CHAPTER 5

Target Response Separation from Mixed Signature

CHAPTER 5

Target Response Separation from Mixed Signature

In previous chapter, *Chapter 4*, the classification of the micro-Doppler (m-D) corresponding to various activities with the help of artificial intelligence method has been carried out. However, it was assumed that only a single target or target class is present in the observing radar channel. But in practical application, there is a huge probability of presence of other moving objects and the occurrence type depends upon the application. In the present chapter, *Chapter 5*, the aim is to extract the desired target's signature from the mixed m-D response recorded by the radar channel when multiple moving objects are present.

5.1 Overview

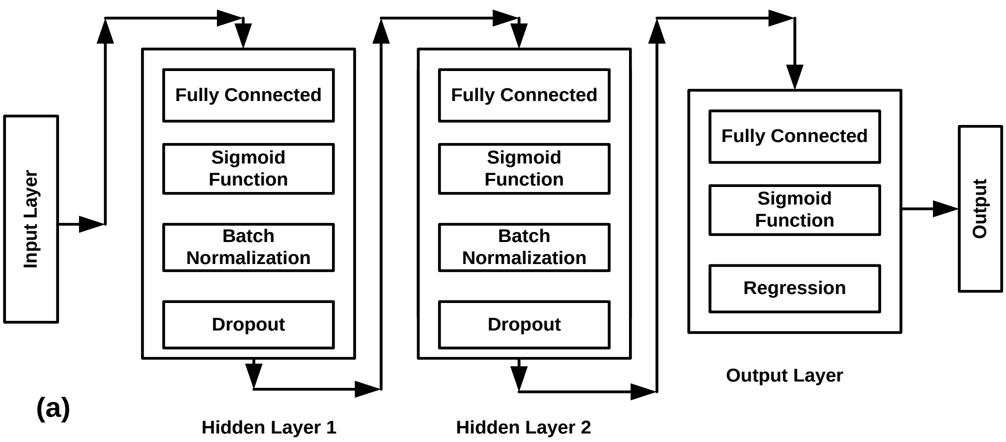
In previously reported work [Vishwakarma and Ram (2018)], the segregation of mixed movement response was addressed using time-domain information but here, a time-frequency based spectrogram governed segregation has been performed which improves the segregated signature correlation with their actual responses before mixing. A training data set has been prepared that includes a mixed m-D and a mask corresponding to the desired mover (target) for supervised training of DNN. The mixed m-D spectrogram has been considered by the regression DNN followed by the proper training of the data set. This yields the extracted signature of the desired moving target. The database prepared by [Vishwakarma and Ram (2017)] is used to implement the work. To check the validation accuracy of the trained model, mean square error is used. After validation, the model is also used to separate the actual field data, consisting of real mixed information of multiple moving targets, and the desired moving target response. The correlation of the disaggregated m-D with the clean condition m-D spectrogram for the two target and three target mixed response cases are achieved as 0.9807 and 0.9633, respectively. The higher value of correlation indicates the higher reliability of the method for m-D separation.

5.2. Methodology

5.2.1 Design of the Network

The masking based single-channel source separation problem falls under the regression category where a continuous prediction output is expected corresponding to the provided inputs. In this work, to define a DNN architecture, one input layer followed by two hidden layers is used. At the output end, one regression layer is used. Since the network is designed to solve a regression problem, the final stage of the network should have a regression layer that must be fully connected with the neuron of the previous layer [Lathuilière and Mesejo (2019)].

The complete network with layer details is shown in Figure 5.1(a). The neurons of the two hidden layers are fully interconnected which transmit information output to the neurons of the next layer through some activation function. The output batches of activation function are normalized before feeding to the next layer. In hidden layers, dropout is also provided to omit a few neurons of fully connected layers to reduce the network complexity and minimize the redundancy of information pass among the layers.



(a)

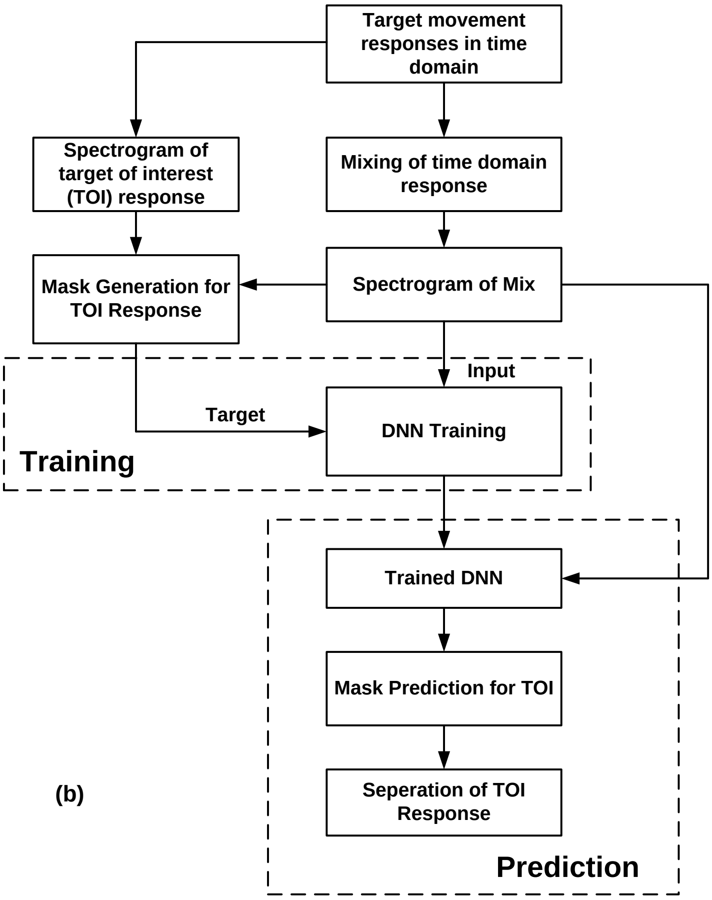


Figure 5.1: (a) Network layer details; (b) Adopted methodology.

5.2.2 Processing Methodology

The detailed adopted methodology of this work is shown in Figure 5.1(b). The data set of reference [Vishwakarma and Ram (2017)] is used here. The data of individual (clean) cases are used to prepare a long series of training spectrogram and training mask. Using these series, the generated mixed information is used as training input while prepared mask was used as the training target as shown in Figure 5.1 (b). Once the training of the network get completed, the trained network is fed with unknown data of mixed responses which provides a separated mask from which the desired target's m-D information can be easily obtained after the processing. The systematic steps of the methodology is explained in Table 5.1. Figure 5.2 shows the architecture comprising input and output data. It further indicates that all the frequency points of the spectrogram at particular instant is fed simultaneously to the first layer of the designed network which corresponds to the mask element at the same time instant on output end.

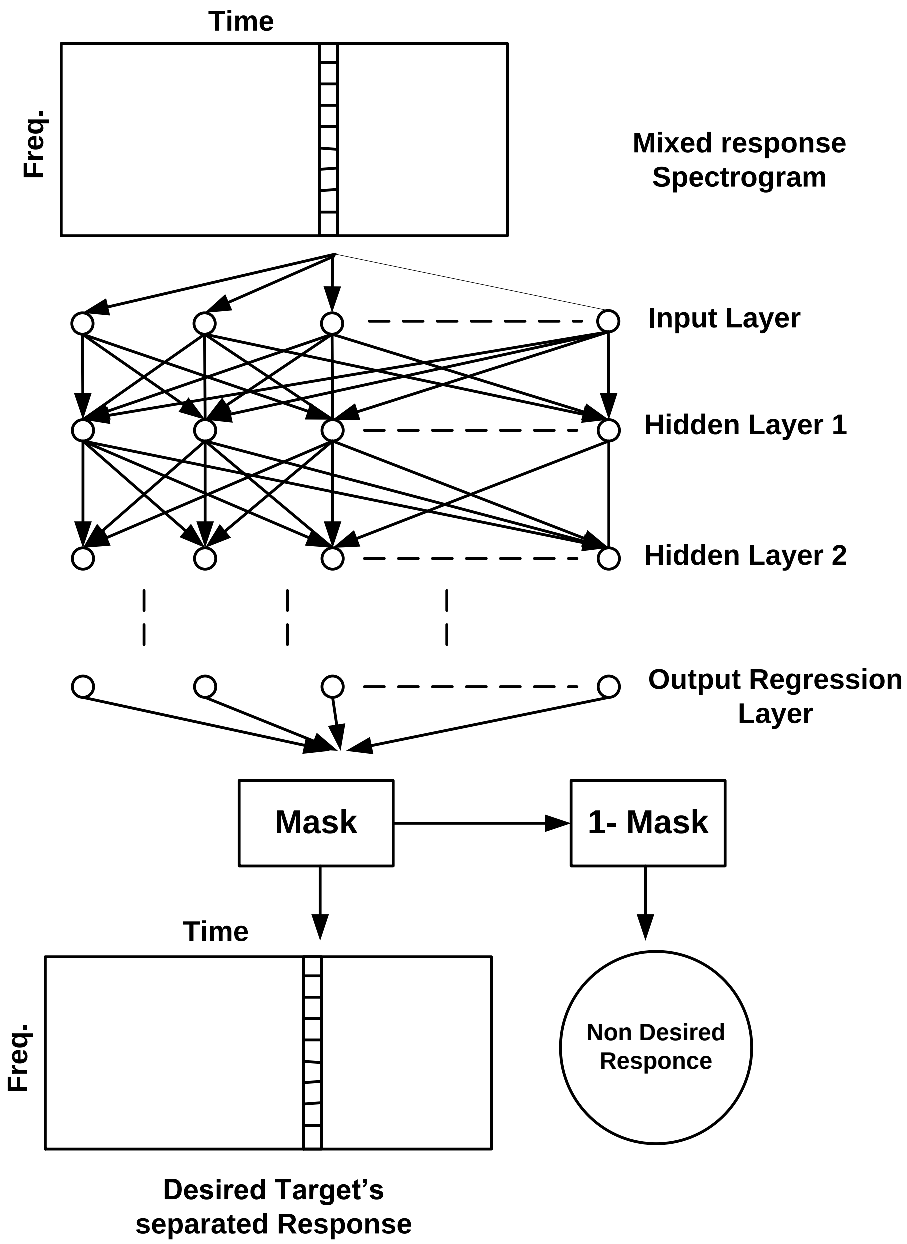


Figure 5.2: Used fully connected DNN regression model.

Table 5.1: Estimation process of desired target response from mixed response of multiple movers present in single channel.

Steps	Processing
1	Data Xi from different targets for $i = 1,2 \ldots I$ moving targets
2	Obtain input mix training data X_m^{train} using superposition of step 1 data.
3	Spectrogram of k^{th} desired target where $1 \le k \le I$ and spectrogram of mixed signal.
4	Soft mask M_k preparation for k^{th} desired moving target
5	Train the regression deep neural network using mix spectrogram as input and mask M_k as the regression training target
6	Give the new mix data to trained deep neural network which provide the estimated mask $\tilde{M}_k$ for k^{th} desired target.
7	Separated desired target response

5.3. Theoretical Background

In this section, the single channel m-D separation approach employing the time-frequency masking is discussed followed by the incorporation of DNN regression model.

5.3.1 Time Frequency Masking Approaches:

The two targets performing activity simultaneously in a single radar channel are represented with their individual radar return responses $x(t)$ and $y(t)$. The mixed signal recorded in the channel is represented by $z(t)$, which can be considered as the resultant of two individual responses $x(t)$ and $y(t)$ shown as:

$$z(t) = x(t) + y(t) \quad . \tag{5.1}$$

Let $X(\omega)$ represent the power spectrum of x(t), i. e., $X(\omega) = |\mathrm{F}(x(t))|^2$ where F is Fourier transform operator and $|.|^2$ represent the component wise squared magnitude. Similarly,

$Y(\omega)$ and $Z(\omega)$ denote the power spectra of y(t) and z(t), respectively. If x (t) and y(t) are uncorrelated, then the resultant power spectrum can be obtained by:

$$Z(\omega) = \mathrm{X}(\omega) + Y(\omega) \quad . \tag{5.2}$$

Assume logarithm of $X(\omega)$, $Y(\omega)$ and $Z(\omega)$ are X, Y, Z , respectively, then expression (5.2) can be written as :

$$Z = \ln(e^X + e^Y) \quad . \tag{5.3}$$

Expression (5.3) can be further rewritten as [Reddy and Raj (2007)]:

$$Z = \max(X, Y) + \chi \quad , \tag{5.4}$$

where $\chi = \ln(1 + e^{\min(\mathrm{X,Y}) - \max(\mathrm{X,Y})})$. This indicates that the maximum value of χ is ln (2) = 0.69 which occurs only when X and Y are equal. In general, it is highly unlikely that the target responses will offer nearly equal amount of backscattered energy at the same frequency within any sufficiently short analysis window. In other words, for a short analysis window, the logarithmic magnitudes of the response generated by two targets are usually significantly different and χ is sufficiently small. Hence, the final observed mixed response in the unit T-F cell of the spectrogram can be written as [Kim and Moon (2015)]:

$$Z \approx \max(X, Y) \quad . \tag{5.5}$$

The probability that the observed log spectral component z_d belongs to target T_x and not to target T_y conditioned on the fact that the entire observed vector is z is given by :

$$\mathrm{P}(x_d = z_d z) = P(\mathrm{x_d} > \mathrm{y_d z}) \quad . \tag{5.6}$$

In other words, the probability that in any T-F cell z_d belongs to T_x is simply the conditional probability that x_d is greater than y_d, i.e. $P(x_d > y_d | z)$. For the mentioned conditions discussed in (5.3) and (5.5), a soft mask based on the spectrogram magnitude can be defined as:

$$M_X(t, f) = \frac{|X(t, f)|}{|Z(t, f)|} . \tag{5.7}$$

Once the mask *Mx* is obtained, the desired signal in time frequency domain is given by [Wang and Chen (2018)]:

$$\tilde{X} = Z.*\mathrm{M_X} , \tag{5.8}$$

where (.*) is bin wise (element wise) multiplication operator. However, equation (5.7) can be further extended to separate more number of targets. In that case the mixed signal z(t) will contain the response of other non-desired targets also and the denominator term will get modify in the undesired signal $y(t) = y_1(t) + y_2(t) + \ldots y_n(t)$.

5.3.1.1 DNN Regression Model

A deep neural network (DNN) is an artificial neural network with multiple hidden layers in between the input and output layers [Mittal (2018)]. Such networks find the correct mathematical manipulation to convert an input into an output. The network moves information through the layers by calculating the probability of each output. Here, to solve a single-channel target m-D response separation problem, a multi-perceptron neural network is incorporated. The network consists of fully connected neurons arranged in a network of one input layer, two hidden layers, and one output layer to generate the output. The neurons of one layer communicate to the neurons of the following layer through some activation function whose outputs are batch normalized before feeding the information to the next stage.

Activation functions are used to propagate the output of a neuron forward in order to introduce non-linearity in the model. It helps network to preserve the relevant information by suppressing the irrelevant data points. The sigmoid function is used as activation function which can be given by:

$$\phi(z) = \frac{1}{1+e^{-z}} \quad . \tag{5.9}$$

The sigmoid function allows reduction in extreme or a typical values in valid data without eliminating them. Moreover, it converts independent variables of almost infinite range into simple probabilities between 0 and 1. Most of its output will be very close to the extremes of 0 or 1.

Batch size is an important hyper-parameter that controls the number of training samples to work through before the models internal parameters are updated. The batch normalization processes of each layer output offers several advantages as it increases the training speed, reduces the requirement of initial weights and a little bit regularize the network model. If batch mean is represented $\mu_B = \frac{1}{m}\sum_i^m x_i$ and the variance of a batch data x_i is specified by $\sigma_B^2 = \frac{1}{m}\sum_i^m (xi - \mu_B)^2$, then normalized batch data can be obtained as [Ioffe and Szegedy (2015)]:

$$\tilde{x}_i = \frac{xi - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad . \tag{5.10}$$

Dropout is a technique where randomly selected neurons are ignored during training. This means that their contribution to the activation of downstream neurons is temporally removed on the forward pass and any weight updates are not applied to the neuron on the backward pass. The dropout is performed at the output of a layer in order to minimize the over fitting and improving the generalization of a DNN.

5.4. Results and Discussion

In this work the data are used from the online available data base resource created by a research group of IIIT Delhi, India [Vishwakarma and Ram (2017)]. Form the mentioned data base, data for three classes of m-D response have been used as illustrated in Table 5.2. The three classes include (i) rotation of a table fan at different fan speeds labeled as TF, (ii) walking of a human along forward direction towards the radar labeled as FH and (iii) human subject walking away from the radar in facing his back to the radar labeled as BH. In human movement cases, to make the data base general for each labeled class, the human test subject was chosen with different body built like variable heights, weights and genders for each data. The data collection duration is for 2.7 seconds with total 1000 samples, which is sufficient for recording movement of human subject [Garcia-Rubia *et al.* (2014)]. For training of the network 100 data is selected randomly while for validation purpose a single data from each labeled classes is chosen as per the adopted methodology. For collection of the data, CW low cost radar is synthesized using a vector network analyzer, mode Keysight N9926A and a pair of horn antennas.

Table 5.2: Used data description.

S.N.	Data class	Symbol used	Used number
1	Rotating Table Fan	TF	101
2	Forward Moving Human	FH	101
3	Backward Moving Human	BH	101
4	Table Fan +Forward Human	TF+FH	1
5	Table Fan + Forward Human + Backward Human	TF+FH+BH	1

The data discussed above have been used in network training as well as in prediction purposes. In the network training processes, the data corresponding to a single object present in the channel are used while for prediction; two scenarios are utilized. In the first case, the data segments corresponding to the single-object cases are artificially mixed using time-domain superposition and on the feed of this merged data in the network, it separates the desired target m-D response. In the later case, the mixed Doppler response collected directly by the radar in the presence of two (or three) targets is fed to the trained network and the separated micro-Doppler response of the desired target is obtained.

The masking-based approach involves the training of DNN, with the known mixed information in the T-F domain as input training data and a T-F mask as the training target. While at the time of prediction, it produces a dissociation mask as an output upon the feed with unknown mixed information. For T-F representation, spectrogram has been obtained by using moving window short time Fourier transform (STFT). Here the generated spectrogram has the dimension of 128x1000 samples in frequency and time dimensions, respectively corresponding to a time domain single data with 1000 samples. The input layer of the designed network consists of 2560 input nodes. The other hidden layers along with along with output layer also consists of 2560 neurons each. The designed network uses the discussed methodology in Table 5.1 to address the separation of desired signals from the mixture of two and three target responses and discussed in details in the next *Sub-Section*.

5.4.1 Case (1): Desired Target Response from a Mixture of Two Target Responses

In this case, using the discussed methodology we separate the mixed response of a rotating table fan (TF) target and a forward-walking human towards the radar system (FH), as illustrated in Figure 5.3. In the training phase, 100 clean condition data from the classes of FH and TH are used to create a long training data series of those classes.

Additive superposition with equal weightage is performed to obtain a mixture of these two data series classes. The spectrograms of the created data series are compared (in each time frequency unit) to prepare the training mask for the desired target response employing equations (5.6) and (5.7).

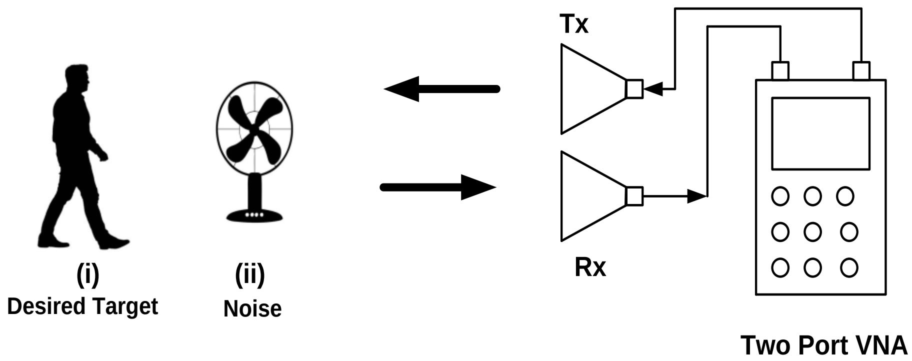
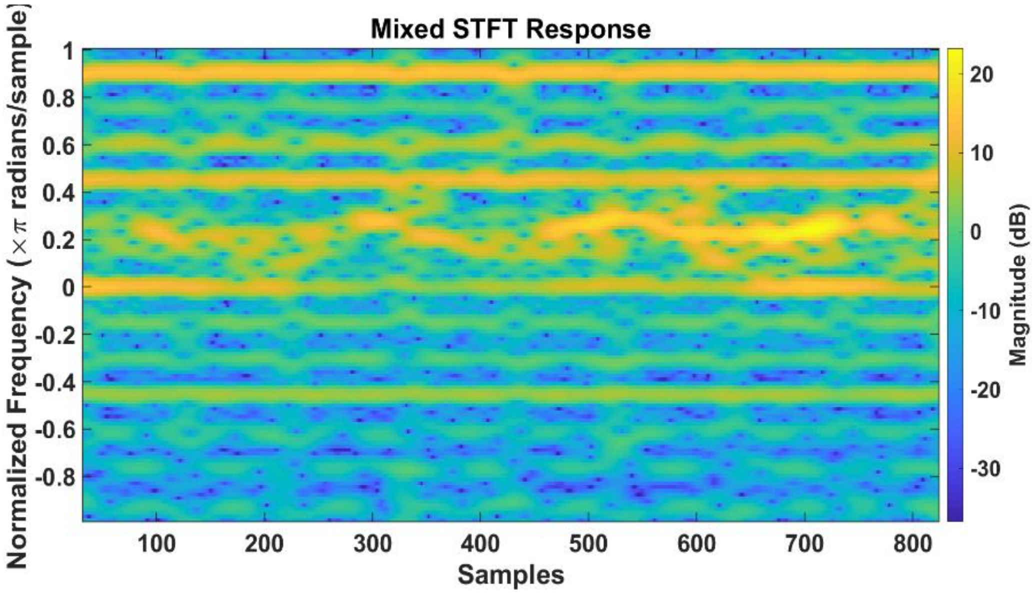


Figure 5.3: Two different types of movers present in the single channel.

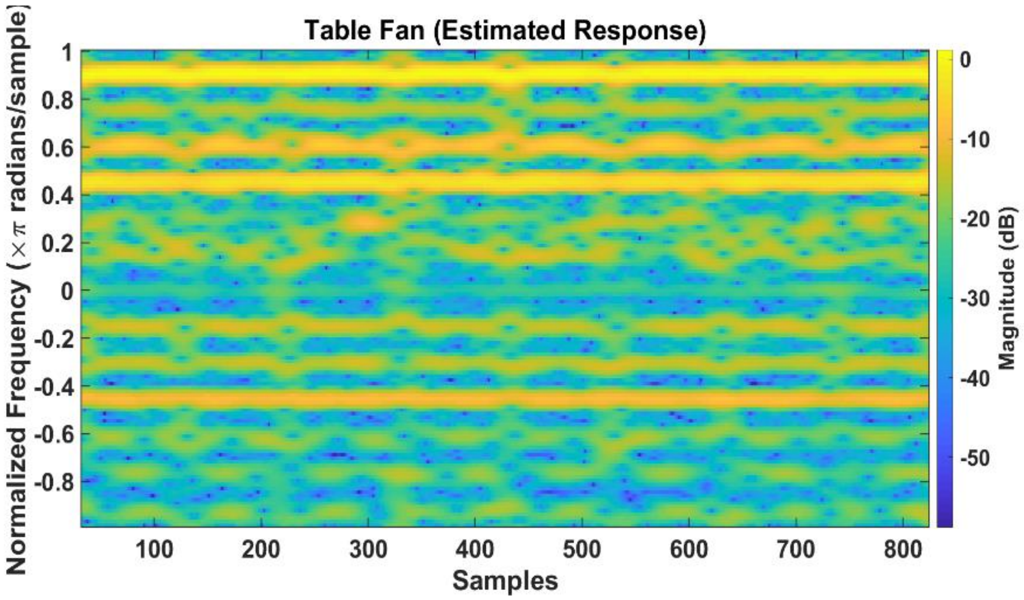
Once the network get trained with mixed series spectrogram and the created mask, the weight and biases of the DNN are adjusted and fixed. The trained network prediction is performed with the known separated clean condition responses of class FH and TF. A random data from each class i.e. FH and TF is chosen from Table 5.2 and their superposition is done to form a mixture whose spectrogram is shown in Figure 5.4(a).

The mixed signals spectrogram are fed to the trained DNN which produces the mask M_{FH} corresponding to the human walking in the forward direction. Since it is a case of two moving target response separation ($1-M_{FH}$) will work as the mask for the second class of target movement i.e., TF rotation, and the corresponding mask will be $M_{TF} = 1 - M_{FH}$. The element wise matrix multiplication with the mixed spectrogram will produce the individual activity spectrogram for the class of forward moving human FH and table fan rotation TF as shown in Figure 5.4(b) and Figure 5.4(c), respectively.

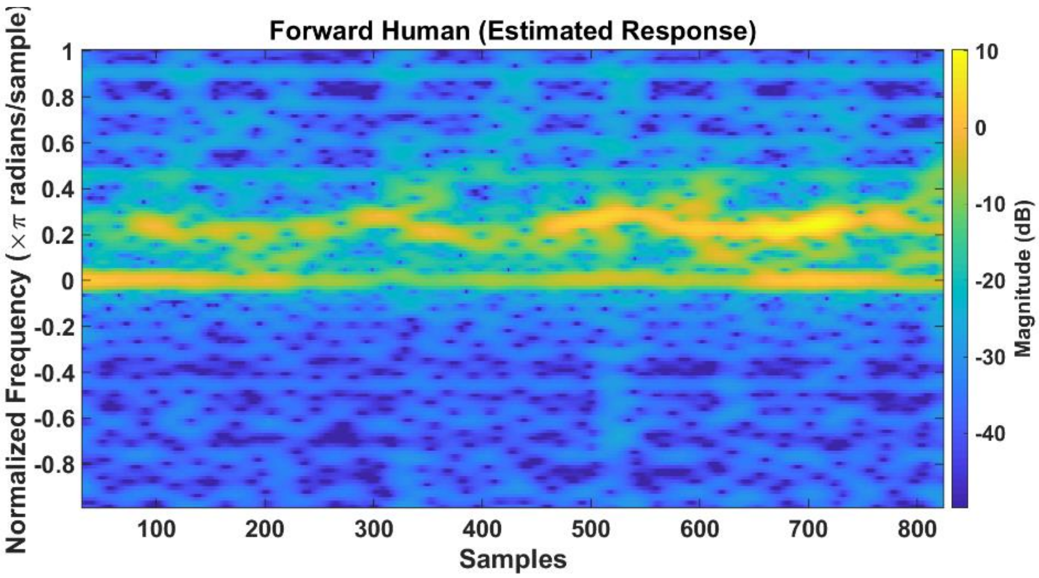
Those are easily identifiable at a monitoring site for recognition purposes. The accuracy of this method can be realized by producing the time-domain responses of the target classes, and comparing with the actual time-domain responses as shown in Figure 5.5(a) and Figure 5.5(b), respectively. This implies that the segregated responses follow almost the identical variations as the original signals before the mixing. By visual inspection the estimated response and actual responses appear quite similar for both the decomposed classes. The correlation of 98.07% value is realized.



(a)

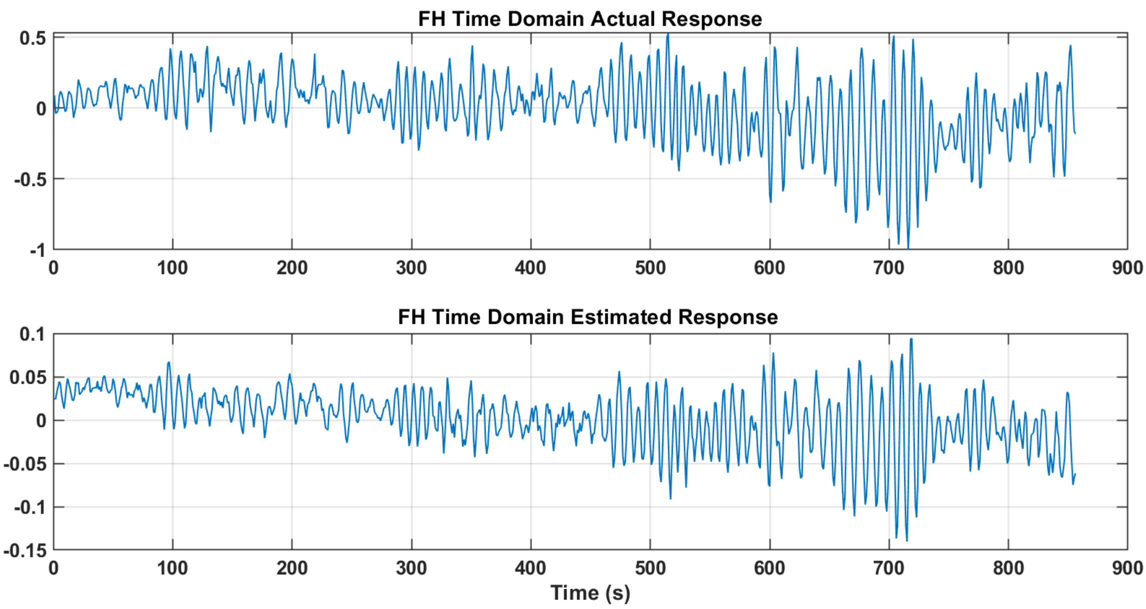


(b)

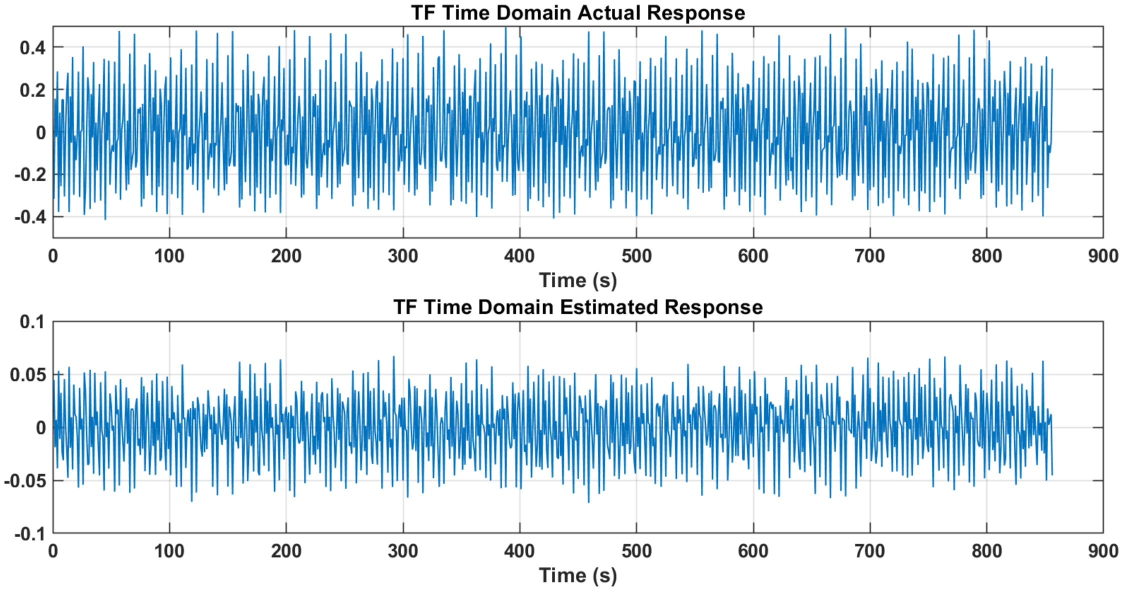


(c)

Figure 5.4: The decomposition of mix micro-Doppler data of a rotating table fan (TF) and forward moving human (FH) (a) Spectrogram of mixed signal in single channel; (b) Separated spectrogram of rotating table fan from the mixed micro-Doppler information; (c) Separated spectrogram of forward moving human target.



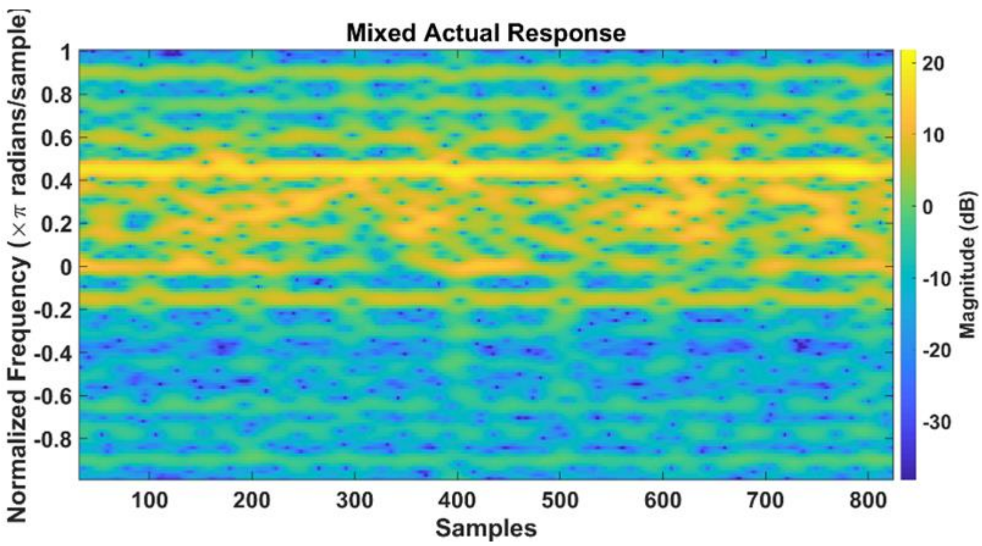
(a)



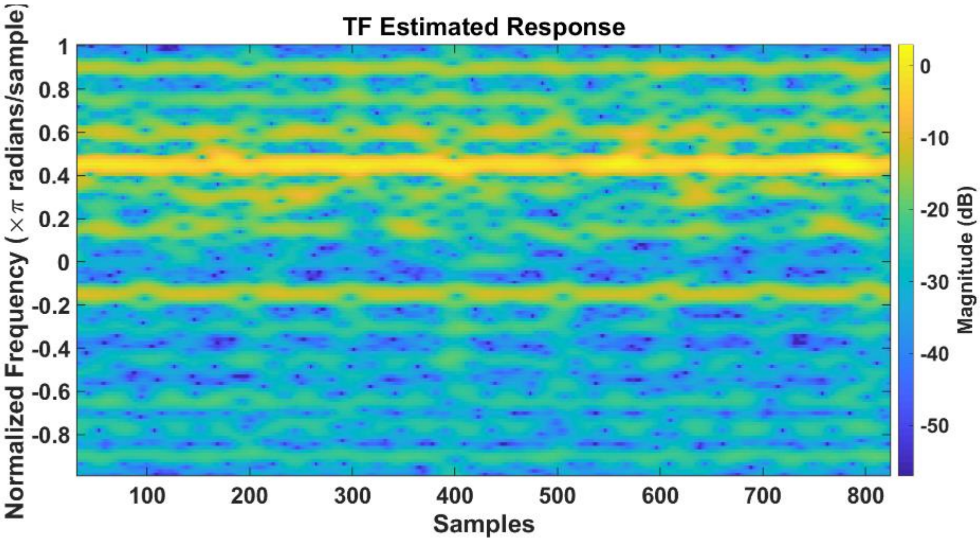
(b)

Figure 5.5: Time domain target responses actual and estimated after separation algorithm (a) Forward moving human; (b) Table fan target.

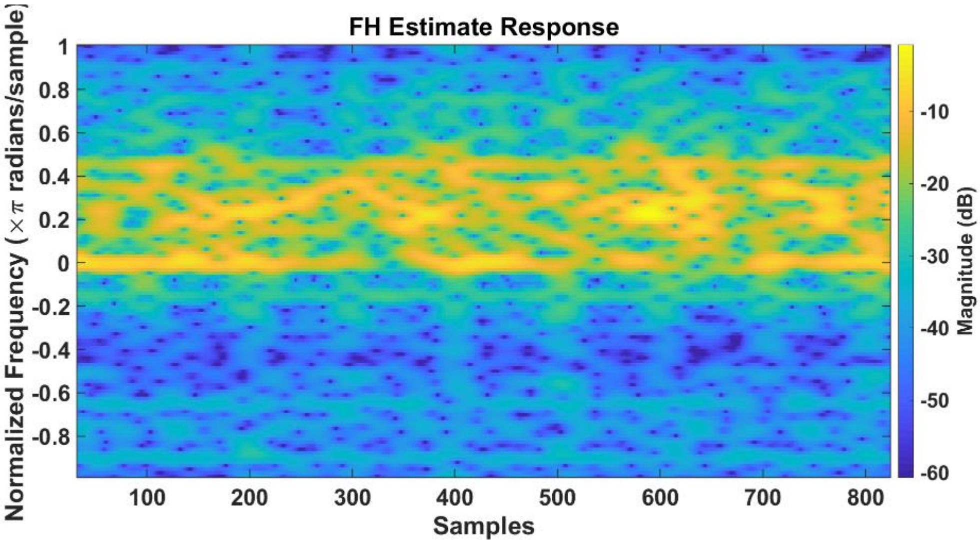
Testing of Model Over Field Data: The trained model is again tested with the actual scenario of multiple targets present in the channel and the mixed information captured by the radar when mounted at a surveillance site for field data capturing. The mixed data (Table 5.2 S.N. 4) are fed to the trained regression network whose spectrogram is given in Figure 5.6 (a); thereby producing two decomposed spectrogram responses as given in Figure 5.6 (b) and Figure 5.6(c).



(a)



(b)



(c)

Figure 5.6: Field test spectrogram of (a) Time domain response when TF and FH two target classes were present; (b) Separated TF class; (c) Separated FH class targets.

5.4.2 Case (2): Desired Target Response from the Mixture of Three Target Responses

The efficiency of the method is further tested for the simultaneous presence of three targets over a single observation channel as indicated in Figure 5.7. The three different target classes are FH, BH and TF (Table 5.2). Here, we assume that our desired target class is forward moving human subject approaching to the radar while the other two cases have been considered to be unwanted or noise case. The training procedure is similar as discussed in Case (1) for the three class mixed information case also. For prediction stage, the spectrogram of the mixed cases is shown in Figure 5.8(a). After application of the separation method, the extracted spectrogram of the desired target class, i. e. FH, is given in Figure 5.8(b) while the remaining portion of the mixed signal containing the BH and TF information as a noise is represented in Figure 8(c). It can be inferred that using this approach, the desired target response can be distinguished from the mixed m-D information of four or more targets provided the networks have been trained for their mixed responses. It is found that the correlation of separated m-D of 96.33% is achieved in comparison with the clean condition m-D response. It is evident from Table 5.3 that an increase in the number of moving targets from two to three in the radar observation channel increases the separation complexity; hence the obtained desired target response correlation with clean condition response reduces.

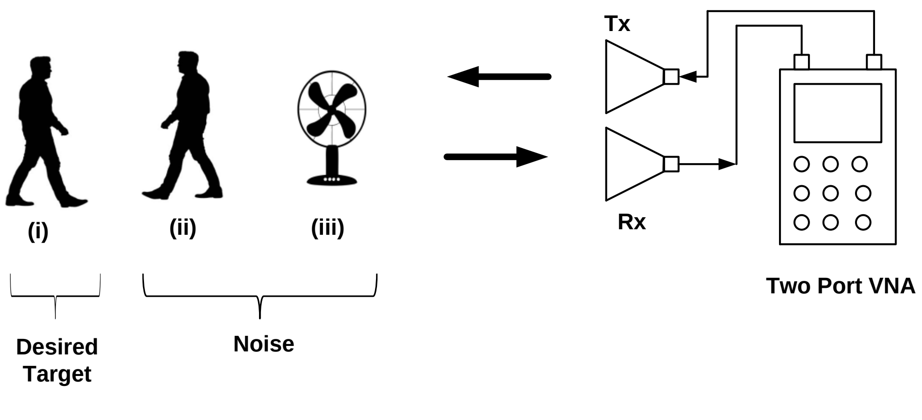
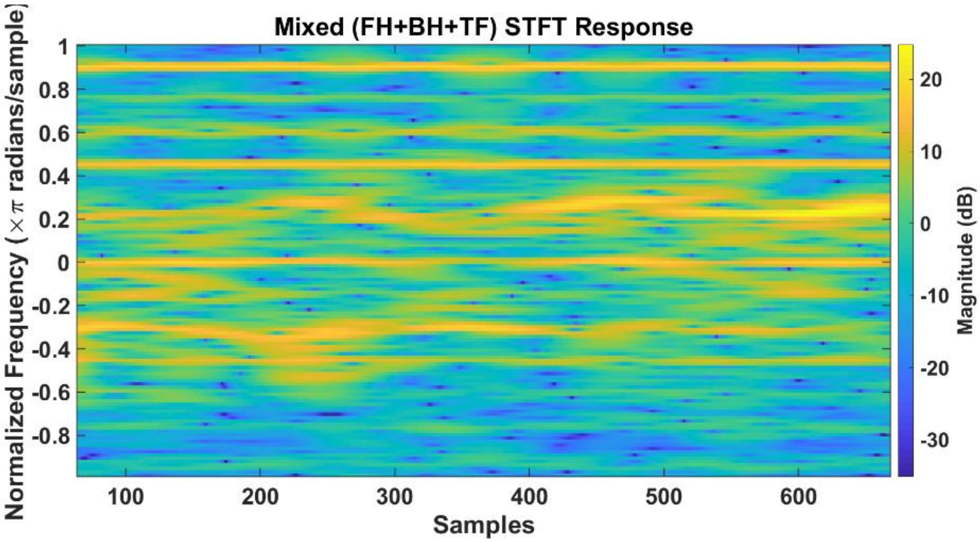
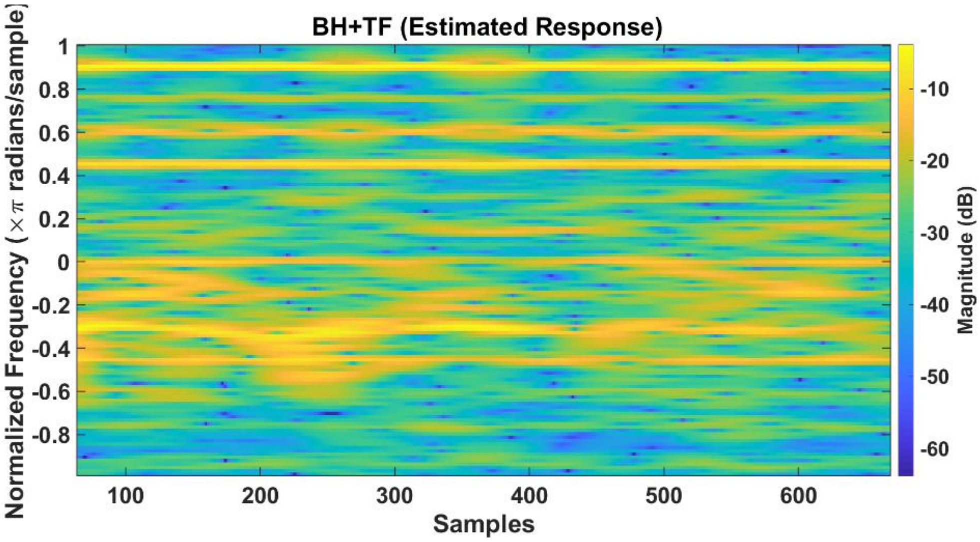


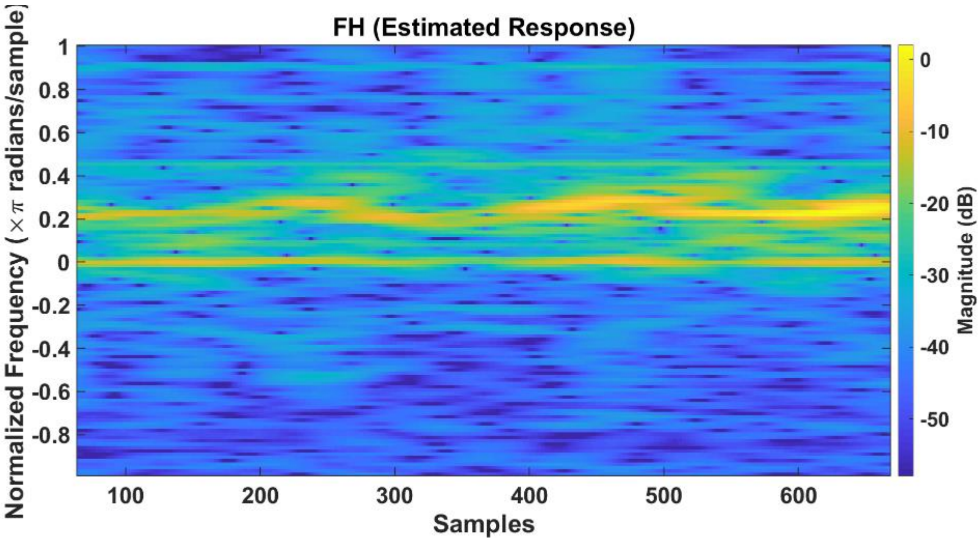
Figure 5.7: Three different types of movers present in the single channel.


Mixed (FH+BH+TF) STFT Response
Normalized Frequency (×π radians/sample)
Samples
Magnitude (dB)

(a)


BH+TF (Estimated Response)
Normalized Frequency (×π radians/sample)
Samples
Magnitude (dB)

(b)


FH (Estimated Response)
Normalized Frequency (×π radians/sample)
Samples
Magnitude (dB)

(c)

Figure 5.8: (a) The spectrogram of the mixed three targets (FH, BH and TF) response; (b) Extracted desired target FH response; (c) Undesired or noise remaining response.

Table 5.3: Estimated disaggregated signal correlation with original signal before mixing.

Case	**Target responses mixture**	**Correlation of desired target with original response**
1	2	98.07 %
2	3	96.33 %

5.5. Conclusions

In a single channel, the desired target m-D signature has been separated from the mixed response of multiple moving targets. The time frequency masking approach has been used for signature separation which require a training of a deep neural network. The predicted m-D response of the desired target has a very high degree of similarity with the original target m-D and can be verified by visual inspection. The used method is the first time showing the segregated micro-Doppler signature of activity correlation with the original signal before mixing and found the correlation of 98.07% and 96.33% for the mixture of two and three mixed activities simultaneously happening in a single channel radar.