MM-MURE: mmWave-Based Multi-Subject Respiration Monitoring via End-to-End Deep Learning

Chandler Bauder, *Member, IEEE*, Abdel-Kareem Moadi, *Graduate Student Member, IEEE*, Vijaysrinivas Rajagopal, Tianhao Wu, *Student Member, IEEE*, Jian Liu, *Member, IEEE*, and Aly E. Fathy, *Life Fellow, IEEE*

***Abstract*—This study presents MM-MURE, a novel method to perform multi-subject contactless respiration waveform monitoring by processing raw multiple-input-multiple-output mmWave radar data with an end-to-end deep neural network. The traditional vital signs monitoring signal processing scheme for mmWave radar involves analog or digital beamforming, human subject localization, phase variation extraction, filtering, and rate or biomarker analysis. This traditional method has many downsides, including sensitivity to selected beamforming weights and over-reliance on phase variation. To avoid these drawbacks, MM-MURE (for MM-wave based MUlti-subject REspiration monitoring) is developed to improve reconstruction accuracy and reliability by taking in unprocessed 60 GHz MIMO FMCW radar data and outputting respiratory waveforms of interest, effectively mimicking an adaptive beamformer and bypassing the need for traditional localization and vital signs extraction techniques. Extensive testing across scenarios differing in range, angle, environment, and subject count demonstrates the network's robust performance, with an average cosine similarity exceeding 0.95. Results are compared to two baseline methods and show more than a 10% average improvement in waveform reconstruction accuracy across single and multi-subject scenarios. Coupled with a rapid inference time of 8.57 ms on a 10 s window of data, MM-MURE shows promise for potential deployment to efficient and accurate near-real-time contactless respiration monitoring systems.**

***Index Terms*—MIMO radar, neural networks, beam steering, telemedicine, biomedical telemetry.**

I. INTRODUCTION

RADAR technology has been demonstrated to enable contactless monitoring of human vital signs, including respiration and heart rate [1]. This technological advancement has numerous critical applications, such as long-term patient health monitoring, rapid triage, sleep studies, through-wall life detection, and exercise health analysis [2], [3]. Recent interest has expanded beyond simple rate measurements to more intricate indicators like interbeat intervals, times of inhalation and exhalation, and arrhythmia detection, all of which demand the radar to have highly accurate reconstructions of vital signs waveforms [4], [5]. Additionally, there's a growing effort to transition from experiments focused on individual subjects to scenarios that include monitoring multiple subjects simultaneously [6].

Many of the state-of-the-art techniques in radar-based vital signs monitoring use millimeter-wave, frequency modulated continuous wave (FMCW) multiple-input multiple-output (MIMO) radars, which have heightened sensitivity to small movements and the ability to focus signals of interest with respect to range and angle [7]. MIMO radars help deal with real-world scenarios where the subject under test is not typically sitting directly in front of the radar, or there are multiple subjects that need to be monitored, by using many transmit-receive elements with relatively wide antenna beamwidths. The traditional multi-subject vital signs signal processing scheme for these applications includes an array of signal analysis techniques to perform digital beamforming and human subject localization, arctangent demodulation (AD), phase variation filtering, and rate or biomarker analysis [8]. In [8], [9], [10], [11], [12], various MIMO radar architectures and beamforming-based processing schemes are used to achieve sufficient angular separation of multiple subjects for simultaneous vital signs monitoring. While accurate results can be achieved with these methods, they suffer from several drawbacks, such as sensitivity to the proper selection of beamforming weights and over-reliance on the phase variation alone, which prevent reliably accurate reconstructed vital sign waveforms when faced with non-idealities such as inaccurate localization or interference from other people.

Addressing the challenge of developing advanced signal processing methods that reliably model the nuanced dynamics between RF reflections and vital signs signals, the use of deep neural networks has been proposed. Though results have shown that neural networks can do an adequate job at approximating this relationship and can outperform traditional methods [14], care must be taken to ensure generalized performance for subjects and scenarios not seen during training. Prior research

Manuscript received 27 March 2024; revised 20 June 2024; accepted 5 August 2024. This work was supported by National Science Foundation under Grant 2052780. *(Corresponding author: Chandler Bauder.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by University of Tennessee Knoxville's Institutional Review Board under Application No. UTK IRB-24-08063-XP.

Chandler Bauder, Abdel-Kareem Moadi, Tianhao Wu, Jian Liu, and Aly E. Fathy are with the Department of Electrical Engineering and Computer Science, Univeristy of Tennessee - Knoxville, Knoxville, TN 37996 USA (e-mail: cbauder@vols.utk.edu; amoadi@vols.utk.edu; twu21@vols.utk.edu; jliu@utk.edu; afathy@utk.edu).

Vijaysrinivas Rajagopal is a Independent Scholar Working in Atlanta, GA 30318 USA (e-mail: vrajagopal28@gmail.com).

Digital Object Identifier 10.1109/JERM.2024.3443782

TABLE I
COMPARISON OF RADAR-BASED VITAL SIGNS MONITORING WORKS UTILIZING NEURAL NETWORKS

Work	Radar	Freq	End-to-End	Multi-Subject
MoRe-Fi[13]	X4M05	10GHz	No(CFAR)	No
Gong *et al.*[14]	AWR1243	77GHz	No(AD)	No
MoVi-Fi[15]	Novelda X4, Position2Go, IWR1443	7.3GHz, 24GHz, 77GHz	No(CFAR)	No
HeRe[16]	BGT60TR13C	60GHz	No(AD)	No
MM-FGRM[17]	IWR1642	77GHz	No(CFAR)	No
MM-MuRe (This work)	**V-MD3**	**60GHz**	**Yes**	**Yes**

has already incorporated neural networks into segments of the traditional single-subject non-contact vital signs processing scheme using single-channel radar implementations [13], [15], [16], but a holistic end-to-end strategy employing the angular focusing strengths of MIMO radar for multi-subject cases has yet to be investigated, as shown in Table I. Thus, the state-of-the-art in multi-subject non-contact vital signs estimation still relies on using traditional processing techniques to either perform localization and separation of subjects with respect to angle or extract the phase variation before the application of further advanced traditional processing techniques or neural networks, which open the door for unreliable performance.

In this study, MM-MuRe is developed, a novel technique to supplement the traditional MIMO-radar based vital sign signal processing scheme with an end-to-end neural network. This multi-functional method automatically handles the job of human subject localization, angular separation, and subsequent signal processing. By taking in unprocessed multi-channel in-phase (I) and quadrature (Q) MIMO-FMCW radar data and training a deep neural network, accurate reconstructed respiration waveforms are output regardless of spatial location or whether there are one or two subjects in the scene, achieving improved performance over baseline methods. To the best of the authors' knowledge, this is the first method to utilize an end-to-end deep neural network on multi-channel radar data to achieve angular separation for multi-subject respiration waveform monitoring.

II. BACKGROUND

A. Single-Subject Vital Signs Signal Model

The use cases of the proposed system involve vital signs monitoring of up to two people within the radar's field of view. The expected vital signs signal of interest can be modeled as reflections from a scatterer derived from FMCW radar theory [18], [19]. The signal transmitted by the radar can be described as:

$$x(t) = e^{j2\pi f_i t}, \tag{1}$$

$$f_i(t) = f_t + St, \quad 0 \le t \le T, \tag{2}$$

where t represents fast-time temporal variations, f_i is the instantaneous frequency, f_t is the carrier frequency, and S is the frequency slope defined by the frequency bandwidth β divided by the chirp period T. The chest displacement of a subject (denoted as Subject 1) seated R^1_{nom} meters from the radar is typically modeled as the superposition of the nominal range of the subject and the variations due to the vital signs across slow-time index v:

$$R_1(v) = R^1_{nom}(v) + R^1_{resp}(v) + R^1_{heart}(v). \tag{3}$$

The radar under test is assumed to be a MIMO system with a virtual uniform linear array (ULA) resulting from M transmitters and N receivers. The round trip time delay at v from a single element of the $M \times N$ element virtual array to Subject 1 seated $R^1_{nom}(v)$ meters from the radar at an angle $\theta_1(v)$ from broadside is denoted as $\Delta t^1_{ij}(v)$. The time delay and the received signal $y^1_{ij}(t, v)$ are calculated as

$$\Delta t^1_{ij}(v) = \frac{2R_1(v)}{c} + \frac{2x_{ij}\sin(\theta_1(v))}{c}, \tag{4}$$

$$y^1_{ij}(t, v) = A_1(v) \cdot x(t - \Delta t^1_{ij}(v)), \tag{5}$$

where x_{ij} is the coordinate of the element formed by the ith transmitter, $i = 1, 2, .., M$, and the jth receiver, $j = 1, 2, .., N$, c is the speed of light, and $A_1(v)$ represents the amplitude of the received signal and encapsulates propagation losses, antenna gain, and the cumulative radar cross-section (RCS) of the subject. Thus, $y^1_{ij}(t, v)$ represents an attenuated and time-delayed copy of the transmitted signal $x(t)$, and the specific time delay to each virtual array element is related to its relative position in space. After I/Q demodulation with the transmitted chirp $x(t)$, the down-converted received signal $z^1_{ij}(t, v)$ for each virtual array element can be described as

$$z^1_{ij}(t, v) = A_1(v) \cdot e^{j(-2\pi S\Delta t^1_{ij} t - 2\pi f_t \Delta t^1_{ij} + \pi S {\Delta t^1_{ij}}^2)}, \tag{6}$$

where the nominal range of the subject $R^1_{nom}(v)$ has been encoded into the frequency value of $z^1_{ij}(t, v)$, which is effectively constant for all transmit-receive channels under the far-field approximation, and can be found through analysis of the inverse fast Fourier transform (IFFT) of the signal. The resolution at which the range can be detected in this way is equal to $\frac{c}{2\beta}$. On the other hand, the first phase term, denoted as $p_n(v)$, corresponding to subject n, is sensitive to the subject's fine movements, and represents complete information about the displacement of the chest over slow-time $R_1(v)$ (along with the amplitude $A_1(v)$ due to RCS variations), as well as the angle of arrival of the reflected signal through differences in $\Delta t^1_{ij}(v)$ between channels. The second phase term is the residual video phase and is sufficiently small compared to $p_n(v)$, and can be ignored.

An azimuth range-angle map $P_1(f, \theta, v)$ can be generated by beamforming the virtual channels through the product of the array steering vector $a_{ij}(\theta)$ and the resulting range profile obtained from the IFFT of $z^1_{ij}(t, v)$, denoted as $U^1_{ij}(f, v)$:

$$a_{ij}(\theta) = e^{-j2\pi \frac{x_{ij}\sin\theta}{\lambda_t}}, \tag{7}$$

$$P_1(f, \theta, v) = \sum_{i=1}^{M}\sum_{j=1}^{N} U^1_{ij}(f, v) \cdot a_{ij}(\theta). \tag{8}$$

The virtual channels are coherently summed to all azimuth angles of interest θ by accounting for phase delays related to the spacing of the array. Each frequency bin f of the beat frequency

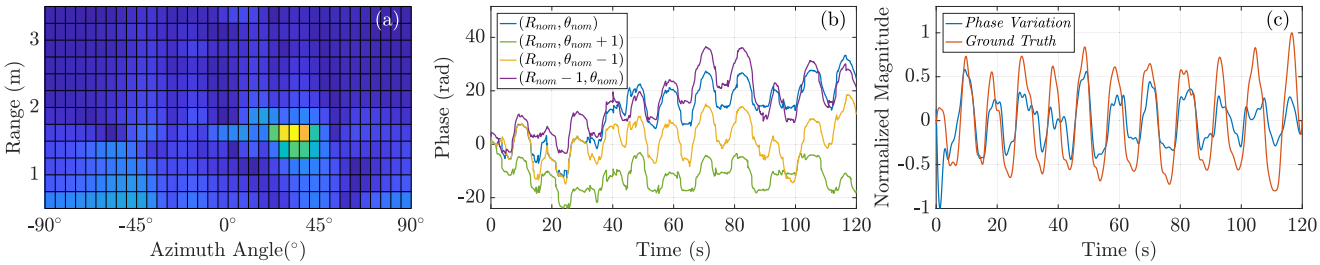

Fig. 1. (a) Magnitude of the range-angle map depicting the spread of a subject seated at approximately (1.5 m, +30deg); (b) Differences in respiration waveform shape across multiple valid range-angle locations; (c) Corrupted phase variation in a simple single seated subject scenario compared with respiration belt ground truth.

representation of the received signal at each element $U^1_{ij}(f, v)$ is directly related to the range from the radar. Thus, utilizing an object detection algorithm such as constant false alarm rate (CFAR) [18] on the range-angle map $P_1(f, \theta, v)$ enables range and angular location estimation of the subject (R^1_{nom}, θ_1). Here we let f_1 represent the beat frequency index corresponding to Subject 1 at range R^1_{nom}.

Conventional radar-based vital signs monitoring techniques use arctangent demodulation of the complex beamformed result at the detected location $P_1(f_1, \theta_1, v)$ to obtain the phase variation $p_1(v)$, which carries the information about the displacement of the chest over time from $R_1(v)$:

$$p_1(v) = \arctan \frac{Im(P_1(f_1, \theta_1, v))}{Re(P_1(f_1, \theta_1, v))} = \frac{-4\pi R_1(v)}{\lambda_t}. \quad (9)$$

Bandpass filtering or another more complex signal denoising technique is then typically done on $p_1(v)$ to clean up the signal before rate or biomarker estimation.

B. Multi-Subject Vital Signs Signal Model

If we assume a second subject is present in the scene located at range R^2_{nom} and angle θ_2, then the total received signal of interest becomes a sum of the reflections from both scatterers:

$$z^{tot}_{ij}(t, v) = z^1_{ij}(t, v) + z^2_{ij}(t, v). \quad (10)$$

And the resulting range-angle map is calculated using the range profile containing information about both subjects:

$$P_{tot}(f, \theta, v) = \sum_{i=1}^{M} \sum_{j=1}^{N} U^{tot}_{ij}(f, v) \cdot a_{ij}(\theta). \quad (11)$$

If the angular resolution condition is met, then CFAR performed on $P_{tot}(f, \theta, v)$ should locate both subjects in the scene and result in the extraction of both $p_1(v)$ and $p_2(v)$. Situations where subjects occupy the same angle are a common problem for radar-based methods, and are not considered in this study. Limitations of angular discrimination will be discussed in Section V-E.

C. Traditional Vital Sign Extraction Method Limitations

The first problem with the formulated traditional vital signs extraction process is that the definition of the subject as a point scatterer is an over-simplification. In reality, depending on the range and angular resolution of the radar under test, the subject likely inhabits multiple range and angle indices of $P_{tot}(f, \theta, v)$, as shown in Fig. 1(a). So while an object detection algorithm such as CFAR may return the most likely approximate location of the subject based on the signal-to-noise ratio (SNR) threshold, denoted as (R_{nom}, θ_{nom}), this does not automatically imply that the range-angle location detected by CFAR is the best precise location to beamform towards to obtain the most accurate vital signs signal. To address this, some works have opted to develop ways to combine multiple range-angle indices to improve the quality of the extracted vital signs signal at the cost of increased computational effort [20].

This leads to the second problem for the traditional method, which is that the resulting vital signs signal after the beamforming operation is relatively sensitive to the chosen range and angle. As seen in Fig. 1(b), changing the beamforming location by small variations in range and angle indices has a significant effect on the shape of the signal. This means that any processing done on this waveform such as breathing rate (BR), interbreath interval (IBI), or inhalation/exhalation (I/E) ratio analysis depends on consistently accurate estimation of the optimal beamforming weights. Additionally, it is possible that CFAR detects the location of the subject wrong altogether, losing all information about the vital signs waveforms. In [21], a search is done on the best power ratio among frequencies within and outside the frequency band of interest to determine the best beamforming weights to address this issue. A camera vision-guided approach was proposed in [22] to inform the radar on the optimal beamforming weights, but requires a complicated calibration process and loses the privacy advantages of radar-only based systems.

Finally, in real-world scenarios the displacement $R_n(v)$ encapsulates not only the linear superposition of the chest motion from (3), but also unwanted artifacts from limb movements and swaying motions as well as nonlinear dependencies [13]. Thus, it has been proposed that operating on the phase variation signal $p_n(v)$ alone is equivalent to reducing the information of interest to a lower-dimensional representation (losing the information from $A_n(v)$) and is thus insufficient to consistently and accurately reconstruct the precise shape of the vital signs signal in a manner that can compete with wearable sensors [13]. Fig. 1(c) shows that even in the simplest case of detecting respiration in

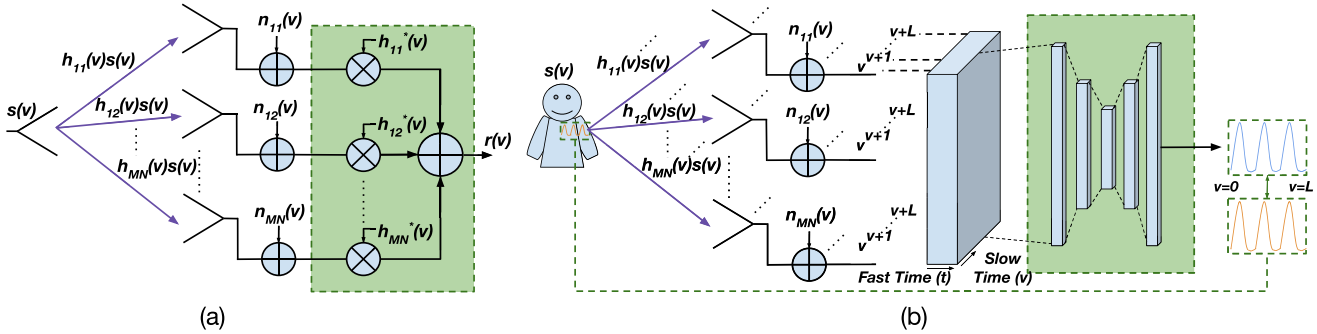

Fig. 2. (a) Summary of the MRC channel combining process in communications where $h_{ij}(v)$ is obtained through a training procedure; (b) Summary of the neural-network-based channel combining process. A ground truth respiratory belt represents the displacement of the chest over slow-time v and allows the network to iteratively learn to extract the respiratory depth information from the raw channel information.

a scenario where there is only one subject in the scene, phase extraction still struggles to maintain a clean waveform.

D. Inspiration for Neural Networks

Ideally, a version of maximal ratio combining (MRC) would be implemented to achieve generalized beamforming and leave behind the need for subject location detection and location-based beamforming [23]. Complex weights $\mathbf{w}(v)$ for slow-time index v that embody the phase differences and channel gain information can be used to beamform the received signals of the virtual array elements in a way that does not align to a specific position in space, but rather works adaptively to cancel out the phases of the channel gain vector $\mathbf{h}(v)$ and extract the reconstructed signal of interest $r(v)$ by maximizing the SNR:

$$r(v) = s(v)\sum_{i=1}^{M}\sum_{j=1}^{N} w_{ij}(v)h_{ij}(v) + n(v), \tag{12}$$

where $s(v)$ is the vital signs signal, $w_{ij}(v)$ is the channel gain of the virtual element, and $n(v)$ is zero-mean Gaussian noise. If the channel gain vector $\mathbf{h}(v)$ is known, then $\mathbf{w}(v)$ can be set to $\mathbf{h}^*(v)$, allowing for coherent summation and preservation of the shape of the vital signs signal $s(v)$. While in the application of communications, a training sequence is used to estimate the complex channel gain information $\mathbf{h}(v)$ over time, this is not applicable in our case due to the ever-changing vital signs signal of interest. Thus, an alternative method to estimate $\mathbf{h}(v)$ must be developed. While the signal of interest $s(v)$ is never known precisely at any given time without information from a wearable sensor, human respiration waveforms have identifiable features that are similar across individuals and time. Motion artifacts from swaying or limb movements also have identifiable features and can be associated with the current state of $\mathbf{h}(v)$. Convolutional neural networks (CNNs) excel at iteratively learning patterns within images or other data representations through a training procedure to pick out objects or features of interest [24]. Work has been done to utilize CNN-based variational autoencoder neural networks as signal source separation tools, specifically for temporal audio signals, by encoding the mixed signal into a lower dimensional latent space and upsampling to keep only the signal of interest [25]. Furthermore, in [13] a variational encoder-decoder was used to extract respiratory waveforms from the complex data belonging to multiple range bins of a single-channel radar.

III. Methods

Here we propose MM-MuRe, a method to utilize a CNN-based variational autoencoder to operate similarly to MRC, operating with the same inputs and outputs. The relationship between the complex multi-channel radar signals and the vital signs signals is learned instead by training a neural network on a large amount of historical data, as depicted in Fig. 2, by using a ground truth waveform from a wearable respiration belt as the reference and unprocessed in-phase (I) and quadrature (Q) experimental data from a MIMO FMCW radar as the input. We aim to demonstrate in the following sections not only that MM-MuRe can reliably reconstruct the respiration waveform of a single subject regardless of location and without using the traditional processing scheme, but also that adaptive beamforming (without using traditional delay-and-sum) is achievable through the network by consistently reconstructing respiration waveforms of two subjects simultaneously, even when they are within the same range from the radar. Respiration waveforms are targeted in this study because wearable chest belts directly measure the mechanical displacement of the chest due to breathing, making them an ideal ground truth sensor for training the radar-based neural network. Wearable sensors that measure the heartbeat waveform, such as ECG sensors or pulse oximeters, indirectly measure the chest's displacement due to the heartbeat by monitoring electrical activity or blood oxygen saturation. This indirect measurement would hinder the goal of direct displacement waveform reconstruction, thus heartbeat reconstruction is left for future research.

The benefits of processing the radar data for multi-subject respiration waveform reconstruction in this way are as follows:

1) Channel combination is not dictated by conventional beamforming equations, thus no information is lost through specific location-based beamforming;
2) The input data is kept in its raw I/Q form, allowing the reconstructed signals to be products of both the amplitude and phase information;

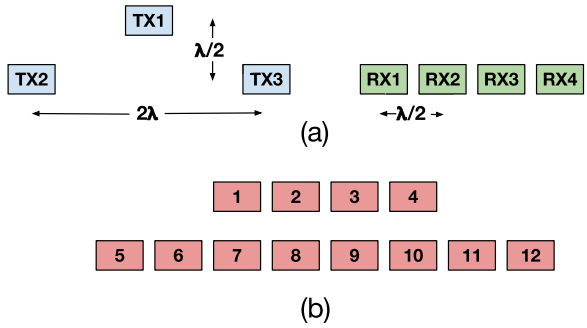

Fig. 3. (a) Relative positioning of the Tx and Rx arrays, with each element having a 60° HPBW in azimuth and a 36° HPBW in elevation; (b) Resulting MIMO virtual array.

3) The ability of neural networks to learn linear and non-linear relationships opens the possibility to have inherent noise removal from unwanted artifacts, promoting more accurate reconstructions;
4) The avoidance of the need for complex traditional processing methods to find the optimal beamforming weights combined with the high parallelization of neural networks means that the inference time of the developed method is fast enough to meet the requirements of efficient near-real time systems.

A. Hardware

To collect data for the proposed method, an RFbeam V-MD3 60 GHz FMCW radar with 3 transmitters (Tx) and 4 receivers (Rx) is used [26]. The Tx/Rx locations, depicted in Fig. 3, result in a 12-element virtual array that provide angular separation in both azimuth and elevation. The elements have a 60° half-power beamwidth (HPBW) in azimuth and a 36° HPBW in elevation, providing enough coverage (especially with MIMO processing) to enable subject monitoring in a wide range of locations off-broadside. Given we are focusing on respiration, a transmitted chirp bandwidth of 600 MHz is used, which corresponds to a range resolution d_{res} of 0.25 m. This allows us to comfortably assume all of the respiration signals of interest appear across only 1-2 range bins, rather than being discontinuously separated across several bins, as is the case with the finer range resolutions typically utilized in heartbeat waveform analysis. A single frame in slow-time consists of a single chirp (counting all chirps transmitted in a TDM-MIMO scheme as one) being received at each Tx/Rx channel and is represented by 64 ADC samples each. Frames are captured at a slow-time sample rate $f_s = 50$ Hz to sufficiently depict respiration waveforms and motion artifacts.

B. Pre-Processing

The raw I/Q radar data matrix must be formed into a representation that allows the developed network to learn appropriate features and reach the desired outcome rapidly and effectively. Based on the number of ADC samples per received chirp and the number of Tx/Rx channels in the virtual array, the I/Q data matrix received for each slow-time index v is of size 64×12.

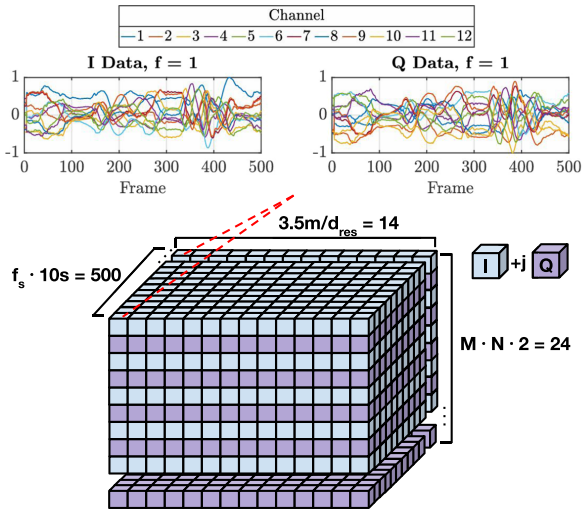

Fig. 4. Depiction of the $24 \times 14 \times 500$ input data matrix. Normalized and concatenated I/Q values form 24 channels, 14 cropped range bins from the range profile form the rows, and a sliding slow-time window of 500 samples make up the columns.

Each column represents the beat frequency representation of the received signal belonging to each Tx/Rx channel. Subsequently, an IFFT must be performed on the beat frequency signals to convert the columns to range profile representation $U_{ij}(f, v)$, as discussed in Section II, where each row index f then corresponds to information present at $d_{res} \cdot f$ away from the radar. Since $d_{res} = 0.25$ m, if it is assumed that subjects in our scene will not be further away than 2 m from the radar, it is more than adequate to crop the data matrix to only the first 14 range bins, corresponding to within a range of $d_{res} \cdot 14 = 3.5$ m away, allowing for more efficient processing. In Section V-E, the 14 range bins fed to the network are shifted to account for measurement distances past 2 m, which will be discussed.

Data is then accumulated on a 10-second sliding window basis to feed into the network. This is to allow the network to learn features that may be more easily interpreted with temporal context. Since $f_s = 50$ Hz, this corresponds to 500 slow-time samples. Due to framework and training limitations, we convert complex I and Q values of the input data matrix to concatenated channels. Thus, for concatenated I data and Q data for 12 Tx/Rx channels, 14 range bins, and 500 slow-time samples, the final input shape of the data matrix is $24 \times 14 \times 500$, as depicted in Fig. 4.

C. Network Architecture

The proposed method builds upon work done in the field of neural network-based signal source separation such as Wave-U-Net [25]. Originally developed for 1-D audio signal source separation, its base upsampling/downsampling convolutional structure is extended here as the backbone to the developed network. Several modifications are made to work with complex-valued radar signals and to perform reliable automatic multi-subject localization and separation of vital signs signals from noise.

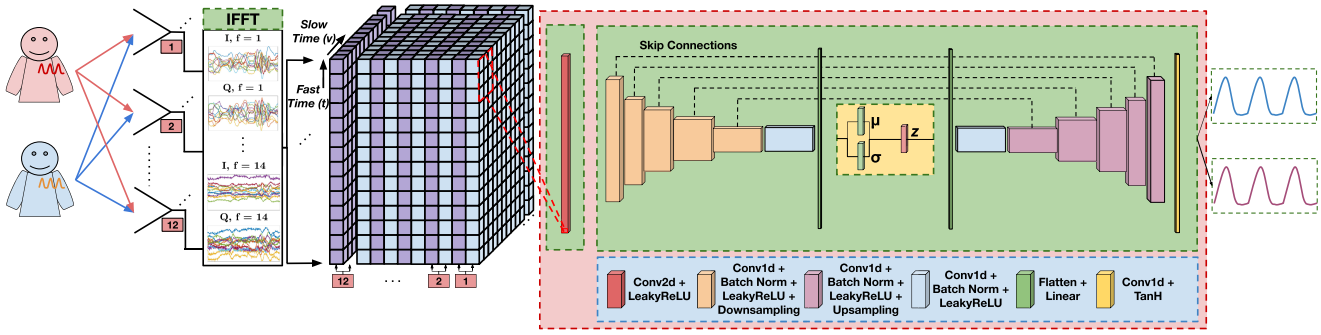

Fig. 5. Breakdown of MM-MURE, the proposed neural network-based channel combining architecture. Raw I/Q data from the FMCW radar is formed into an input data matrix over time and fed into MM-MURE. The number of output waveforms depends on the number of subjects in the scene.

The $24 \times 14 \times 500$ input data matrix is comprised of 500 instances of frames containing concatenated complex information belonging to all 14 range bins for all Tx/Rx channels. It is expected that human subjects belong to only 1 or 2 range bins at once. At the same time, the network architecture operates by downsampling and upsampling the input signal with multiple 1-D convolutional layers, expecting a 1-D time-varying signal (independent of number of channels). Thus, a series of 2-D convolutional layers is implemented on the input to effectively scan over the range profile input and transform the overall size to $24 \times 1 \times 512$ (512 being a power of 2 for proper compatibility with the time halving procedure throughout the network), as shown in Fig. 5.

The base encoder-decoder structure, skip connections, and feature decimation techniques of Wave-U-Net are then utilized with 5 levels, a channel interval of 32, a downsampling filter size of 15, and an upsampling filter size of 5 [25]. However, the middle layer is modified to become a variational bottleneck to help improve performance on subjects and scenarios not seen during the training process. It is assumed that the experimental scenario under test (single or multi-subject) is known to the user. Thus, a 1D-convolutional output layer with a hyperbolic tangent (TanH) activation function [27] is added to the end that reduces the channel dimension to either 1 for single-subject or 2 for multi-subject scenarios. This allows for specialized versions of the network capable of handling both scenarios. The total reconstructed waveform signal size is either 1×500 or 2×500 after cropping the extra padding, the same size as the ground truth respiration waveform corresponding to the same 10-second window. In the multi-subject case, the first output channel is trained to reconstruct the respiration of the left-most person and vice versa. Thus, our method is dependent on subjects being separated in azimuth.

D. Dataset Normalization and Augmentation

Normalization is performed on the radar dataset by considering the polar coordinate representation of every I/Q datapoint recorded across the entire training set. Each I/Q datapoint is normalized in its polar form to a magnitude between 0 and 1 based on the maximum polar magnitude present in the dataset and then converted back to cartesian I/Q coordinates. On the other hand, values recorded by the ground truth respiration belt depend on the tightness at which the belt is worn. Thus, ground truth waveforms are normalized between a value of -1 and $+1$ based on the maximum recorded value on a per-experiment basis.

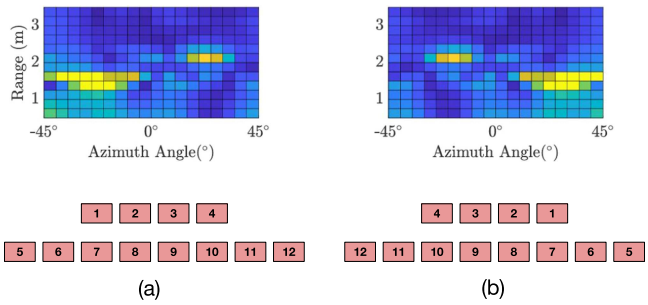

Fig. 6. (a) Range-angle heatmap resulting from the default array order and two subjects in the scene seated at $(1.5\,\text{m}, -30\,\text{deg})$ and $(2\,\text{m}, +30\,\text{deg})$; (b) Resulting range-angle heatmap after the mirroring augmentation.

Subsequently, several steps are taken to augment the original recorded experiments to increase our dataset size. Firstly, after splitting into training and testing sets, as will be discussed in Section IV, each sample in the training set is generated by using a 10-second window with a 9.5-second overlap between consecutive windows on the raw data. Secondly, each 10-second window is augmented by rotating the I/Q points from 0° to 360° by steps of 15° in the I/Q plane, as proposed in [13]. The information carried within the I/Q datapoints is preserved with rotation, with the relative phase of each point depending only on the distance to the subject.

Lastly, we propose to further augment the dataset through manipulation of the order of the I/Q channels as presented to the network. The I/Q values of the 12-element virtual array layout depicted in Fig. 3(b) can be organized in the data input matrix to reflect their relative position in the array with respect to the line of symmetry. Thus, reversing the order of the elements in the dimension belonging to the 24 I/Q values of the radar data matrix effectively mirrors the experimental data in azimuth with respect to the axis perpendicular to the plane formed by the array, as shown in Fig. 6. On each mirrored multi-subject dataset,

the ground truth value assigned to the left-most and right-most person is subsequently switched.

E. Training Hyperparameters

For training, all samples are gathered randomly in batches of 32 within a single epoch. The loss function $\mathcal{L}_{tot}$ is calculated as the addition of the mean squared error (MSE) loss and a scaled Kullback-Leibler (KL) divergence term to promote both accurate reconstructed waveforms and regularization of the latent space normal distribution $\boldsymbol{z}$ represented by mean μ and standard deviation σ, shown in Fig. 5:

$$\mathcal{L}_{tot} = \mathcal{L}_{MSE} + \gamma \cdot D_{KL}(\mathcal{N}(\mu, \sigma), \mathcal{N}(0, I)), \tag{13}$$

where γ is found empirically to be 0.001. NAdam optimization with a learning rate of 1e-5, decay rates of $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and momentum decay of 0.004 were used [28].

IV. Experimental Methodology

The developed model requires thorough training to effectively address a broad spectrum of targeted scenarios. To achieve this, a series of experiments are conducted, enabling the network to meet and exhibit the following objectives:

1) MM-MuRe is expected to have the capability to precisely reconstruct respiration waveforms for up to two individuals located within $\pm 60°$ from the radar's broadside, extending well beyond the main beam, and at a distance of up to 2 meters from the radar (unless trained for more);
2) Robust performance should be demonstrated by maintaining accuracy when analyzing individuals not included in the training set;
3) The presence of clutter with a significant radar cross-section (RCS) should not compromise reconstruction accuracy;
4) The network should exhibit angular separation by accurately and reliably reconstructing the respiration waveforms of two individuals located at the same distance from the radar.

To systematically work towards these objectives, experiments[1] are taken by placing subjects along the vertices of a grid formed between $\pm 60°$ from broadside in increments of $15°$ and at a distance of 0.75–2 m in increments of 0.25 m, as shown in Fig. 7. Although not every potential scenario is tested for clutter and multiple subjects, a sufficient variety of configurations are chosen to ensure the network's ability to account for minor variations in position. In all experiments, the precisely synchronized ground truth respiration signals are recorded using a Vernier Go-Direct wearable respiration belt [29]. Unless specified otherwise, approximately 70% of data recorded is reserved for training, while the remaining data is reserved for validation and testing.

A. Experimental Setup

Single-Subject Experiments: To enable the network to reconstruct the respiration waveforms of a single subject in the scene,

[1]This research has been approved by our institution's IRB.

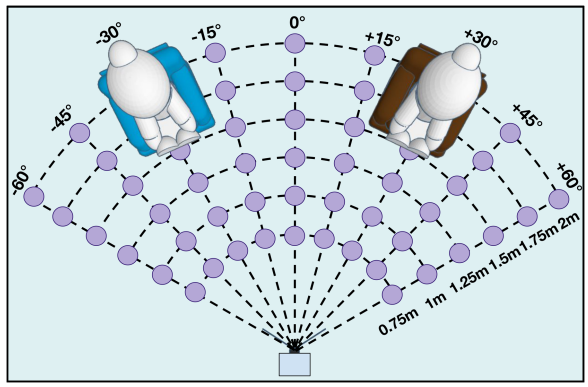

Fig. 7. Ranges and angular locations used to train and test the network. Example of 2 subjects located at $(1.75\ \text{m}, -30\ \text{deg})$ and $(1.75\ \text{m}, +30\ \text{deg})$.

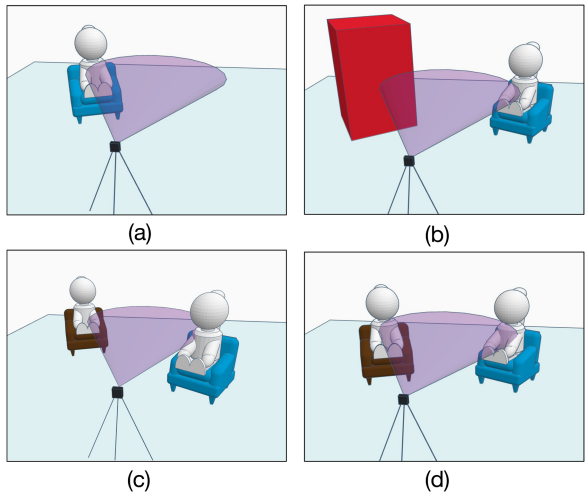

Fig. 8. Depiction of the different experimental scenarios: (a) Single-subject. (b) Single-subject with clutter. (c) Multi-subject at different ranges. (d) Multi-subject with the same range.

data at each location in the grid is taken with 3 separate subjects, as visualized in Fig. 8(a). Each dataset ranges between 1–2 minutes and consists of one of the subjects sitting and breathing normally. Data is taken at $\pm 60°$ from broadside to showcase the ability of MIMO radar to focus on subjects well outside the $\pm 30°$ HPBW. To validate the network's applicability to subjects not included in the training set, only portions of the data collected from Subjects 1 and 2 are used for training, while all data from Subject 3 is reserved for testing. The dataset comprises a total of 136 minutes of unaugmented single-subject data.

Another case considered is that of objects in the scene that cause significant RF reflections and create sources of multipath. While radar-based vital signs studies are typically taken in a lab environment with minimal clutter, the real-world applications of such a system include locations such as hospital rooms that have numerous objects and machines that could cause stronger reflections than the subject under test. Rather than using traditional preprocessing techniques to filter out static objects, we opt here to include several cases of single-subject datasets with significant clutter in the scene to train the network to filter

these out, as depicted in Fig. 8(b). The object used to represent worst-case clutter is a large reflective metal cabinet. In total, approximately 20 minutes of single-subject data with clutter at various locations is included in the dataset.

Multi-Subject Experiments: While it can be argued that the single-subject experiments could be performed using a single-channel radar with a wide beamwidth, multi-subject analysis typically requires the use of multi-channel radar and beamforming techniques in order to cover various cases where the subjects could lay within the same range bin. Simultaneous data of Subjects 1 and 2 are taken across 50 combinations of locations from the defined grid. A wide variety of challenging combinations of locations are taken, such as when one subject is located at broadside and another subject is located outside the main beam. A total of 17 of the 50 locations in the dataset are allocated to scenarios where the two subjects are within different ranges from the radar, as shown in Fig. 8(c).

Additionally, 33 of the 50 locations represented in the dataset are ones where Subjects 1 and 2 are within the same range from the radar, as depicted in Fig. 8(d), in order to properly train the network to beamform to each respiration signal of interest. A total of 2 minutes of data are taken at each location, totaling to 100 minutes of multi-subject data for this subset.

To show that the network can maintain performance over time, an additional 20 seconds of data at each of the same 50 locations are taken of Subjects 1 and 2 several weeks after the initial data collection. Additionally, the applicability to subjects outside the training set is further tested by taking 40 seconds of data at each location several weeks after the training set collection of two more subjects, Subjects 3 and 4. Both of these subsets are reserved for the testing set only.

B. Evaluation Metrics

The main performance metrics used to compare the similarity between the reconstructed $\mathbf{r}$ and ground truth $\mathbf{g}$ respiration waveforms are cosine similarity and MSE, which are defined as:

$$\text{sim}(\mathbf{r}, \mathbf{g}) = \frac{\mathbf{r} \cdot \mathbf{g}}{|\mathbf{r}||\mathbf{g}|}, \tag{14}$$

$$\text{MSE}(\mathbf{r}, \mathbf{g}) = \frac{1}{N} \sum_{v=1}^{N} (r(v) - g(v))^2. \tag{15}$$

Cosine similarity is applied where both signals range between -1 to 1. A cosine similarity value of 1 is desired to maximize proportionality. A high cosine similarity implies that analyses for BR, IBI, and I/E ratio (defined in Section II) can be extracted with high accuracy with respect to the ground truth. Average IBI is calculated through basic peak detection of the valleys of the reconstructed waveforms, while average I/E ratio is estimated by calculating the ratio between the respective peaks and valleys belonging to inhalation and exhalation within individual cycles. Breathing rate is predicted using the maximum value in the frequency spectrum within the range of expected values (0.1–1 Hz).

C. Baseline Systems

To benchmark MM-MuRe against other state-of-the-art systems, we have adopted Ahmad et al.'s method [8], denoted here as CFAR-AD, a conventional radar signal processing approach for multi-subject respiration monitoring that aligns with the process detailed in Section II, as a baseline. Furthermore, to demonstrate our learning-model's superior performance, we have also modified a deep-learning-based method that operates on single-channel radar data and requires prior localization, MoRe-Fi [13], to enable its application in multi-subject respiration monitoring. Specifically, for CFAR-AD, traditional processing is used to locate each subject by generating the range-angle plot with an angular resolution of approximately 12° through delay-and-sum beamforming and utilizing the cell-averaging constant false alarm rate (CA-CFAR [18]) algorithm with 1 guard cell and 1 training cell. The phase variation is then extracted using AD over time for each subject, followed by bandpass filtering between 0.1-1 Hz. For MoRe-Fi, CA-CFAR is performed on the range-angle map, and MoRe-Fi is subsequently applied separately to the groupings of hot-zones corresponding to each identified subject in both single-subject and multi-subject scenarios, and trained until convergence.

V. Experimental Results

A. Single Subject Results

To verify the location-agnostic performance of MM-MuRe with a single subject in the scene, the cosine similarities between the reconstructed waveforms and the ground truth waveforms for the entire test dataset are compared with respect to angle and range, as shown in Fig. 9(a) and 9(b), respectively. The results depict the individual reconstruction accuracies of Subjects 1, 2, and 3 as they take turns being the lone subject under test. The network demonstrates robust performance, with an average cosine similarity of 0.955, showing no significant link between the subject's location and the accuracy of waveform reconstruction. Results are broken down between the subjects present in the training set (Subjects 1 and 2) and the one not present in the training dataset (Subject 3). While Subject 3's results are marginally lower than Subjects 1 and 2 in some cases, an average cosine similarity of greater than 0.923 is maintained across all scenarios, indicating the network maintains performance on new subjects as well. The addition of clutter to the scene does not appear to impact MM-MuRe's ability to reconstruct Subject 1's respiration waveforms, maintaining a cosine similarity of 0.96 or better, even when clutter is located at broadside and close distances, indicating robustness against environmental interferences.

The reconstruction performance of MM-MuRe across all subjects is also compared with the baseline methods with respect to angle and range, as seen in Fig. 10. MM-MuRe's cosine similarity is approximately 10% higher than that of the next-best performing baseline (MoRe-Fi, 0.872 average cosine similarity) across all scenarios. Additionally, MM-MuRe showcases high accuracy regardless of location, validating the advantages of bypassing the traditional CFAR-based localization procedure

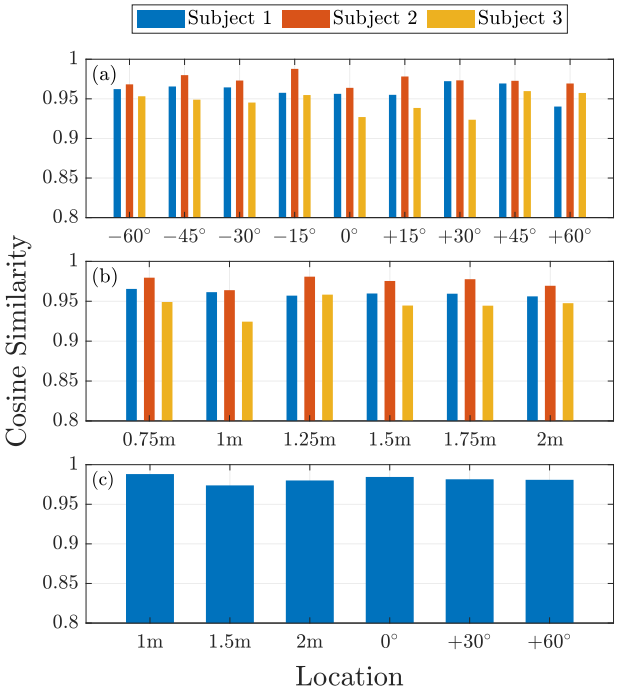

Fig. 9. (a) Single-subject reconstruction accuracy for subjects 1, 2, and 3 across angle and (b) range (subjects 1 and 2 are in the training set, while subject 3 is not). (c) Effect of significant clutter.

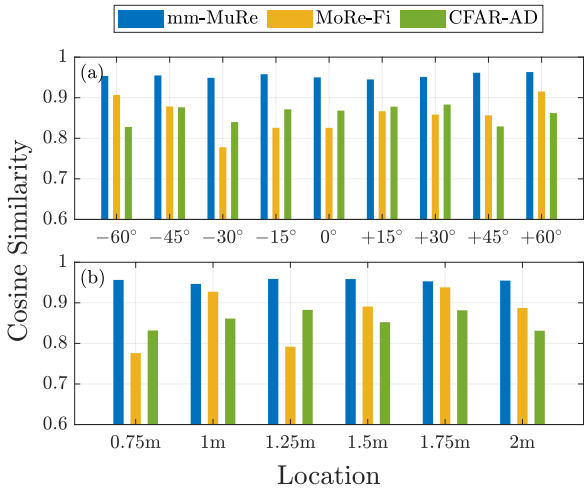

Fig. 10. (a) Single-subject waveform reconstruction performance comparison to baselines across angle and (b) range.

by operating on multi-channel data with a variational U-Net architecture.

B. Multi-Subject Results

Fig. 11(a) and 11(b) show the reconstruction performance of the network for all subjects with respect to angle and range. High cosine similarity for Subjects 1 and 2 is achieved for all scenarios (0.953 average), with only slight degradation on subjects outside the training set (0.919 average), confirming the network's generalizability. Additionally, no significant performance is lost

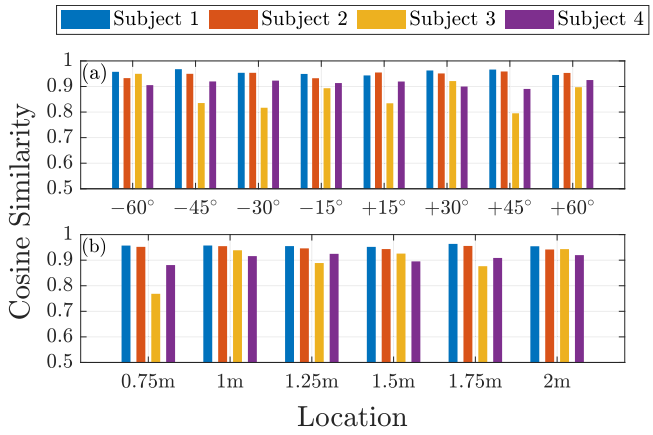

Fig. 11. (a) Multi-subject reconstruction similarity for subjects 1-4 across angle and (b) range (subjects 1 and 2 are in the training set, while subjects 3 and 4 are not).

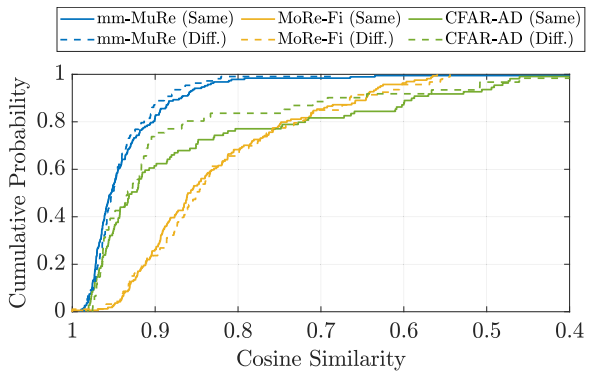

Fig. 12. CDF curve comparison between MM-MURE and baselines for multi-subject cases where two subjects are located in the same vs. different range bins, demonstrating angular separation.

when compared to the single subject scenarios of Fig. 9, even while testing on data collected several weeks after the initial training set collection, promoting confidence in the network's stability over time.

The improvement compared to baseline methods is even more stark in the multi-subject case. Fig. 12 shows a cumulative distribution function (CDF) curve of multi-subject performance comparisons for each method between cases of subjects being in the "same" range bin vs. "different" range bins. The proposed method shows consistent performance across the two scenarios, with 87% and 81% of estimations achieving a cosine similarity of 0.9 or better in different-range and same-range bin scenarios, respectively, proving that the network successfully achieves adequate angular separation. These same values drop to 73%/61% for CFAR-AD and 24%/26% for MoRe-Fi, revealing the significant improvement for our system. An example of a same range bin multi-subject reconstruction from all methods is shown in Fig. 13. While baselines successfully locate both subjects using the range-angle heatmap, the reconstructed respiration waveform contains artifacts that can negatively impact further waveform analysis.

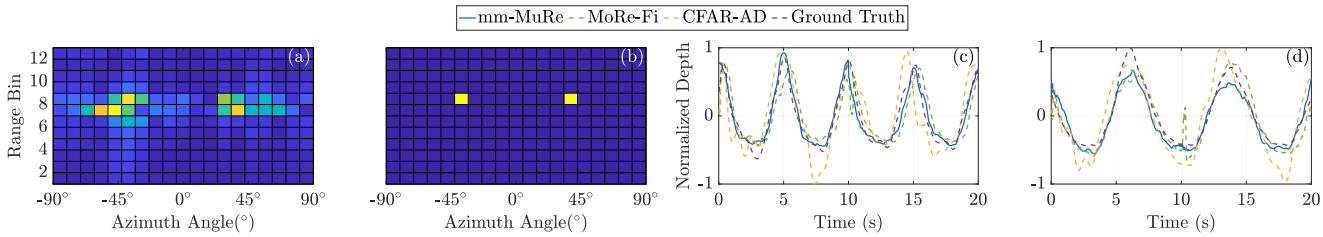

Fig. 13. (a) Range-angle heatmap for use by CFAR-AD and MoRe-Fi. (b) Detections made by CFAR. (c) Reconstructed waveforms from MM-MuRe vs. baselines for subject 1 (MM-MuRe - 0.96, MoRe-Fi - 0.85, CFAR-AD - 0.85 cosine similarity) and (d) subject 2 (MM-MuRe - 0.95, MoRe-Fi - 0.94, CFAR-AD - 0.90 cosine similarity).

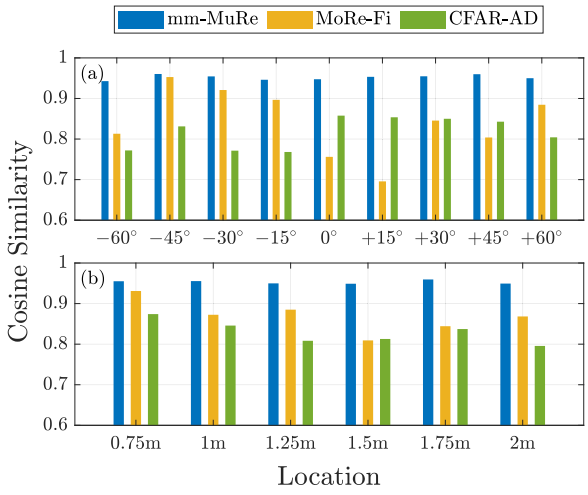

Fig. 14. (a) Multi-subject waveform reconstruction performance for MM-MuRe vs. baselines across angle and (b) range.

A comparison with respect to location is also demonstrated in Fig. 14 for the subjects in the training set. The reconstruction accuracy of MM-MuRe remains consistent across all locations yet again, with an average multi-subject cosine similarity of 0.953. The signal mixing in multi-subject scenarios makes proper CFAR-based identification of beamforming weights for each individual crucial, with improper selection of weights causing waveform degradation, shown by the decrease in performance of the baselines as compared to single-subject experiments (0.863 and 0.822 average multi-subject cosine similarity for MoRe-Fi and CFAR-AD, respectively). On the other hand, MM-MuRe maintains similar performance in both single-subject and multi-subject scenarios, showing the strengths of not having traditional localization preprocessing steps before the implementation of the network.

C. Biomarker Estimation

To demonstrate how a high cosine similarity between the predicted and ground truth waveforms is an indicator for more accurate biomarker analysis, other performance metrics such as MSE, BR, IBI, and I/E ratio (see Section IV-B) are calculated across the entire test set for MM-MuRe, MoRe-Fi, and CFAR-AD, as shown in Table II. MM-MuRe offers improved

TABLE II
COMPARISON OF RESULTS BETWEEN THE PROPOSED AND BASELINE METHODS

Method (Scenario)	Cos. Sim	MSE	BR (bpm) (MAE)	IBI (s) (MAE)	I/E Ratio (MAE)
CFAR-AD (Single)	0.864	0.070	0.204	0.355	0.180
MoRe-Fi (Single)	0.872	0.054	0.134	0.327	0.159
MM-MuRe (Single)	**0.955**	**0.041**	**0.071**	**0.315**	**0.147**
CFAR-AD (Multi)	0.822	0.087	0.403	0.426	0.137
MoRe-Fi (Multi)	0.863	0.055	0.220	0.365	0.151
MM-MuRe (Multi)	**0.953**	**0.044**	**0.066**	**0.321**	**0.131**

performance over baselines in each analysis. Additionally, performance is not degraded between single- and multi-subject cases, showing the network's consistency in performing similar to wearable respiration sensors.

D. Computational Analysis

The applications of mmWave radar-based vital signs monitoring involve scenarios where having a wearable sensor is either non-ideal such as long-term patient health monitoring or sleep studies, or impossible such as some through-wall life detection [30], [31]. In such applications, near-real time analysis is crucial and a requirement for developing advanced processing methods. The average inference time of MM-MuRe across each 10-second window in the test dataset is found to be 8.57 ms on an Intel i7-8665 U CPU @ 2.11 GHz, which shows its potential for deployment to near-real time systems. The input data can be fed to the network on a sliding-window basis to provide updated waveform analysis at the desired frame rate. On the same CPU, the average inference times for MoRe-Fi and CFAR-AD on the test dataset are 20.42 ms and 48.3 ms, respectively.

E. Generalizability Studies

In this work, sensible limitations are set on the expected functionality of MM-MuRe for both narrowing the focus to target indoor monitoring applications and setting boundaries on the range of data to be collected for training. However, it is important to test how generalizable the system is to conditions not well represented in the training dataset, more than to just

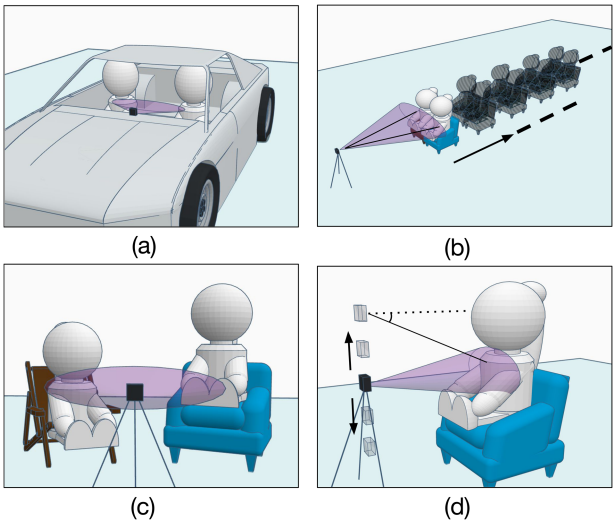

Fig. 15. Depiction of the extended experimental scenarios: (a) Inside car (both front and back). (b) Angular discrimination limitation test. (c) Multi-subject with different elevations. (d) Single-subject at varying elevations.

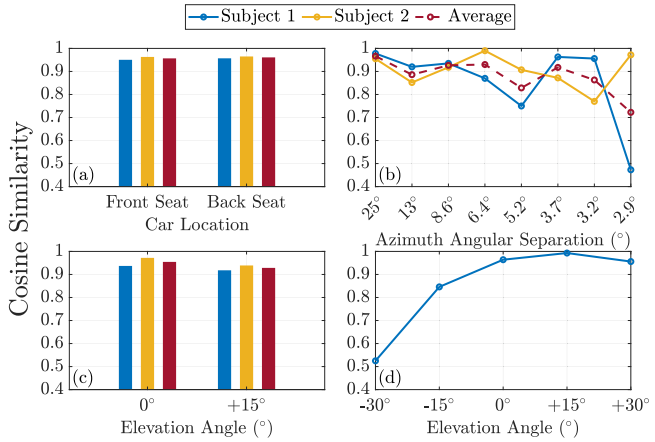

Fig. 16. (a) Multi-subject performance vs. car location. (b) Performance vs. angle between two subjects seated shoulder-to-shoulder (obtained by continually increasing distance). (c) Two subjects seated in the same range bin at different elevations. (d) Single-subject performance vs. elevation angle.

new test subjects, such as a different environment or variations in elevation. Similarly, the limitations of the system are of interest, such as the angular discrimination capabilities and how far away the subjects can go before reconstruction accuracy significantly degrades. Thus, additional data is taken in a variety of extended scenarios, as shown in Fig. 15. For each individual additional scenario, a total of two minutes of data is taken, 80 seconds of which is added to the original training set and 40 seconds of which is used to analyze the capabilities.

To test MM-MURE's performance in a different environment, the radar was positioned on the dashboard inside a car, as shown in Fig. 15(a). This scenario is similar to the majority of the training dataset, with two subjects in view of the radar, separated in angle but located at approximately the same range. This setup is particularly relevant for radar-based vital signs monitoring applications [32]. Tests were conducted with subjects in both the front and back seats. For the back seat dataset, the radar was moved to the middle console to maintain a line of sight to the chests. As seen in Fig. 16(a), the network achieved average cosine similarities of 0.957 and 0.961 for the front and back seat scenarios, respectively. This demonstrates that performance can be maintained in a different environment, even with minimal representation in the training dataset.

The angular discrimination capabilities are also tested, as seen in Fig. 15(b). First, we determined the maximum distance from the radar at which the network could maintain high single-subject reconstruction accuracy. A single subject was placed in front of the radar at 0° off-broadside and moved from 1 m to 10 m away, with 10 m being the limit of our testing capability. By adjusting the 14 range bins fed into the network to include those where the subject is present, we found no significant drop in MM-MURE's single-subject performance up to the 10 m limit. Thus, by repeating the experiment with two subjects shoulder-to-shoulder, angular separation values (measured from chest-to-chest) of 2.9°–25° are achievable, and drops in performance can safely be attributed to angular separation limitations rather than the effects of increasing distance. The results, shown in Fig. 16(b), indicate that the network tends to favor the reconstruction of one subject over the other past approximately 6.4°, but maintains relatively high performance until approximately 2.9°, where the reconstruction of one subject completely dominates (0.972 cosine similarity versus 0.473).

As seen in Fig. 3, the virtual array in use has elements that provide some resolution in elevation, each with a 36° HPBW. While the majority of the experiments in this work aim to test performance versus azimuth angle, it is also of interest to test slight variations in elevation between subjects. There are slight variations in elevation present in the original multi-subject dataset due to differences in height between subjects, but a more exaggerated difference in elevation should be tested, as depicted in Fig. 15(c). Therefore, an additional dataset is taken where one subject is seated at approximately 0° elevation to the center of the chest while the other is located at approximately $+15^\circ$ elevation, with both being 0.75 m from the radar and separated by 60° in azimuth, and another where they switch places. The results of Fig. 16(c) show high average reconstruction cosine similarities of 0.955 and 0.929 are achieved for Subjects 1 and 2, respectively.

The effects of changes in elevation are further analyzed by varying the location of a single subject in elevation to well outside the main beam, as shown in Fig. 15(d). The height of the radar is adjusted up and down relative to a stationary subject located 0.75 m away, in order to make the angle to the center of the chest relative to broadside vary from -30° to $+30^\circ$ in steps of 15°. As seen in Fig. 16(d), high performance is achieved for elevation angles greater than 0° (worst case 0.956 cosine similarity), and and performance gradually decreases as the elevation angle moves to -30° (0.846 and 0.525 cosine similarity for -15° and -30°, respectively). This is explained by the fact that the majority of the data in the training set contains information where the subjects' chests are at or above 0° elevation. Thus, the system performs best when the radar is

positioned lower rather than higher. The network would likely account for significant decreases in elevation if provided with sufficient training data for these cases.

This leads to the most notable limitation of MM-MuRe, which is that the end-to-end development of the method causes it to be tailored to the application it is trained for, which is multi-subject monitoring of up to 2 people. So while it has been shown that performance can be maintained for variations in environment and measurement angle, extending its capability to monitoring 3 or more subjects simultaneously would warrant significantly more training data. So while it offers increased performance and efficiency compared to baselines when monitoring two or less people, more research is needed to achieve improved scalability.

VI. Conclusion

In this work, multi-subject respiration waveform reconstruction is achieved using an end-to-end neural network rather than traditional FMCW radar processing techniques for the first time. The novelty of the approach lies within the representation of the raw in-phase and quadrature radar data of a MIMO array as an input data matrix to an integrated neural network. This approach successfully achieves adequate angular separation of subjects and extraction of their respiratory waveforms, similar to the workings of an adaptive beamformer. MM-MuRe offers a 10% improvement in average reconstructed waveform similarity for both single- and multi-subject cases compared to baselines, as well as improved biomarker analysis accuracy. Results showcase the generalizability of the network to subjects not in the training set, scenarios where there is clutter in the scene, variations in environment and measurement angle, and to datasets taken with a significant time delay to those taken for training. Interesting next steps to this work would be investigating if the proposed network can be extended to reconstruct both fine respiration and heartbeat waveforms, analyzing its motion mitigation capabilities, and evaluating the methodology on other radar systems.

Acknowledgment

Authors report no conflict of interest.

References

[1] M. Kebe, R. Gadhafi, B. Mohammad, M. Sanduleanu, H. Saleh, and M. Al-Qutayri, "Human vital signs detection methods and potential using radars: A review," *Sensors*, vol. 20, no. 5, 2020, Art. no. 1454. [Online]. Available: https://www.mdpi.com/1424-8220/20/5/1454

[2] B. Walid, J. Ma, M. Ma, A. Qi, Y. Luo, and Y. Qi, "Recent advances in radar-based sleep monitoring—A review," in *2021 IEEE Int. Conf. Dependable, Autonomic Secure Comput., Int. Conf. Pervasive Intell. Comput., Int. Conf. Cloud Big Data Comput., Int. Conf. Cyber Sci. Technol. Congr.*, 2021, pp. 759–766.

[3] Harikesh, S. S. Chauhan, A. Basu, M. P. Abegaonkar, and S. K. Koul, "Through the wall human subject localization and respiration rate detection using multichannel Doppler radar," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1510–1518, Jan. 2021.

[4] A. Rahman, V. M. Lubecke, O. Boric-Lubecke, J. H. Prins, and T. Sakamoto, "Doppler radar techniques for accurate respiration characterization and subject identification," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 8, no. 2, pp. 350–359, Jun. 2018.

[5] V. L. Petrovic, M. M. Jankovic, A. V. Lupsic, V. R. Mihajlovic, and J. S. Popovic-Bozovic, "High-accuracy real-time monitoring of heart rate variability using 24 GHz continuous-wave Doppler radar," *IEEE Access*, vol. 7, pp. 74721–74733, 2019.

[6] A. Singh, S. U. Rehman, S. Yongchareon, and P. H. J. Chong, "Multi-resident non-contact vital sign monitoring using radar: A review," *IEEE Sensors J.*, vol. 21, no. 4, pp. 4061–4084, Feb. 2021.

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

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

[9] C. J. Bauder, A. Moadi, M. Joshi, and A. E. Fathy, "Multi-subject heart rate estimation and real-time tracking using a mmWave radar and trace carving algorithm," in *2022 IEEE/MTT-S Int. Microw. Symp.*, 2022, pp. 955–958.

[10] C. Feng et al., "MultiTarget vital signs measurement with chest motion imaging based on MIMO radar," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 11, pp. 4735–4747, Nov. 2021.

[11] Y. Zhao, V. Sark, M. Krstic, and E. Grass, "Multi-target vital signs remote monitoring using mmWave FMCW radar," in *2021 IEEE Microw. Theory Techn. Wireless Commun.*, 2021, pp. 290–295.

[12] S. Ahmed, J. Park, and S. H. Cho, "Effects of receiver beamforming for vital sign measurements using FMCW radar at various distances and angles," *Sensors*, vol. 22, no. 18, 2022, Art. no. 6877. [Online]. Available: https://www.mdpi.com/1424-8220/22/18/6877

[13] 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. 19th ACM Conf. Embedded Netw. Sensor Syst.*, Nov. 2021, pp. 111–124, doi: 10.1145/3485730.3485932.

[14] J. Gong, X. Zhang, K. Lin, J. Ren, Y. Zhang, and W. Qiu, "RF vital sign sensing under free body movement," *ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 5, no. 3, Sep. 2021, Art. no. 101, doi: 10.1145/3478090.

[15] Z. Chen, T. Zheng, C. Cai, and J. Luo, "Movi-Fi: Motion-robust vital signs waveform recovery via deep interpreted RF sensing," in *Proc. 27th Annu. Int. Conf. Mobile Comput. Netw.*, 2021, pp. 392–405, doi: 10.1145/3447993.3483251.

[16] H. Wang et al., "Here: Heartbeat signal reconstruction for low-power millimeter-wave radar based on deep learning," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 4004515.

[17] S. Wang, C. Han, J. Guo, and L. Sun, "MM-FGRM: Fine-grained respiratory monitoring using MIMO millimeter wave radar," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 4000913.

[18] M. A. Richards, *Fundamentals of Radar Signal Processing*, 3rd ed. New York, NY, USA: McGraw-Hill, 2022. [Online]. Available: https://www.accessengineeringlibrary.com/content/book/9781260468717

[19] Y. Huang, P. V. Brennan, D. Patrick, I. Weller, P. Roberts, and K. Hughes, "FMCW based MIMO imaging radar for maritime navigation," *Prog. Electromagn. Res.*, vol. 115, pp. 327–342, 2011.

[20] B. Zhang, B. Jiang, R. Zheng, X. Zhang, J. Li, and Q. Xu, "Pi-VIMO: Physiology-inspired robust vital sign monitoring using mmWave radars," *ACM Trans. Internet Things*, vol. 4, no. 2, May 2023, Art. no. 15, doi: 10.1145/3589347.

[21] U. Ha, S. Assana, and F. Adib, "Contactless seismocardiography via deep learning radars," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.*, 2020, doi: 10.1145/3372224.3419982.

[22] S. Jiang, A. Alkhateeb, D. W. Bliss, and Y. Rong, "Vision guided MIMO radar beamforming for enhanced vital signs detection in crowds," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 4, pp. 4640–4649, Aug. 2024.

[23] S. Roy and P. Fortier, "Maximal-ratio combining architectures and performance with channel estimation based on a training sequence," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1154–1164, Jul. 2004.

[24] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[25] D. Stoller, S. Ewert, and S. Dixon, "Wave-U-Net: A multi-scale neural network for end-to-end audio source separation," in *Proc. 19th Int. Soc. Music Inf. Retrieval Conf.*, Paris, France, September 23-27, 2018, pp. 334–340.

[26] RFBeamMicrowave, "V-MD3 radar transceiver." Accessed: 27 Mar. 2024. [Online]. Available: https://rfbeam.ch/product/v-md3-radar-transceiver/

[27] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, vol. 503, pp. 92–108, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222008426
[28] T. Dozat, "Incorporating nesterov momentum into adam," in *Proc. ICLR*, 2016, pp 1–4.
[29] Vernier, "Go direct respiration belt," 2023. [Online]. Available: https://www.wardsci.com/store/product/24015448/go-directtm-respiration-belt/
[30] M. Joshi, "Non-contact vital sign detection using mm-Wave radar," M.S. thesis, Univ. of Tennessee, Knoxville, TN, USA, 2022.
[31] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Monitoring vital signs using millimeter wave," in *Proc. 17th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2016, pp. 211–220, doi: 10.1145/2942358.2942381.
[32] M. Yang, X. Yang, L. Li, and L. Zhang, "In-Car multiple targets vital sign monitoring using location-based VMD algorithm," in *2018 10th Int. Conf. Wireless Commun. Signal Process.*, 2018, pp. 1–6.

Chandler Bauder (Member, IEEE) received the B.S. *(summa cum laude)* degree in electrical engineering in 2018 from the University of Tennessee, Knoxville, TN, USA, where he is currently working toward the Ph.D. degree in electrical engineering with a concentration in electromagnetics and RF circuits. Since 2018, he has been a Graduate Research Assistant with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville. He has worked on projects involving ground penetrating radar, microstrip filter design, multipactor breakdown simulation of RF components, and deep learning-based radar signal processing for contactless vital sign monitoring. In 2021, he was an Intern with the U.S. Naval Research Laboratory Radar Division's Advanced Concepts Group on a project involving microwave power beaming.

Abdel-Kareem Moadi (Graduate Student Member, IEEE) received the B.S. *(summa cum laude)* degree in electrical engineering in 2019 from the University of Tennessee, Knoxville, TN, USA, where he is currently working toward the Ph.D. degree in electrical engineering with a concentration in electromagnetics and RF circuits. Since 2020, he has been a Graduate Research Assistant with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville. He has worked on projects involving array signal processing and radar signal processing for contactless vital sign monitoring.

Vijaysrinivas Rajagopal received the B.S. and M.S. degrees in computer science from the University of Tennessee, Knoxville, TN, USA, in 2021 and 2022 respectively. During his collegiate career, he conducted research into optimizing neural networks for real-time, and edge-based applications. His current body of work is enabling industrial computer vision applications with a focus on automation and safety.

Tianhao Wu (Graduate Student Member, IEEE) received the B.S. degree in computer science from Northeast Agricultural University, Harbin, China, in 2022. He is currently working toward the Ph.D. degree in computer science with the University of Tennessee, Knoxville, TN, USA, under the supervision of Dr. Jian Liu, focusing on mobile sensing and machine learning for smart fitness and health monitoring. Since 2023, he has been a Graduate Research Assistant with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville.

Jian Liu (Member, IEEE) is currently an Assistant Professor with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA. He leads Mobile Sensing and Intelligence Security (MoSIS) Lab, University of Tennessee. His research work has been published at top-tier security/mobile-computing/HCI/AI venues, such as ACM MobiCom, IEEE S&P, ACM CCS, CVPR, AAAI, ECCV, ACM MobiSys, ACM SenSys, ACM UbiComp, and ICASSP, and has been regularly featured in the media including BBC News, Yahoo News, MIT Technology Review, NBC New York, IEEE Spectrum, WCBS TV, and Voice of America TV. He has also filed seven U.S. patents, two of which have been licensed to industrial companies. His research interests include robust and trustworthy machine learning, computational sensing, system security, and smart healthcare. He was the recipient of multiple awards, including two Best Paper Awards at IEEE SECON 2017 and IEEE CNS 2018, ACM SigMobile Research Highlights 2022, and 2023 IChemE Biochemical Engineering Award.

Aly E. Fathy (Life Fellow, IEEE) received the B.S.E.E. degree, the B.S. degree in pure and applied mathematics, and the M.S.E.E. degree from Ain Shams University, Cairo, Egypt, in 1975, 1979, and 1980, respectively, and the Ph.D. degree from the Polytechnic Institute of New York, Brooklyn, NY, USA, in 1984. In 1985, he joined RCA Research Laboratory (Sarnoff Corporation, now SRI International), Princeton, NJ, USA, as a Member of the Technical Staff. At Sarnoff Corporation, he was involved in the research and development of various enabling technologies, such as high-Tc superconductors, low-temperature cofired ceramic, and reconfigurable holographic antennas. In 2003, he joined the University of Tennessee, Knoxville, TN, USA, where he is currently a James W. McConnell Professor and Head of the Antenna Laboratory. He has authored or coauthored numerous transactions and conference papers, and holds 12 U.S. patents. His research interests include GNSS, vital sign detection, gait analysis, DBS antennas, wireless reconfigurable antennas, see-through walls radar, UWB systems, and high-efficiency combining structures.