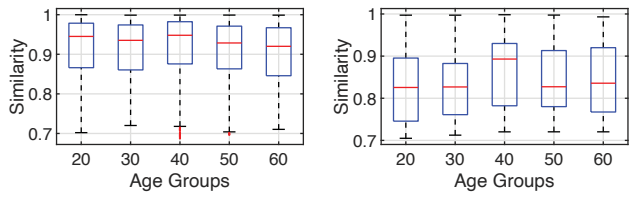
Figure 26: Cosine similarity of (a) breath and (b) heartbeat under different number of rx antennas.

different distances. As body movements may disperse the vital signs excited signals to several neighboring bins, involving more bins should be beneficial, as shown in Figure 25 (from 3 bins to 5 bins). Nonetheless, further involving more bins may only benefit relatively large-scale movements (e.g., WT and SS) but not others, as only interference can be introduced if the range covered by these bins exceed those of the movements. This later effect explains the overall negative trend, particularly to heartbeat, of involving 7 bins in Figure 25.

We further evaluate the spatial diversity gain using TI's FMCW radar (with a 3×4 antenna array) in Figure 26, where we fix one tx antenna and varying the number of rx antennas. We observe that, while using more rx antennas brings marginal improvement to breath sensing, it appears to affect heartbeat sensing negatively (according to the medians, but averages are slightly increased). As explained in Section 3.1, involving more antennas should narrow the beam, improving the SNR but also reducing the FoV. Therefore, spatial diversity does help under body movements (e.g., TW) that barely affect the neck position, but it hinders otherwise. In the following, we switch back to the IR-UWB radar to report MoVi-Fi's performance under other factors, as the corresponding results on other radars are similar.

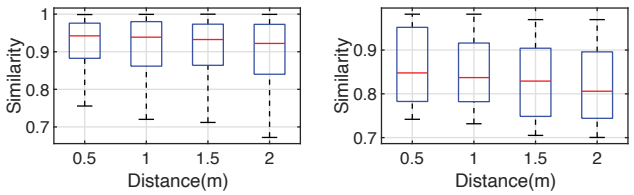
5.2.6 Impact of Different People. We separate the 12 subject into five groups: 20 (15-24), 30 (25-34), 40 (35-44), 50 (45-54), and 60 (55-64), and each group contains 2 subjects except 30 and 40 that both have one more subject. The results in Figure 28 demonstrate the ability of MoVi-Fi in generalizing across different age groups (only samples from 20 and 40 groups are used for training), but also indicate a curious "peak" for the 40 group. Though the limited number of subjects may cause a bias, we suspect that the relatively stabilized constitution in that age group could be a factor.

5.2.7 Impact of Sensing Distance. We ask two subjects to perform SW (a movement that may require a different sensing range) at different distances from 0.5m to 2m, as reported in Figure 28. As



(a) Breath. (b) Heartbeat.

Figure 27: Cosine similarity of (a) breath and (b) heartbeat under different age groups.

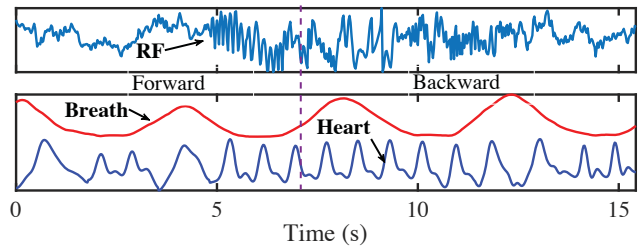


(a) Breath. (b) Heartbeat.

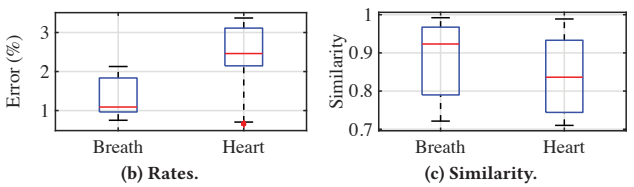
Figure 28: Cosine similarity of (a) breath and (b) heartbeat under different distances.

expected, while both vital signs are negatively affected by an increasing sensing distance, heartbeat suffers more, because the micro-vibrations excited by these two signs have different strengths.

Since we keep recording RF signals when these subjects change their distances, we may also report a continuous monitoring when subjects move within the range of 0.3 m to 2.5 m. Though this monitoring violates the motion scope defined in Section 3.2, we apply two makeshifts to get around the violated constraints. On one hand, as MoVi-Fi cannot monitor a subject turning back to it, we let the subjects walk backward when moving away. On the other hand, as the range spans well beyond the scope of ±30 cm during the whole process, we compensate this by involving about 30 bins/observations in total but sampling at 6 sequential time points to obtain 6 partially overlapped hot-zones each containing 7 bins. As a results, MoVi-Fi operates as usual to extract waveforms from



(a) Examples of vital signs waveform recovered during walking.



(b) Rates. (c) Similarity.

Figure 29: Heartbeat waveform recovery when subjects move towards and away from the radar.

individual hot-zones and to merge the outcome using VS-Net. The example output of MoVi-Fi in Figure 29a is chosen to emphasize the impact of distance change on the waveform quality, though typical results (especially in terms of heartbeat waveform) often degrade less evident in distance. Also, it is reasonable to observe in both Figures 29b and 29c that MoVi-Fi's overall performance during actual walking (albeit within a short distance) is similar to the average performance of SW shown in Figure 28.

5.3 Discussions

As the first software RF-sensing system achieving motion-robust vital signs monitoring, MoVi-Fi certainly leaves quite a few directions to be further explored. First of all, we are yet to extract important cardiac events from the recovered vital signs waveform; this should be readily achievable with sufficient clinical data and the labelling technique proposed in [17]. Second, we have not made use of the fine-grained waveforms to infer related human physical conditions (e.g., breath volume, blood pressure, and blood oxygen level). However, we believe that latest research outcome (e.g., [7]) may certainly help MoVi-Fi to close the gap between vital signs and the related physical conditions. Third, MoVi-Fi's motion-robustness is currently confined by the sensing scope of a single radar, so it is worth investigating how the capabilities of multiple radars can be synthesized so as to extend the coverage of MoVi-Fi. Finally, we have only focused on a single subject so far, and monitoring multiple subjects is certainly a more challenging issue, especially with the motion-robustness requirement. We are planning to exploit the spatial diversity offered by large-scale antenna arrays [11, 63] to approach this topic.

6 CONCLUSION

Although contact-free vital signs monitoring have been studied for years, how to achieve it in a motion-robust manner is still an open problem. In order to close this gap, we have designed and implemented MoVi-Fi as a contact-free RF-sensing prototype. In order to forge MoVi-Fi into a software system that delivers motion-robustness by leveraging pure algorithmic analytics, we have made several relevant contributions. First, we have unified the sensing data captured by various radars, making MoVi-Fi independent of underlying hardware platforms. Second, we have conducted a serious study on the composition between body movements and the micro-vibrations excited by vital signs; the outcome is a motion categorization that assists further algorithm designs. Third, we have explored and innovatively extended the recently developed deep contrastive learning framework, so as to separate the nonlinear signal composition in a self-supervised manner without the need for ground truth labels. Our extensive evaluations have clearly demonstrated the strong competence of MoVi-Fi in achieving motion-robust vital signs monitoring under real-life scenarios.

ACKNOWLEDGMENTS

We are grateful to the anonymous reviewers for their valuable and constructive comments. We would like to thank WiRUSH [50] for providing fund to develop MoVi-Fi, and we further thank Energy Research Institute @ NTU (ERI@N) and NTU-IGP for supporting the PhD scholarship of Tianyue Zheng.

REFERENCES

[1] Heba Abdelnasser, Khaled A. Harras, and Moustafa Youssef. 2015. UbiBreathe: A Ubiquitous Non-invasive WiFi-based Breathing Estimator. In *Proc. of the 16th ACM MobiHoc*. 277–286.

[2] D. A. Abraham. 2018. Performance Analysis of Constant-False-Alarm-Rate Detectors Using Characteristic Functions. *IEEE Journal of Oceanic Engineering* 43, 4 (2018), 1075–1085.

[3] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. 2015. Smart Homes that Monitor Breathing and Heart Rate. In *Proc. of the 33rd ACM CHI*. 837–846.

[4] Infineon Technologies AG. 2018. Industrial Radar Sensing. https://www.infineon.com/dgdl/Infineon-Presentation_24GHz+Sensing-PPT-v01_00-EN.pdf?fileId=5546d4625debb399015e0a4773e042e7. Accessed: 2020-07-23.

[5] Novelda AS. 2017. Single-Chip Radar Sensors with Sub-mm Resolution - XETHRU. https://www.xethru.com/. Accessed: 2021-03-03.

[6] David Blumenthal, Elizabeth Malphrus, and J. Michael McGinnis. 2015. *Vital Signs: Core Metrics for Health and Health Care Progress*. National Academies Press.

[7] Yetong Cao, Huijie Chen, Fan Li, and Yu Wang. 2021. Crisp-BP: Continuous Wrist PPG-based Blood Pressure Measurement. In *Proc. of the 27th ACM MobiCom*. 1–14.

[8] CCTV 5. 2019. Personal Body Exercises. https://www.youtube.com/watch?v=0TbpMBMSDCI. Online; accessed 28 June 2021.

[9] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *Proc. of the 37th ICML*. 1597–1607.

[10] Weixuan Chen and Daniel McDuff. 2018. DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks. In *Proc. of the 15th ECCV*. 349–365.

[11] Zhe Chen, Tianyue Zheng, and Jun Luo. 2021. Octopus: A Practical and Versatile Wideband MIMO Sensing Platform. In *Proc. of the 27th ACM MobiCom*. 1–14.

[12] Jennifer Gonik Chester and James L Rudolph. 2011. Vital Signs in Older Patients: Age-related Changes. *Journal of the American Medical Directors Association* 12, 5 (2011), 337–343.

[13] Shuya Ding, Zhe Chen, Tianyue Zheng, and Jun Luo. 2020. RF-Net: A Unified Meta-Learning Framework for RF-Enabled One-Shot Human Activity Recognition. In *Proc. of the 18th ACM SenSys*. 517–530.

[14] S. Dong, Y. Zhang, C. Ma, Q. Lv, C. Li, and L. Ran. 2020. Cardiogram Detection with a Millimeter-wave Radar Sensor. In *Proc. of the IEEE RWS*. 127–129.

[15] Pedro Fonseca, Ronald M Aarts, Xi Long, Jérôme Rolink, and Steffen Leonhardt. 2016. Estimating Actigraphy from Motion Artifacts in ECG and Respiratory Effort Signals. *Physiological Measurement* 37, 1 (2016), 67–82.

[16] Google. 2019. TensorFlow 2.0. https://www.tensorflow.org/. Accessed: 2021-03-03.

[17] Unsoo Ha, Salah Assana, and Fadel Adib. 2020. Contactless Seismocardiography via Deep Learning Radars. In *Proc. of the 26th ACM MobiCom*. Article 62, 14 pages.

[18] Aapo Hyvärinen and Hiroshi Morioka. 2016. Unsupervised Feature Extraction by Time-Contrastive Learning and Nonlinear ICA. In *Proc. of the 30th NeurIPS*. 3772–3780.

[19] Aapo Hyvärinen and Hiroshi Morioka. 2017. Nonlinear ICA of Temporally Dependent Stationary Sources. In *Proc. of the 20th AISTATS*. 460–469.

[20] Aapo Hyvärinen and Petteri Pajunen. 1999. Nonlinear Independent Component Analysis: Existence and Uniqueness Results. *Neural Netw.* 12, 3 (1999), 429–439.

[21] IEFT. 2017. Precision Time Protocol Version 2 (PTPv2). Accessed: 2021-03-13.

[22] Texas Instruments. 2017. DCA1000EVM. https://www.ti.com/tool/DCA1000EVM. Accessed: 2021-03-03.

[23] Texas Instruments. 2017. IWR1443BOOST. https://www.ti.com/store/ti/en/p/product/?p=IWR1443BOOST. Accessed: 2021-03-03.

[24] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. Image-to-Image Translation with Conditional Adversarial Networks. In *Proc. of the IEEE CVPR*. 1152–1134.

[25] Ossi Kaltiokallio, Hüseyin Yiğitler, Riku Jäntti, and Neal Patwari. 2014. Non-invasive Respiration Rate Monitoring Using a Single COTS TX-RX Pair. In *Proc. of the 13th ACM IPSN*. IEEE, 59–69.

[26] Te-Yu J Kao, Yan Yan, Tze-Min Shen, Austin Ying-Kuang Chen, and Jenshan Lin. 2013. Design and Analysis of a 60-GHz CMOS Doppler Micro-Radar System-in-Package for Vital-Sign and Vibration Detection. *IEEE Transactions on Microwave Theory and Techniques* 61, 4 (2013), 1649–1659.

[27] Rajet Krishnan, Balasubramaniam Natarajan, and Steve Warren. 2010. Two-Stage Approach for Detection and Reduction of Motion Artifacts in Photoplethysmographic Data. *IEEE Transactions on Biomedical Engineering* 57, 8 (2010), 1867–1876.

[28] C. Li and J. Lin. 2008. Random Body Movement Cancellation in Doppler Radar Vital Sign Detection. *IEEE Transactions on Microwave Theory and Techniques* 56, 12 (2008), 3143–3152.

[29] Feng Lin, Chen Song, Yan Zhuang, Wenyao Xu, Changzhi Li, and Kui Ren. 2017. Cardiac Scan: A Non-Contact and Continuous Heart-Based User Authentication System. In *Proc. of the 23rd ACM MobiCom*. 315–328.

[30] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. 2015. Tracking Vital Signs during Sleep Leveraging Off-the-Shelf WiFi. In *Proc. of the 16th ACM MobiHoc*. 267–276.

[31] G. Lu, F. Yang, J. A. Taylor, and J. F. Stein. 2009. A Comparison of Photoplethysmography and ECG Recording to Analyse Heart Rate Variability in Healthy Subjects. *Journal of Medical Engineering & Technology* 33, 8 (2009), 634–641.

[32] Q. Lv, L. Chen, K. An, J. Wang, H. Li, D. Ye, J. Huangfu, C. Li, and L. Ran. 2018. Doppler Vital Signs Detection in the Presence of Large-Scale Random Body Movements. *IEEE Transactions on Microwave Theory and Techniques* 66, 9 (2018), 4261–4270.

[33] J. Muñoz-Ferreras, Z. Peng, R. Gómez-García, and C. Li. 2016. Random Body Movement Mitigation for FMCW-radar-based Vital-sign Monitoring. In *Proc. of IEEE BioWireleSS*. 22–24.

[34] NeuLog. 2017. Heart Rate & Pulse logger sensor NUL-208. https://neulog.com/heart-rate-pulse/. Accessed: 2021-03-12.

[35] NeuLog. 2017. Respiration Monitor Belt Logger Sensor NUL-236. https://neulog.com/respiration-monitor-belt/. Accessed: 2021-03-21.

[36] Phuc Nguyen, Xinyu Zhang, Ann Halbower, and Tam Vu. 2016. Continuous and Fine-Grained Breathing Volume Monitoring from Afar Using Wireless Signals. In *Proc. of the 35th IEEE INFOCOM*. 1–9.

[37] A Pai, A Veeraraghavan, and A. Sabharwal. 2021. HRVCam: Robust Camera-based Measurement of Heart Rate Variability. *J. Biomed Opt* 26 (2021), 1–23.

[38] RA Payne, CN Symeonides, DJ Webb, and SRJ Maxwell. 2006. Pulse Transit Time Measured from the ECG: An Unreliable Marker of Beat-to-Beat Blood Pressure. *Journal of Applied Physiology* 100, 1 (2006), 136–141.

[39] Yu Rong. 2018. *Remote Sensing For Vital Signs Monitoring Using Advanced Radar Signal Processing Techniques*. Ph.D. Dissertation. Arizona State University, Tempe, AZ, USA.

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proc. of MICCAI*. 234–241.

[41] Fred Shaffer and J.P. Ginsberg. 2017. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health* 5 (2017), 258.

[42] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. 2020. SpiroSonic: Monitoring Human Lung Function via Acoustic Sensing on Commodity Smartphones. In *Proc. of the 26th ACM MobiCom*. 691–704.

[43] George Takla, John H Petre, D John Doyle, Mayumi Horibe, and Bala Gopakumaran. 2006. The Problem of Artifacts in Patient Monitor Data During Surgery: a Clinical and Methodological Review. *Anesthesia & Analgesia* 103, 5 (2006), 1196–1204.

[44] M. Tang, F. Wang, and T. Horng. 2017. Single Self-Injection-Locked Radar With Two Antennas for Monitoring Vital Signs With Large Body Movement Cancellation. *IEEE Transactions on Microwave Theory and Techniques* 65, 12 (2017), 5324–5333.

[45] J. Tu, T. Hwang, and J. Lin. 2016. Respiration Rate Measurement Under 1-D Body Motion Using Single Continuous-Wave Doppler Radar Vital Sign Detection System. *IEEE Transactions on Microwave Theory and Techniques* 64, 6 (2016), 1937–1946.

[46] Anran Wang, Jacob E. Sunshine, and Shyamnath Gollakota. 2019. Contactless Infant Monitoring Using White Noise. In *Proc. of The 25th MobiCom*. 52:1–16.

[47] F. Wang, T. Horng, K. Peng, J. Jau, J. Li, and C. Chen. 2011. Single-Antenna Doppler Radars Using Self and Mutual Injection Locking for Vital Sign Detection With Random Body Movement Cancellation. *IEEE Transactions on Microwave Theory and Techniques* 59, 12 (2011), 3577–3587.

[48] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW Based Contactless Respiration Detection Using Acoustic Signal. *Proc. of the 20th ACM UbiComp* 1, 4 (2018), 170.

[49] Christoph Will, Kilin Shi, Sven Schellenberger, Tobias Steigleder, Fabian Michler, Jonas Fuchs, Robert Weigel, Christoph Ostgathe, and Alexander Koelpin. 2018. Radar-based Heart Sound Detection. *Nature Scientific Reports* 8, 1 (2018), 1–14.

[50] WiRUSH. 2019. Guangxi Wanyun Technology Co., Ltd. https://www.wirush.ai/.

[51] Chenshu Wu, Zheng Yang, Zimu Zhou, Xuefeng Liu, Yunhao Liu, and Jiannong Cao. 2015. Non-Invasive Detection of Moving and Stationary Human with WiFi. *IEEE Journal on Selected Areas in Communications* 33, 11 (2015), 2329–2342.

[52] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. BreathListener: Fine-Grained Breathing Monitoring in Driving Environments Utilizing Acoustic Signals. In *Proc. of the 17th ACM MobiSys*. 54–66.

[53] Zhicheng Yang, Parth H Pathak, Yunze Zeng, Xixi Liran, and Prasant Mohapatra. 2017. Vital Sign and Sleep Monitoring using Millimeter Wave. *ACM Transactions on Sensor Networks* 13, 2 (2017), 1–32.

[54] Zi-Kai Yang, Heping Shi, Sheng Zhao, and Xiang-Dong Huang. 2020. Vital Sign Detection during Large-Scale and Fast Body Movements Based on an Adaptive Noise Cancellation Algorithm Using a Single Doppler Radar Sensor. *Sensors* 20, 15 (2020), 4183:1–17.

[55] Zitong Yu, Wei Peng, Xiaobai Li, Xiaopeng Hong, and Guoying Zhao. 2019. Remote Heart Rate Measurement From Highly Compressed Facial Videos: An

End-to-End Deep Learning Solution With Video Enhancement. In *Proc. of the IEEE/CVF ICCV*. 151–160.

[56] Shichao Yue, Hao He, Hao Wang, Hariharan Rahul, and Dina Katabi. 2018. Extracting Multi-Person Respiration from Entangled RF Signals. *Proc. of the 18th ACM UbiComp* 2, 2 (2018), 86:1–22.

[57] J. M. Zanetti and K. Tavakolian. 2013. Seismocardiography: Past, Present and Future. In *Proc. of the 35th IEEE EMBC*. 7004–7007.

[58] Youwei Zeng, Dan Wu, Ruiyang Gao, Tao Gu, and Daqing Zhang. 2018. FullBreathe: Full Human Respiration Detection Exploiting Complementarity of CSI Phase and Amplitude of WiFi Signals. *Proc. of the 18th ACM UbiComp* 2, 3 (2018), 148:1–19.

[59] Zhilin Zhang. 2015. Photoplethysmography-Based Heart Rate Monitoring in Physical Activities via Joint Sparse Spectrum Reconstruction. *IEEE Transactions on Biomedical Engineering* 62, 8 (2015), 1902–1910.

[60] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2018. Emotion Recognition Using Wireless Signals. *Commun. ACM* 61, 9 (2018), 91–100.

[61] Tianyue Zheng, Zhe Chen, Chao Cai, Jun Luo, and Xu Zhang. 2020. V^2iFi: in-Vehicle Vital Sign Monitoring via Compact RF Sensing. In *Proc. of the 22nd ACM UbiComp*. 70:1–27.

[62] Tianyue Zheng, Zhe Chen, Shuya Ding, and Jun Luo. 2021. Enhancing RF Sensing with Deep Learning: A Layered Approach. *IEEE Communications Magazine* 59, 2 (2021), 70–76.

[63] Tianyue Zheng, Zhe Chen, Jun Luo, Lin Ke, Chaoyang Zhao, and Yaowen Yang. 2021. SiWa: See into Walls via Deep UWB Radar. In *Proc. of the 27th ACM MobiCom*. 1–14.