Catch Your Breath: Simultaneous RF Tracking and Respiration Monitoring with Radar Pairs

Tianyue Zheng, *Student Member, IEEE*, Zhe Chen, *Member, IEEE*, Shujie Zhang, *Student Member, IEEE*, and Jun Luo, *Senior Member, IEEE*

Abstract—Continuous respiration monitoring is significant for real-life healthcare applications, but realizing it is extremely hard as wearable sensors are cumbersome and contact-free sensors largely fail to tolerate user movements. Meanwhile, tracking users indoors mostly demands user-held devices, while device-free localization can barely tell what and who it tracks. Fortunately, as both contact-free respiration monitoring and device-free localization may rely on Radio-Frequency (RF) sensing, fusing them together creates a novel system capable of continuously tracking users while recovering their fine-grained respiratory waveforms. To this end, we propose BreathCatcher as a continuous human respiration tracking system for indoor applications. To build this system, we employ commercial-grade compact radar pairs to capture RF reflections containing respiratory signals. We then propose a hybrid human respiration and position tracking algorithm to locate and identify respiratory signals from complex RF reflection mixtures. Finally, we design an encoder-decoder deep neural network driven by variational inference to recover fine-grained respiratory waveforms. Essentially, BreathCatcher cannot only obtain respiratory waveforms from multiple walking users, but also identify each user according to the latent properties of the respiratory signals. We evidently demonstrate the accuracy of both tracking and respiration monitoring via experiments involving 12 subjects and 80 man-hour data.

Index Terms—RF contact-free sensing, device-free sensing, respiration monitoring, user tracking, deep learning, IR-UWB radar.

1 Introduction

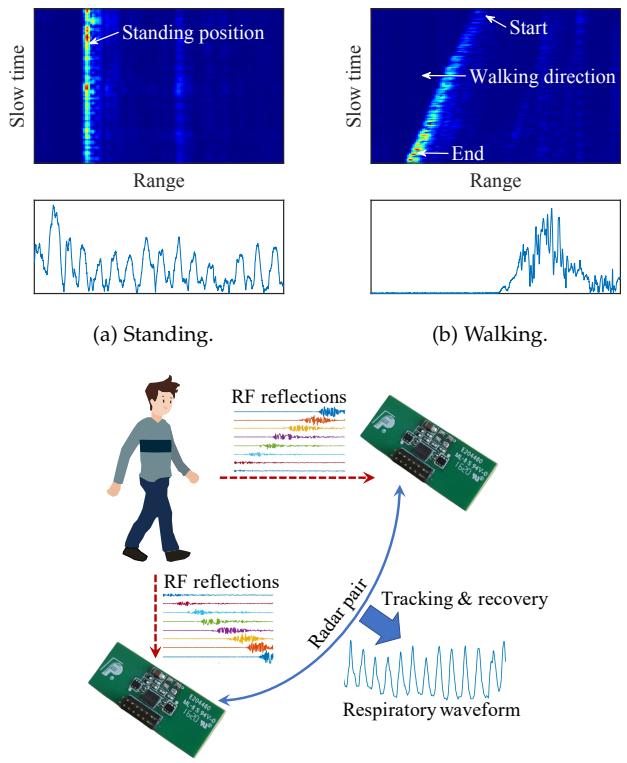
Continuous respiration monitoring plays a significant role in both human daily life and clinical settings, since respiration is one of the extremely representative vital signs that could indicate underlying physical and mental health conditions [1], [2], and be used for diagnosing related diseases [3], [4]. Conventionally, continuous respiration monitoring demands cumbersome wearable sensors such as chest bands [5] and headset [6] to be applied to human bodies. Unfortunately, such wearable sensors can make human subjects uncomfortable and may thus affect monitoring performance [7]. In the past few years, to alleviate the aforementioned issues with wearable sensors, both industry and academia have made substantial research efforts to realize *contact-free* respiration sensing leveraging Radio-Frequency (RF) signals [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. Although these systems estimate respiration rate correctly when a human subject remains static or only moves around a spot, none of them can recover respiration signals when the subject is walking, because human walking may overwhelm respiration-induced micro-activities and hence complicates the analysis.

In the meantime, plenty of research works have focused on active tracking demanding a subject to hold a device. Although these tracking schemes can achieve very high accuracy [20], [21], [22], their *device-based* nature can be a potential source of inconvenience. Recent years have witnessed a trend toward *device-free* passive tracking, which locates and tracks moving human subjects without involving them [23], [24], [25], [26], [27]. However, passive tracking results in a critical issue: as it only senses objects and/or motions, telling what and who it is tracking becomes a challenge. For example, a moving robot may be tracked as a human, and two tracked human subjects may not be clearly differentiated; this issue has rendered passive tracking systems less useful in human-centric scenarios.

Fortunately, since both contact-free respiration monitoring and device-free tracking can be enabled by RF sensing, it makes sense to fuse them together and to create a novel system capable of performing both tasks. On one hand, the information embedded in fine-grained respiratory waveforms can help exclude non-human subjects and differentiate human subjects. On the other hand, tracking results may assist the signal analytics aiming to recover respiratory waveforms of a human subject under movements. Whereas existing proposals [8], [24] do validate the potential of subject differentiation via respiratory waveforms, they are only concerning static human subjects; our envisioned solution should instead be robust to body movements inevitably introduced by human walking.

However, realizing such a solution faces two major challenges. First, identifying respiration within RF signal reflected from a walking human subject has never been put under scrutiny. Existing solutions have only studied static cases [8], [10], [28] and on-the-spot movements [19], [29]. As such, no existing framework is available to guide the design of such a hybrid solution for simultaneous tracking and respiration monitoring. Secondly, walking interferes

- *T. Zheng, S. Zhang, and J. Luo are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. T. Zheng is also with the Energy Research Institute, Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore. E-mail: {tianyue002,shujie002,junluo}@ntu.edu.sg*
- *Z. Chen is with China-Singapore International Joint Research Institute and AIWiSe Technology Co., Ltd, Guangzhou, China. E-mail: chenz@ssijri.com*
- *T. Zheng and Z. Chen contributed equally to this research.*



(a) Standing. (b) Walking.

(c) An example of BreathCatcher.

Fig. 1. The heatmaps and raw RF signals containing respiration (a) without and (b) with walking. In (c), we use a radar pair to track and recover fine-grained respiratory waveforms simultaneously.

with the received respiratory signals nonlinearly. To better understand this damaging effect, we draw a straightforward comparison between static and walking cases shown Fig. 1a and 1b. In both cases, we place an IR-UWB radar [30] in front of a human subject, but the subject is asked to stand still in Fig. 1a and the same subject walks towards the radar in Fig. 1b. In the former case, the RF signal at the range of that subject clearly reflects respiratory patterns, yet no valid respiratory signals can be observed if one focuses only on RF reflections at a fixed distance in the latter case.

To tackle these challenges, we propose BreathCatcher to enable simultaneous device-free tracking and continuous and fine-grained human respiration monitoring. As shown in Fig. 1c, BreathCatcher relies on a pair of commercial-grade compact radars [30] to capture RF reflections bounced off human subjects: these RF reflections contain nonlinear mixture of signals induced by both walking and respiration. The collaborative sensing from the radar pair provides higher spatial coverage and diversity, and also facilitates tracking the walking subjects on a 2D plane by observing from a higher dimension. The key to tracking position and respiration simultaneously is an algorithm that can jointly acquire them. Essentially, BreathCatcher consists of two tightly coupled modules: position tracking and respiration recovery. The position tracking module relies on adapting a Kalman filter [31] to locate human subjects while identifying relevant signal mixtures to facilitate further respiration recovery. The respiration recovery module outputs fine-grained respiratory waveforms containing features specific to each subject, which in turn helps differentiate the subjects and improve the tracking functionality. To recover respiration from the identified nonlinear signal mixtures, we design a novel IQ Variational Encoder-Decoder (IQ-VED) neural network. Leveraging the generalizability brought by variational inference, IQ-VED achieves fine-grained waveform recovery under body movements of walking human subjects. In summary, our major contributions are as follows:

- We propose a novel algorithm for simultaneous human tracking and respiration recovery. To our best knowledge, BreathCatcher is the first system that tracks human position and respiration continuously.
- We design an algorithm driven by both adapted Kalman filter and signal fusion, in order to achieve multi-target multi-radar tracking. This algorithm helps locate subjects while identifying subject-specific signals, thus facilitating respiration recovery.
- We propose IQ-VED for recovering and refining the respiratory waveform from nonlinear mixture in a motion-robust manner. The recovered feature-rich waveform can further help identify each human subject, hence improving the functionality of tracking.
- We conduct extensive evaluations on BreathCatcher with an 80 man-hour dataset; the results strongly confirm the excellent abilities of BreathCatcher of tracking, identifying human subjects, and recovering their fine-grained respiration waveforms.

The rest of the paper is organized as follows. Sec. 2 introduces the background and preliminaries for simultaneous tracking and respiration monitoring with IR-UWB radar. Then we provide a detailed exposition on constructing BreathCatcher from scratch in Sec. 3. We provide implementation details in Sec. 4 and report performance evaluations in Sec. 5. We survey the literature in Sec. 6 before finally concluding our paper in Sec. 7.

2 Background and Preliminaries

In this section, we carefully study the principles of ranging and respiration monitoring with an IR-UWB radar. We focus on modeling the radar signal to reflect how it represents human subjects and respiration, hence laying a theoretical foundation for the system design discussed in Sec. 3.

2.1 Ranging with IR-UWB Radar

We first explain the working principles of IR-UWB radar on how it performs ranging. Each frame of the transmitted signal $x(t)$ is formed by a baseband Gaussian pulse $s(t)$ modulated by a cosine carrier:

$$x(t) = s(t)\cos(2\pi f_c t), \tag{1}$$

where f_c is the carrier frequency. The signal is then transmitted and reflected by different objects in the space, which can be modeled as the channel. In a typical indoor environment, the Channel State Information (CSI) $h(t)$ with multipaths can be represented as follows:

$$h(t) = \sum_{p=1}^{P} \alpha_p \delta\left(t - \tau_p\right), \tag{2}$$

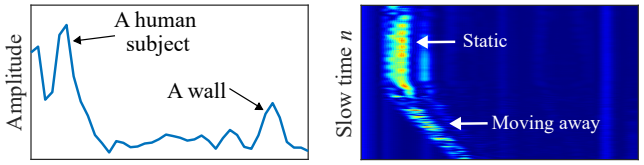
where α_p and τ_p are the attenuation and the time delay due to signal propagation of the p-th reflection path, respectively. Therefore the received signal can be expressed as:

$$y(t) = h(t) * x(t) = \sum_{p=1}^{P} \alpha_p \cos(2\pi f_c(t-\tau_p))s(t-\tau_p). \quad (3)$$

By employing IQ downconversion and lowpass filter, the received baseband signal frame $r(t)$ can be obtained as:

$$\begin{aligned} r(t) =& \mathbf{LP}\left[e^{j2\pi f_c t} y(t)\right] \\ =& \mathbf{LP}\left[\sum_{p=1}^{P} e^{j2\pi f_c t}\alpha_p[e^{j2\pi f_c(t-\tau_p)} + e^{-j2\pi f_c(t-\tau_p)}]\cdot s(t-\tau_p)\right] \\ =& \sum_{p=1}^{P} e^{j2\pi f_c \tau_p}\alpha_p s(t-\tau_p). \end{aligned} \quad (4)$$

Objects at different distances from the radar exhibit different τ_p. By taking the amplitude of $r(t)$, we can observe several peaks corresponding to the objects at different distances. For example, in Fig. 2a, a human subject, and a wall can be observed in the signal.



(a) Received signal $|r(t)|$. (b) Signal matrix $|\boldsymbol{r}(t,n)|$.

Fig. 2. Amplitude of single-frame radar signal $r(t)$ and signal matrix $\boldsymbol{r}(t,n)$ composed of multiple frames.

To enable tracking of subjects, the radar transmits frames at a regular interval, and then stacks the received frames to form a signal matrix $\boldsymbol{r}(t) = [r_1(t), \cdots r_n(t), \cdots r_N(t)]^T$, where t and n are respectively the *fast-time* and *slow-time* indices, and N is the number of slow-time frames [32], [33]. We hereafter slightly abuse the terminology by writing $\boldsymbol{r}(t)$ as $\boldsymbol{r}(t,n)$ to clearly indicate its matrix nature. One example of the signal matrix is illustrated in Fig. 2b, in which the trajectory of a walking subject can be observed. However, a single radar could only enable 1D ranging of walking subjects, at least two radars would be required to track the subjects on a 2D plane, which will be further explained in Sec. 3.1 and 3.2.

2.2 Capturing Respiration with IR-UWB Radar

One may think that we can use the radar to record the distance change between the human chest and the radar, thus achieving respiration monitoring. However, a 10 cm range resolution[1] of the radar is insufficient for monitoring fine movements of human respiration, which is usually between 4 mm and 12 mm [35]. Therefore, we have to find another "ruler" to measure the distance change caused by human respiration. Notice that the fine displacements induced by human chest movements cause τ_p to change periodically, thus changing the term $e^{j2\pi f_c \tau_p}$ in Eqn. (4) that represents the phase of a received signal frame $r(t)$.

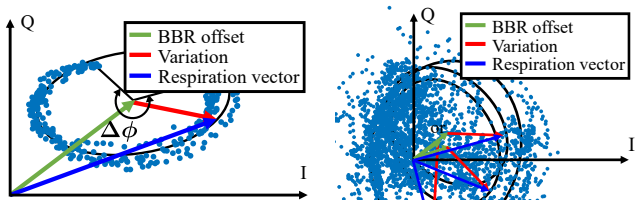
Nonetheless, we will get severe distortions if we recover the respiration signal by taking the phases of $\boldsymbol{r}(t,n)$ for a fixed t directly. To understand why, we consider a static human subject located at a fixed fast-time t index with slow-time signal $\boldsymbol{r}_t(n)$ (abbreviated as $r(n)$ in the following for brevity), whose IQ components can be represented as:

$$r_{\mathrm{I}}(n) = \alpha(n)\cos\left(\frac{4\pi d(n)}{\lambda} + \frac{4\pi d^{\mathrm{h}}(n)}{\lambda}\right) + o_{\mathrm{I}}^{\mathrm{BBR}}, \quad (5)$$

$$r_{\mathrm{Q}}(n) = \alpha(n)\sin\left(\frac{4\pi d(n)}{\lambda} + \frac{4\pi d^{\mathrm{h}}(n)}{\lambda}\right) + o_{\mathrm{Q}}^{\mathrm{BBR}}, \quad (6)$$

where $\alpha(n)$ is the strength of the reflected signal from the human chest, $d(n)$ is the distance from the radar to the human subject, λ is carrier wavelength, and $d^{\mathrm{h}}(n)$ denotes human chest movement. In both equations, the first terms (varying with the slow-time n) are caused by respiration, and the second terms $o_{\mathrm{I}}^{\mathrm{BBR}}$ and $o_{\mathrm{Q}}^{\mathrm{BBR}}$ are the offsets caused by *body background reflection* (BBR).

To visualize $r_{\mathrm{I}}(n)$ and $r_{\mathrm{Q}}(n)$, we take the signal of an unmoving breathing person $r(n)$ (as bounded by the red box in Fig. 2b, $d(n)$ is fixed) as an example, and display it as a constellation diagram in Fig. 3a. The blue *respiration vector* in the graph corresponds to the complex signal $r_{\mathrm{I}}(n) + jr_{\mathrm{Q}}(n)$; they are the sum of BBR offset (green vector) and respiration-induced variation (red vector). One may observe that, as the human subject breathes, the red vector rotates and the trace of the blue vector forms an elliptic arc. The arc may not be circular because $\alpha(n)$, the radius, is time-varying due to a varying radar cross-section [36]. Now Fig. 3a clearly



(a) human subject is static. (b) human subject is moving.

Fig. 3. Constellation diagrams of $r(n)$.

explains why recovering respiration by taking the phases of $r(n)$ directly will cause distortion: the existence of BBR offset makes 1-D respiration 2-D movements in the IQ space, thus taking the phases of $r(n)$ is equivalent to projecting the respiration vector trace onto a lower dimension; this surely causes loss of information.

2.3 Interference Caused by Body Movements

If the human subject of interest is moving, the previous analysis no longer holds true for the following reasons. First, the BBR offset (caused by reflections from limbs and torso other than the chest) previously presumed to be static is no longer so, resulting in a random shift of the elliptic center $\left(o_{\mathrm{I}}^{\mathrm{BBR}}, o_{\mathrm{Q}}^{\mathrm{BBR}}\right)$. Second, the distance from the radar to the human subject d also varies and further changes the signal phase. Being in the previous case of the unmoving human subject, the distance $d(n)$ becomes a variable now,

1. Suppose the radar bandwidth B is 1.5GHz, then the range resolution can be written as $\Delta r = \frac{c}{2B}$ [34], where c is the speed of light.

thus asking for human subject tracking. Last but not least, the radar cross-section changes with body movements as well, causing a varying reflected signal strength and hence unpredictable changes in the long/short axes of the ellipse. By tracking the human subject and showing $r(n)$ in the IQ diagram, the detrimental effects of a moving human subject are illustrated in Fig. 3b. We can see that the elliptic arcs in the static case disappear and get blended altogether, rendering the respiration pattern not analyzable by conventional signal processing and geometric methods.

3 System Design

This section introduces the design of BreathCatcher, whose block diagram is shown in Fig. 4. Upon obtaining the IQ signal matrices $\boldsymbol{r}(t,n)$ from the IR-UWB radar pair, BreathCatcher finds signal peaks in the matrices corresponding to the human subjects. It then tracks the signal in each matrix and fuses signals from the radar pair by their properties. After data augmentation, BreathCatcher leverages IQ-VED neural network to distill a latent representation of the respiratory waveform. Based on the respiratory features in the latent space, BreathCatcher decides if the tracked signal is related to humans and if the respiratory waveform should be recovered. Next, our IQ-VED trained by the ground truth waveform obtained from wearable sensors [5], recovers fine-grained respiratory waveform under severe interference produced by walking human subjects. Finally, human identification is performed by leveraging the waveform features unique to each human subject.

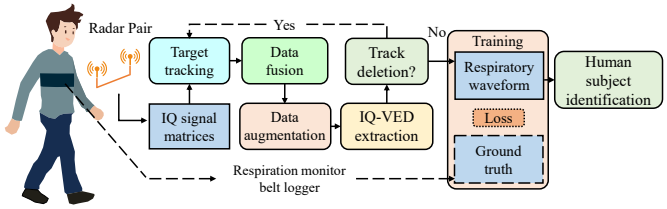


Fig. 4. System diagram of BreathCatcher.

3.1 Multi-target Tracking

We begin by defining the scope of multiple target locating and tracking, as shown in Fig. 5. Basically, the monitoring scenario should satisfy two conditions: i) the human subjects should stay within the overlapping area of the Field of Views (FoV) of the radar pair, and ii) both radars can see the chest of the human subjects. Note that a single pair of radar is only the basic setting, more radars can be used to expand the coverage. As introduced in Sec. 2.1, the radar performs the ranging of different subjects in the space by measuring the time delay τ_p of reflected signals. To refine the received signals and locate human subjects, we first remove clutters, i.e., unwanted echoes from the background, using an exponential moving average method [37], and then employ CFAR algorithm [34] to adaptively detect signal peaks whose level is above the average adjacent power level, and the detected signal peaks are deemed to correspond to different subjects. After locating subjects in each frame, we further process and combine these measurements for multi-target tracking. With each new incoming signal frame,

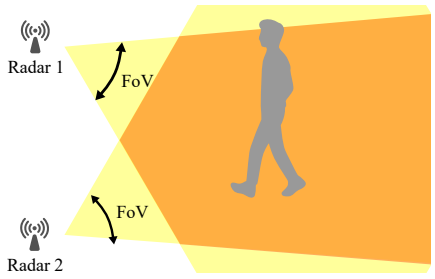


Fig. 5. The human subject should stay within the FoV of the radar pair, and the radars should see the chest of the human subject.

the tracking procedure follows three steps: state prediction, state correction, and measurement association.

Suppose the subject is moving following a trajectory $d(n) = [d_1, \cdots, d_n, \cdots, d_N]$ with a constant velocity. At each moment n, the distance measurement (containing error) of the subject is b_n, and the state of the subject is $\boldsymbol{s}_n = \begin{bmatrix} d_n \\ \dot{d}_n \end{bmatrix}$, where $\dot{d}_n$ is the derivative of d_n with respect to the slow time n, i.e., the velocity of the subject. Then the movement of the subject can be modeled as:

$$\boldsymbol{s}_n = \boldsymbol{F}_{n-1}\boldsymbol{s}_{n-1} + \boldsymbol{w}_n, \tag{7}$$

$$b_n = \boldsymbol{H}_n\boldsymbol{s}_n + \boldsymbol{v}_n. \tag{8}$$

where $\boldsymbol{F}_n$ is the state transition matrix, and $\boldsymbol{F}_n = \begin{bmatrix} 1 & \Delta n \\ 0 & 1 \end{bmatrix}$ (Δn is the interval between two frames), $\boldsymbol{w}_n$ is the process noise, whose covariance is $\boldsymbol{Q}_n$, $\boldsymbol{H}_n$ is the observation matrix defined by $\boldsymbol{H}_n = [1, 0]$, $\boldsymbol{v}_n$ is the observation noise, whose covariance is $\boldsymbol{R}_n$.

Based on the constant velocity model, we elaborate on the details of these steps of multi-target tracking using Kalman filter as follows:

State prediction. Suppose there is only one subject. With each new measurement of the subject's position, we predict an a priori state $\bar{\boldsymbol{s}}_n$ and its corresponding covariance $\bar{\boldsymbol{P}}_n$:

$$\bar{\boldsymbol{s}}_n = \boldsymbol{F}_{n-1}\hat{\boldsymbol{s}}_{n-1}, \tag{9}$$

$$\bar{\boldsymbol{P}}_n = \boldsymbol{F}_{n-1}\hat{\boldsymbol{P}}_{n-1}\boldsymbol{F}_{n-1}^T + \boldsymbol{Q}_{n-1}. \tag{10}$$

State correction The next step is correcting the predicted values by using historical data. The Kalman filter gain can be calculated as:

$$\boldsymbol{K}_t = \bar{\boldsymbol{P}}_n\boldsymbol{H}_n^T\left(\boldsymbol{H}_n\bar{\boldsymbol{P}}_n\boldsymbol{H}_n^T + \boldsymbol{R}_n\right)^{-1}. \tag{11}$$

And the corrected state and covariance of the tracked subject can be described as:

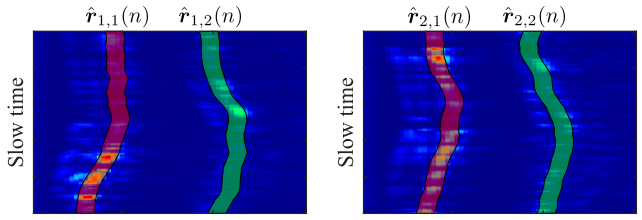
$$\hat{\boldsymbol{s}}_n = \bar{\boldsymbol{s}}_n + \boldsymbol{K}_n\left(b_n - \boldsymbol{H}_n\bar{\boldsymbol{s}}_n\right), \tag{12}$$

$$\hat{\boldsymbol{P}}_n = \left(\boldsymbol{I} - \boldsymbol{K}_n\boldsymbol{H}_n\right)\bar{\boldsymbol{P}}_n. \tag{13}$$

Measurement association. The above tracking method only applies to the case where there is only one subject. If there are multiple subjects to track, we have to associate the measurements of the current frame to existing tracks. The Global Nearest-Neighbor (GNN) algorithm [38] is used for this association. The algorithm calculates the distance between the measurement b_n to the predicted measurement according to the following equation:

$$\gamma_n^2 := \left(b_n - \boldsymbol{H}_n\bar{\boldsymbol{s}}_n\right)^T\boldsymbol{S}_n^{-1}\left(b_n - \boldsymbol{H}_n\bar{\boldsymbol{s}}_n\right), \tag{14}$$

where $\boldsymbol{S}_n$ is the covariance matrix defined by $\boldsymbol{S} = \boldsymbol{H}_n \bar{\boldsymbol{P}}_n \left(\boldsymbol{H}_n\right)^T + \boldsymbol{R}_n$. The first step of measurement association is gating, which means that for each subject, only measurements that are sufficiently "close" to the current estimate of the subject's state is considered. The algorithm sets a threshold ζ_{g}, and any $\gamma_n^2 \leq \zeta_{\mathrm{g}}$ is considered inside the gating region. Among the measurements inside the region, the one with the smallest distance value to the predicted measurement, is used to update the state of the subject's trajectory $d(n)$. For each radar i and subject k, after following the trajectory $d_{i,k}(n)$ of the subject in that radar, BreathCatcher selects multiple fast-time indices adjacent to the detected trajectory $d_{i,k}(n)$ to form a slow-time signal matrix $\hat{\boldsymbol{r}}_{i,k}(n)$, as shown in Fig. 6.



(a) Radar 1. (b) Radar 2.

Fig. 6. Walking trajectories of two human subjects captured by a radar pair.

Note that tracks related to non-human subjects will be removed utilizing the information obtained during respiration recovery, and the details are in Sec. 3.4.5.

3.2 Radar Pair Data Fusion

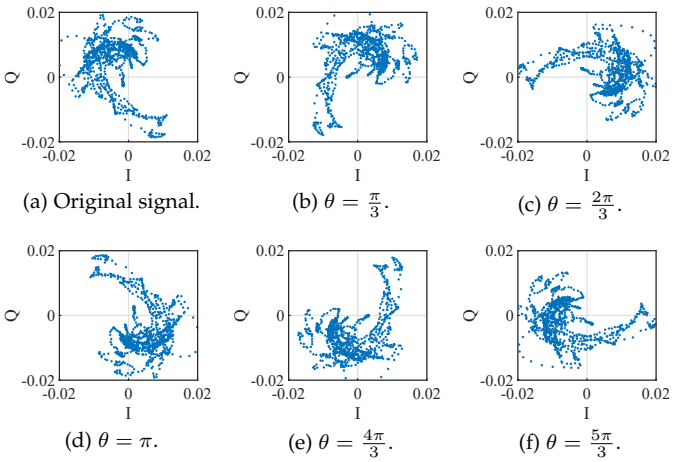
The aforementioned multi-target tracking is only for a single radar. If there is a radar pair, we have to associate the k-th signal from the first radar $\hat{\boldsymbol{r}}_{1,k}(n)$ and l-th signal from the second radar $\hat{\boldsymbol{r}}_{2,l}(n)$ of the same human subject for later respiration extraction. In order to achieve this, we assume i) signals of the same human subject from the two radars are statistically dependent, and ii) signals of different human subjects from the two radars are statistically independent. As such, for the k-th signal from the first radar, we find its corresponding l-th signal from the second radar by selecting the one that maximizes their absolute correlation.

$$l = \underset{m \in \mathcal{M}}{\arg\max} \left| \rho \left(\hat{\boldsymbol{r}}_{1,k}\left(n\right), \hat{\boldsymbol{r}}_{2,m}\left(n\right) \right) \right|, \tag{15}$$

where $\mathcal{M}$ is the set of all human-related signals observed by the second radar. An example association result is shown Fig. 6, and the signals are color-coded to indicate the association. With the associated signals, we can then employ basic trilateration algorithm [39] to locate the human subject in the space (the ambiguous localization point can be elimimated by considering the radar FoV). For each human subject, we concatenate the signal pair $\hat{\boldsymbol{r}}_{1,k}(n)$ and $\hat{\boldsymbol{r}}_{2,l}(n)$ from two radars to form $\hat{\boldsymbol{r}}(n)$, which will be later analyzed by the deep learning network to extract human respiration.

3.3 Data Augmentation

Before putting the I/Q signal contained in $\boldsymbol{r}(n)$ ($\hat{\boldsymbol{r}}(n)$ in Sec. 3.2 for brevity) into deep analytics, it is necessary to perform data augmentation. On one hand, data collection is highly non-trivial because one has to coordinate among human subjects, data recording of the IR-UWB radar, and the wearable ground truth sensors. Therefore, it is desirable to increase the diversity of a dataset by applying certain transformations. On the other hand, data augmentation often helps a deep neural network comprehend intrinsic structures of the raw data. For BreathCatcher, this I/Q-induced intrinsic structure is non-trivially preserved only by rotation but not other transforms such as cropping and scaling. Therefore, we propose to augment $\hat{\boldsymbol{r}}(n)$ by rotating every complex element in the I/Q domain:



(a) Original signal. (b) $\theta = \frac{\pi}{3}$. (c) $\theta = \frac{2\pi}{3}$.

(d) $\theta = \pi$. (e) $\theta = \frac{4\pi}{3}$. (f) $\theta = \frac{5\pi}{3}$.

Fig. 7. Augmenting data by rotating a slow-time signal of $\hat{\boldsymbol{r}}(n)$ in the I/Q domain.

$$\begin{bmatrix} r_{\mathrm{I}}^{\mathrm{aug}} \\ r_{\mathrm{Q}}^{\mathrm{aug}} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r_{\mathrm{I}} \\ r_{\mathrm{Q}} \end{bmatrix}, \tag{16}$$

where θ specifies a rotation angle and it is varied to achieve data augmentation. Fig. 7 illustrates five versions of augmented $\hat{\boldsymbol{r}}(n)$: the rotation preserves the respiration traces, because it affects only the distance d (which is anyway varying drastically during walking, according to Sec. 2.3) but not the respiration-induced periodic motions $d^{\mathrm{h}}(n)$ of the chest. As shown in Fig. 7, the overlapped respiration ellipses maintain the overall distribution despite varying rotations. In practice, BreathCatcher may choose to employ more rotation angles for better enriching a dataset.

3.4 Fine-Grained Respiration Recovery via Variational Encoder-Decoder

3.4.1 *Design Rationale*

Extracting certain signals from a nonlinear signal mixture is highly non-trivial [40], [41]; the deep learning community has been employing an Encoder-Decoder (ED) network for this task [42], [43]. Unfortunately, the latent space of a regular ED network is not continuous given limited training data, so it lacks sufficient generalization ability when dealing with unseen data. Inspired by the idea of *variational inference* [44], we tackle the problem of latent space irregularity by forcing the encoder to return probability distributions rather than discrete vectors, and we name the modified network Variational Encoder-Decoder (VED). It is worth noting that our VED is fundamentally different from Variational AutoEncoder (VAE) [45], [46]: whereas VED aims to extract

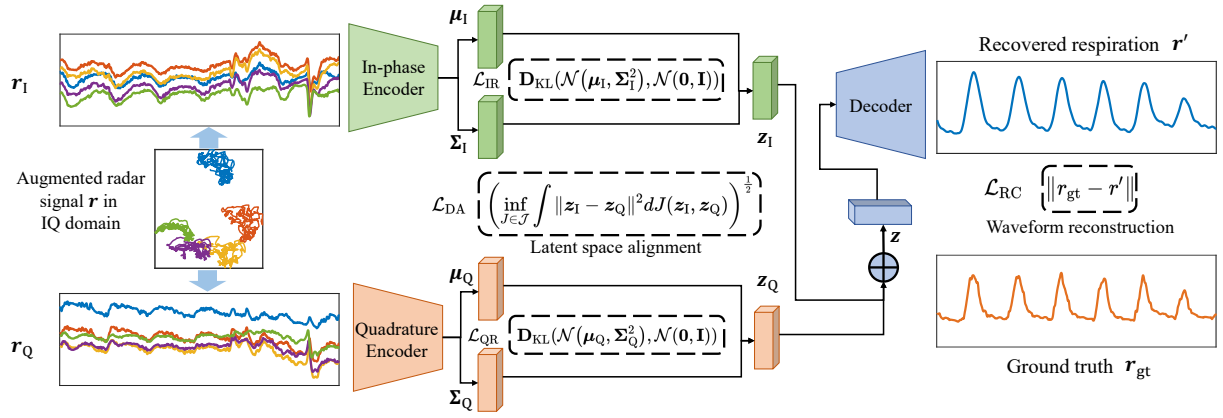


Fig. 8. IQ-VED architecture: it takes in I/Q data as two streams and encodes them separately. The resulting latent representations are aligned and fed to the decoder to recover the respiratory waveform by minimizing the reconstruction error.

signal from a nonlinear mixture, VAE intends to learn an efficient representation of the input.

To achieve the goal of respiratory waveform recovery, a regular ED learns an *encoder* $q_\phi(\boldsymbol{z}|\boldsymbol{r})$ mapping input data $\boldsymbol{r}$ to a latent representation $\boldsymbol{z}$, and generates output $\boldsymbol{r'}$ (i.e., respiratory waveform) by a *decoder* $p_\psi(\boldsymbol{r'}|\boldsymbol{z})$. In other words, $\boldsymbol{z}$ represents the partial features extracted from $\boldsymbol{r}$ to characterize only $\boldsymbol{r'}$. VED shares the pair of encoder ϕ and decoder ψ with ED, but it maps $\boldsymbol{r}$ to a Gaussian distribution parameterized by a mean and variance. Essentially, the generative process of the VED is enabled by maximizing the variational lower bound (VLB) [45]:

$$\begin{aligned}\log p_\psi(\boldsymbol{r'}) \geq\ & \mathrm{VLB}_{\mathrm{VED}}(\boldsymbol{r},\boldsymbol{r'};\psi,\phi)\\ =\ & \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{r})}\left[\log p_\psi(\boldsymbol{r'}|\boldsymbol{z})\right] - \mathbf{D}_{\mathrm{KL}}\left(q_\phi(\boldsymbol{z}|\boldsymbol{r})\|p_\psi(\boldsymbol{z})\right), \quad (17)\end{aligned}$$

where $p_\psi(\boldsymbol{z}) = \mathcal{N}(\mathbf{0},\mathbf{I})$ is a Gaussian prior on the latent representation $\boldsymbol{z}$ and $\mathbf{D}_{\mathrm{KL}}(\cdot)$ denotes the Kullback-Leibler (KL) divergence [47]; it works as a regularizer by minimizing the difference between $q_\phi(\boldsymbol{z}|\boldsymbol{r})$ and $p_\psi(\boldsymbol{z})$. In this way, VED gets around the hardness in estimating the (distribution) of $\boldsymbol{r'}$ directly from $\boldsymbol{r}$, by using the latent representation $\boldsymbol{z}$ as an intermediate relay. Moreover, representing $\boldsymbol{z}$ as a probability distribution rather than a discrete vector set, VED has a continuous latent space. Upon unseen inputs, this latent space will be sampled in a more meaningful manner than that of a conventional ED. Essentially, the continuous property of the latent space helps avoid overfitting and hence better handle out-of-range inputs.

There is yet one link missing before applying VED to separate respiration from I/Q-represented RF signal mixture: most building blocks for deep learning are based on real-valued operations and representations, how to reform VED to handle complex I/Q signals remains a problem. Previously, deep complex networks [48], [49] have been proposed to handle complex numbers, but they require redefining calculus operations including differentiation [48], so they are in general hard to train (with super slow convergence) and hence not widely adopted. The same convergence and complexity issues also apply to neural networks for sequential processing, such as general RNNs that include LSTM [50]. Consequently, our IQ-VED performs a bivariate analysis of the I/Q signal, as explained in Sec. 3.4.2.

3.4.2 IQ-VED Encoder

The encoder of IQ-VED takes in the I/Q signal matrix $\boldsymbol{r}(n)$ and encodes it to a latent representation $\boldsymbol{z}$. Specifically, IQ-VED adopts a two-stream design, where the I/Q components $\boldsymbol{r}_{\mathrm{I}}(n)$ and $\boldsymbol{r}_{\mathrm{Q}}(n)$ are fed into two separate encoders, as shown in Fig. 8.

Each encoder consists of i) multiple layers of One-Dimensional Convolutional Neural Network (1D-CNN) [51] for feature extraction, ii) instance norm layers [52] for normalization, and iii) leaky ReLU [53] layers for adding nonlinearity. Both $\boldsymbol{r}_{\mathrm{I}}(n)$ and $\boldsymbol{r}_{\mathrm{Q}}(n)$ are treated as multi-channel 1-D sequences, with each channel corresponding to one fast-time index in $\boldsymbol{r}(n)$. Since conventional convolutional layers can only look at an input sequence size linear to the network depth, we employ dilated convolutions to enable an exponentially large receptive field [54], thus facilitating respiration monitoring. In Fig. 9, one can see that at the n-th layer of the encoder/decoder, the dilation factor is 2^{n-1}, and the corresponding receptive field size is 2^{n+1}.

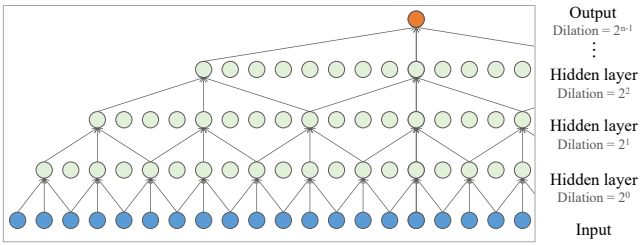


Fig. 9. Encoder and decoder of IQ-VED employ dilated convolution to increase receptive field.

Essentially, the IQ-VED encoder decomposes the input I/Q signals and filters out motion interference. The resulting respiration-induced signal is compressed and mapped to the latent distribution, which will then be sampled to drive the decoder so as to recover the desired respiratory waveform. The overall convolutional filter aims to extract useful features, so it can be deemed as a demixing function [55] to reverse the entanglement between respiration signal and nonlinear motion interference. It is well known that processing the I/Q components of complex signals separately, though substantially lowering the training complexity, may cause misalignment. To overcome this problem, we specifically align their respective latent spaces in Sec. 3.4.3.

3.4.3 Latent Space Alignment

The outputs of the encoder are two Gaussians $\boldsymbol{z}_{\mathrm{I}} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{I}}, \boldsymbol{\Sigma}_{\mathrm{I}}^2)$ and $\boldsymbol{z}_{\mathrm{Q}} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{Q}}, \boldsymbol{\Sigma}_{\mathrm{Q}}^2)$ parameterized by respective means and variances, according to Eqn. (17). Since both latent distributions are integral parts of the complex signal representation, IQ-VED should guarantee that their processing (via individual encoders) has been conducted in a coordinated manner. Fortunately, since both IQ signals are 1-D perspectives of the same complex signal, they share common structures sufficient to align their corresponding latent representations. To this end, we choose to minimize the 2-Wasserstein distance [56] between them:

$$\mathbf{W}_{\mathrm{IQ}} = \left(\inf_{J \in \mathcal{J}(\mathcal{N}_{\mathrm{I}}, \mathcal{N}_{\mathrm{Q}})} \int \|\boldsymbol{z}_{\mathrm{I}} - \boldsymbol{z}_{\mathrm{Q}}\|^2 dJ(\boldsymbol{z}_{\mathrm{I}}, \boldsymbol{z}_{\mathrm{Q}}) \right)^{\frac{1}{2}}, \quad (18)$$

where $\mathcal{J}$ denotes the set of all joint distributions J that has $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{I}}, \boldsymbol{\Sigma}_{\mathrm{I}}^2)$ and $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{Q}}, \boldsymbol{\Sigma}_{\mathrm{Q}}^2)$ as respective marginals. The reason for employing the Wasserstein distance is twofold. On one hand, minimizing the distance shifts the two distributions "close" to each other, enforcing them to encode the same respiration features. On the other hand, unlike KL divergence, Wasserstein distance is able to provide a useful gradient when the distributions are not overlapping [56]. As a result, while the two distributions are mostly aligned, some discrepancies inherent to the I/Q components, e.g., amplitude and phase are allowed to be maintained.

For multivariate Gaussian distributions, a closed-form solution of Eqn. (18) can be obtained according to [57]:

$$\mathbf{W}_{\mathrm{IQ}} = \Big[\|\boldsymbol{\mu}_{\mathrm{I}} - \boldsymbol{\mu}_{\mathrm{Q}}\|_2^2 + \mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathrm{I}}\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathrm{Q}}\right) - 2\,\mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathrm{I}}^{\frac{1}{2}} \boldsymbol{\Sigma}_{\mathrm{Q}} \boldsymbol{\Sigma}_{\mathrm{I}}^{\frac{1}{2}}\right)^{\frac{1}{2}} \Big]^{\frac{1}{2}}. \quad (19)$$

Since the covariance matrices obtained by IQ-VED are of diagonal form, Eqn. (19) can be simplified as follows:

$$\mathbf{W}_{\mathrm{IQ}} = \|\boldsymbol{\mu}_{\mathrm{I}} - \boldsymbol{\mu}_{\mathrm{Q}}\|_2 + \left\| \boldsymbol{\Sigma}_{\mathrm{I}}^{\frac{1}{2}} - \boldsymbol{\Sigma}_{\mathrm{Q}}^{\frac{1}{2}} \right\|_{\mathrm{Frob}}, \quad (20)$$

where $\|\cdot\|_{\mathrm{Frob}}$ is the Frobenius norm, defined as the square root of the sum of the squares of the matrix elements.

3.4.4 IQ-VED Decoder and Loss Function

The decoder can be deemed as the reverse of the encoder. To this end, we replace 1D-CNN in the decoder with 1-D transposed convolutional layers [58] to upsample the latent representation and map them to a longer sequence, so as to finally derive respiratory waveform $\boldsymbol{r}'(n)$. Note that the encoder and decoder are not exactly symmetric: at the last stage of the decoder, a single-channel signal is recovered, instead of a multi-channel one as the input to the encoder. To train IQ-VED, we employ three loss functions, namely the reconstruction loss, the I/Q regularizing loss, and the distribution alignment loss.

- **Reconstruction Loss.** To correctly recover the respiratory waveform, the reconstruction loss $\mathcal{L}_{\mathrm{RC}}$ aims to make the decoder output similar to the ground truth r_{gt} via an L^2 norm to measure the sum of all the squared differences between the two waveforms. This loss practically implements the term $\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{r})}\left[\log p_\psi(\boldsymbol{r'}|\boldsymbol{z})\right]$ in Eqn. (17):

$$\mathcal{L}_{\mathrm{RC}} = \|r_{\mathrm{gt}} - r'\|_2. \quad (21)$$

- **I/Q Regularizing Loss.** In Sec. 3.4.1, it is pointed out that VED regularizes the latent distribution according to a standard Gaussian prior. For IQ-VED, two distributions from the I/Q encoders should be regularized. The I/Q regularizing losses are defined as:

$$\mathcal{L}_{\mathrm{IR}} = \mathbf{D}_{\mathrm{KL}}\left(\mathcal{N}\left(\boldsymbol{\mu}_{\mathrm{I}}, \boldsymbol{\Sigma}_{\mathrm{I}}^2\right), \mathcal{N}(\mathbf{0}, \mathbf{I})\right), \quad (22)$$

$$\mathcal{L}_{\mathrm{QR}} = \mathbf{D}_{\mathrm{KL}}\left(\mathcal{N}\left(\boldsymbol{\mu}_{\mathrm{Q}}, \boldsymbol{\Sigma}_{\mathrm{Q}}^2\right), \mathcal{N}(\mathbf{0}, \mathbf{I})\right). \quad (23)$$

- **Distribution Alignment Loss.** As described in Sec. 3.4.3, the misalignment between the two distributions from the I/Q encoders can be measured by a Wasserstein distance. Therefore, the distribution alignment loss is defined according to Eqn. (20):

$$\mathcal{L}_{\mathrm{DA}} = \mathbf{W}_{\mathrm{IQ}} = \|\boldsymbol{\mu}_{\mathrm{I}} - \boldsymbol{\mu}_{\mathrm{Q}}\|_2 + \left\| \boldsymbol{\Sigma}_{\mathrm{I}}^{\frac{1}{2}} - \boldsymbol{\Sigma}_{\mathrm{Q}}^{\frac{1}{2}} \right\|_{\mathrm{Frob}}. \quad (24)$$

Combining these loss functions, the overall loss function for training IQ-VED can be obtained as follows:

$$\mathcal{L}_{\mathrm{IQ-VED}} = \mathcal{L}_{\mathrm{RC}} + \gamma(\mathcal{L}_{\mathrm{IR}} + \mathcal{L}_{\mathrm{QR}}) + \eta\mathcal{L}_{\mathrm{DA}}, \quad (25)$$

where γ and η are the respective weights for regularizing losses and distribution alignment loss. $\mathcal{L}_{\mathrm{IR}}$ and $\mathcal{L}_{\mathrm{QR}}$ share the same weight γ because the I/Q data and encoders are symmetric and of equal importance.

3.4.5 Non-human Track Deletion

Unlike previous tracking systems that have no knowledge of the tracking subjects, BreathCatcher leverages information obtained during respiration recovery to exclude non-human subjects in its analysis. Since IQ-VED is trained by only human-related data, we formulate the non-human track deletion as an anomaly detection problem [59]. To be specific, BreathCatcher first clusters the latent representations of input signals yielded by the encoder using the K-means algorithm [60]. Since "normal" latent representations are associated with human subjects and reflect properties of respiration, "abnormal" representations of non-human subjects (e.g., robots and fans) will fall far from existing clusters. BreathCatcher then calculates the distance between a new instance and clustering center, if the distance exceeds a threshold ζ_{a}, the corresponding track will be classified as an anomaly, hence be deleted.

3.4.6 Waveform Recovery

With a well-trained IQ-VED, respiratory waveform can be recovered from radar signal even under motion interference. One immediate way of applying IQ-VED to waveform recovery is to re-sample the latent vectors $\boldsymbol{z}_{\mathrm{I}}$ and $\boldsymbol{z}_{\mathrm{Q}}$ from $\mathcal{N}\left(\boldsymbol{\mu}_{\mathrm{I}}, \boldsymbol{\Sigma}_{\mathrm{I}}^2\right)$ and $\mathcal{N}\left(\boldsymbol{\mu}_{\mathrm{Q}}, \boldsymbol{\Sigma}_{\mathrm{Q}}^2\right)$. However, this leads to non-deterministic outputs that may cause problem in practice. To tackle this problem, we perform a deterministic inference without sampling $\boldsymbol{z}_{\mathrm{I}}$ and $\boldsymbol{z}_{\mathrm{Q}}$ as follows:

$$\boldsymbol{r'}^* = \arg\max_{\boldsymbol{r'}} p_\psi\left(\boldsymbol{r'}|\boldsymbol{r}_{\mathrm{I}}, \boldsymbol{r}_{\mathrm{Q}}, \boldsymbol{z}_{\mathrm{I}}^* + \boldsymbol{z}_{\mathrm{Q}}^*\right), \quad (26)$$

where the deterministic latent vectors $\boldsymbol{z}_{\mathrm{I}}^*$ and $\boldsymbol{z}_{\mathrm{Q}}^*$ are obtained as $\boldsymbol{z}_{\mathrm{I}}^* = \mathbb{E}[\boldsymbol{z}_{\mathrm{I}}|\boldsymbol{r}_{\mathrm{I}}]$ and $\boldsymbol{z}_{\mathrm{Q}}^* = \mathbb{E}[\boldsymbol{z}_{\mathrm{Q}}|\boldsymbol{r}_{\mathrm{Q}}]$. The recovered respiratory waveform contains subject-specific features which can be used for identification, hence improving the tracking functionality of BreathCatcher.

4 Implementation

In this section, we elaborate on the implementation details of BreathCatcher.

4.1 Hardware Implementations

BreathCatcher uses IR-UWB signals for simultaneous respiration and position tracking. The core components are a pair of compact and low-cost Novelda X4M05 [61] IR-UWB radar transceivers. The radar operates at a center frequency of 7.29 with a bandwidth of 1.5 GHz. The sampling rate of the radar is 23.328 GHz, and the frame rate is set to 50 fps. The radars have a pair of tx-rx antennas with an FoV of $65°$. A Raspberry Pi single-board computer [62] is used to control the transceiver and to interface with a desktop computer; this computer has an Intel Xeon W-2133 CPU, 16 GB RAM, and a GeForce RTX 2080 Ti graphics card. A laser tape measure is used to measure the ground truth positions of the human subjects, and multiple NeuLog respiration monitor belt logger sensors NUL-236 [5] are used to collect ground truth respiratory waveforms, the sampling rate of the NeuLog sensors is also set to 50 fps, the same as the radar.

4.2 Software Implementations

We implement BreathCatcher using Python 3.7 and C/C++, with the neural network components built upon PyTorch 1.7.1 [63]. To align the ground truth respiration signal from the respiration monitor belt logger and radar signals, the Precision Time Protocol [64] relying on message exchanges over Ethernet is used to synchronize the clocks between hardware components. In the data augmentation process, each signal $\boldsymbol{r}(n)$ is rotated from 0 to 2π with an interval of $\pi/30$ for 60 times. The parameters of IQ-VED are set as follows: γ and η in Eqn. (25) are set to 1 and 2e-4, respectively. For the encoder, 8 consecutive 1-D convolutional layers are used, whose kernel size is set to 3, and the dilation factors of these convolutional layers are set to 1, 2, 4, 6, 8, 16, 32, 64, and 128. As for the decoder, 8 consecutive transposed convolutional layers are used, their kernel size is set to 3, and the dilation factors of these layers are set to 128, 64, 32, 16, 8, 4, 2, and 1. All weights are initialized by the Xavier uniform initializer [65]. The collected dataset is divided into training and test sets. The training set contains 10,000 pairs of radar signal matrices and respiration ground truths obtained from the NeuLog sensors, and the test set contains 4,400 pairs. The size of the raw training radar signal matrix is 1000×138. For the training process, the batch size is set to 64, the IQ-VED loss in Eqn. (25) is adopted, and the learning rate and momentum of the Stochastic Gradient Descent optimizer [66] are respectively set to 0.01 and 0.9. To prove that the features embedded in the recovered waveforms can be used for human subject identification, a 5-layer 1-D CNN network (with a kernel size of 5) is used to classify the recovered waveforms of the 12 human subjects in the evaluation.

5 Evaluation

In this section, we evaluate BreathCatcher given several real-life scenarios and under various parameter settings.

5.1 Experiment Setup

The experiment setup is shown in Fig. 10. The monitoring area of the radar pair is approximately 21 m^2 (3 m × 7 m). We recruit 12 subjects (6 females and 6 males), aged from 15 to 64, and weighing from 50 to 80 kg. All subjects are healthy, and their respirations are measured in natural states without consciously controlling breathing or undergoing external forceful intervention. We ask 1 to 3 subjects to walk in the monitoring area randomly to simulate everyday indoor scenarios. The IR-UWB radar pair is placed to face the subjects, and on the same height as the subjects' chests. All experiments have strictly followed the standard procedures of IRB of our institute. Our data collection leads to an 80 man-hour dataset of RF and ground truth recordings, including approximately 70,000 respiration cycles and roughly the same amount of data from each subject. After collection, both the raw and ground truth data are sliced into 20 s samples. Two-thirds of the collected data (of 8 subjects) are used for training IQ-VED, and the remaining one-third (of 4 subjects) are used for testing the performance of BreathCatcher in recovering respiratory waveform. These two datasets are made such that the subjects are disjoint, so the testing subjects are "unseen" to our trained model.

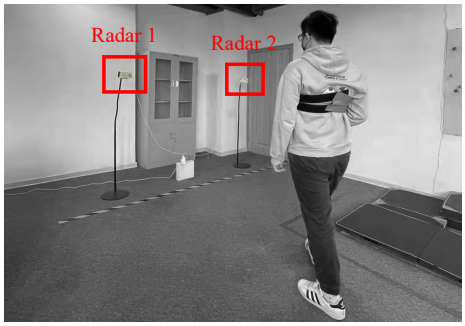


Fig. 10. A snapshot of our experiment setting, with two radars and one walking subject.

We further define the following metrics:

- **Tracking distance error.** As explained in Sec. 3.1, subjects in the radar FoV can be tracked by an algorithm leveraging range information from the radar pair. To study the tracking performance of BreathCatcher, we calculate the tracking distance error, i.e., the distance between the estimated subjects' positions and the true positions measured by a laser measure $|\boldsymbol{x}^{\mathrm{e}} - \boldsymbol{x}^{\mathrm{a}}|$.
- **Cosine similarity.** The cosine similarity $\mathcal{S}(\boldsymbol{r}', \boldsymbol{r}_{\mathrm{gt}})$ between the IQ-VED recovered waveform $\boldsymbol{r}'(n)$ and the ground truth $\boldsymbol{r}_{\mathrm{gt}}(n)$ is used to measure the recovering performance of IQ-VED. Specifically, the cosine similarity is measured by the cosine of the angle between two vectors $\boldsymbol{r}'(n)$ and $\boldsymbol{r}_{\mathrm{gt}}(n)$, and then determines to what extent the two vectors point to the same "direction" in a high dimensional space. It is defined as follows:

$$\mathcal{S}(\boldsymbol{r}', \boldsymbol{r}_{\mathrm{gt}}) = \frac{\sum_{n=1}^{N} \boldsymbol{r}'(n)\boldsymbol{r}_{\mathrm{gt}}(n)}{\sqrt{\sum_{n=1}^{N} \boldsymbol{r}'^{2}(n)}\sqrt{\sum_{n=1}^{N} \boldsymbol{r}_{\mathrm{gt}}^{2}(n)}}. \tag{27}$$

- **Time estimation error.** Several time-related biomarkers, including total cycle time, inspiratory and expiratory time can be calculated from peak and valley times in the waveforms. Therefore, we study estimation errors in terms of

the peak and valley times. Suppose the timestamps of the peak and valley are t_{p} and t_{v}, then the errors are defined as absolute differences between the estimated and actual values, i.e., $|t_{\mathrm{p}}^{\mathrm{e}} - t_{\mathrm{p}}^{\mathrm{a}}|$ and $|t_{\mathrm{v}}^{\mathrm{e}} - t_{\mathrm{v}}^{\mathrm{a}}|$.

- **Respiratory rate estimation error.** After recovering the respiratory waveforms, we can obtain the respiratory rate ρ_{R} by finding the strongest component in the frequency spectrum. The error of respiratory rate is defined as the absolute difference between the estimated respiratory rate $\rho_{\mathrm{R}}^{\mathrm{e}}$ and the actual respiratory rate $\rho_{\mathrm{R}}^{\mathrm{a}}$, namely, $|\rho_{\mathrm{R}}^{\mathrm{e}} - \rho_{\mathrm{R}}^{\mathrm{a}}|$.

We choose BreathListener [67] as the baseline for comparison. BreathListener is designed for respiration monitoring in driving environments, so it claims to be able to handle motion interference (albeit only small-scale ones) caused by steering vehicles. However, we have to port BreathListener to radar as it was designed for acoustic sensing. Essentially, BreathListener adopts a two-stage processing pipeline. It first employs EEMD [68] to separate respiration from interference. Since EEMD can only recover coarse-grained waveform (if not incorrect one), BreathListener further applies a Generative Adversarial Network (GAN) [69] for adding details to the recovered waveform.

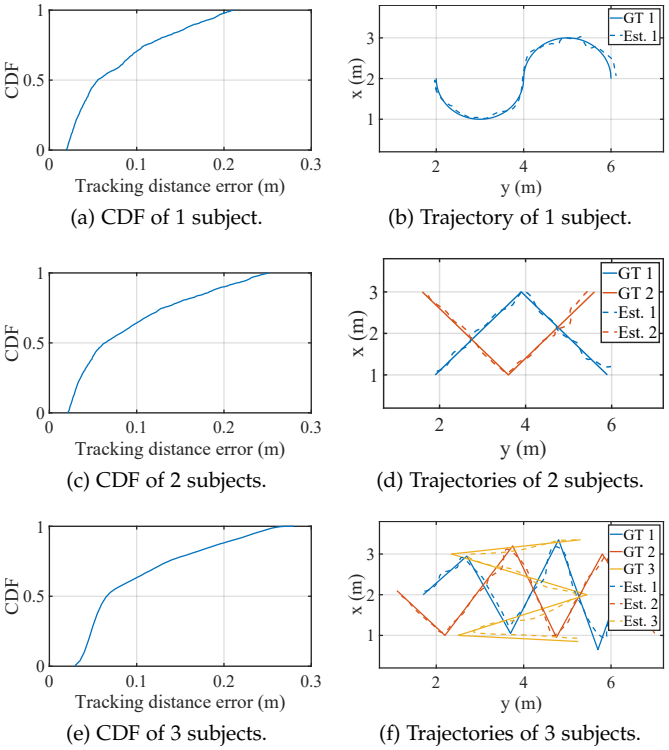
5.2 Overall Performance

We evaluate the overall performance of BreathCatcher in this section, focusing on subject tracking, waveform recovery, and subject identification, respectively.

5.2.1 Human Subject Tracking

The tracking results of up to 3 human subjects are shown in Fig. 11. For a different number of subjects, we first show the overall tracking distance errors as CDF, then demonstrate the performance by letting the several subjects walk wavy and zig-zag trajectories simultaneously. In Fig. 11a, 11c, and 11e, it can be seen that the median tracking distance errors are 0.06 m, 0.06 m, and 0.07 m for 1, 2, and 3 subjects, respectively. The small discrepancies among the tracking distance errors prove that BreathCatcher is insensitive and robust to the number of human subjects. Fig. 11b, 11d, and 11f show some tracking examples with the ground truth and estimated trajectories denoted as "GT" and "Est.", respectively. The trajectories are also color-coded and numbered for each subject. It can be seen that trajectories of different human subjects are clearly differentiated, and the estimated ones are very close to the ground truths, confirming the capability of BreathCatcher in trajectory tracking.

Due to the sensing range limitation imposed by RF power regulations, we have only tested BreathCatcher in a limited area. Nonetheless, as shown in Fig. 11b, 11d, and 11f, we strive to use the radar pair to cover the largest possible area of 21 m^2 (3m × 7m), which is sufficient for typical indoor usage. Also, we believe it is reasonable to test BreathCatcher on a relatively smaller scale, as the current prototype with a pair of radar is only used for conceptual verification. As a consequence of the limited radar coverage, we employ at most 3 human subjects in the experiment to avoid overcrowding the area and complicating data collection. With that being said, we are also actively exploring possibilities of using more radars to increase the coverage. While more radars and subjects can be readily incorporated into the current algorithm framework, certain engineering details still need to be addressed, so we leave such extensions to our future work.

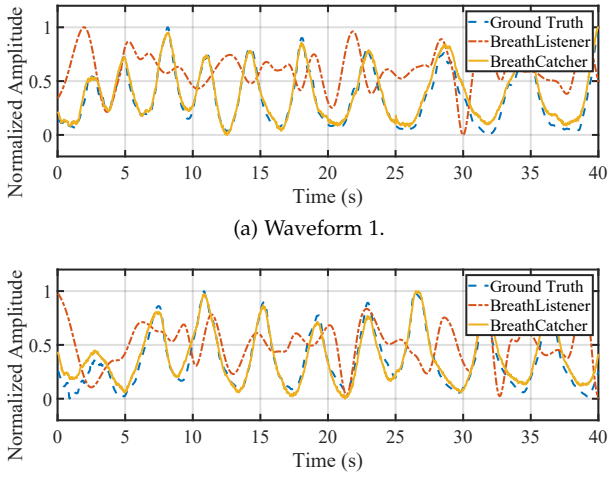


(a) CDF of 1 subject. (b) Trajectory of 1 subject.
(c) CDF of 2 subjects. (d) Trajectories of 2 subjects.
(e) CDF of 3 subjects. (f) Trajectories of 3 subjects.

Fig. 11. CDF of tracking distance errors and example tracking results of zig-zag walking.

5.2.2 Waveform Recovery

After tracking signals associated with human subjects, we show some intuitive recovery results. Fig. 12a and 12b show the respiratory waveform generated by IQ-VED and the baseline method BreathListener, compared to the corresponding ground truth versions during two 40 s periods. In general, BreathCatcher recovers respiratory waveform accurately due to the robust design, whereas BreathListener tends to generate distorted and even erroneous waveforms when the human subjects are walking.

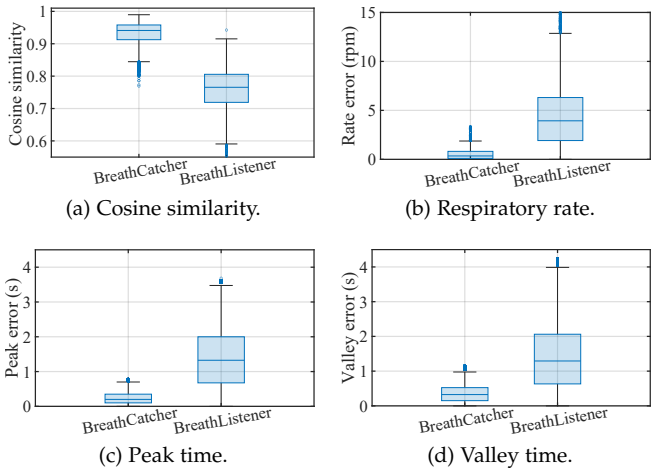


(a) Waveform 1.
(b) Waveform 2.

Fig. 12. Two examples of waveform recovery results.

5.2.3 Estimation Errors of Indicators

Given the waveform results of BreathCatcher reported in the previous sub-section, we hereby pay special attention to the estimation errors of several indicators. Naturally, the cosine similarities of the recovered waveforms are shown in Fig. 13a. Overall, the average cosine similarity between the recovered and ground truth respiratory waveform is 0.929, indicating a very successful recovery, as a similarity greater than 0.8 suggests a strong positive correlation. The performance of BreathCatcher in estimating instantaneous respiratory rate is then evaluated in Fig. 13b, showing a very consistent accuracy with the mean respiratory rate errors being under 0.4 bpm. To evaluate time-related biomarkers, we inspect the estimation errors of the peak and valley times on respiratory waveform, and the results are shown in Fig. 13c and 13d. It can be observed that most of the mean errors are below 0.3 s, indicating high accuracy of BreathCatcher's event time estimation. An interesting phenomenon is that for BreathCatcher, the errors of valley time are noticeably larger than those of the peak time; this can be attributed to the fact that the valleys in the waveform, as shown in Fig. 12a, are relatively "flatter" than the peaks, thus making it harder for IQ-VED to capture and recover the exact times of the valleys. One may observe that Breath-



(a) Cosine similarity. (b) Respiratory rate.

(c) Peak time. (d) Valley time.

Fig. 13. Estimation errors of several indicators.

Catcher outperforms the baseline method BreathListener on all benchmarks as shown in Fig. 13a, 13b, 13c, and 13d.

The inferior performance of BreathListener can be attributed to the mismatch between EEMD and GAN adopted by it. The EEMD algorithm is incapable of handling complex I/Q signals, so one has to first project the I/Q signals to a 1-D sequence in an information-lossy manner. A consequence is that motion interference cannot be correctly separated, as illustrated in Fig. 12a. Given the potentially erroneous decomposition of EEMD, GAN that already suffers from instability during training [70] becomes even harder to converge. For those converged cases, the EEMD decomposed waveform is already close to ground truth, though possibly with wrong features (e.g., phase) that GAN barely helps to correct. Consequently, the biomarkers inferred from BreathListener recovered waveforms can have very large errors. On the contrary, IQ-VED is trained and operates in an integrated manner: it uses the encoder to decompose signal and the decoder to reconstruct the respiratory waveform. As a result, BreathCatcher is far more effective than the baseline, as demonstrated by these comparisons.

5.2.4 Human Subject Identification

The features embedded in the recovered respiratory waveforms contain information unique to each human subject, hence can be used to identify and track different subjects. We employ a 5-layer 1-D CNN network to perform classification, and the confusion matrix is shown in Fig. 14. It

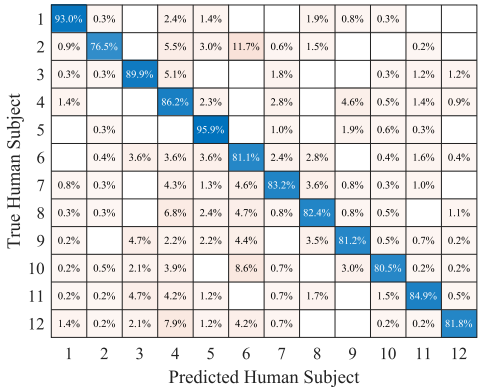


Fig. 14. Confusion matrix of human subject identification.

can be seen that classification results are consistent among all human subjects, and the average classification accuracy is 84.72%, indicating that subject-specific features are preserved in the respiratory waveform recovered by BreathCatcher, thus leading to successful human identification.

5.3 Impact of Practical Factors

In this section, we study the impact of different practical factors. Because all biomarkers can all be inferred from respiratory waveforms, we hereafter only focus on evaluating the cosine similarity of the waveforms.

5.3.1 Human Subjects

We show the cosine similarities of BreathCatcher recovered respiratory waveform for all 12 subjects in Fig. 15. Based on the figure, one may readily conclude that the mean cosine similarities are always around 0.95, and more than 75% of all similarities are above 0.85. These results show that respiratory waveform recovery of BreathCatcher remains accurate across all involved subjects, largely insensitive to physical discrepancies among them.

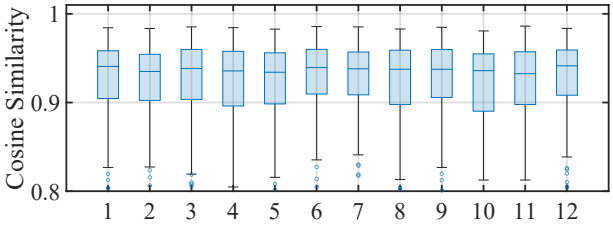
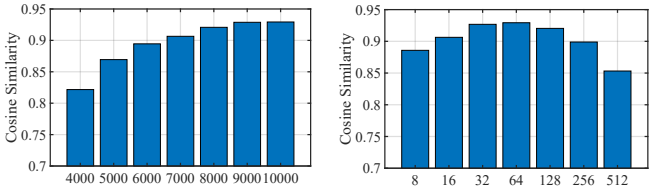


Fig. 15. Impact of different human subjects on the cosine similarity.

5.3.2 Training Set Size

As stated in Sec. 4, we collect 10,000 data samples of human subjects walking for training the IQ-VED network of BreathCatcher. Fig. 16a shows the impact of training set size on the cosine similarity between the recovered and ground truth waveforms. One may observe that, as the training set size increases, the cosine similarity first increases and then comes to saturation. Specifically, BreathCatcher achieves a cosine similarity greater than 0.9 with 7,000 training samples, which corresponds to 39 hours of activity data. Because more training data improve the waveform recovery performance only by a negligible margin, our selection of 10,000 training samples is sufficient.



(a) Training set size. (b) Latent space dimension.

Fig. 16. Impact of training set size (a) and latent space dimension (b).

5.3.3 Latent Space Dimension

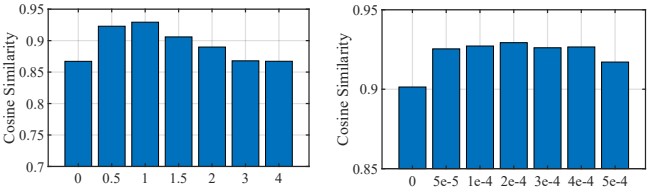
Another property of IQ-VED that affects the recovery performance is the number of latent space dimensions. On one hand, a small latent space dimension may limit the capacity of the latent representation and potentially prevent the loss function from converging to a sufficiently small value. On the other hand, as most practical signals are sparse, overly increasing the dimension of the latent space can be unnecessary while causing slow convergence in training. Consequently, a competent system should strike a balance between expressiveness and compactness of the latent space. According to Fig. 16b that shows the impact of latent space dimension on the cosine similarity between the recovered and ground truth waveforms, the performance first improves with the dimension thanks to a better expressiveness but degrades after the dimension reaches 64 due to the increased hardness in training. Therefore, 64 is chosen as the dimension of the latent space for IQ-VED.

5.3.4 Weights of the Loss Function

The weights in Eqn. (25) are crucial parameters to be tuned for IQ-VED. In theory, a larger γ encourages continuity and disentanglement of the latent space, potentially improving the generalization capability of IQ-VED. A larger η improves the alignment of the I/Q representations but may restrict their expressiveness of the underlying I/Q signal. To determine the optimal weights, Fig. 17 shows the cosine similarity between the recovered and ground truth waveforms as the functions of individual weights; one can clearly observe that $\gamma = 1$ and $\eta = 2 \times 10^{-4}$ allow IQ-VED to achieve the best performance in waveform recovery.

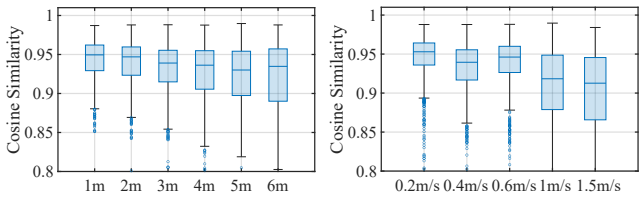
5.3.5 Sensing Distance

Sensing distance is a major limiting factor of the RF respiration monitoring system. We ask the human subjects to be away from the center of the radar pair at 1 m, 2 m, 3 m, 4 m, 5 m, and 6 m to study the impact of sensing distance. Not surprisingly, the cosine similarities reported in Fig. 18a clearly demonstrate a negative effect of sensing distance on performance. Nonetheless, the average cosine similarity remains above 0.9 even at the furthest distance, firmly proving the effectiveness of BreathCatcher.



(a) γ (when $\eta = 2 \times 10^{-4}$). (b) η (when $\gamma = 1$).

Fig. 17. Impact of different weights of the loss function.



(a) Sensing distance. (b) Walking speed.

Fig. 18. Impact of sensing distance (a) and walking speed (b).

5.3.6 Walking Speed

The moving speed of the walking human subject also affects the sensing performance. We ask the human subjects to walk at speeds of 0.2 m/s, 0.4 m/s, 0.6 m/s, 1 m/s, and 1.5 m/s to study the impact of walking speed. We find that the walking speed negatively correlates with the recovery performance. Fig. 18b demonstrates this effect. However, even if the human subject is walking at the highest walking speed, 75% of the cosine similarities remain above 0.85 indicating successful respiration recovery. This result demonstrates that BreathCatcher can be used for everyday respiration monitoring of walking human subjects.

6 Related Work

Respiration Monitoring. A substantial amount of research on contact-free respiration monitoring has been developed in the past decade. Although there are some technologies such as computer vision [71], [72]) and sound [67], [73], [74] that can achieve contract-free respiration monitoring, they all face various challenges such as privacy leakage and insufficient robustness. Recently, RF becomes the most popular mainstream sensing medium for respiration monitoring. RF-sensing technologies can be roughly categorized into three types according to different kinds of platforms: Wi-Fi, SDR, and commodity-grade radars.

Wi-Fi-based respiration monitoring [11], [12], [13], [14] estimate respiratory rates of subjects located between the transmitter and receiver by CSI variations obtained from Wi-Fi cards, such as Intel 5300 [75]. Due to their narrow bandwidth, Wi-Fi systems have coarse spatial resolution [18]. Consequently, the systems cannot resolve multiple

paths and multiple subjects, and has to employ multiple additional Wi-Fi cards to achieve improved spatial diversity [76]. Moreover, traditional Wi-Fi systems are designed for communication but not for sensing applications, and as a result, Wi-Fi sensing-based systems may not always work well, since the packets of sensing Wi-Fi cards are usually interrupted in a crowded Wi-Fi channel [77].

Some proposals rely on SDR platforms such as USRP [8], [9], [10] for their system development. Due to the flexibility of SDR, researchers can design arbitrary modulated signals (e.g. FMCW) with a large bandwidth, and employ a large-scale antenna array to achieve higher spatial resolution. Therefore, the authors in [8], [10] distinguish interference and respiration signals readily, and get good performance. But the SDR solutions are very expensive, and often large in size. It is hard to apply in our daily life.

Given the unsatisfactory progress from the above two solutions, researchers find potentials in commodity-grade radars for respiration monitoring. They have utilized commodity-grade radars to estimate respiration rate via conventional digital signal processing algorithms [15], [16], [17], [18]. More importantly, more fine-grained vital sign waveforms have been recovered in recent papers [78], [79], [80] by taking advantage of deep learning techniques. Most of the above works claim their systems make big progress towards even clinic-level applications. However, all of them focus only on relatively static subjects, and cannot track the respiration of the moving subjects in a large-scale area.

Passive Tracking. There is a substantial amount of research work studying passive tracking in the recent decade. Wi-Fi-based passive tracking systems leverage the CSI or RSSI to track the human subjects, but they also need to deploy high-density Wi-Fi anchors [26], [27], [81], [82]. The authors in [25] have proposed Wi-Vi based on beamforming nulling technique to cancel static objects reflection signals, and estimate the direction of the subjects' movement, but Wi-Vi does not have any location information. To combat this drawback, they have also designed WiTrack [23], [24] to cover a 1.79 GHz bandwidth, thus obtaining a very high distance resolution. Similar to WiTrack, some other systems such as [83], [84], [85] also utilize more than 1 GHz to track the subjects. Although the above works may track the subjects precisely, none of them can recover the respiration signals during the subject's movement such as walking.

7 Conclusion

Taking an important step toward continuous and ubiquitous health care, we have proposed BreathCatcher, an RF sensing system for simultaneously tracking users indoors while recovering their fine-grained respiratory waveform. Built upon a commercial-grade compact radar pair, BreathCatcher cannot only discover respiratory signals of multiple human subjects when they are walking, but also identify and track each subject by associating signals from the radar pair according to their properties. To further refine the signals, we design a novel IQ-VED architecture to disentangle the nonlinear RF mixture and recover fine-grained respiratory waveforms. Via extensive experiments on healthy subjects, we have demonstrated the promising performance of BreathCatcher in simultaneous RF tracking and fine-grained waveform recovery. As this line of work progresses, we are planning to evaluate BreathCatcher's performance in real-life clinical scenarios, as we believe this work has significant implications to various medical applications including pulmonary disease diagnosis.

Acknowledgments

We would like to thank WiRUSH [86] for providing funds to develop BreathCatcher.

References

[1] D. Blumenthal, E. Malphrus, and J. M. McGinnis, *Vital Signs: Core Metrics for Health and Health Care Progress*. National Academies Press, 2015.

[2] J. G. Chester and J. L. Rudolph, "Vital Signs in Older Patients: Age-related Changes," *Journal of the American Medical Directors Association*, vol. 12, no. 5, pp. 337–343, 2011.

[3] L. Brochard, G. S. Martin, L. Blanch, P. Pelosi, F. J. Belda, A. Jubran, L. Gattinoni, J. Mancebo, V. M. Ranieri, J.-C. M. Richard *et al.*, "Clinical Review: Respiratory Monitoring in the ICU-A Consensus of 16," *Critical Care*, vol. 16, no. 2, pp. 1–14, 2012.

[4] M. Folke, L. Cernerud, M. Ekström, and B. Hök, "Critical Review of Non-Invasive Respiratory Monitoring in Medical Care," *Medical and Biological Engineering and Computing*, vol. 41, no. 4, pp. 377–383, 2003.

[5] NeuLog, "Respiration Monitor Belt Logger Sensor NUL-236," https://neulog.com/respiration-monitor-belt/, 2017, accessed: 2021-10-12.

[6] GIGAOM, "Could a Breath-monitoring Headset Improve Your Health?" https://gigaom.com/2013/09/20/could-a-breath-monitoring-headset-improve-your-health/, 2013, accessed: 2021-10-12.

[7] H. Gokalp and M. Clarke, "Monitoring Activities of Daily Living of the Elderly and the Potential for its Use in Telecare and Telehealth: a Review," *Telemedicine and e-Health*, pp. 910–923, 2013.

[8] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart Homes that Monitor Breathing and Heart Rate," in *Proc. of the 33rd ACM CHI*, 2015, pp. 837–846.

[9] A. M. Ashleibta, Q. H. Abbasi, S. A. Shah, A. Khalid, N. A. AbuAli, and M. A. Imran, "Non-Invasive RF Sensing for Detecting Breathing Abnormalities using Software Defined Radios," *IEEE Sensors Journal*, 2020.

[10] S. Yue, H. He, H. Wang, H. Rahul, and D. Katabi, "Extracting Multi-Person Respiration from Entangled RF Signals," in *Proc. of the 20th ACM UbiComp*, 2018, pp. 86:1–22.

[11] Y. Zeng, D. Wu, R. Gao, T. Gu, and D. Zhang, "FullBreathe: Full Human Respiration Detection Exploiting Complementarity of CSI Phase and Amplitude of WiFi Signals," *Proc. of the 18th ACM UbiComp*, p. 148, 2018.

[12] X. Li, D. Zhang, J. Xiong, Y. Zhang, S. Li, Y. Wang, and H. Mei, "Training-free Human Vitality Monitoring using Commodity Wi-Fi Devices," *Proc. of the 18th ACM UbiComp*, pp. 1–25, 2018.

[13] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang, "Monitoring Vital Signs and Postures During Sleep using WiFi Signals," *IEEE Internet of Things Journal*, pp. 2071–2084, 2018.

[14] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking Vital Signs during Sleep Leveraging Off-the-shelf WiFi," in *Proc. of the 16th ACM MobiHoc*, 2015, pp. 267–276.

[15] O. Kaltiokallio, H. Yiğitler, R. Jäntti, and N. Patwari, "Non-Invasive Respiration Rate Monitoring using a Single COTS TX-RX Pair," in *Proc. of the 13th ACM IPSN*. IEEE, 2014, pp. 59–69.

[16] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Vital Sign and Sleep Monitoring using Millimeter Wave," *ACM Transactions on Sensor Networks*, vol. 13, no. 2, pp. 1–32, 2017.

[17] P. Nguyen, X. Zhang, A. Halbower, and T. Vu, "Continuous and Fine-Grained Breathing Volume Monitoring from Afar Using Wireless Signals," in *Proc. of the 35th IEEE INFOCOM*, 2016, pp. 1–9.

[18] T. Zheng, Z. Chen, C. Cai, J. Luo, and X. Zhang, "V^{2}iFi: in-Vehicle Vital Sign Monitoring via Compact RF Sensing," in *Proc. of the 22nd ACM UbiComp*, 2020, pp. 70:1–27.

[19] T. Zheng, Z. Chen, S. Zhang, C. Cai, and J. Luo, "MoRe-Fi: Motion-robust and Fine-grained Respiration Monitoring via Deep-Learning UWB Radar," in *Proc. of the 19th ACM SenSys*, 2021, pp. 111–124.

[20] C. Zhang, F. Li, J. Luo, and Y. He, "iLocScan: Harnessing Multipath for Simultaneous Indoor Source Localization and Space Scanning," in *Proc. of the 12nd ACM SenSys*, 2014, pp. 91–104.

[21] Z. Chen, Z. Li, X. Zhang, G. Zhu, Y. Xu, J. Xiong, and X. Wang, "AWL: Turning Spatial Aliasing from Foe to Friend for Accurate WiFi Localization," in *Proc. of the 13th ACM CoNEXT*, 2017.

[22] Z. Chen, G. Zhu, S. Wang, Y. Xu, J. Xiong, J. Zhao, J. Luo, and X. Wang, "M^3: Multipath Assisted Wi-Fi Localization with a Single Access Point," *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 588–602, 2019.

[23] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D Tracking via Body Radio Reflections," in *Proc. of the 10th USENIX NSDI*, 2014, pp. 317–329.

[24] F. Adib, Z. Kabelac, and D. Katabi, "Multi-Person Localization via RF Body Reflections," in *Proc. of the 11th USENIX NSDI*, 2015, pp. 279–292.

[25] F. Adib and D. Katabi, "See Through Walls with WiFi!" in *Proc. of the ACM SIGCOMM*, 2013, pp. 75–86.

[26] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "Widar: Decimeter-level Passive Tracking via Velocity Monitoring with Commodity Wi-Fi," in *Proc. of the 18th ACM MobiHoc*, 2017, pp. 1–10.

[27] J. Wang, H. Jiang, J. Xiong, K. Jamieson, X. Chen, D. Fang, and B. Xie, "LiFS: Low Human-effort, Device-free Localization with Fine-grained Subcarrier Information," in *Proc. of the 22nd ACM MobiCom*, 2016, pp. 243–256.

[28] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "FarSense: Pushing the Range Limit of WiFi-based Respiration Sensing with CSI Ratio of Two Antennas," in *Proc. of the 21st ACM UbiComp*, 2019, pp. 1–26.

[29] Z. Chen, T. Zheng, C. Cai, and J. Luo, "MoVi-Fi: Motion-robust Vital Signs Waveform Recovery via Deep Interpreted RF Sensing," in *Proc. of the 27th ACM MobiCom*, 2021, p. 392–405.

[30] Z. Chen, T. Zheng, and J. Luo, "Octopus: A Practical and Versatile Wideband MIMO Sensing Platform," in *Proc. of the 27th ACM MobiCom*, 2021, p. 601–614.

[31] G. Bishop, G. Welch *et al.*, "An Introduction to the Kalman Filter," in *Proc of SIGGRAPH, Course*, vol. 8, no. 27599-23175, 2001, p. 41.

[32] T. Zheng, Z. Chen, S. Ding, and J. Luo, "Enhancing RF Sensing with Deep Learning: A Layered Approach," *IEEE Communications Magazine*, vol. 59, no. 2, pp. 70–76, 2021.

[33] S. Ding, Z. Chen, T. Zheng, and J. Luo, "RF-Net: A Unified Meta-Learning Framework for RF-Enabled One-Shot Human Activity Recognition," in *Proc. of the 18th ACM SenSys*, 2020, p. 517–530.

[34] N. Levanon, *Radar Principles*. Wiley, 1988.

[35] A. De Groote, M. Wantier, G. Chéron, M. Estenne, and M. Paiva, "Chest Wall Motion during Tidal Breathing," *Journal of Applied Physiology*, vol. 83, no. 5, pp. 1531–1537, 1997.

[36] A. Lee, X. Gao, J. Xu, and O. Boric-Lubecke, "Effects of Respiration Depth on Human Body Radar Cross Section Using 2.4 GHz Continuous Wave Radar," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 4070–4073.

[37] A. Lazaro, D. Girbau, and R. Villarino, "Techniques for Clutter Suppression in the Presence of Body Movements during the Detection of Respiratory Activity through UWB Radars," *Sensors*, vol. 14, no. 2, pp. 2595–2618, 2014.

[38] R. Popoli and S. S. Blackman, *Design and Analysis of Modern Tracking Systems*. Artech House Radar Library, 1999.

[39] F. Thomas and L. Ros, "Revisiting Trilateration for Robot Localization," *IEEE Transactions on Robotics*, vol. 21, no. 1, pp. 93–101, 2005.

[40] A. Hyvärinen and P. Pajunen, "Nonlinear Independent Component Analysis: Existence and Uniqueness Results," *Neural Netw.*, vol. 12, no. 3, p. 429–439, 1999.

[41] C. Jutten and J. Karhunen, "Advances in Nonlinear Blind Source Separation," in *Proc. of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, 2003, pp. 245–256.

[42] Y. Luo and N. Mesgarani, "TasNet: Time-Domain Audio Separation Network for Real-Time, Single-Channel Speech Separation," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 696–700.

[43] S. I. Mimilakis, K. Drossos, T. Virtanen, and G. Schuller, "A Recurrent Encoder-Decoder Approach with Skip-Filtering Connections for Monaural Singing Voice Separation," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2017, pp. 1–6.

[44] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational Inference: A Review for Statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[45] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *Proc. of ICLR*, 2013, pp. 1–14.

[46] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic Backpropagation and Approximate Inference in Deep Generative Models," in *Proc. of ICML*, 2014, pp. 1278–1286.

[47] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[48] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep Complex Networks," in *Proc. of ICLR*, 2018, pp. 1–19.

[49] T. E. Potok, C. Schuman, S. Young, R. Patton, F. Spedalieri, J. Liu, K.-T. Yao, G. Rose, and G. Chakma, "A Study of Complex Deep Learning Networks on High-performance, Neuromorphic, and Quantum Computers," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 14, no. 2, pp. 1–21, 2018.

[50] T. Trinh, A. Dai, T. Luong, and Q. Le, "Learning Longer-Term Dependencies in RNNs with Auxiliary Losses," in *Proc. of ICML*. PMLR, 2018, pp. 4965–4974.

[51] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object Recognition with Gradient-Based Learning," in *Shape, Contour and Grouping in Computer Vision*. Springer, 1999, pp. 319–345.

[52] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance Normalization: The Missing Ingredient for Fast Stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[53] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," in *Proc. of the 14th AISTATS*, 2011, pp. 315–323.

[54] F. Yu and V. Koltun, "Multi-scale Context Aggregation by Dilated Convolutions," *Proc. of ICLR*, 2016.

[55] I. Khemakhem, D. Kingma, R. Monti, and A. Hyvarinen, "Variational Autoencoders and Nonlinear ICA: A Unifying Framework," in *Proc. of the 20th AISTATS*. PMLR, 2020, pp. 2207–2217.

[56] I. Olkin and F. Pukelsheim, "The Distance between Two Random Vectors with Given Dispersion Matrices," *Linear Algebra and its Applications*, vol. 48, pp. 257–263, 1982.

[57] C. R. Givens, R. M. Shortt *et al.*, "A Class of Wasserstein Metrics for Probability Distributions," *The Michigan Mathematical Journal*, vol. 31, no. 2, pp. 231–240, 1984.

[58] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional Networks," in *Proc. of the 23rd IEEE CVPR*. IEEE, 2010, pp. 2528–2535.

[59] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[60] S. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[61] N. AS, "Single-Chip Radar Sensors with Sub-mm Resolution - XETHRU," https://www.xethru.com/, 2017, accessed: 2021-10-12.

[62] Raspberry Pi Foundation, "Teach, Learn and Make with Raspberry Pi - Raspberry Pi," https://https://www.raspberrypi.org/, 2021, accessed: 2021-04-28.

[63] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," *arXiv preprint arXiv:1912.01703*, 2019.

[64] IEFT, "Precision Time Protocol Version 2 (PTPv2)," 2017, accessed: 2021-04-30.

[65] S. K. Kumar, "On Weight Initialization in Deep Neural Networks," *arXiv preprint arXiv:1704.08863*, 2017.

[66] L. Bottou, "Stochastic Gradient Descent Tricks," in *Neural Networks: Tricks of the Trade*. Springer, 2012, pp. 421–436.

[67] X. Xu, J. Yu, Y. Chen, Y. Zhu, L. Kong, and M. Li, "BreathListener: Fine-Grained Breathing Monitoring in Driving Environments Utilizing Acoustic Signals," in *Proc. of the 17th ACM MobiSys*, 2019, p. 54–66.

[68] Z. Wu and N. E. Huang, "Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method," *Advances in Adaptive Data Analysis*, vol. 1, no. 01, pp. 1–41, 2009.

[69] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," in *Proc. of The 27th NIPS*, 2014, pp. 1–9.

[70] K. Shmelkov, C. Schmid, and K. Alahari, "How Good is My GAN?" in *Proc. of the 15th ECCV*, 2018, pp. 213–229.

[71] W. Chen and D. McDuff, "DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks," in *Proc. of the 15th IEEE ECCV*, 2018, pp. 349–365.

[72] M. Bartula, T. Tigges, and J. Muehlsteff, "Camera-based System for Contactless Monitoring of Respiration," in *Proc. of the IEEE EMBC*. IEEE, 2013, pp. 2672–2675.

[73] H. Wan, S. Shi, W. Cao, W. Wang, and G. Chen, "RespTracker: Multi-user Room-scale Respiration Tracking with Commercial Acoustic Devices," in *Proc. of 40th IEEE INFOCOM*, 2021, pp. 1–10.

[74] A. Wang, J. E. Sunshine, and S. Gollakota, "Contactless Infant Monitoring Using White Noise," in *Proc. of The 25th ACM MobiCom*, 2019, pp. 52:1–16.

[75] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool Release: Gathering 802.11 n Traces with Channel State Information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, 2011.

[76] P. Wang, B. Guo, T. Xin, Z. Wang, and Z. Yu, "TinySense: Multi-user Respiration Detection using Wi-Fi CSI Signals," in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 2017, pp. 1–6.

[77] D. C. Salyers, A. D. Striegel, and C. Poellabauer, "Wireless Reliability: Rethinking 802.11 Packet Loss," in *2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks*. IEEE, 2008, pp. 1–4.

[78] F. Lin, C. Song, Y. Zhuang, W. Xu, C. Li, and K. Ren, "Cardiac Scan: A Non-Contact and Continuous Heart-Based User Authentication System," in *Proc. of the 23rd ACM MobiCom*, 2017, p. 315–328.

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

[80] U. Ha, S. Assana, and F. Adib, "Contactless Seismocardiography via Deep Learning Radars," in *Proc. of the 26th ACM MobiCom*, 2020, pp. 1–14.

[81] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free Location-oriented Activity Identification using Fine-grained WiFi Signatures," in *Proc. of the 20th ACM MobiCom*, 2014, pp. 617–628.

[82] M. Youssef, M. Mah, and A. Agrawala, "Challenges: Device-free Passive Localization for Wireless Environments," in *Proc. of the 13th ACM MobiCom*, 2007, pp. 222–229.

[83] Y. Jia, L. Kong, X. Yang, and K. Wang, "Through-wall-radar Localization for Stationary Human based on Life-sign Detection," in *Proc. of the IEEE RadarCon*, 2013, pp. 1–4.

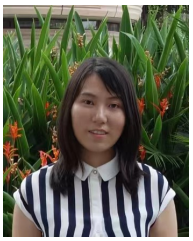
[84] N. Maaref, P. Millot, C. Pichot, and O. Picon, "A Study of UWB FMCW Radar for the Detection of Human Beings in Motion Inside a Building," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1297–1300, 2009.

[85] Y. Zhou, C. L. Law, Y. L. Guan, and F. Chin, "Localization of Passive Target based on UWB Backscattering Range Measurement," in *Proc. of the IEEE UWB*, 2009, pp. 145–149.

[86] WiRUSH/AIWiSe, "Guangxi Wanyun and Guangzhou AIWiSe Technology Co., Ltd," https://www.wirush.ai and https://aiwise.wirush.ai, 2019.

Tianyue Zheng (tianyue002@e.ntu.edu.sg) received his B.Eng. degree in Telecommunication Engineering from Harbin Institute of Technology, China, and M.Eng. degree in Computer Engineering from the University of Toronto, Canada. He is currently a Ph.D. candidate at the School of Computer Science and Engineering and the Energy Research Institute, Nanyang Technological University, Singapore. His research interests include mobile and pervasive computing, the Internet of Things, and machine learning. More information can be found at https://tianyuez.github.io/index.html.

Zhe Chen (chenz@ssijri.com) is the co-founder of AIWiSe. He obtained his Ph.D. degree in Computer Science from Fudan University, China, with a 2019 ACM SIGCOMM China Doctoral Dissertation Award. Before joining AIWiSe, he worked as a research fellow in NTU for two years, and his research achievements, along with his efforts in launching products based on them, have thus earned him 2021 ACM SIGMOBILE China Rising Star Award recently. His current research interests include wireless networking, deep learning, mobile and pervasive computing, and embedded systems. More information can be found at https://rabbitnick.github.io/index.html.



Shujie Zhang (shujie002@e.ntu.edu.sg) received the B.Eng. degree in Electrical and Electronic Engineering and the M.S. degree in Signal Processing from Nanyang Technological University, Singapore, in 2017 and 2019 respectively. She is currently working towards her Ph.D. degree in Computer Science and Engineering from Nanyang Technological University, Singapore. Her research interests include contact-free sensing and deep learning.

Jun Luo received his BS and MS degrees in Electrical Engineering from Tsinghua University, China, and the Ph.D. degree in Computer Science from EPFL (Swiss Federal Institute of Technology in Lausanne), Lausanne, Switzerland. From 2006 to 2008, he has worked as a postdoctoral research fellow in the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. In 2008, he joined the faculty of the School Of Computer Science and Engineering, Nanyang Technological University in Singapore, where he is currently an Associate Professor. His research interests include mobile and pervasive computing, wireless networking, machine learning and computer vision, as well as applied operations research. More information can be found at https://personal.ntu.edu.sg/junluo.