Fall 2007

Device signal detection methods and time frequency analysis

Narayana Ravirala

Department:

DEVICE SIGNAL DETECTION METHODS

AND

TIME FREQUENCY ANALYSIS

by

NARAYANA RAVIRALA

A THESIS

Presented to the Faculty of the Graduate School of the

UNIVERSITY OF MISSOURI-ROLLA

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

2007

Approved by

Dr. Steve Grant, Advisor

Dr. Joe Stanley

Dr. Randy Moss

ABSTRACT

In this thesis the application of independent component analysis (ICA) to a multiple-source detection problem is studied. Specifically, a detection problem that is somewhat of a blend of the classical "signals with unknown parameters in noise" and "random signals in noise" problems is investigated. The idea is to detect the presence of a class of devices from their unintended radio frequency (RF) radiation. Because the signal is not an intended radiation, each device will emit a slightly different pattern. Moreover, these patterns may change under varying conditions such as battery voltage and ambient temperature. In addition, there may be several devices present simultaneously. When only a single device is present the classical matched filtering (MF) approach works well. However, when several devices are present they interfere with each other, especially when two or more are from the same class. To overcome this problem independent component analysis prior to MF is used to separate the mixture of signals into their statistically independent components. Then, MF may be performed without interference. In a further refinement, ICA-R is employed to considerably reduce the required number of MF's and thus the overall computational complexity.

Another approach emphasizing the detection of the devices by their patterns in the Time-frequency domain was also investigated. While theoretically promising, this approach was more computationally demanding and the results were less robust.

ACKNOWLEDGMENTS

I wish to express my gratitude to my advisor Dr. Steve Grant for providing excellent guidance and support during the course of my MS thesis work. I thank Dr. Joe Stanley and Dr. Randy Moss for serving as my committee members and being a part of my academic learning process. My sincere thanks to Dr. Daryl Beetner and Sarah Seguin for providing me the data and helping me in my research work. I wish to acknowledge the useful insights I recieved from my discussion on part of my research work with my friends, Brad Ashley, Peng Xie, Amit Shah. I wish to thank Dr. Sanjeev Agarwal for his suggestions regarding my research as well. I would also like to thank all my friends for their kindness and support.

I shall be forever grateful and indebted to my loving parents, R. Sadanandam and R. Krishna Kumari, and my sister Ramani Vutukuri for their constant encouragement and affection.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

Page

LIST OF TABLES

Page

NOMENCLATURE

Symbol	Description
MF	Matched Filtering
SVD	Singular Value Decomposition
ICA	Independent Component Analysis
TFD	Time-Frequency Distribution

1. INTRODUCTION

1.1. DETECTION SYSTEM

The basic purpose of this thesis is to classify and detect the presence or absence of electronic devices from their unintended radio frequency (RF) emissions. Figure 1.1 shows the basic block diagram of a detection system. A typical example of such a radiation source is the receiver of a remote controlled toy truck, though it could be any RF device which needs to be detected. The signal is attenuated by the channel which in this case is the air and ambient noise of the third floor hallway of the Electrical Engineering building at the University of Missouri-Rolla. Most of the data was collected by keeping the toy truck in the middle of the hallway (as shown in the figure) at varying distances, denoted by d.

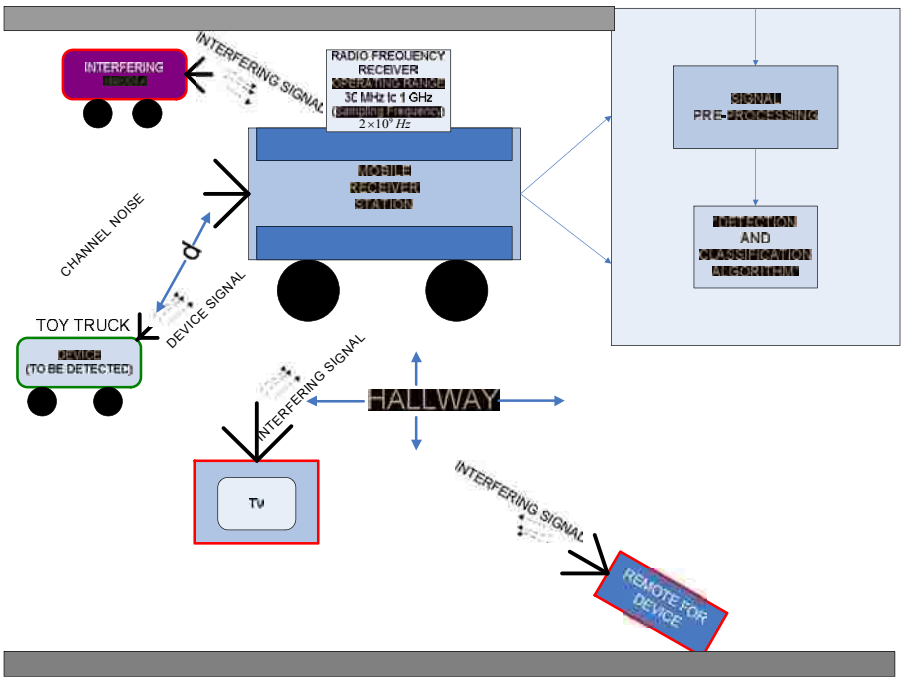


Figure 1.1.

Block Diagram of RF Device Detection System

Additionally, interference from unknown signals adds to the noise level. The received signal is $r(t) = \gamma s(t-\tau) + n(t);$ where γ the attenuation factor, τ is the time delay, $s(t)$ is the transmitted signal, and $n(t)$ is the combined noise from the channel and interfering unknown sources. The receiver detection system has two main blocks. One for pre-processing and another for detection and classification. In pre-processing the objective is to increase the SNR if the device signal is present. One way of doing this is to filter the signal by a bandpass filter. The detection methods will be discussed later.

1.2. RECEIVER OPERATING CHARACTERISTICS

In this section the basic theory about the detection problem and the characteristics of the detection system are introduced.

1.2.1. Detection Theory. It is assumed that the device's signal has some unique characteristics in the time and frequency domains that can be used for its detection. The detection problem can be classified into three main categories as follows [1]:

Case 1) Known Signals in Noise. This is the simplest form of detection. The signal is $s(t)$, and the noise is $n(t)$. The received signal is given as

$$r(t) = s(t) + n(t) \tag{1}$$

and, other than noise, there are no unknown factors affecting the signal; hence, the detection of signals in such cases is very simple.

Case 2) Signals with Unknown Parameters in Noise. Consider the example where RADAR is used for detection of an object. The radar transmits a pulse at a certain frequency given as

$$S(t) = \sin(\omega_c t) \tag{2}$$

If the object is present, then the reflected signal from the target is given by

$$R(t) = \xi \sin(\omega_c (t-\tau) + \theta) \tag{3}$$

where ξ is the attenuation factor, τ is the time delay, and θ is the phase shift. All of these listed parameters are unknown [1].

Case 3) Random Signals in Noise. An example for this case would be a passive SONAR detection system, which tries to identify the random noise generated by engines, propellers, etc [1]. Here, the received also includes environment noise [1]. Therefore, a

decision boundary has to be carefully selected to classify different random signals. In the framework of this thesis the device detection problem can be broadly classified within cases 2 and 3.

1.2.2. Receiver Operating Characteristics. The Receiver Operating Characteristic or the ROC curve is the standard performance evaluation criterion for a detection system. Figure 1.2 [1] depicts the signal generator of an RF device that is being detected. Consider two Hypotheses, $H1$ and *H0*. Hypothesis *H1* is true when the device is present and active, and hypothesis $H0$ is true when the device is not present or inactive. The objective is to detect the presence or activity of the device.

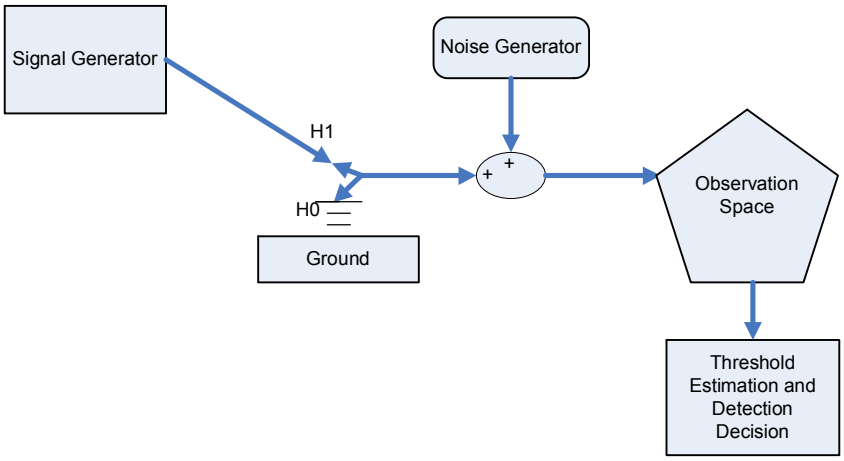


Figure 1.2. Illustration of a Detection System for ROC

Then, under the two hypotheses $H1: r = s + n$ and $H0: r = n,$ where n is the noise which, for this example, is assumed to be Gaussian and s is the signal. A decision rule is made such that each observation is assigned to one of the hypotheses [1]. As shown in Figure 1.3 [1], the decision rule divides the observation space into two sections *Z0* and *Z1*. The regions *Z0* and *Z1* correspond to the two hypotheses *H0* and *H1*.

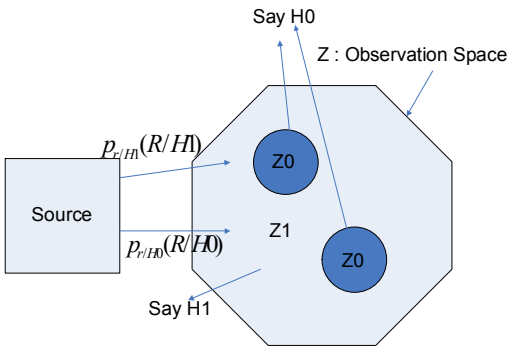


Figure 1.3. Observation Space and Classification Regions

If R falls into *Z0*, *H0* is declared. Likewise, if R falls into *Z1*, *H1* is declared. Here $P_{r/H1}(R/H1)$ is the probability distribution of the observed random variable R when the source is present and $P_{r/H0}(R/H0)$ is the probability distribution of the observed random variable R when the source is not present. Within the scope of the observation space,

$$P_D = \int_{Z1} P_{r/H1}(R/H1)dR \tag{4}$$

where P_D is the probability of detection and

$$P_F = \int_{Z1} P_{r/H0}(R/H0)dR \tag{5}$$

where P_F is the probability of false detection. Let the noise added to the signal being detected is Gaussian with mean 0 and variance σ and the signal, s = 1, if the device is present and 0, if the device is not present.

If the signal is present, then in observation space probability distribution of the received signal with hypothesis *H1* is Gaussian distributed with a mean 1 and variance σ. If the signal is not present, then its probability distribution under the hypothesis *H0* will be Gaussian with a mean 0 and variance σ. Figures 1.4 and 1.5 shows the detection region and false alarm region, respectively. The distribution of the received signal under the hypothesis *H1* and *H0* is given as:

$$P_{r/H1}(R \mid H1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(R-\mu)^2}{2\sigma^2}}; \mu = 1; s(t) = 1 \quad (6)$$

$$P_{r/H0}(R \mid H0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(R-\mu)^2}{2\sigma^2}}; \mu = 0; s(t) = 0 \quad (7)$$

and the probability of detection P_D is the area under $P_{r/H1}(R \mid H1)$, in the detection region

$$P_D = \int_{Threshold}^{\infty} P_{r/H1}(R \mid H1)\, dR \quad (8)$$

while the probability of false alarm, P_F, is the area under $P_{r/H0}(R \mid H0)$ in the false alarm region,

$$P_F = \int_{Threshold}^{\infty} P_{r/H0}(R \mid H0)\, dR \quad (9)$$

The receiver operating characteristics (ROC) curve plots the probability of detection versus the probability of false alarm. Let d be the difference between the two means. As shown in Figure 1.6, d increases, P_D for a given P_F also increases. Generally, the closer the ROC curve approaches the left and upper sides of the graph, the better the detection performance.

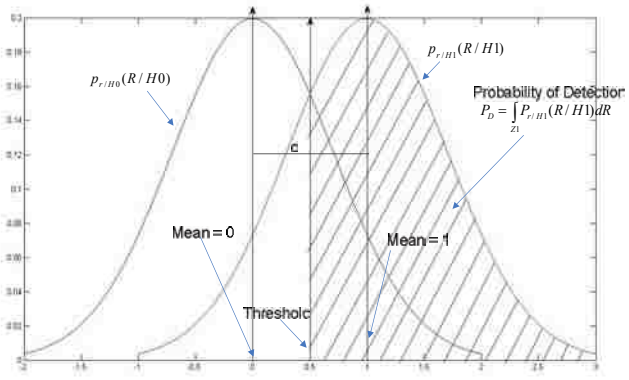


Figure 1.4. Detection Region

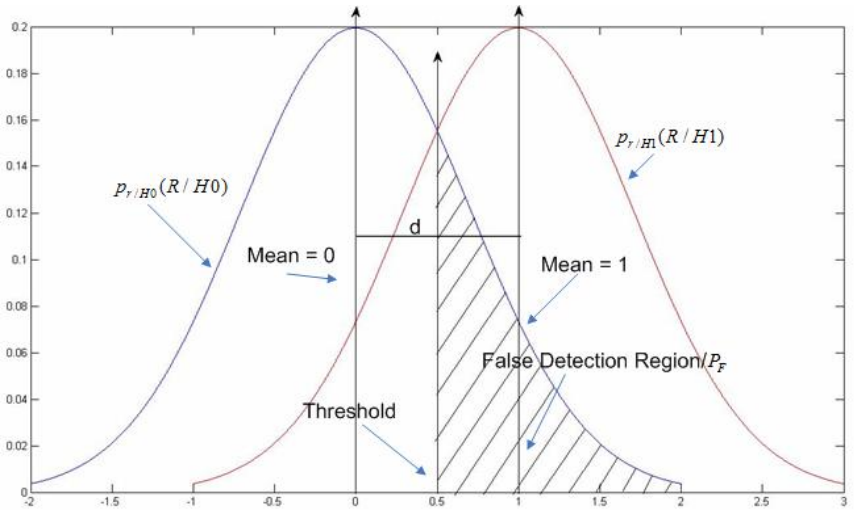


Figure 1.5. False Detection Region

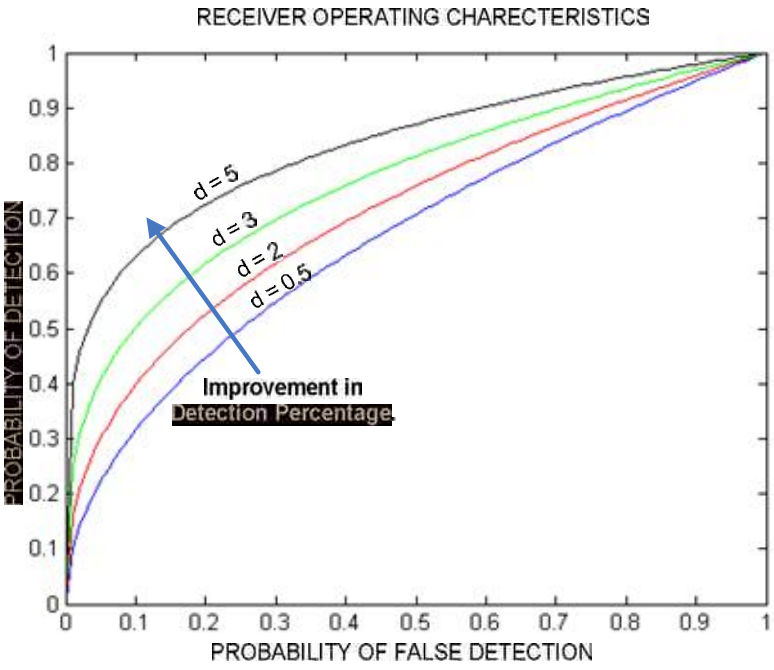


Figure 1.6. ROC Curve

1.3. SNR MEASUREMENT

Consider a signal, $x(t)$, shown in Figure 1.7 and a noisy signal, as shown in Figure 1.8, denoted by

$$y(t) = Ax(t) + n(t) \tag{10}$$

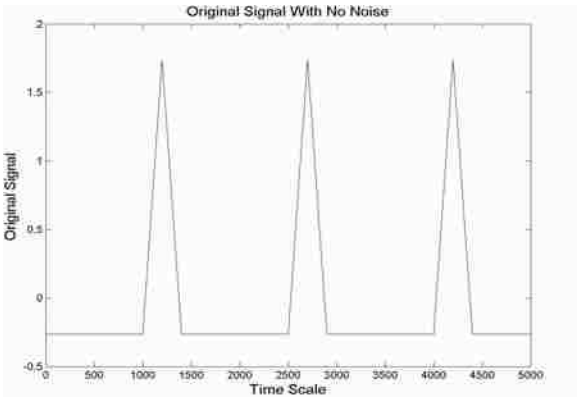


Figure 1.7. Original Clean Signal Without Noise

Now, the objective is to estimate the SNR of $y(t)$ with the knowledge of $x(t)$. The autocorrelation of the $x(t)$ is shown in Figure 1.9. The energy of the signal is given by

$$E_x = E[x(t)^2]. \tag{11}$$

The correlation of $y(t)$ with $x(t)$ is shown in Figure 1.10 and the zeroth lag correlation of $x(t)$ and $y(t)$ is given by

$$R_{yx}(0) = E[\{Ax(t) + n(t)\}\{x(t)] = AE_x. \tag{12}$$

$$A = \frac{R_{yx}(0)}{E_x} \tag{13}$$

which is the device signal amplitude present in $y(t)$ can be obtained. The energy of the signal is given by

$$S = A^2 E_x \tag{14}$$

and the energy of the noise is

$$N = E_y - A^2 E_x, \tag{15}$$

where

$$E_y = E[\{y(t)\}\{y(t)\}] \tag{16}$$

and so

$$\text{SNR} = 10\log10(\text{S/N}) \tag{17}$$

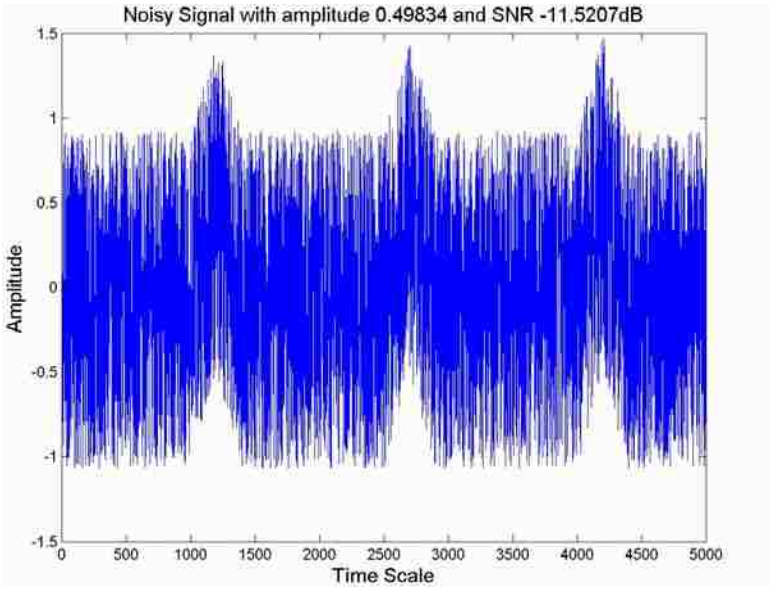


Figure 1.8. Noisy Test Signal

1.4. OUTLINE OF SECTIONS

In the next section, the basics of matched filtering (MF) are explained; its application with respect to the device data is also discussed. The performance of MF based on the ROC curve will be evaluated. Here, most of the results from MF are confined to signals measured in environments with a low level of interference. MF may

under-perform if there are other strong interfering signals or noise. Hence, in section 3 the concept of Independent Component Analysis (ICA) which is a method available to extract independent signal components from mixed signals, and ICA with Reference (ICA-R) are studied. These methods are used for extracting the signal when noise presence is high. A further modification of the ICA-R is discussed with respect to the device detection.

In Section 4 some of the methods of obtaining Time-frequency spectrums are compared. In Section 5, the SVD (Singular Value Decomposition) of the TFD is presented for signal detection.

SVD is one of the methods available for dimensionality reduction of the large TFD matrix. The ROC curve with a TFD-SVD detection method will be discussed and compared with the one obtained from MF.

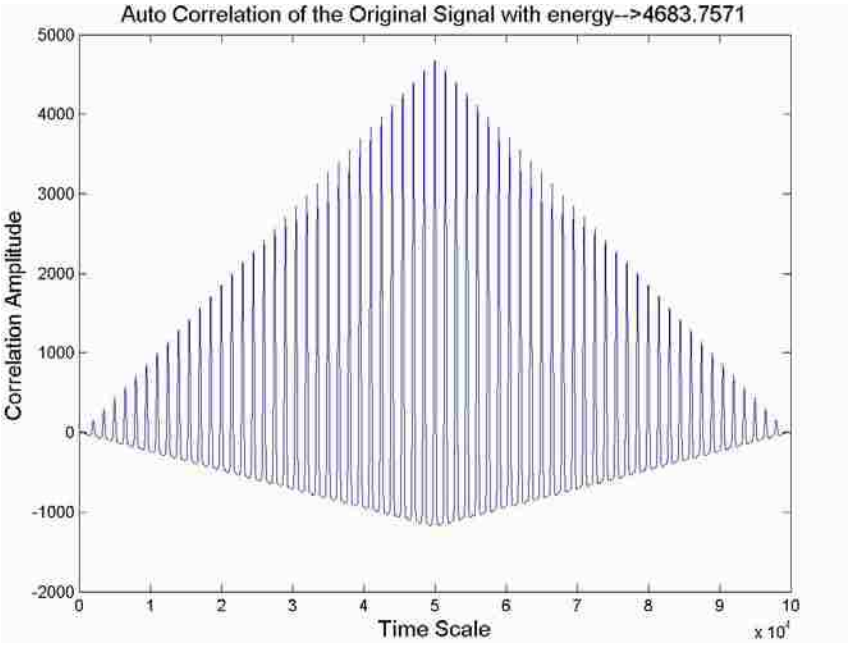


Figure 1.9. Auto Correlation of the Original Signal

As in the MF, a TFD-SVD based method performs poorly if the signal has significant noise interference. Hence, in section 6, ICA-R is applied over the SV's

obtained from SVD of the TFD to extract the device signal component, if present.

All the above mentioned results are compared and discussed in the conclusion.

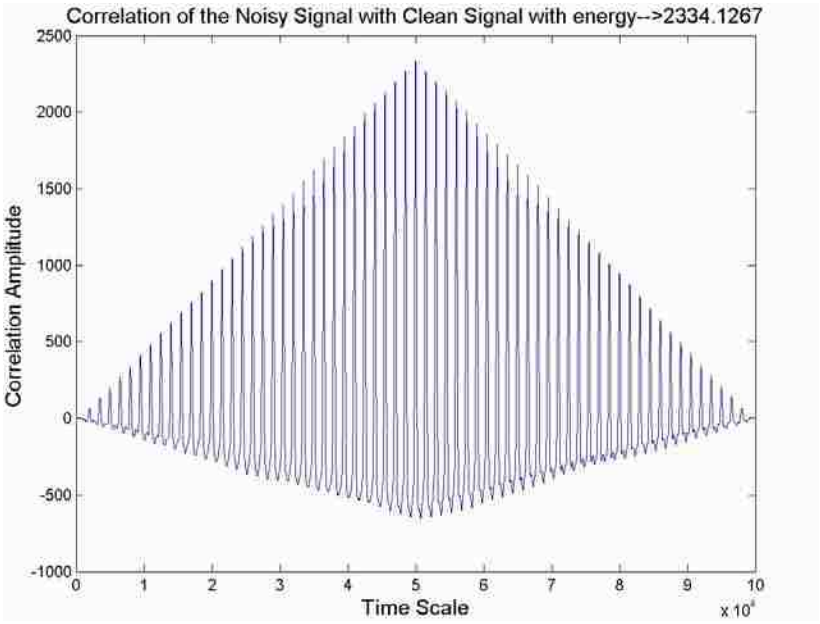


Figure 1.10. Correlation of Noisy Signal with Clean Signal

2. MATCHED FILTERING

MF is one of the most common methods for the detection of known signals. For the unintended RF signal detection problem, a template signal is created by averaging the periodic bursts of the signal from different measurements. The difficulty in template formation with such signals is caused by 1) noise presence and 2) template inconsistency. For testing the performance of this type of MF, the signals are measured at different distances and under different conditions creating different noise levels.

2.1. MATCHED FILTERING BASICS

MF is one of the standard methods used for the identification and detection of the known type of signals.

The basic algorithm of the MF is illustrated in Figure 2.1. In the diagram the performance of the MF is given for two cases, i.e., when the signal is present and when the signal is not present. Only one period of the signal is used as the matched filter. When the signal is present and is convolved with the matched filter the output has periodic peaks. The threshold can be set on these periodically occurring peaks to detect the presence of the signal. As shown in Figure 2.1, the peaks are significantly above the threshold when the signal is present. When the signal is not present then there are no periodic peaks and the peaks that occur are well below the threshold.

For the device signals, the template that would be required for the MF would not be consistent at each measurement. However, using different sets of measurements taken from a device (for example, a toy truck) toy truck kept in the middle of the hallway at a distance of 3m from the receiver, the template is averaged. Empirically, we find this approach to yield good template.

2.2. SIGNAL PRE-PROCESSING

The pre-processing of the signal is an important step in device detection.

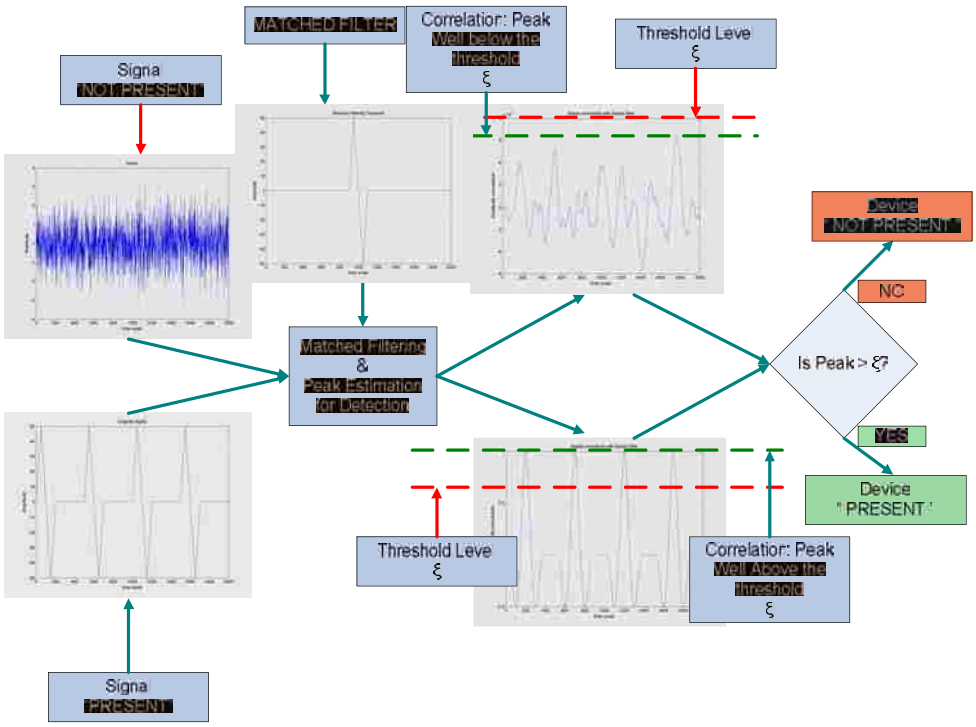


Figure 2.1. Matched Filtering Outline

One can improve the SNR by having prior knowledge about the frequency content of the signal so that it can be appropriately filtered. Consider the signal shown in Figure 2.2 with $SNR = -16.89$ dB. The signal has been measured in the middle of the hallway at a distance of 3m from the receiver. The SNR is -16.89 dB. The TFD (Time Frequency Distribution) of the signal is shown in Figure 2.3, in which the periodicity of the signal is prominent in the time domain. Also, the signal power is high in the frequency range of 33-66MHz. Hence, if a bandpass filter is designed to allow only these frequency ranges, then the signal to noise ratio can be boosted. Figures 2.4 and 2.5, shows the high pass filter and a low pass filter. The high pass filter has it's cut off frequency at 33MHz and the low pass has its cut off frequency at 66MHz. So, together they effectively act as a bandpass filter allowing frequencies between 33MHz and 66MHz, respectively.

Figure 2.6 shows that the *SNR* is improved from -16.79dB to - 4.79dB with a well designed bandpass filter. This further will improve the signal detection performance.

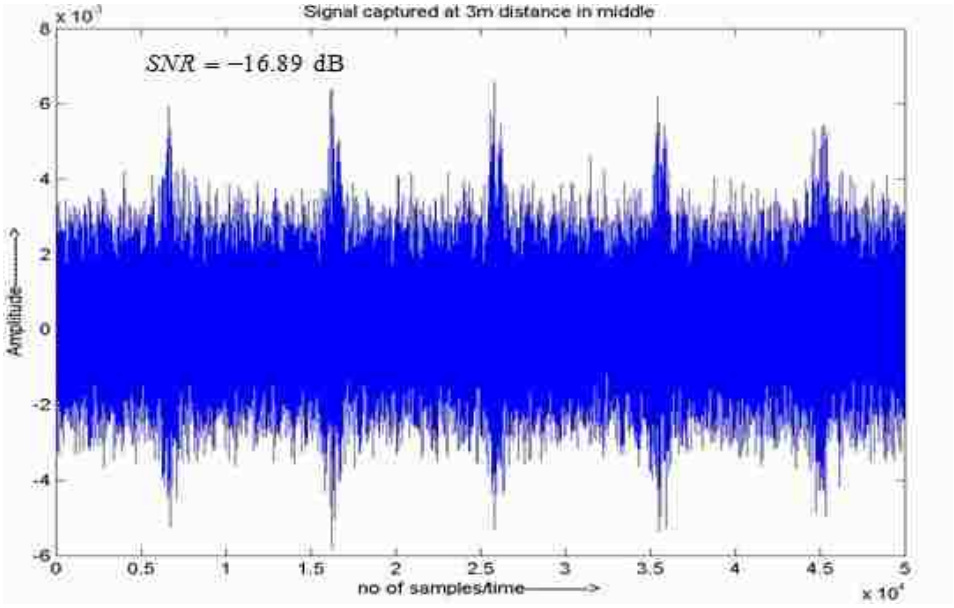


Figure 2.2. Signal

2.3. MATCHED FILTERING RESULTS AND ROC CURVE

The MF technique for signal detection can be made robust by creating a template that on average represents the periodic part of the signal. The template is obtained by averaging the periodic part of the signal over a few measurements taken keeping the toy truck in the middle of the hallway at a distance of 3m from the receiver. The template for the signal to be used for MF is shown in Figure 2.7. As shown in Figure 2.8, when the signal is convolved with this matched filter template, peaks occur with the same periodicity as the template in the signal. Hence, by measuring an approximate distance between the peaks, the signal can be detected.

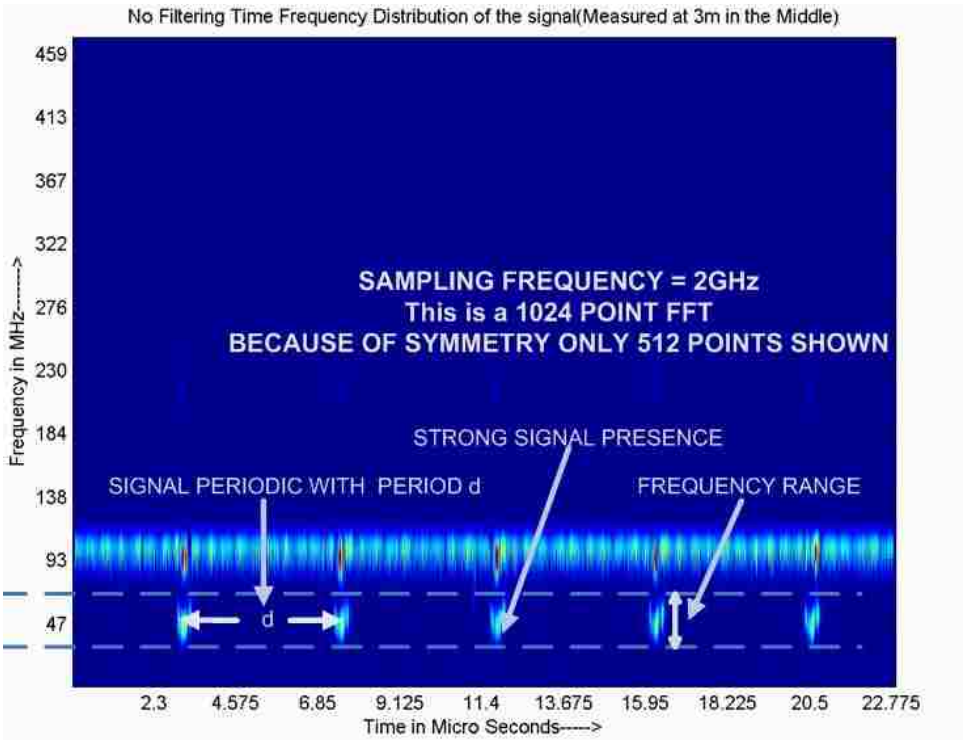


Figure 2.3. Time Frequency Distribution of the Signal

Figure 2.4. High Pass Filter

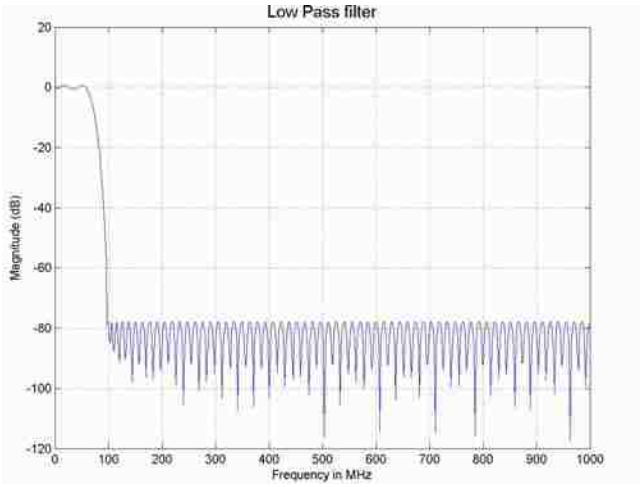


Figure 2.5. Low Pass Filter

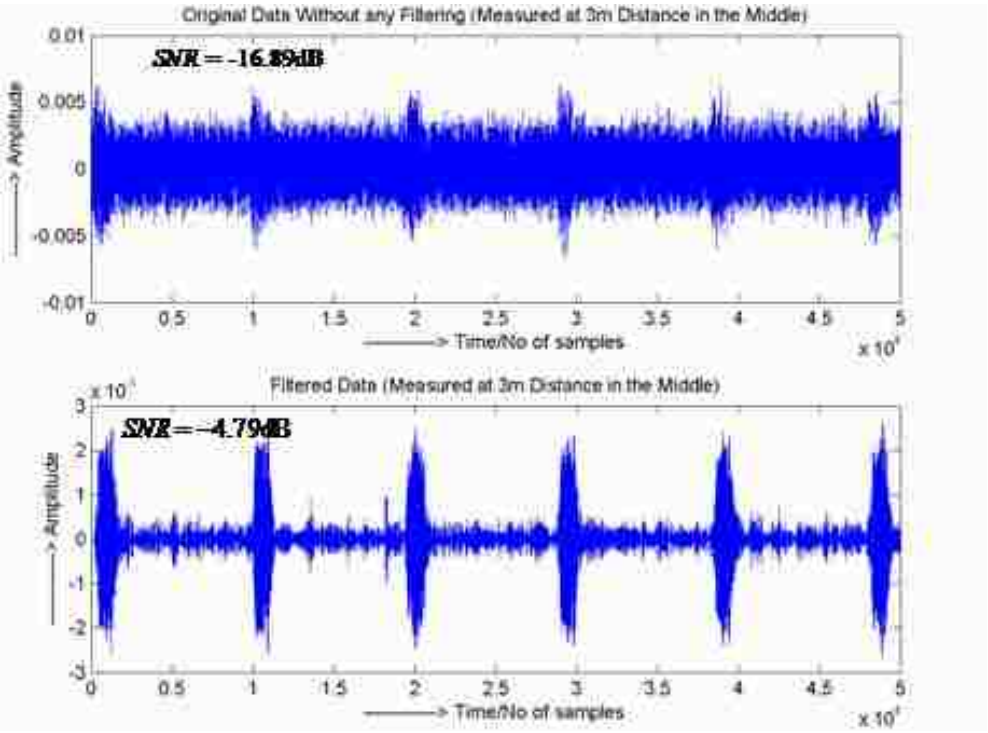


Figure 2.6. Improvement from Pre-Processing

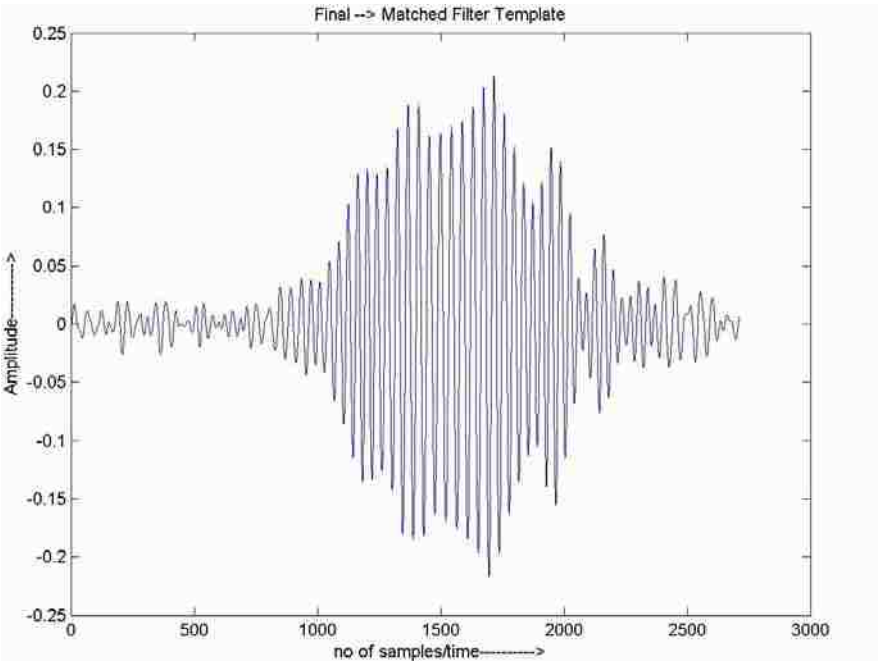


Figure 2.7. Template for Matched Filter

In this particular signal's case, the average distance between the peaks was approximately 10,000 samples. The algorithm flow for the detector is shown in Figure 2.9. The signal is bandpass filtered, then convolved with the matched filter. Then peaks in the convoluted signal with the bandpass filter are detected. Approximate distance between the peaks is estimated, which should equal d as shown in Figure 2.8. The percentage tolerance in the distance between the peaks is given by Δ. Figure 2.10 shows the ROC curve using MF. The varying parameter for the detector is Δ. Generating the ROC curve involved the use of 242 data files, in which the data was measured at varying distances from the device. The signals were measured at distances of 3m, 5m, 10m, 15m, 20m, 25m, 30m, and 35m in the middle of the hallway. The number of data files for checking the false detection percentage is 289, which included some of the other toy truck signals and noise signals. Figure 2.11 shows the variation of the SNR for all data files in which the data was present and was detected.

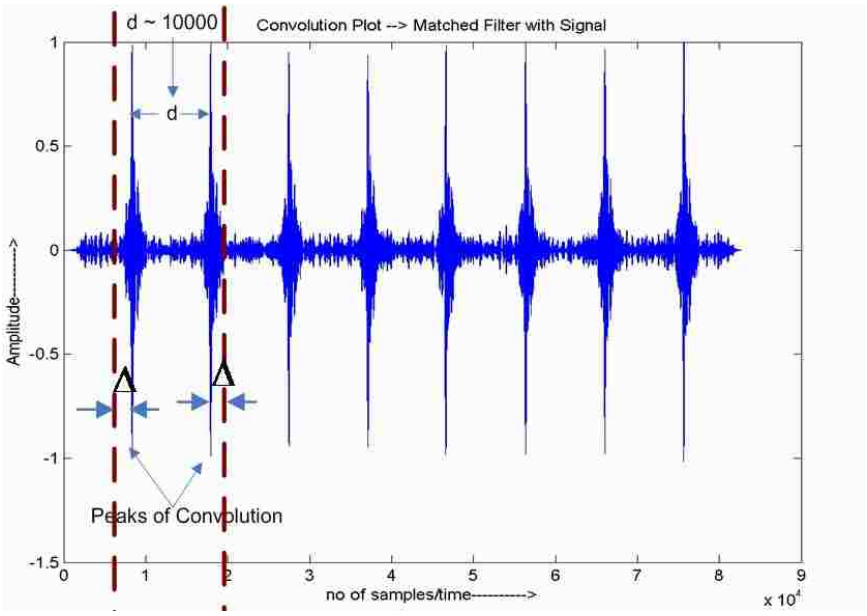


Figure 2.8. Signal Convolved with Matched Filter

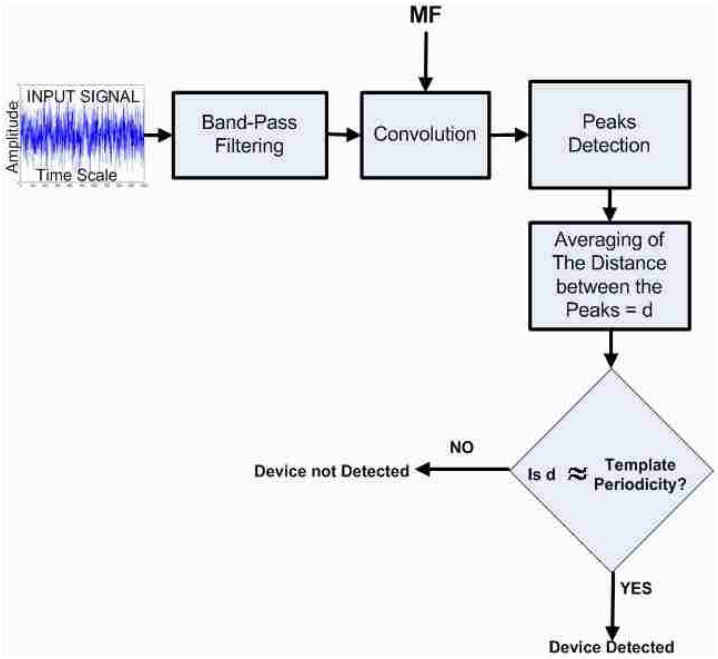


Figure 2.9. Matched Filtering Algorithm

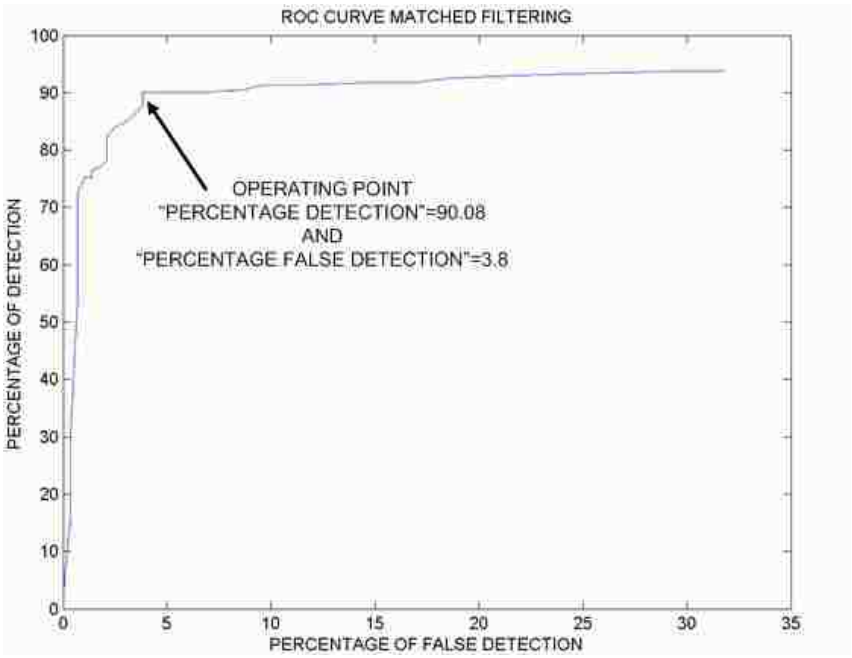


Figure 2.10. ROC Curve for Matched Filter

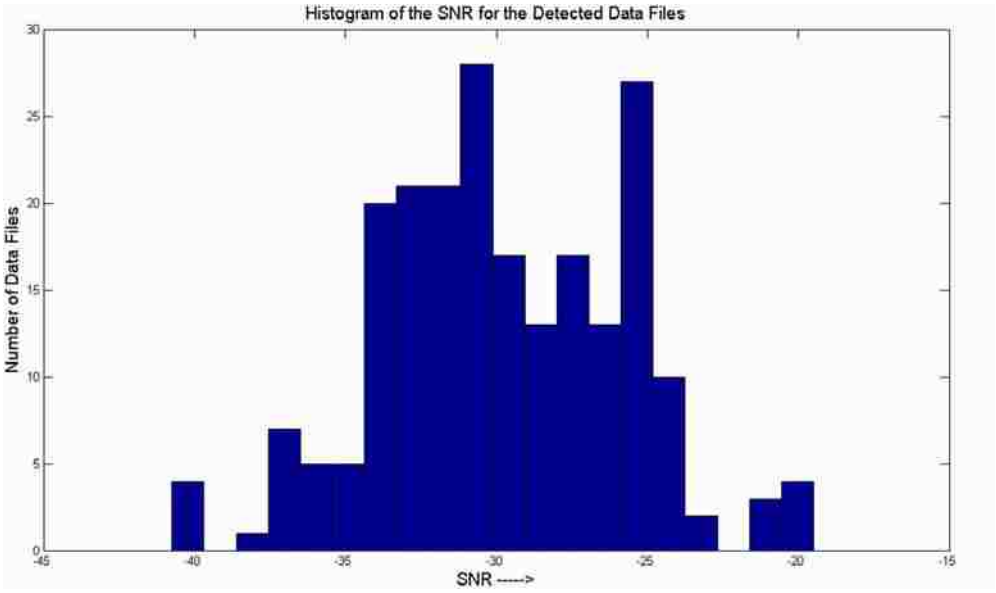


Figure 2.11. SNR Histogram of the Detected Data Files

2.4. CONCLUSION ON MATCHED FILTERING

The ROC curve illustrated that MF was able to classify a large percentage of the data files correctly. And the MF algorithm can be considered robust because, even with large SNR variation, the correct detection percentage is 90.08, and false alarm is 3.8 percent.

3. INDEPENDENT COMPONENT ANALYSIS

In the MF based detection the signal remains undetected if its strength is very low compared to the noise or when there is considerable interference from other RF sources operating in a similar frequency range. For example, consider a signal which is the sum of a low energy device signal and other device signals as shown in Figure 3.1. If one used only MF, then there is a good possibility of device not being detected. It is reasonable to assume that the device signal and the noise are considered independent. If they can be separated, then the device detection can be conducted easily. Hence, the concept of ICA, i.e. Independent Component Analysis is introduced in this section. ICA extracts the statistically independent components present in the signal.

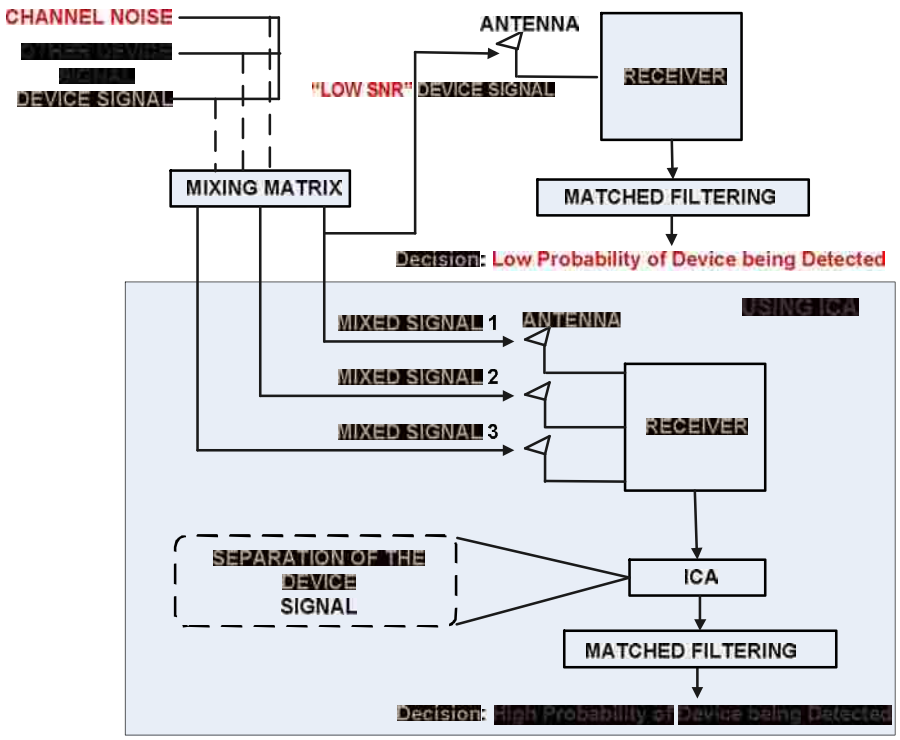


Figure 3.1. Block Diagram of ICA on Matched Filtering

3.1. OBJECTIVE OF ICA

Using Figure 3.2, the idea of ICA will be explained. Consider one of the classical examples used for ICA [2]: a cocktail party, where many people are talking to each other. Each person's speech can be labeled as a separate signal. Signal one denoted by s_1 and signal two denoted by s_2, respectively. Microphones record the mixtures of the signals from both sources.

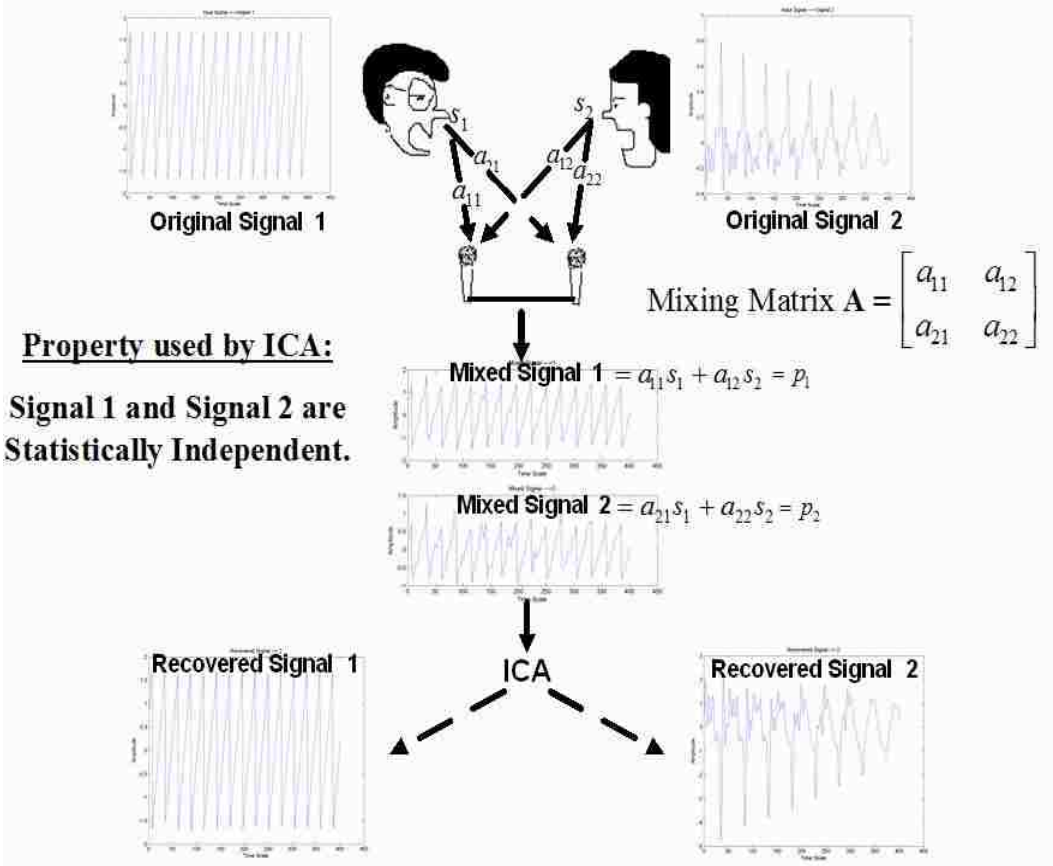


Figure 3.2. Illustration of ICA

Given the recorded signal, one cannot interpret what has been spoken by each person.

The mixing matrix, denoted by **A,** determines the manner in which the signals are mixed.

$$\mathbf{p} = \mathbf{As} \tag{18}$$

where

$$\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \tag{19}$$

is the source vector,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{20}$$

and

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \tag{21}$$

is the mixed signal observed at the sensors. Generally the sources **s** and the mixing matrix **A** are unknown.

$$\mathbf{y} = \mathbf{Wp} \tag{22}$$

$$\text{where } \mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 \end{bmatrix} \tag{23}$$

is the "de-mixing matrix" and

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{24}$$

is the separated signal, also **y** may be permuted version of **s** with each individual element scaled differently. Hence, for instance we could have two cases: Case 1:

$$\mathbf{p} = \mathbf{WA} = \begin{bmatrix} 0 & C_2 \\ C_1 & 0 \end{bmatrix}, \tag{25}$$

and Case 2:

$$\mathbf{p} = \mathbf{WA} = \begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix}, \tag{26}$$

where C_1 and C_2 are arbitrary constants. Since,

$$\mathbf{y} = \mathbf{WAs}, \tag{27}$$

for Case 1:

$$\mathbf{y} = \begin{bmatrix} C_2 s_2 \\ C_1 s_1 \end{bmatrix} \tag{28}$$

and for Case 2:

$$\mathbf{y} = \begin{bmatrix} C_1 s_1 \\ C_2 s_2 \end{bmatrix}. \tag{29}$$

If the de-mixing matrix denoted by **W** can be determined, then the original signals can be recovered. ICA uses the idea that the signals are statistically independent. Consider the two random variables s_1 and s_2 (this example is as explained in [2]). Both variables have a uniform probability distribution function (PDF), varying between 0 and 1 such that,

$$P(s_1) = p(s_2) = 1/n;\ s_i \in \{0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}\}. \tag{30}$$

The 2-D scatter plot of s_1 and s_2 is shown in Figure 3.3. As shown in Figure 3.3 consider a point (s_1, s_2), such as s_1= 0.82, the range of s_2 is from 0 to 1 as is true for all values of s_1 (the blue dots are the points). This implies that s_2 and s_2 are independent of each other, which means

$$P(s_1, s_2) = P(s_1)P(s_2). \tag{31}$$

The mixing matrix **A**, combines the two inputs s_1 and s_2.

If, $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ we have

$$\mathbf{p} = \sum_{i=1}^{2} \mathbf{a_i} s_i \tag{32}$$

where $\mathbf{a_1}$ and $\mathbf{a_2}$ are the column vectors of vector matrix **A**. The output vector is

$$\mathbf{p} = \mathbf{As} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \tag{33}$$

where p_1 and p_2 are the outputs. The scatter plot of p_1 and p_2 is shown in Figure 3.4. Unlike in the scatter plot of the independent random variables s_1 and s_2, when a point in p_1 and p_2 scatter space is considered, such as, p_1 = 2.25, the range of p_2 is not from 0 to 7 (its actual entire range) but is from 4.5 to 5.5. This shows that p_1 and p_2 are not independent of each other.

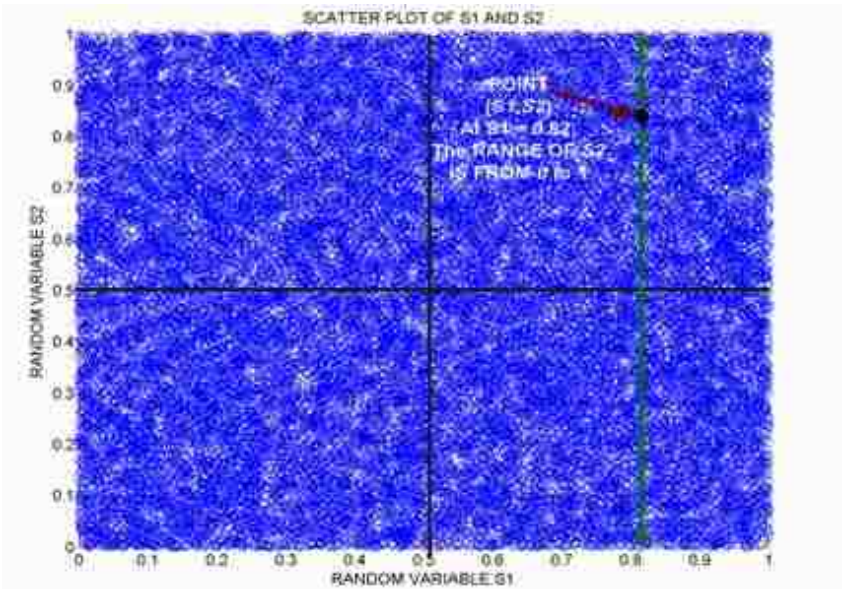


Figure 3.3. 2D Scatter Plot of s_1 and s_2

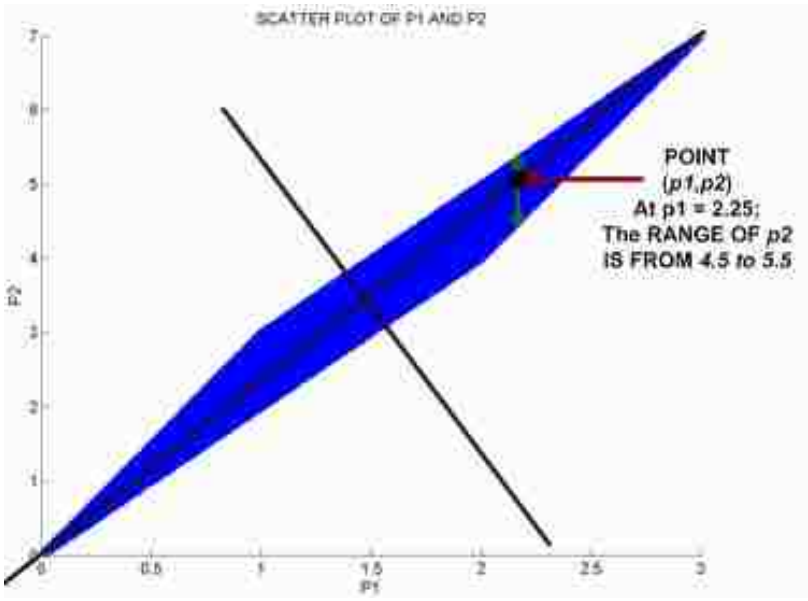


Figure 3.4. Scatter Plots of Mixed Random Variables p_1 and p_2

The Central Limit Theorem states that:

> "If there are N random variables given by X1, X2, X3…XN-1 and XN, each having a probability distribution function fi(xi), standard deviation σ, and mean of μ. Then, the PDF of the sum of random variables given by S = X1 + X2 + X3…+ XN will tend to a normal distribution which will have mean Nμ and standard deviation σ√N [2]."

Most of the following equations for ICA are from reference [2]. Recalling one of the equations mentioned above,

$$y_i = \mathbf{z}^T \mathbf{s}. \tag{34}$$

intuitively, one can say when y_i is simply s_1, then the PDF of y_i will be less Gaussian compared to the case when y is a linear combination of s_1 and s_2 [2]. Given the mixed vector **p,** the unknown that needs to be determined is the $\mathbf{w}_i$ vector [2]. The solution provided by ICA, generates outputs close to the original signals by estimating the $\mathbf{w}_i$ vectors. The $\mathbf{w}_i$ vectors are obtained in such a manner that the output produced is least Gaussian. Thus, the question now to be answered is how to measure the Gaussianity of the output signal.

The Gaussian signal is associated with maximum randomness which in turn means randomness means maximum entropy [2]. The term called "negentropy" refers to a function that is generally negative of entropy. So, the Gaussian signal has minimum negentropy. If the output signal obtained is least Gaussian, then it will have minimum entropy and maximum negentropy. The entropy of a random variable y_i, is given by

$$H(y_i) = -\int f(y_i) \log[f(y_i)] dy_i \tag{35}$$

where $f(y_i)$ is the PDF of y_i. The negentropy is the negative of the entropy it can be defined as:

$$J(y_i) = H(y_{gauss}) - H(y_i) \tag{36}$$

It is, however, difficult to obtain the PDF of y_i based on the data for y_i. Therefore, an approximation is used for computing the negentropy of the extracted random variable. This approximation is

$$J(y_i) \approx [E\{G(y_i)\} - E\{G(v)\}]^2 \tag{37}$$

where $G(\cdot)$ is practically any non-quadratic function [3]. Now, define the cost function as [6]:

$$J = \rho[\{E(G(y_i))\} - \{E(G(v))\}]^2 \tag{38}$$

in which

$$y_i = \mathbf{w}_i^{\mathrm{T}}\mathbf{p} \tag{39}$$

and v is a random distribution with mean 0 and variance 1, ρ is a scaling parameter. Suitable choices available for $G(\cdot)$ are as follows:

$$G_1(t) = \frac{\log(\cosh(a_1 t))}{a_1} \tag{40}$$

$$G_2(t) = -\exp(-t^2/2) \tag{41}$$

In the cost function, $\mathbf{p}$ is known and $\mathbf{w}_i$ is to be determined. Hence, J is a function of $\mathbf{w}_i$ and can be written as $J(\mathbf{w}_i)$. So, when the cost function is maximized, the solution of $\mathbf{w}_i$ is obtained. Once $\mathbf{w}_i$ is obtained, one can easily determine $y_i = \mathbf{w}_i^T\mathbf{p}$. The cost function $J(\mathbf{w}_i)$ is maximized when

$$\frac{\partial J(\mathbf{w}_i)}{\partial \mathbf{w}_i} = J'(\mathbf{w}_i) = 0. \tag{42}$$

When the solution is constrained by $\left\|\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}}\right\| = 1$, the Lagrangian cost function is given as

$$J(\mathbf{w}_i) = E\{[G(y_i)]^2\} + \beta(1 - \mathbf{w}_i\mathbf{w}_i^T) \tag{43}$$

and according to the Kuhn-Tucker conditions [2], the solution is obtained at a point where

$$J_m'(\mathbf{w}_i) = E\{\mathbf{p}g(\mathbf{w}_i^T\mathbf{p})\} - \beta\mathbf{w}_i = 0 \tag{44}$$

and β is the of Lagrangian constant.

In the first part of the equation, the function $g(\cdot)$ is a derivative of $G(\cdot)$. $g(\cdot)$ for the functions of G defined in equations (11) and (12), are:

$$g_1(t) = \tanh(a_1 t) \tag{45}$$

$$g_2(t) = t \times \exp(-t^2/2) \tag{46}$$

Next, $J_m(\mathbf{w}_i)$ is a modified form of the original cost function $J(\mathbf{w}_i)$. By Newton's method of linear optimization [2], the vector $\mathbf{w}_i$ determined at iteration $k+1$ is given by

$$\mathbf{w}_i(k+1) = \mathbf{w}_i(k) - \frac{\mathbf{J}_{\mathbf{m}}^{'}[\mathbf{w}_i(k)]}{J_m^{''}[\mathbf{w}_i(k)]} \tag{47}$$

where

$$J_m^{''}([\mathbf{w}_i(k)]) = E\{\mathbf{p}\mathbf{p}^T g'([\mathbf{w}_i(k)]^T \mathbf{p})\} - \beta\mathbf{I} = E\{g'([\mathbf{w}_i(k)^T]\mathbf{p})\} - \beta, \tag{48}$$

since

$$\mathrm{E}\{\mathbf{p}\mathbf{p}^T\} = \mathbf{I}; \tag{49}$$

The reason for equating $\mathrm{E}\{\mathbf{p}\mathbf{p}^T\}$ to Identity Matrix will be explained in Section 3.2 where the concept of whitening will be introduced [2]. Simplifying the above equation

$$\mathbf{w}_i(k+1) = \mathbf{w}_i(k) - \frac{\{E\{\mathbf{p}g\{[\mathbf{w}_i(k)^T]\mathbf{p}\}\} - \beta[\mathbf{w}_i(k)]\}}{\{E\{\mathbf{p}\mathbf{p}^T g'\{[\mathbf{w}_i(k)^T]\mathbf{p}\}\} - \beta\mathbf{I}\}}$$

$$\mathbf{w}_\mathbf{i}(k+1) = \frac{\mathbf{w}_i(k) \times \{E\{\mathbf{p}\mathbf{p}^T g'\{[\mathbf{w}_i(k)^T]\mathbf{p}\}\} - \beta\mathbf{I}\} - \{E\{\mathbf{p}g\{[\mathbf{w}_i(k)^T]\mathbf{p}\}\} - \beta\mathbf{w}_i(k)\}}{\{E\{\mathbf{p}\mathbf{p}^T g'\{[\mathbf{w}_i(k)^T]\mathbf{p}\}\} - \beta\mathbf{I}\}}$$

$$\mathbf{w}_\mathbf{i}(k+1) = \frac{\mathbf{w}_\mathbf{i} E\{g'\{[\mathbf{w}_i(k)^T]\mathbf{p}\}\} - E\{\mathbf{p}g\{[\mathbf{w}_i(k)^T]\mathbf{p}\}\}}{\{E\{\mathbf{p}\mathbf{p}^T g'\{[\mathbf{w}_i(k)^T]\mathbf{p}\}\} - \beta\mathbf{I}\}};$$

Approximating for fastICA[6], multiplying by $\{\beta\mathbf{I} - E\{\mathbf{p}\mathbf{p}^T g'\{[\mathbf{w}_i(k)^T]\mathbf{p}\}\}\}$ as normalization step

$$\mathbf{w}_\mathbf{i}(k+1) = E\{\mathbf{p}g\{[\mathbf{w}_i(k)^T]\mathbf{p}\}\} - \mathbf{w}_i(k)E\{g'([\mathbf{w}_i(k)^T]\mathbf{p})\} \tag{50}$$

The main steps are summarized in Figure 3.5.

ICA finds the de-mixing vector $\mathbf{w}_i$ such that the cost function is maximized. The cost function here is, as explained, the measure of independence of the extracted signal component and thus an independent component is extracted from the mixed signals.

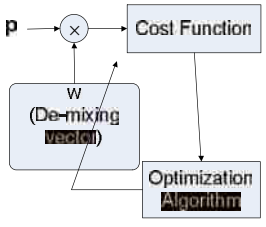


Figure 3.5. Basic Module for ICA Algorithm

3.2. STEPS OF INDEPENDENT COMPONENT ANALYSIS

The main steps of ICA are Pre-Processing and Cost Function Optimization.

1) Pre-processing Steps

The pre-processing required for the ICA is called "whitening". Whitening the data in ICA involves two steps [2].

a) Removal of the Mean

The input signal, i.e. the mixed signal to ICA is **p**. Its mean is removed by replacing **p** with $\mathbf{p} = \mathbf{p} - E(\mathbf{p})$.

b) Whitening

The input data is transformed in such a manner that the data are orthogonal to each other [2]. By making them orthogonal to each other, the input data to the ICA, the whitened mixed signals now denoted by **p** are un-correlated. As explained in [2],

$$E(\mathbf{p}\mathbf{p}^T) = \mathbf{E}\mathbf{D}\mathbf{E}^T \tag{51}$$

where ***E*** is the Eigen matrix and ***D*** is the diagonal matrix, in which the diagonal elements represent the Eigen values of $E(\mathbf{p}\mathbf{p}^T)$. So, the new vector is $\tilde{\mathbf{p}} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{p}$. Now, it can be shown that the whitened data $\tilde{\mathbf{p}}$ has an identity covariance matrix. $\Rightarrow E(\tilde{\mathbf{p}}\tilde{\mathbf{p}}^T) = \mathbf{I}$ is used in equation 4 in the previous section. For convenience sake, from now onwards $\tilde{\mathbf{p}}$ shall be denoted by **p**.

2) Cost Function Optimization

The cost function concept is as defined in [2]. The cost function for the ICA defines the evolution of the weight vector, which is used to extract the independent component in the signal.

A summary of steps to extract an independent component is as given in [2]. Start with any random weight vector $\mathbf{w}_i$.

i) Substitute $\mathbf{J}'[\mathbf{w}_i(k)]$ and $J''[\mathbf{w}_i(k)]$ in the updating equation

$$\mathbf{w}_i(k+1) = \mathbf{w}_i(k) - \mu \frac{\mathbf{J}'[\mathbf{w}_i(k)]}{J''[\mathbf{w}_i(k)]}. \tag{52}$$

Choose a small stepping parameter μ, since this can affect the performance of the convergence algorithm.

ii) The update equations are

$$\mathbf{w}_i(k+1) = E\{\mathbf{p}g[\mathbf{w}_i(k)^T\mathbf{p}]\} - E\{g'[\mathbf{w}_i(k)^T\mathbf{p}]\}\mathbf{w}_i(k) \quad (53)$$

after simplification [2]. We, then normalize $\mathbf{w}_i$ as

$$\mathbf{w}_i(k+1) \Leftarrow \mathbf{w}_i(k+1)/\left\|\mathbf{w}_i(k+1)\right\| \quad (54)$$

iii) Check for convergence, i.e., at (k+1)'th iteration if

$$\left\|[\mathbf{w}_i(k+1) - \mathbf{w}_i(k)]\right\| < \xi; \quad (55)$$

where ξ is a small number, $\mathbf{w}_i$ is said to have converged. If it has not converged, go back to step (ii).

The above steps are for obtaining the i'th component [2]. So the algorithm needs to be run as many times as the number of independent components. To prevent the convergence of the weight vectors to the same maxima $\mathbf{w}_1{}^T\mathbf{p}$, $\mathbf{w}_2{}^T\mathbf{p}$ ..,$\mathbf{w}_n{}^T\mathbf{p}$ must be de-correlated for the 1st, 2nd …nth iteration respectively [2]. One of the ways of doing this [2], is to use a Grahm-Schmidt procedure. Thus, we calculate the weights (while computing the (p+1)th component),

$$\mathbf{w}_{i+1}(k+1) \Leftarrow \mathbf{w}_{i+1}(k) - \sum_{j=1}^{i} \mathbf{w}_{i+1}^{T}(k)\mathbf{w}_j\mathbf{w}_j \quad (56)$$

$$\mathbf{w}_{p+1}(k+1) \Leftarrow \mathbf{w}_{p+1}(k+1)/\sqrt{\mathbf{w}_{p+1}^{T}(k+1)\mathbf{w}_{p+1}(k+1)} \quad (57)$$

ICA Simulation Results: This section presents simulation results of the ICA algorithm. The three input signals are shown in Figure 3.6: a speech signal, a Gaussian random signal, and a saw-tooth signal. They are mixed by random mixing matrix **A**. The mixtures of the signals are shown in Figure 3.7. Figure 3.8 illustrates the performance of the ICA algorithm in the signal separation. It shows that ICA was able to extract all three original signals. From the mixed signal, the original signal's actual signature cannot be estimated. But by the using the Central limit theorem, the de-mixing matrix can be easily estimated which enables the recovery of the original signals.

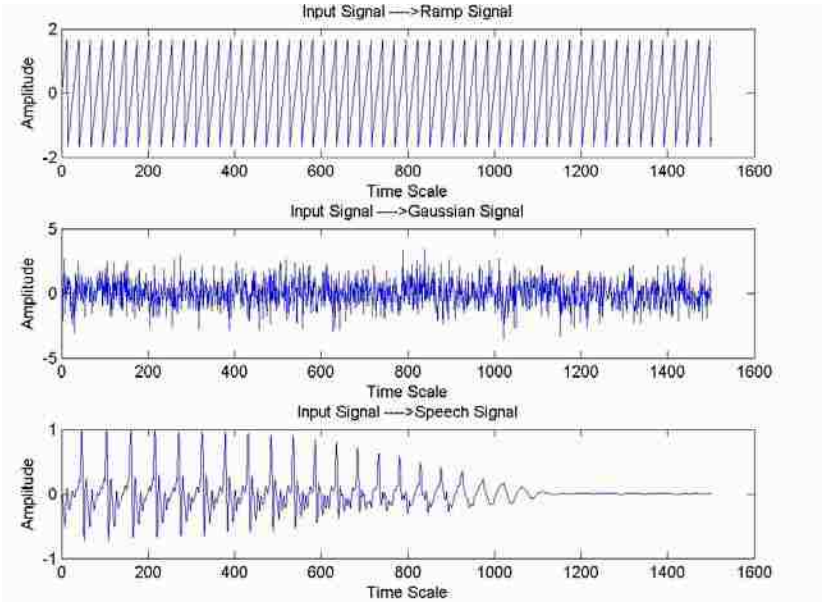


Figure 3.6. Original Input Signals

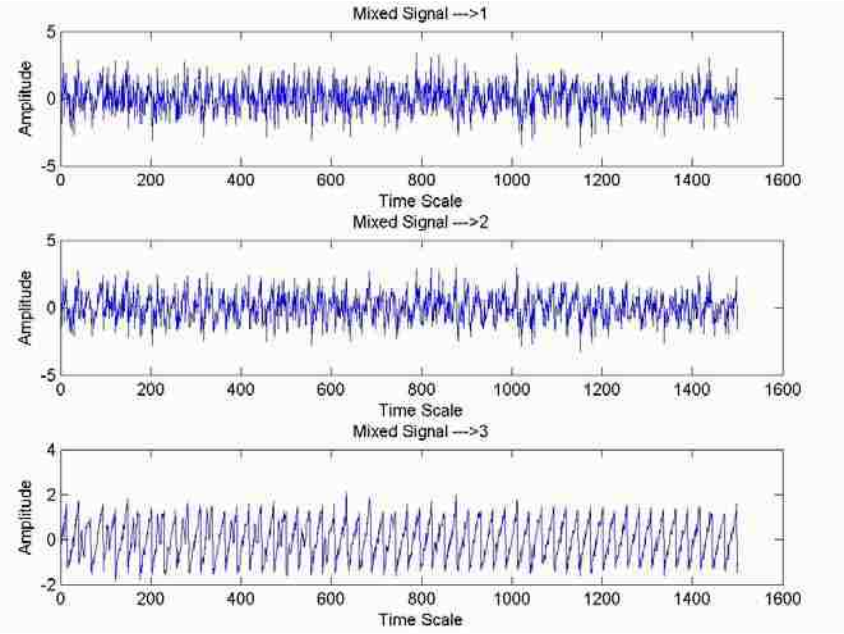


Figure 3.7. Mixed Signals

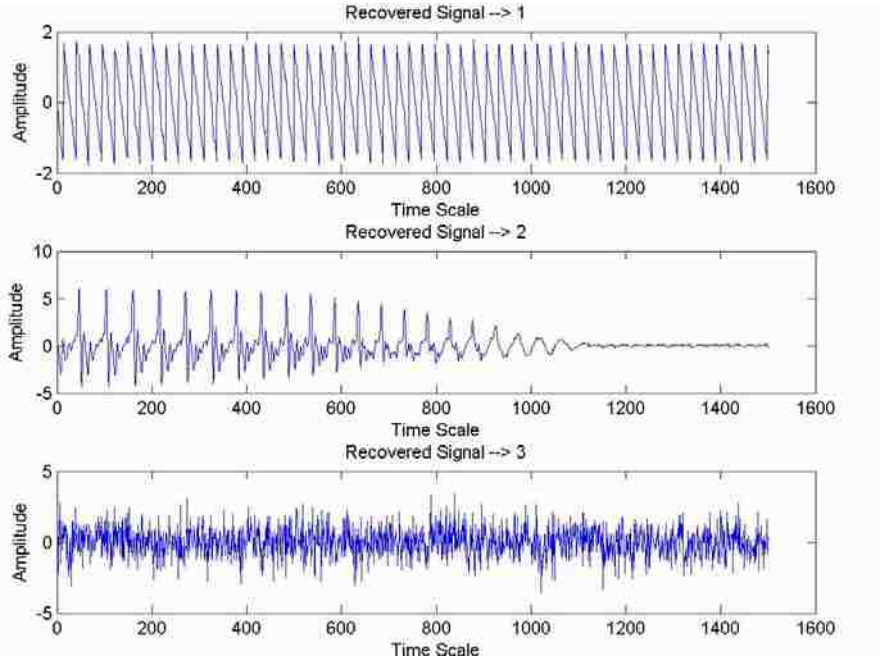


Figure 3.8. Recovered Signals from ICA

1) Advantages of ICA

The advantages of ICA are:

a) It is computationally efficient.

b) It's, performance is fairly reliable when the number of receivers is greater than the number of sources.

2) Disadvantages of ICA

The disadvantages of ICA are:

a) The signal component being extracted during a particular iteration is random.

b) Like many of the optimization algorithms, global maxima cannot be guaranteed.

c) The number of receivers needs to be, greater than or equal to the number of source signals present. Otherwise, a good solution is not obtained.

d) Computational complexity. The computational complexity is illustrated in

Figure 3.9. Suppose there are N original signals and the ICA algorithm would give N outputs. The idea is to detect M devices, where $M \leq N$. Therefore, at each ICA output, M MF's are required to detect which signal is at that particular ICA output. So, the total number of computations required to detect the M devices is M×N matched filters. In the next section we show how the number of matched filters may be reduced by a factor of M by using a slightly different algorithm ICA-R.

3.3. APPLICATION OF ICA-R FOR DEVICE DETECTION

Some of the failure cases of MF are due to the strong presence of noise in the signal. By noise we mean not only randomness added to the signal in the form of Gaussian noise, but also in the form of strong interfering signals in the same frequency range as the desired signal.

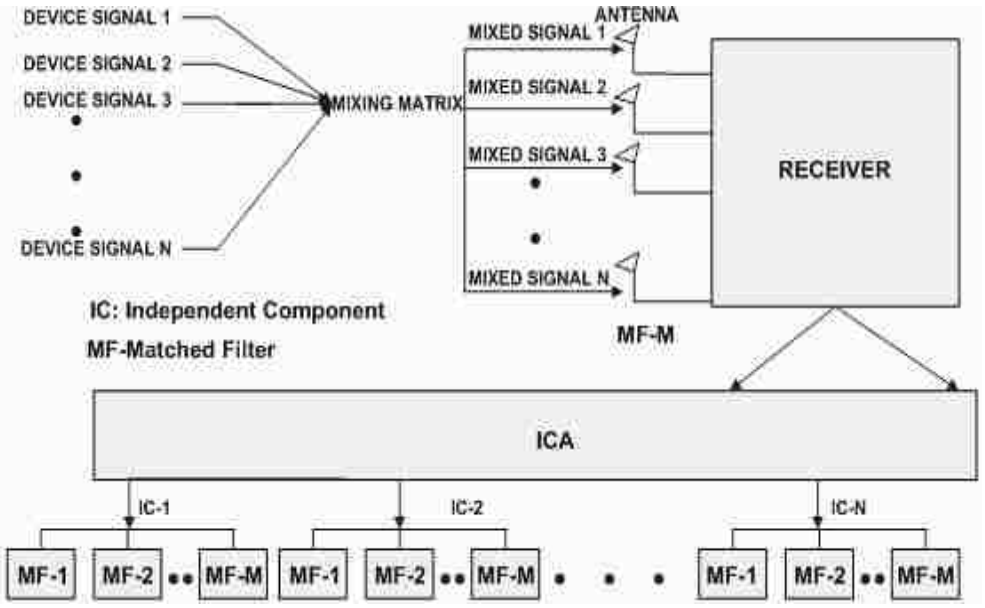


Figure 3.9. Computations of ICA for Matched Filtering

In the device detection problem, the device signal to be detected is generally known. For the extraction of the known signals, the concept of ICA with reference (ICA-R), introduced by Rajaphakse [4], can be used for device detection to reduce the computational complexity. As explained when using the ICA-MF algorithm, the number of MF's involved are M×N; using the ICA-R MF algorithm, the number of MF's are just M as shown in Figure 3.10.

3.3.1. ICA with Reference. ICA-R [4] enables extraction of a particular signal component from the mixed signal using the basic ideas of ICA, and by taking advantage of a priori knowledge of the particular independent component we desire to detect. The ICA-R algorithm is explained in [4] and [5]. Before ICA-R is studied the basic concept of penalty functions is introduced.

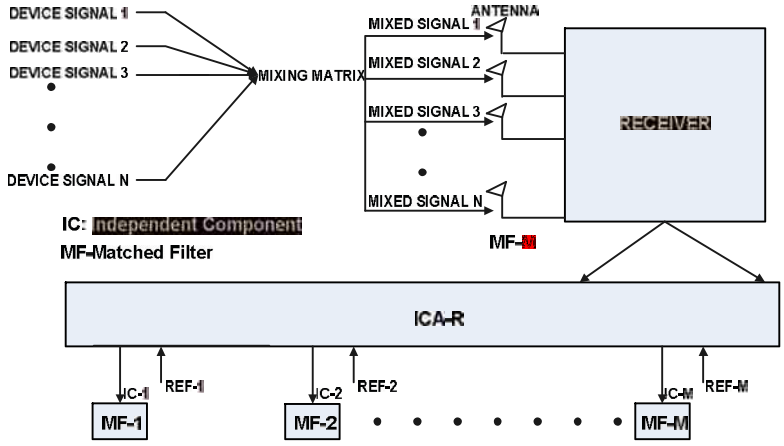


Figure 3.10. Computations of ICA-R for Matched Filtering

The concept of penalty functions: Consider the following problem:

Minimize $f(\mathbf{x})$ subject to $g(\mathbf{x}) = 0$.

The Lagrangian function for this simple form of equality constraints is given by

$$L(\mathbf{x}) = f(\mathbf{x}) + \mu g^2(\mathbf{x}) \tag{58}$$

in which μ is > 0 and is called the Lagrangian constant.

So, when $g(\mathbf{x}) > 0$ or $g(\mathbf{x}) < 0$ a penalty is added to the cost function. This type of problem has a well defined solution.

However, if the problem has inequality constraint [6],

Minimize $f(\mathbf{x})$

subject to $g(\mathbf{x}) < 0$.

the same Lagrangian function defined as i.e.

$$L(\mathbf{x}) = f(\mathbf{x}) + \mu g^2(\mathbf{x}) \tag{59}$$

is inappropriate, because the cost function is punished even if $g(\mathbf{x}) < 0$, which should not happen. Instead we define a new penalty function, as $\mu[\max\{0, g(\mathbf{x})\}]^2$, by which the cost function is reformulated as

$$L(\mathbf{x}) = f(\mathbf{x}) + \mu[\max\{0, g(\mathbf{x})\}]^2 \tag{60}$$

So, when $g(\mathbf{x}) < 0$, then $\mu[\max\{0, g(\mathbf{x})\}]^2 = 0$, and, hence, no punishment is induced on the cost function but when $g(\mathbf{x}) < 0$ a penalty is introduced to the cost function.

ICA-R uses a similar form of constrained optimization and is set up as:

Maximize $: J(\mathbf{w}) = \rho[\{E(G(y))\} - \{E(G(v))\}]^2$ where ρ is a constant.

Subject to $: h(\mathbf{w}) = E\{(y - r)^2 - \varsigma\}$.

Here, y is the extracted signal component, r is the reference signal and ς is a measure of the desired closeness. From the expression of $h(\mathbf{w})$ we see interpreted that the closer the extracted signal component is to the reference, the less the punishment is in the cost function. The new cost function, which is now a case of constrained optimization, is defined by the Lagrangian function as:

$$L(\mathbf{w}, \mu) = J(\mathbf{w}) - \frac{1}{2\gamma}[\max^2\{\mu + \gamma h(\mathbf{w}), 0\} - \mu^2] \tag{61}$$

where γ is the scaling penalty parameter.

The steps for obtaining the IC (independent component) are detailed as follows:

1) Whiten the input data **p** as discussed in Section 3.3.

2) Choose penalty parameter γ and an initial Lagrangian multiplier μ.

3) Start with a random weight vector and make sure that it is normalized.

4) So, at iteration (k+1), compute the sign of $E\{G(y)\} - E\{G(v)\}$ defined as $\overline{\rho}$.

5) Calculate the new Lagrangian multiplier given by

$$\mu_i(k+1) = \max\{0, \mu_i(k) + \gamma h[\mathbf{w}_i(k)]\} \tag{62}$$

and at update the weight vector

$$\mathbf{w}_i(k+1) = \mathbf{w}_i(k) - \eta \frac{\mathbf{L}'_{\mathbf{w}_i}[\mathbf{w}_i(k)]}{L''_{\mathbf{w}_i}[\mathbf{w}_i(k)]} \tag{63}$$

where,

$$\mathbf{L}'_{\mathbf{w}_i}[\mathbf{w}_i(k)] = \overline{\rho} E\{\mathbf{p}G'(y)\} - 0.5\mu_i(k+1)E\{\mathbf{p}g'[\mathbf{w}_i(k)]\} \text{ and} \tag{64}$$

$$\mathrm{L}''_{\mathbf{w}_i}[\mathbf{w}_i(k)] = \overline{\rho} E\{G''(y)\} - 0.5\mu_i(k+1)E\{g''[\mathbf{w}_i(k)]\} \tag{65}$$

6) Normalize the updated weight vector by computing

$$\mathbf{w}_i(k+1) \Leftarrow \frac{\mathbf{w}_i(k+1)}{\|\mathbf{w}_i(k+1)\|} \tag{66}$$

7) Test whether

$$\|\mathbf{w}_i(k+1) - \mathbf{w}_i(k)\| < \varepsilon \tag{67}$$

where $\varepsilon \cong 0.000001$. If this condition is not satisfied, go back to step 4.

3.3.2. Results from ICA-R. Two sets of results will be discussed. One using test signals and another with measured device signal.

3.3.2.1. Simulated results ICA-R. In this section, the performance of ICA-R will be illustrated.

Example 1: Consider five source signals as shown in Figure 3.11. They are mixed by a random mixing matrix **A**. The initial parameter set is:

$\varepsilon = 0.000001; \mathrm{N}_{MAX} = 300; \eta = 0.1; \mu = 0.5; \gamma = 0.7; \varsigma = 0.001,$ where $\mathrm{N_{MAX}}$ is the maximum number of iterations in ICA-R. The mixed signals are shown in Figure 3.12. In the convergence graphs shown in Figure 3.13 and Figure 3.14, the error has been minimized and the cost function has been maximized. As shown in Figure 3.15, given a

suitable reference signal, ICA-R was able to extract the similar independent component from the mixture. The value of ς is preferably small for optimal convergence [5]. As shown in Figure 3.15, the reference signal is in time sync with the extracted signal and the extracted signal component is close to the reference signal. However, in a practical scenario, one frequently cannot have a reference in time sync with the similar component in the mixed signal. This time shift can be dealt with by a small modification in the error function computation. The new error function is

$$g(\mathbf{w}) = E[y(t)y(t+\varsigma)] - E[r(t)r(t+\varsigma)] \tag{68}$$

which is the difference between the auto correlations of the signals. By computing the auto-correlations of the extracted component, the reference time shift can be negated and an error function similar to the one implemented in [5] can be used. As illustrated in Figure 3.16, for a shifted reference signal the extracted signal component is as desired and close to the reference signal.

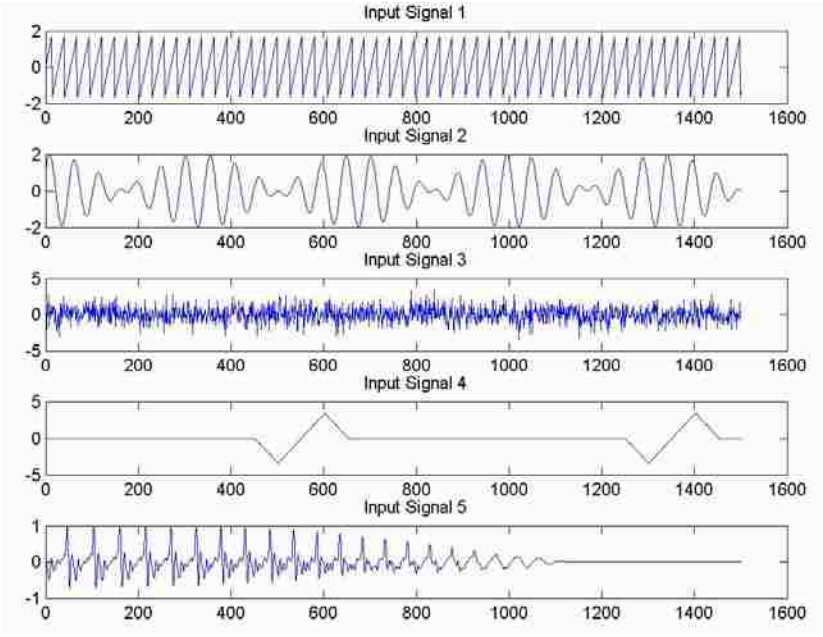


Figure 3.11. Input Signals for ICA-R

Unfortunately, this type of error function does not perform consistently with large shifts. A possible reason for this is that the information on the large, time shifted reference signal is not close to the signal that needs to be extracted.

Now, consider a case where the reference signal is not present. The original signals and the mixed signals are still the same as shown in Figures 3.11 and 3.12, respectively. If the reference signal is given as the sine signal as shown in Figure 3.17, the extracted signal is the speech signal. Hence, it is shown that the signal component extracted is mainly dominated by the part of the cost function that helps extract the independent component. Also, the reference signal error only helps in providing constraints to extract the required independent component.

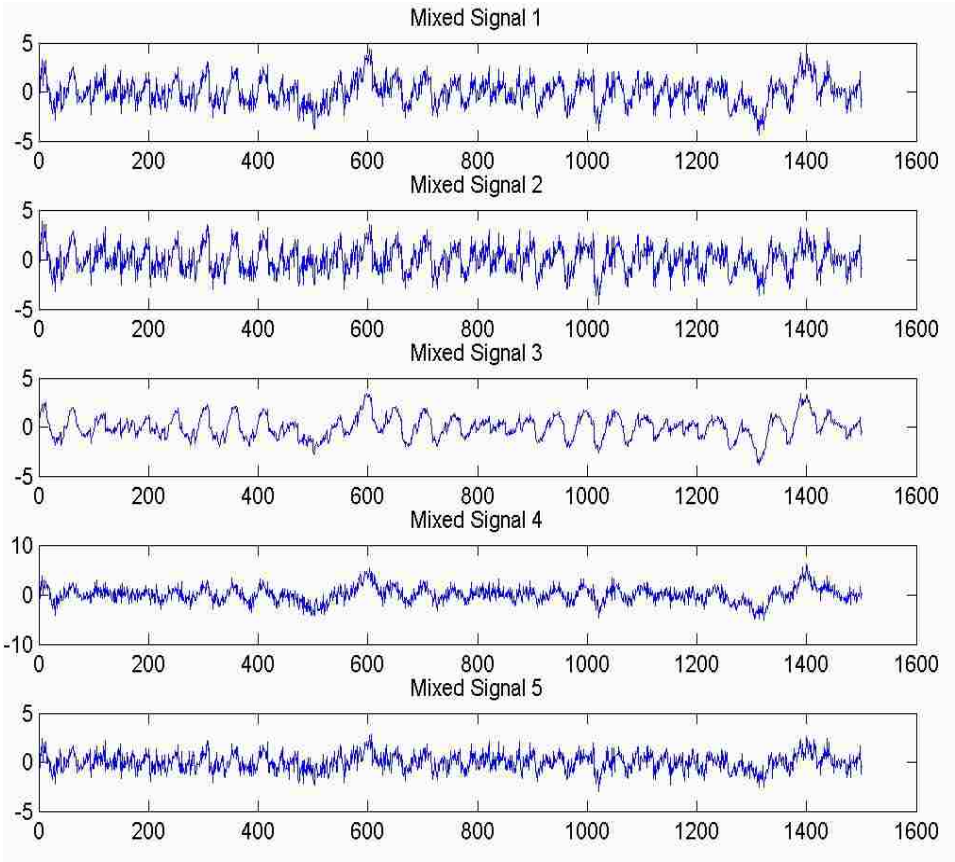


Figure 3.12. Mixed Signals for ICA-R

3.3.2.2. Measured device signal results for ICA-R. Some of the ICA-R results will be illustrated in this section.

Example 1: Consider the device signal shown in Figure 3.18. As discussed in section 2, in the preprocessing steps, the signal is bandpass filtered in a frequency range of 33MHz to 66MHz.

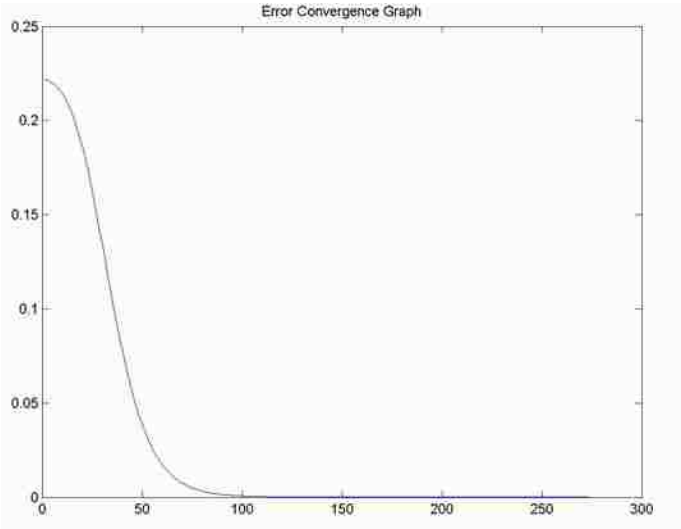


Figure 3.13. Error Convergence

First the MF algorithm where a 45MHz sinusoidal jamming signal has been added to the devices signal is tested. The sinusoid signal is shown in the Figure 3.19. The mixture of the device signal and the jamming signal is shown in Figure 3.20. The detection of the signal is not possible with only the matched filtering step. In a simulation result of a two channel receiver with ideal mixing as shown in Section 3.3.2.1, the ICA-R is able to extract the signal close to the reference.

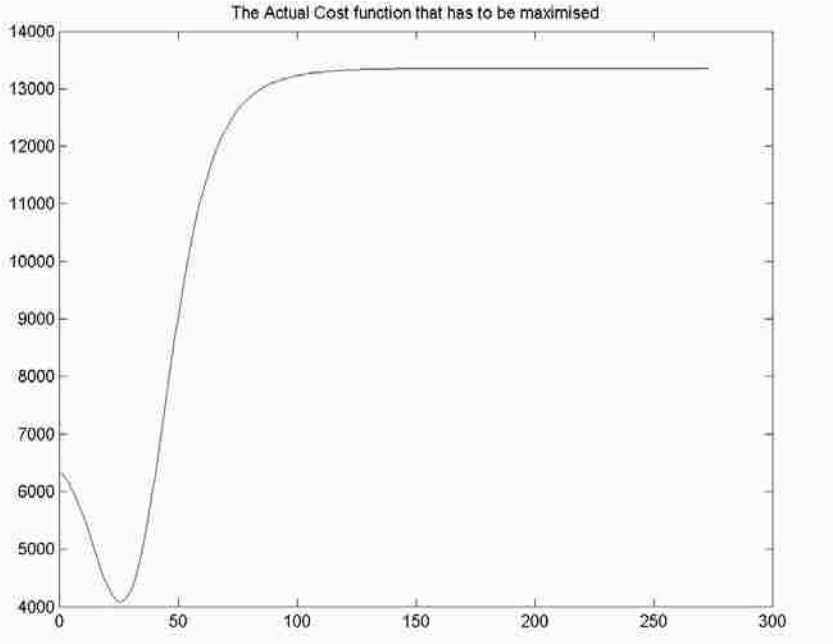


Figure 3.14. Cost Function Convergence

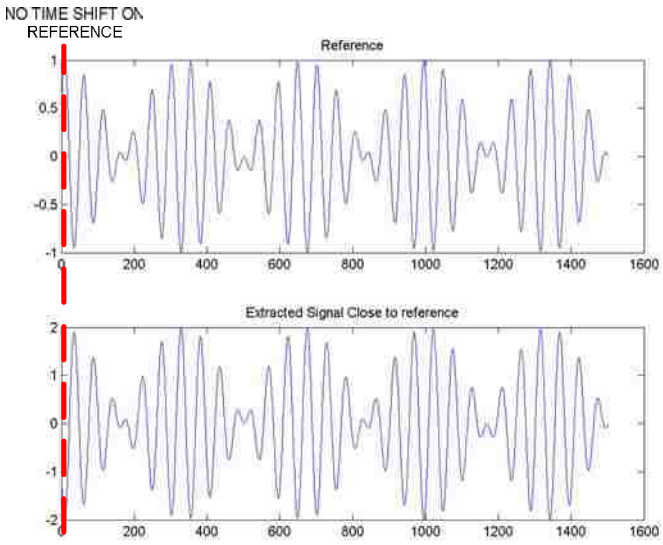


Figure 3.15. Reference and Extracted Signal

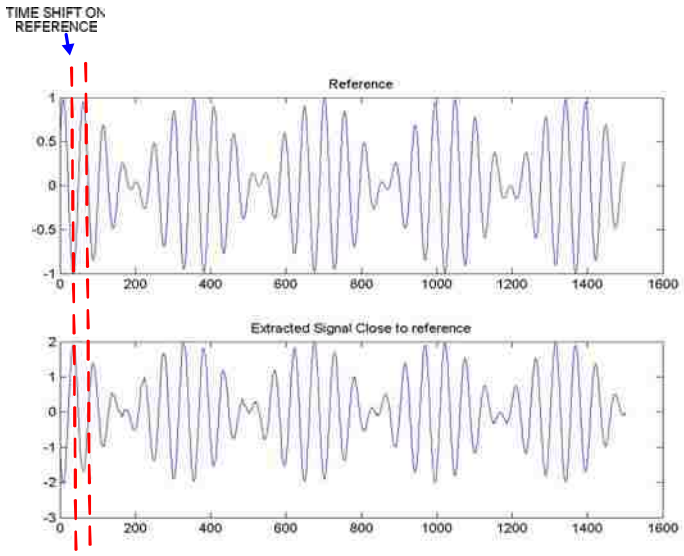


Figure 3.16. Shifted References and Extracted Signal

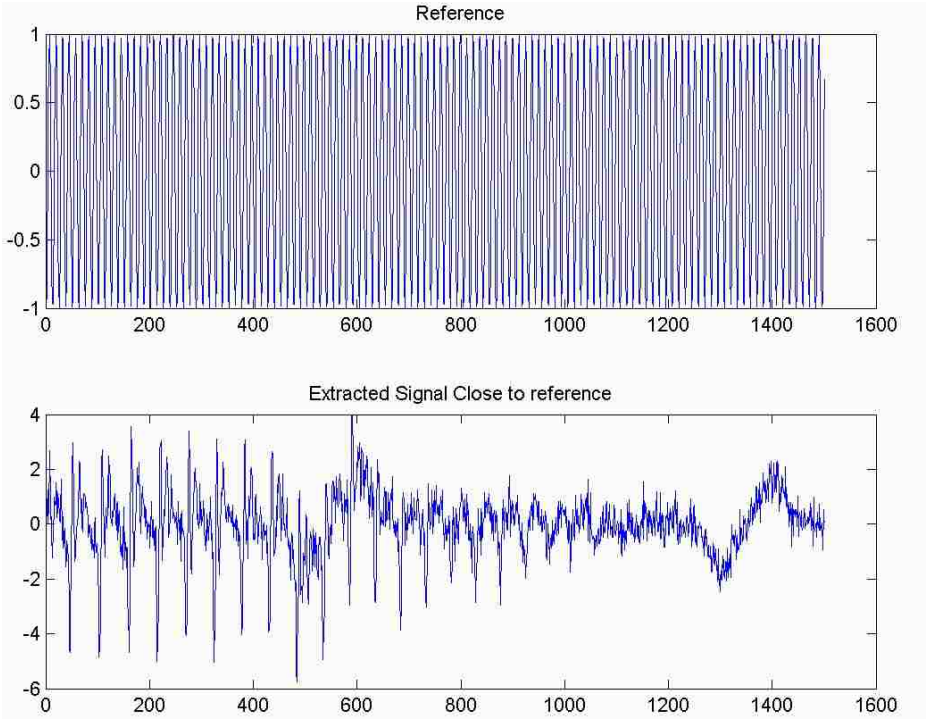


Figure 3.17. ICA-R Testing when Reference Not Present

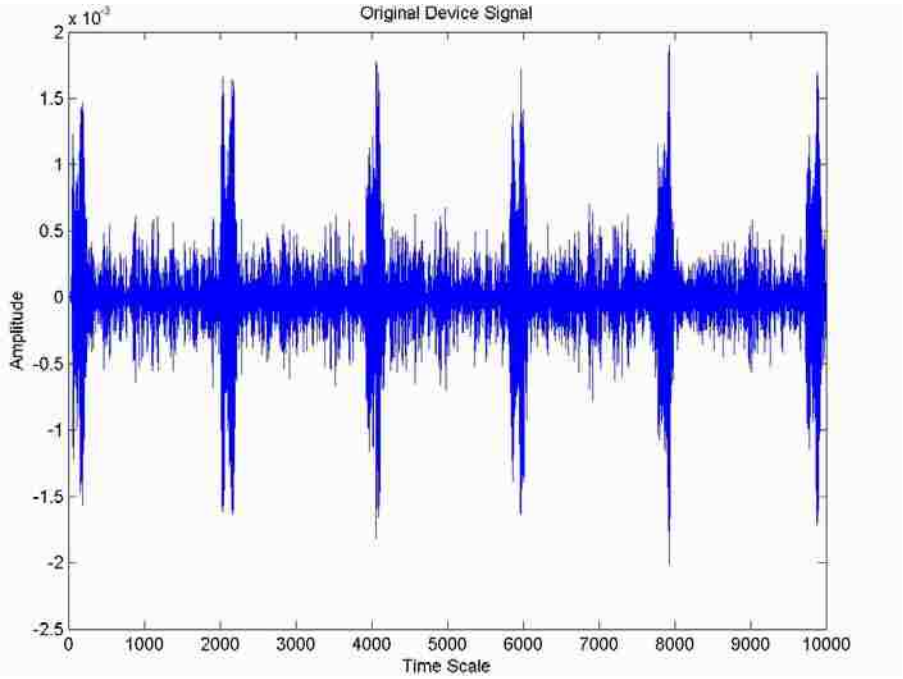


Figure 3.18. Device Signal

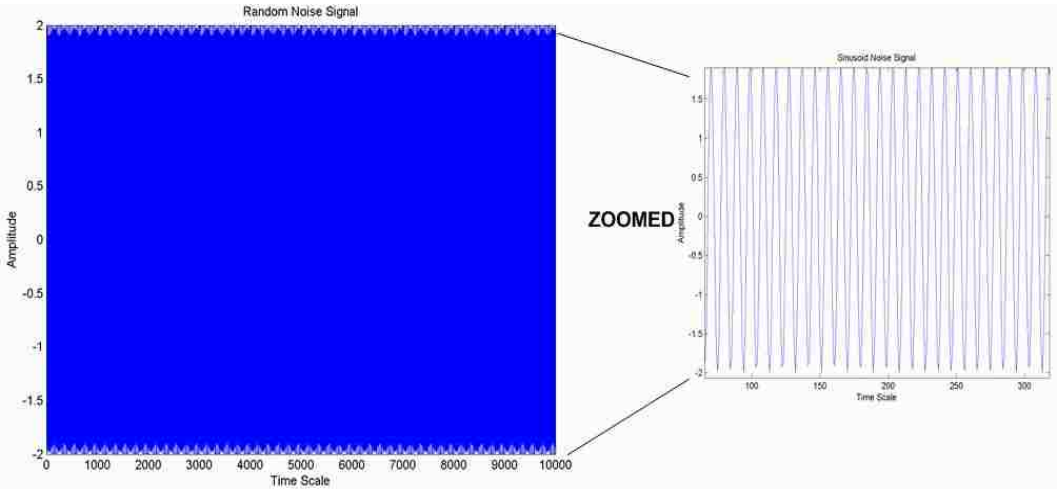


Figure 3.19. Noise Added is Sinusoid at *45 MHz*

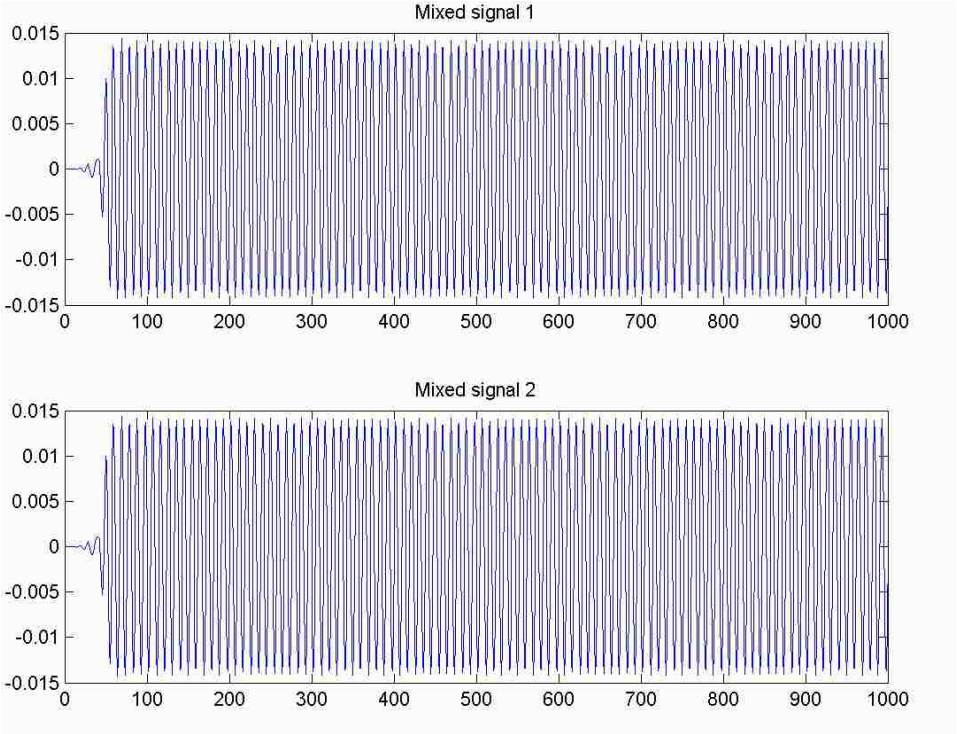


Figure 3.20. Mixed Device Signal and Sinusoid Noise

As shown in Figure 3.21, in the case of mixed signal one, the MF done is unable to extract the periodicity of the device signal. Hence, in the mixed signal the presence of the device signal is not detected. ICA-R however successfully extracts of the device signal, as shown in Figure 3.22, which could be further undergo MF for detection.

Example 2: In this example, the performance of the ICA-R is tested when high energy is added to the signal. The noise shown in Figure 3.23. The signal plus noise is shown in Figure 3.24. The maximum signal amplitude after filtering is 2.5×10^{-3}. The MF result of the mixed signal is shown in Figure 3.25. As seen in the Figure 3.25, there is no periodicity of the template occurrence in the output from the MF; hence, the device signal remains un-detectable. Alternately, the mixed signals are processed by ICA-R to extract the reference signal, as shown in Figure 3.26.

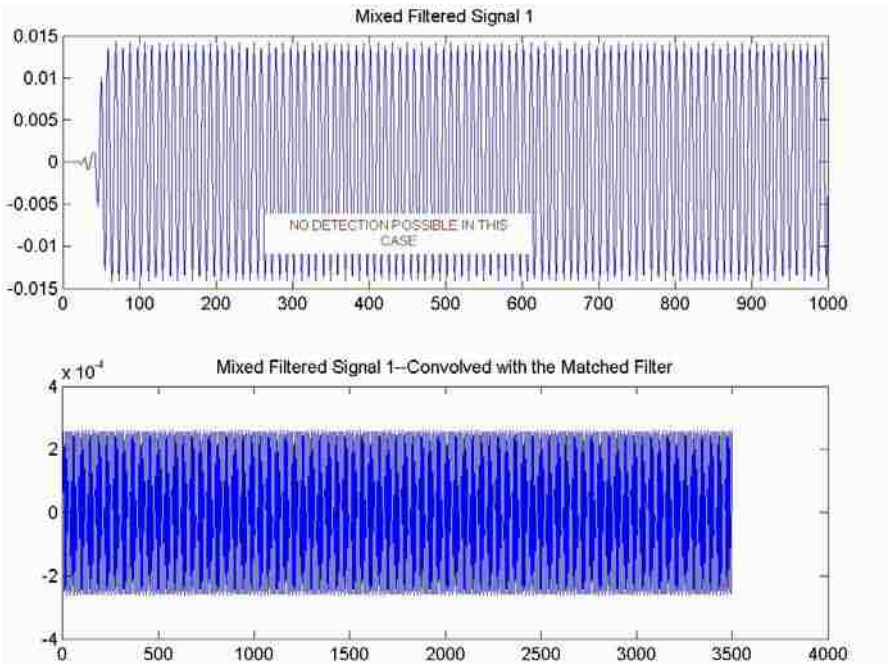


Figure 3.21. Mixed Signal Convolved with Matched Filter for Detection

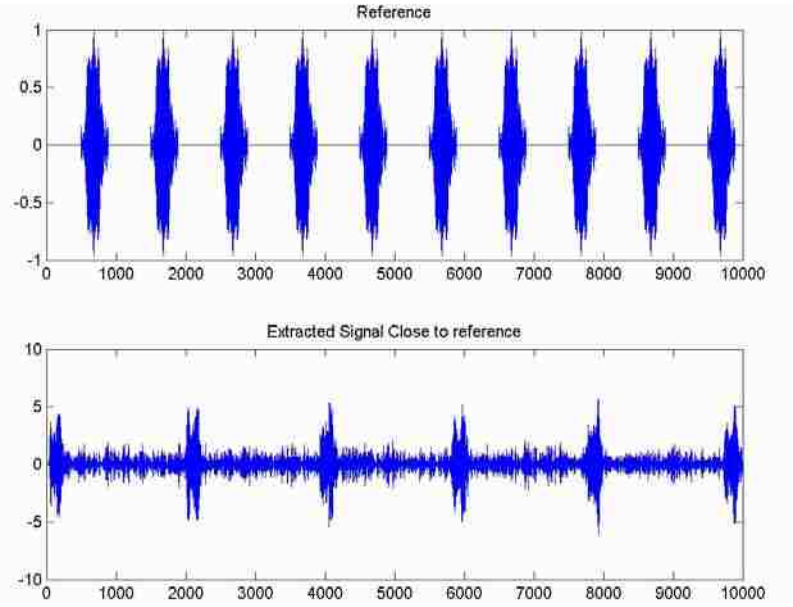


Figure 3.22. Device Signal Extracted from Original Mixed Signals using ICA-R

This extracted signal is then convolved with the matched filter and the periodicity of the signal is distinct as shown in Figure 3.27. Using example 1 and example 2 it has been illustrated that there is potential improvement in the performance of MF in the case of very low SNR signals if ICA-R is used before the MF.

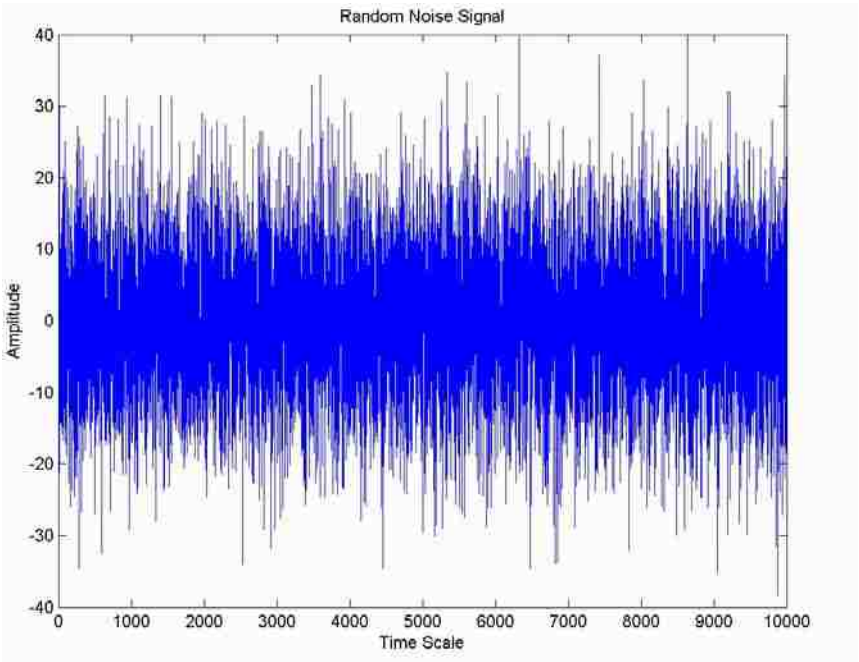


Figure 3.23. Random Noise Added to the Signal

3.3.2.3. Conclusion ICA-R (for matched filtering). The examples illustrate the performance of ICA-R for signals with low SNR a strong interfering signals of similar frequency content. The results show that ICA-R is able to extract the required signal component close to the reference which can be further processed by MF for detection. As illustrated, simply using MF over the mixed signal would not give appropriate results for detection. Computational complexity is reduced if ICA-R is used instead of ICA.

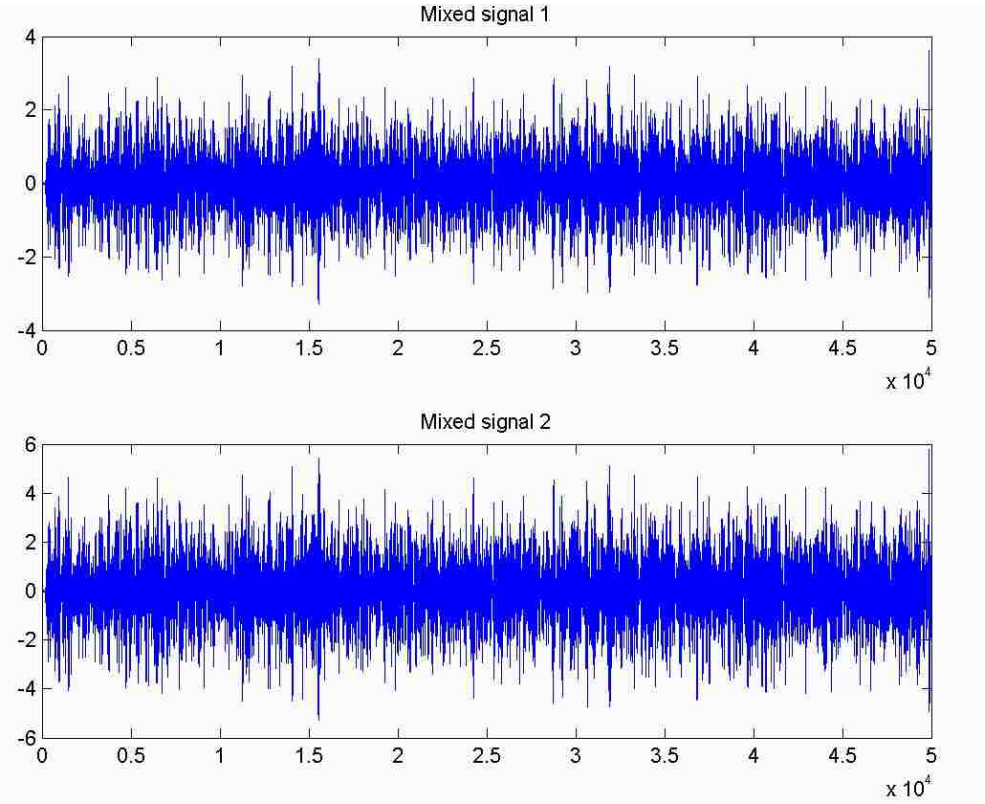


Figure 3.24. Device Signal Added with Random Noise

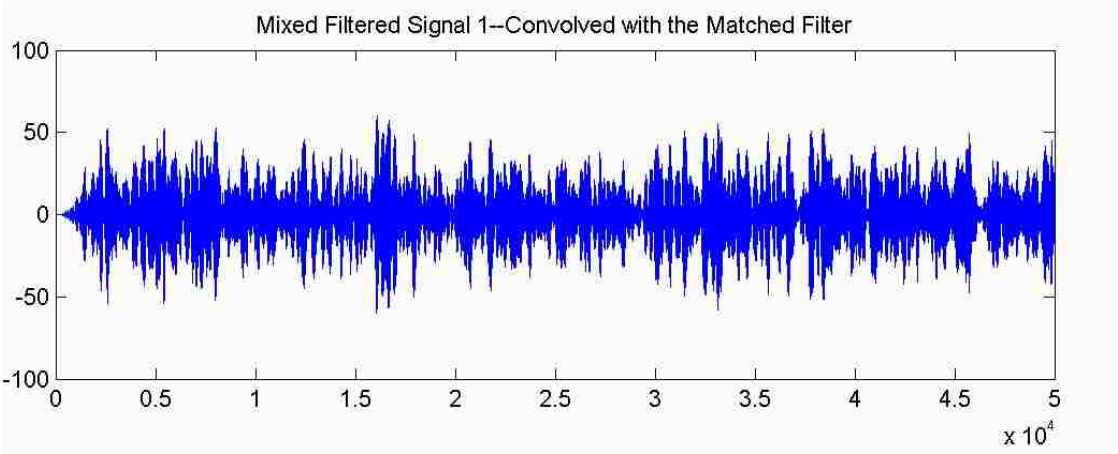


Figure 3.25. Mixed Signal Convolved for Matched Filtering

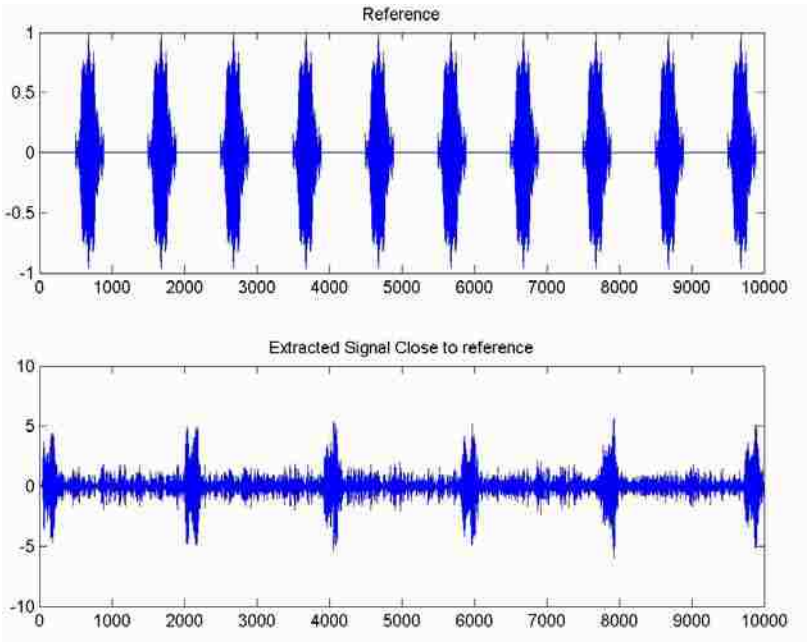


Figure 3.26. Reference Component and Extracted Component from ICA-R

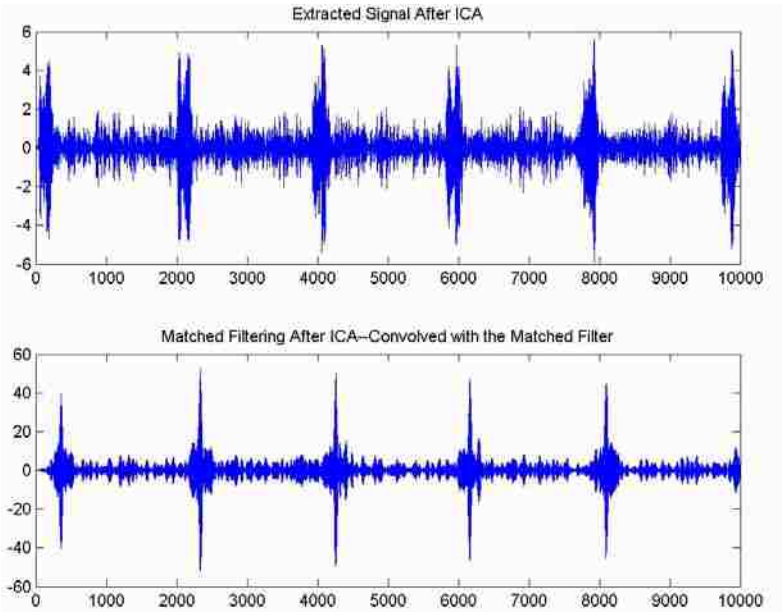


Figure 3.27. Extracted Component Convolved with Matched Filter

4. TIME FREQUENCY ANALYSIS

In Section 3, a time domain matched filtering technique was used for device detection. Sometimes, analyzing the signal in the time-frequency domain is also useful since for non-stationary signals, time domain periodicity does not always hold. So, with matched filtering, identification of such signals is not possible.

Most signals are normally characterized by their frequency spectrum, but only when the signal is stationary and the window length considered for the signal analysis is large enough can the frequency content of the signal can be correctly analyzed. The time-frequency distribution is an important tool for data analysis and it finds its application in the fields of earth science, RADAR, speech and many more.

Time-frequency distribution is a two dimensional representation of the signal given as $TFD(n, f)$, where n = sample and f = normalized frequency.

Information and potential characteristics can be inferred from these distributions. Hence, it is important to get a time-frequency spectrum that has good resolution both in time and frequency. It should be possible to determine how spectral energy changes with time and how temporal energy changes with frequency.

For the time-frequency distribution Heisenberg's uncertainty principle has to be understood. Heisenberg's uncertainty principle expresses a fundamental relationship between the standard deviation of the signal s(n) and the standard deviation of its Fourier transforms S(f) [7]. Thus, when resolving a local temporal event of a signal, a sacrifice in frequency resolution is expected and vice versa [7].

Time-frequency distribution basically represents the power of the frequencies present in the signal and, hence, it should always be positive [7], i.e., $TFD(n, f) \geq 0$.

Two commonly used methods for time-frequency spectral analysis are 1) STFT (Short Time Fourier Transform), also called the spectrogram, and 2) the Wigner-Ville Distribution.

4.1. STFT (SHORT TIME FOURIER TRANSFORM)

The STFT is basically a sliding window technique over the data from which the

Fourier transform is computed. There are several smoothing windows that are used for STFT, including Hanning, Hamming, Gaussian, Hermite, etc. The computation of the STFT is shown in Figure 4.1.

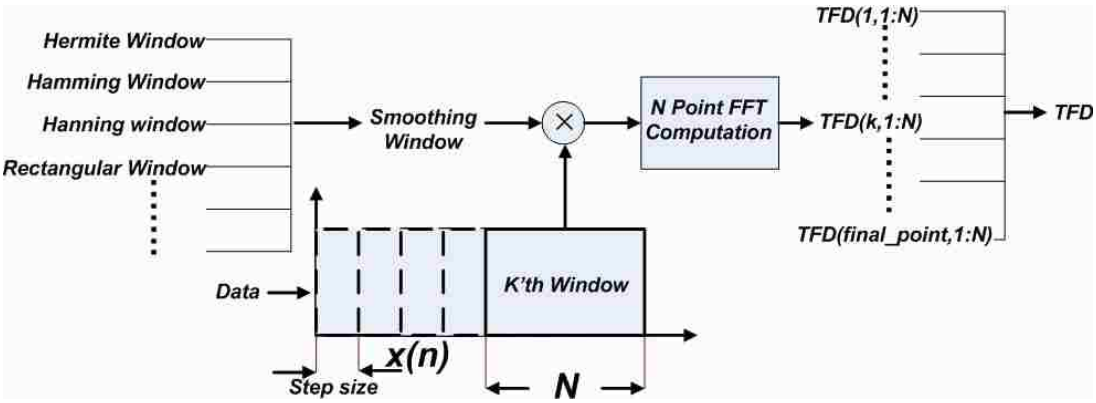


Figure 4.1. STFT Computations

$$\mathbf{x}_w(n) = \mathbf{W}\mathbf{x}(n) \tag{69}$$

where

$$\mathbf{x}(n) = \begin{bmatrix} x(n-\frac{N}{2}) \\ x(n-\frac{N}{2}+1) \\ . \\ x(n) \\ . \\ . \\ . \\ x(n+\frac{N}{2}-1) \end{bmatrix} \tag{70}$$

and $\mathbf{W}$ is a diagonal matrix where the diagonal elements correspond to the elements of

the smoothing window. Where x(n) is the signal, N = length of the window being considered and n is the sampling point at the k'th window. So,

$$TFD(n,f) = \sum_{n=0}^{N-1} X(n) e^{-i2\pi fn/N} \tag{71}$$

One of the advantages of the STFT is that it is fast to compute and easily implemented. However, it must be noted that the window size chosen will always be a compromise between the time and the frequency resolution, as shown Figure 4.2 in which, this is illustrated with a test signal. Suppose, a window of length N is taken, then by taking an N point FFT, the frequency resolution is given by $(f_s / 2)/(N-1)$, where f_s is the sampling frequency of the signal. N also happens to be the time resolution of the analysis. So, as N increases the frequency resolution increases.

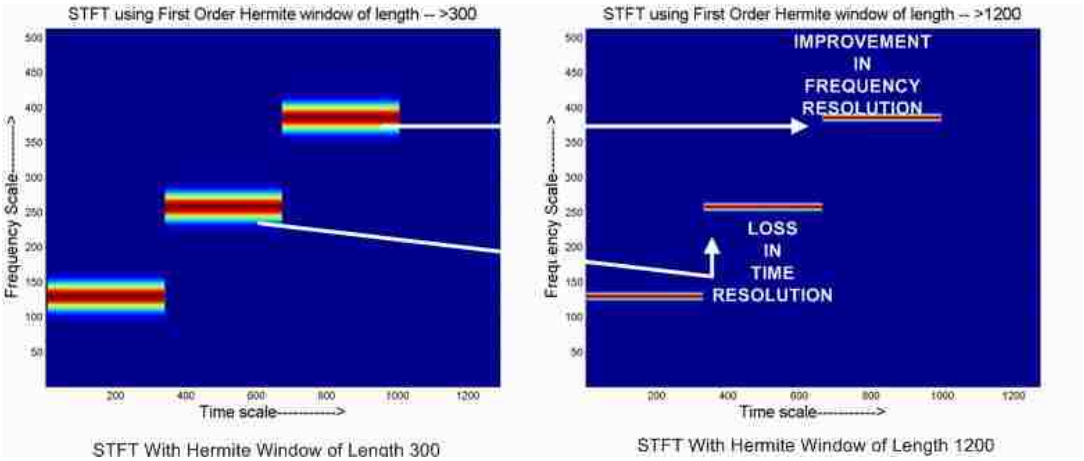


Figure 4.2. Windows Size Affect With Test Signal

4.2. WIGNER-VILLE DISTRIBUTION

The Wigner-Ville is another form of time-frequency distribution. The Wigner-Ville distribution of signal $x(n)$ is given by [8]

$$WV_x(t,f) = \int_{-\infty}^{+\infty} x(t+\frac{\varsigma}{2})\, x^*(t-\frac{\varsigma}{2})\, e^{-2\pi f\varsigma} d\varsigma \tag{72}$$

where t = time and f = frequency. One of the advantages of the Wigner-Ville distribution is that it provides excellent spectral and temporal resolution. Unfortunately though, the Wigner-Ville distribution has cross terms because of its quadratic nature. Suppose there are two sinusoids at f_1 and f_2 occurring at some time interval corresponding to the signal representation of x and y. Then, the Wigner-Ville distribution of the signal $z = x + y$ is given by

$$WV_{x+y}(t,f) = \int_{-\infty}^{+\infty} [x(t+\frac{\varsigma}{2}) + y(t+\frac{\varsigma}{2})][\, x^*(t-\frac{\varsigma}{2}) + y^*(t-\frac{\varsigma}{2})]\, e^{-2\pi f\varsigma} d\varsigma \tag{73}$$

$$WV_{x+y}(t,f) = WV_x(t,f) + WV_y(t,f) + WV_{xy}(t,f) + WV_{yx}(t,f) \tag{74}$$

and will thus produce a significant cross term of $(f_1 + f_2)/2$ as shown in Figure 4.3. However it should be noted that the resolution is very good compared to the STFT; hence, it is widely used for signal analysis. When there are multiple events occurring, then the Wigner-Ville distribution is not preferred because of the excessive cross terms it produces, due to which signal interpretation is difficult.

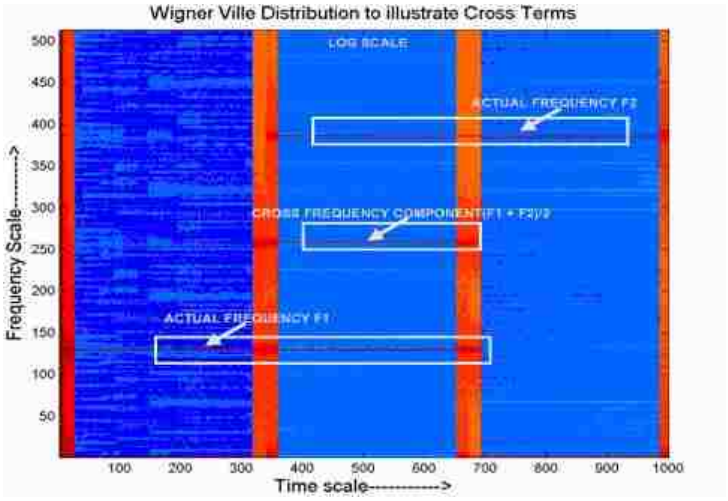


Figure 4.3. Wigner-Ville Distribution to Illustrate the Cross Term

4.3. MULTIPLE WINDOW WIGNER-VILLE DISTRIBUTION [9]

The STFT temporal and frequency resolution is subject to the chosen window size and the artifact control for the type of chosen window. STFT does not produce spectral details as fine as the Wigner-Ville distribution, as one can see by comparing Figures 4.3 and 4.4 which correspond to the STFT and Wigner-Ville, respectively. The Wigner-Ville distribution on the other hand, suffers from cross terms if multiple signal components are present, but it has the desirable property of highlighting fine spectral and temporal changes [9]. In this section we propose generating a family of STFT's using a set of orthogonal windows. These STFT's can then be appropriately combined to obtain optimal time and frequency resolution offered by the Wigner-Ville distribution [9]. The Hermite window family is used for obtaining the family of STFT's. The weights for the spectrogram are then obtained by comparing it to the Wigner-Ville distribution [9] in least squares formulation. This procedure also minimizes the cross terms which appears as artifacts in the Wigner-Ville distribution. In the method of multiple windows up to fourth order Hermite windows are used. The four Hermite windows used are shown in Figures 4.5 and 4.6. The time-frequency distribution thus obtained has minimum cross term contents like the STFT. A masking operation is also done to remove the side lobe

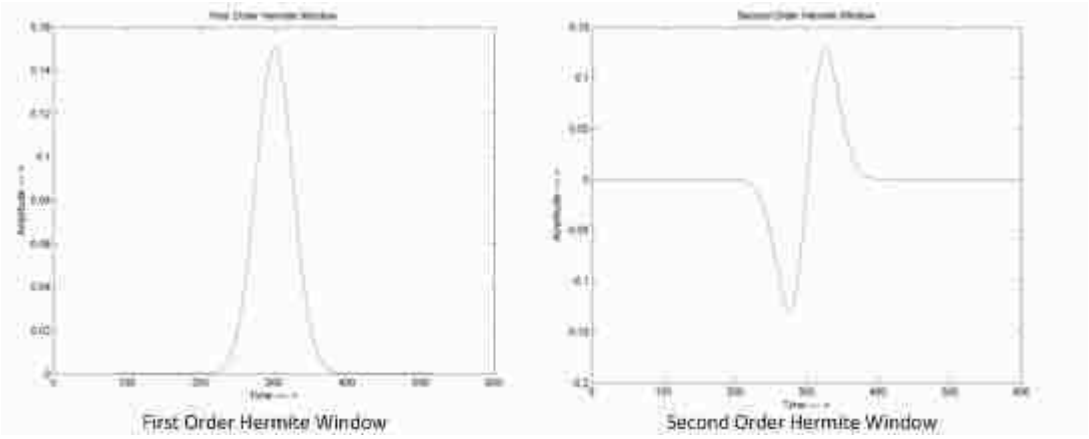


Figure 4.4. First and Second Order Hermite Window

artifacts by the higher order Hermite windows. Thus, overall the method is a non-linear operation for obtaining the TFD [9].

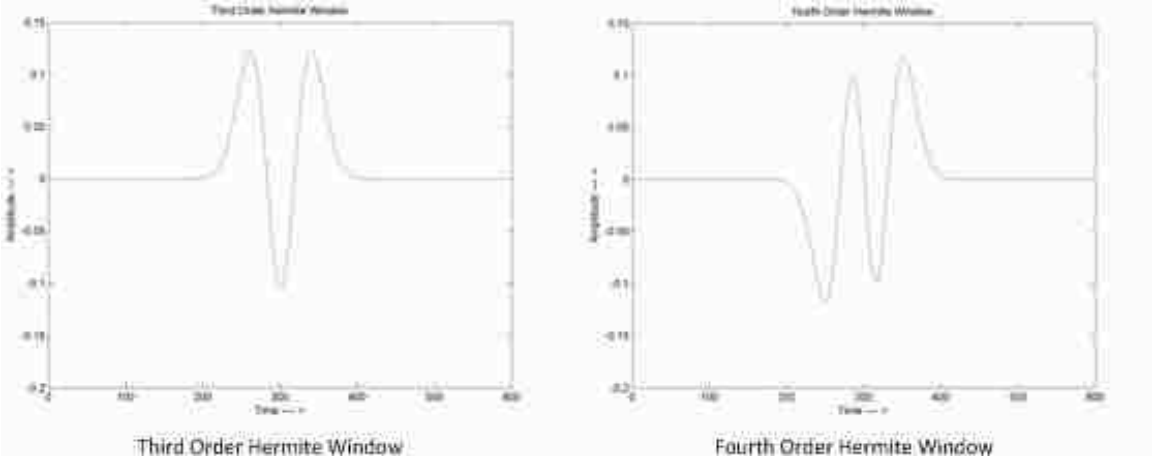


Figure 4.5. Third and Fourth Order Hermite Window

Steps for obtaining multiple window Wigner-Ville distribution [9]

Obtain the Hermite windows at varying orders, as given by

$$h_k(t) = \frac{((-1)^k e^{t^2/2}(\frac{d}{dt})^k e^{-t^2/2})}{\sqrt{(\sqrt{\pi})2^k k!}}. \tag{75}$$

Then find the spectrograms from the Hermite windows as given by

$$STF_K(m,n) = \left| \sum_{l=-L/2}^{L/2-1} x(n+l)\ h_k(l) e^{-j2pilm/L} \right|^2. \tag{76}$$

Next, we find the Wigner-Ville distribution from

$$WV(m,n) = \int_{-\infty}^{+\infty} x(m-\varsigma)\ x^*(m+\varsigma)\ e^{-2\pi n\varsigma} d\varsigma. \tag{77}$$

As mentioned, the Multiple Window Wigner-Ville distribution is obtained by weighting the spectrograms from different Hermite windows. And the objective is to obtain a resulting spectrogram as close as possible to the spectrogram from the Wigner-Ville distribution. So the cost function is set up as:

$$J = \sum_{n=1}^{N}\sum_{m=1}^{M}\left| P(m,n) - \sum_{k=0}^{K-1} d_k S_k(m,n)\right|^2 \tag{78}$$

Where $P(m,n)$ is the reference TFD, and in this case is WV, and $S_k = STF_k$ is the k'th order Hermite window spectrogram. M and N are the row and column dimensions of the TFD matrix. The weights d_k are the weights or the unknown's which can be estimated by minimizing the above equation. The solution to the above mentioned equation is given by

$$\mathbf{d}^{OPT} = \arg\{min(\mathbf{d})\|\mathbf{P} - \mathbf{S}\mathbf{d}\|^2\} \tag{79}$$

where **P** is a vector formed from columns of P(m,n). S is a matrix whose k^{th} column is formed from the columns of $S_k(m,n)$, and $d = [d_1, d_2 \ldots d_{K-1}]^T$. Then the least square solution is given by

$$\mathbf{d}^{OPT} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{P} \tag{80}$$

The TFD is then obtained by combining the k weights and the spectrogram is given by

$$P_N(m,n) = \sum_{K=0}^{N-1} d_k^{OPT} STF_K(m,n) \tag{81}$$

where N is the number of Hermite windows used. In this particular case, it is 4.

The higher order spectrogram often has its spectral energy at undesirable locations. This is due to the presence of more than one positive side lobe. Hence, a masking function denoted as $\zeta(m,n)$ is used. The spectral information from the first two Hermite windows is below an acceptable noise level. Hence, the combination of the first two spectrograms is used as a reference to cancel the noise or the undesirable spectral content arising from the higher order Hermite windows (of an order more than 3). So,

$$P_N(m,n) = \sum_{K=0}^{K=N-1} d_k^{OPT} S_K(m,n) \tag{82}$$

If N = 2, $P_2(m,n)$ is obtained.

The masking function is defined as:

$$\zeta(m,n) = 1 \text{ If } \quad P_2(m,n) \geq 0 \tag{83}$$

$$\zeta(m,n) = 0 \text{ If } \quad P_2(m,n) < 0 \tag{84}$$

So, the final TFD is given by

$$P_N(m,n) = (\sum_{K=0}^{K=N-1} d_k^{OPT} S_K(m,n)) \times \varsigma(m,n)), \text{ where N = 4} \tag{85}$$

4.4. RESULTS AND CONCLUSION TIME-FREQUENCY DISTRIBUTIONS

Three different signals are analyzed to evaluate the performance of each time-frequency distribution, i.e., i) Short Time Fourier Transform, ii) Wigner-Ville Distribution and iii) Multiple Hermite Windows.

CASE 1

Consider the test signal which has frequency variation with time given as $f=k \times t^2$, where k is a constant. The signal in the time domain is shown in Figure 4.6. The ideal time-frequency template is shown in Figure 4.7. It can be observed from Figures 4.8 and 4.9 that the STFT offers poorer resolution than the Wigner-Ville distribution. From Figure 4.10 it is apparent that the Multiple Window Wigner-Ville (MWWV) distribution offers better resolution as compared to STFT, and is comparable to the Wigner-Ville distribution.

CASE 2

The chirp signal is used as a bench mark for evaluating the time-frequency distribution. The signal is repeated after a certain silent time frame. The signal in the time domain is shown in Figure 4.11 and the ideal time-frequency template is shown in Figure 4.12. From Figures Figure 4.13, 4.14 and 4.16 it is apparent that the STFT again offers poorer resolution as compared to Wigner-Ville distribution and Multiple Window Wigner-Ville distribution. The gaps on the TFD are because the signal is repeating after certain time gap.

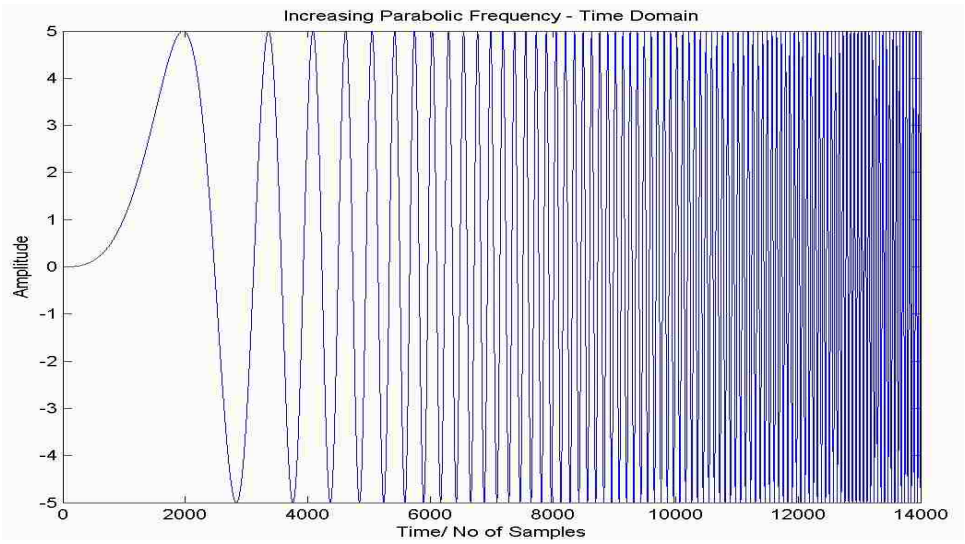


Figure 4.6. Increasing Parabolic Frequency Signal

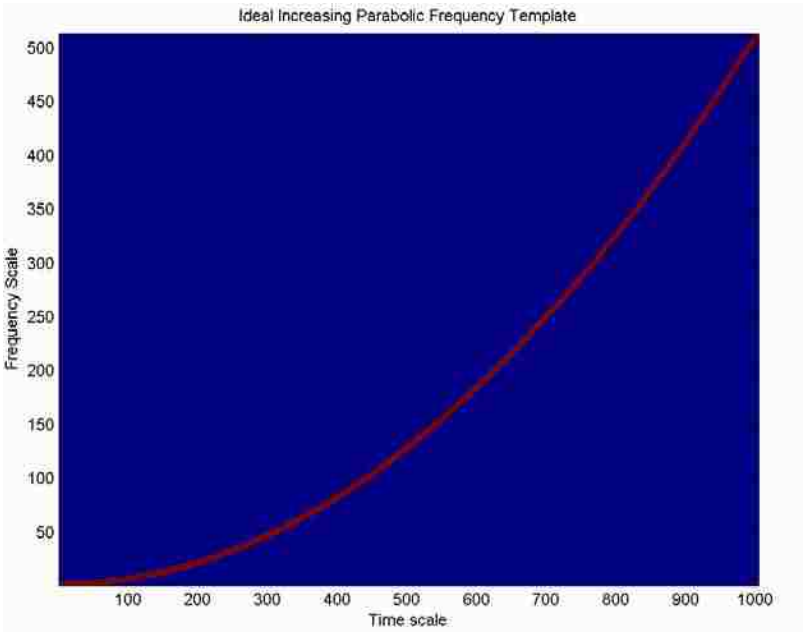


Figure 4.7. Template Ideal TFD for Increasing Parabolic Frequency

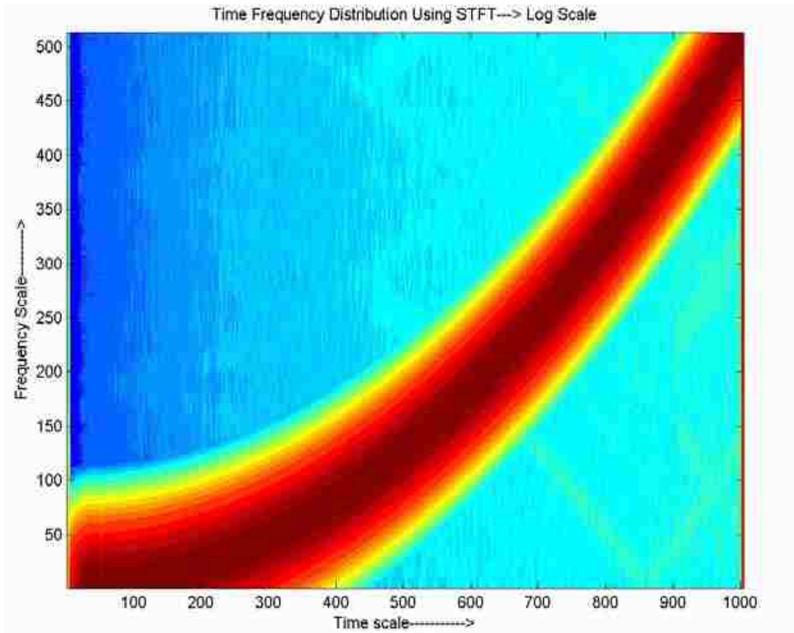


Figure 4.8. STFT for Increasing Parabolic Frequency

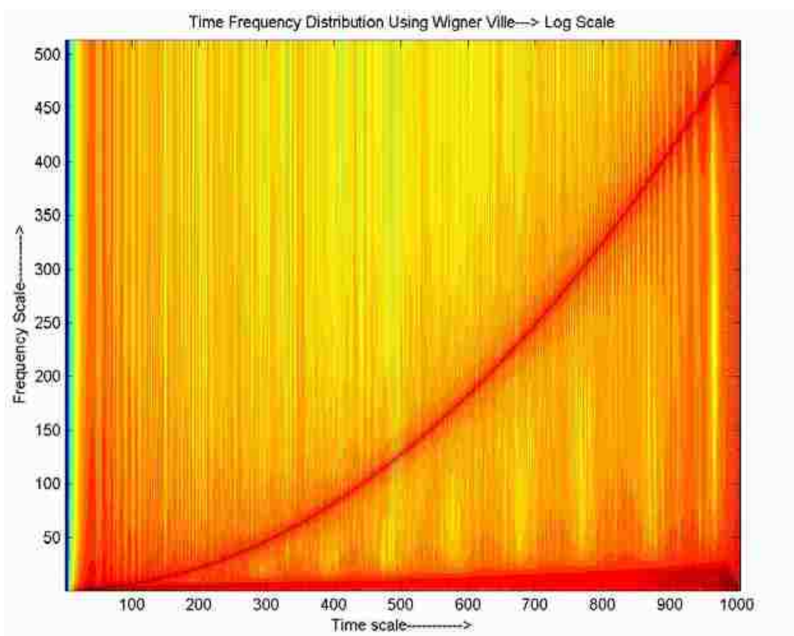


Figure 4.9. Wigner-Ville TFD for Increasing Parabolic Frequency

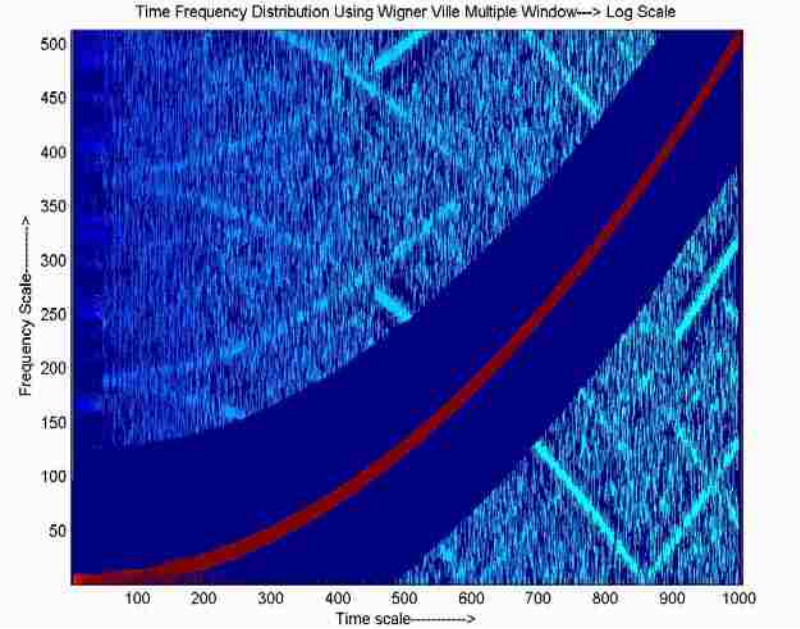


Figure 4.10. MWVV TFD for Increasing Parabolic Frequency

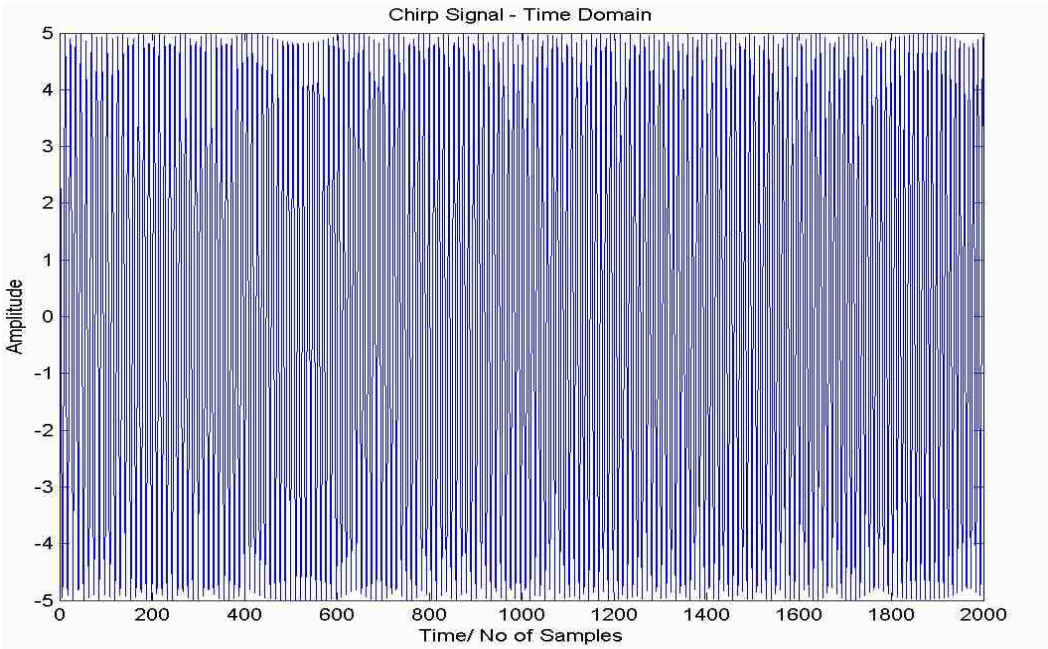


Figure 4.11. Chirp Signal

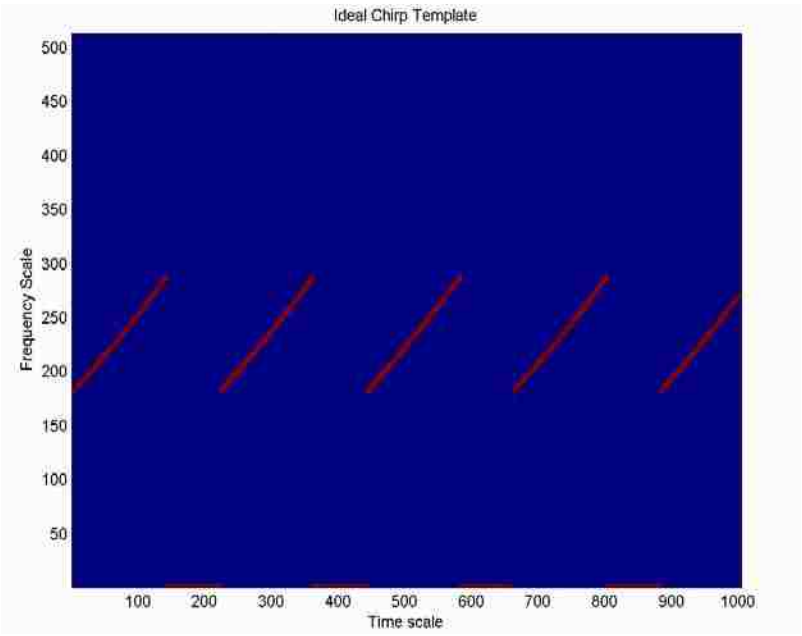


Figure 4.12. Template TFD for Chirp Signal

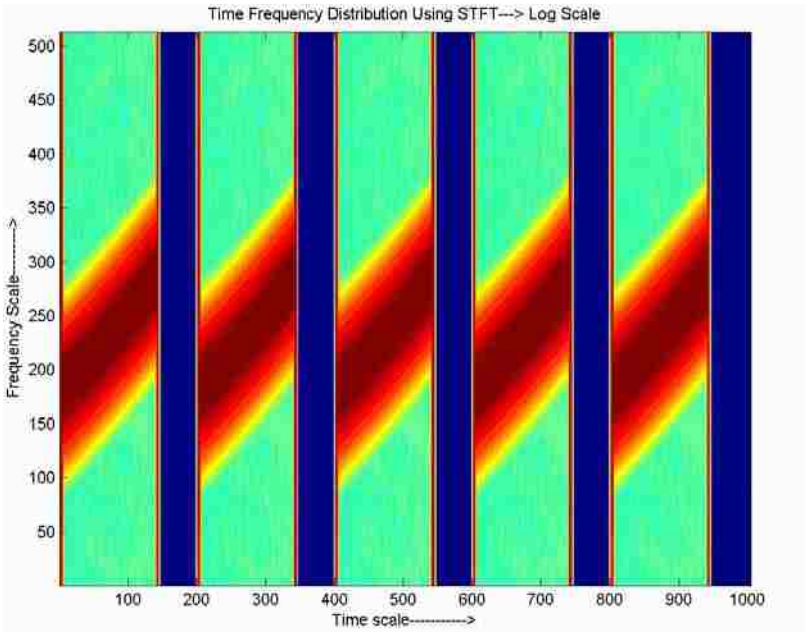


Figure 4.13. STFT for Chirp Signal

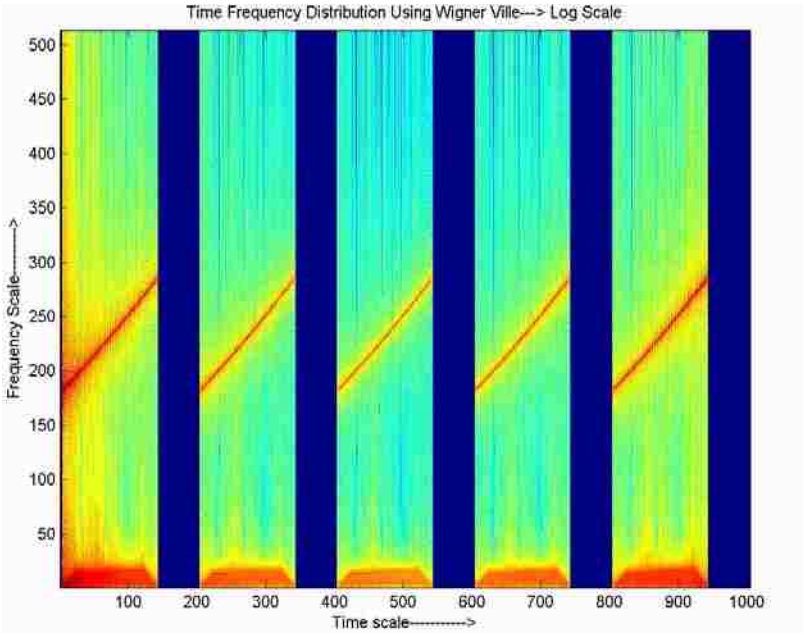


Figure 4.14. Wigner-Ville TFD for Chirp Signal

Figure 4.15. MWWV TFD for Chirp Signal

CASE 3

Consider this test signal to which both decreasing and increasing parabolic frequency signals have been added. The signal in the time domain is shown in Figure 4.16 and the ideal time-frequency template is shown in Figure 4.17. It is observed from the Figures 4.18 and 4.19, that this time STFT is much better than the Wigner-Ville distribution (Figure 4.19). The cross-components are present in Wigner-Ville TFD. The cross component appears when ever there is more than one signal component is present in the signal. Figure 4.20 shows that the Multiple Hermite window technique has a much better SNR compared to the STFT and the Wigner-Ville distribution.

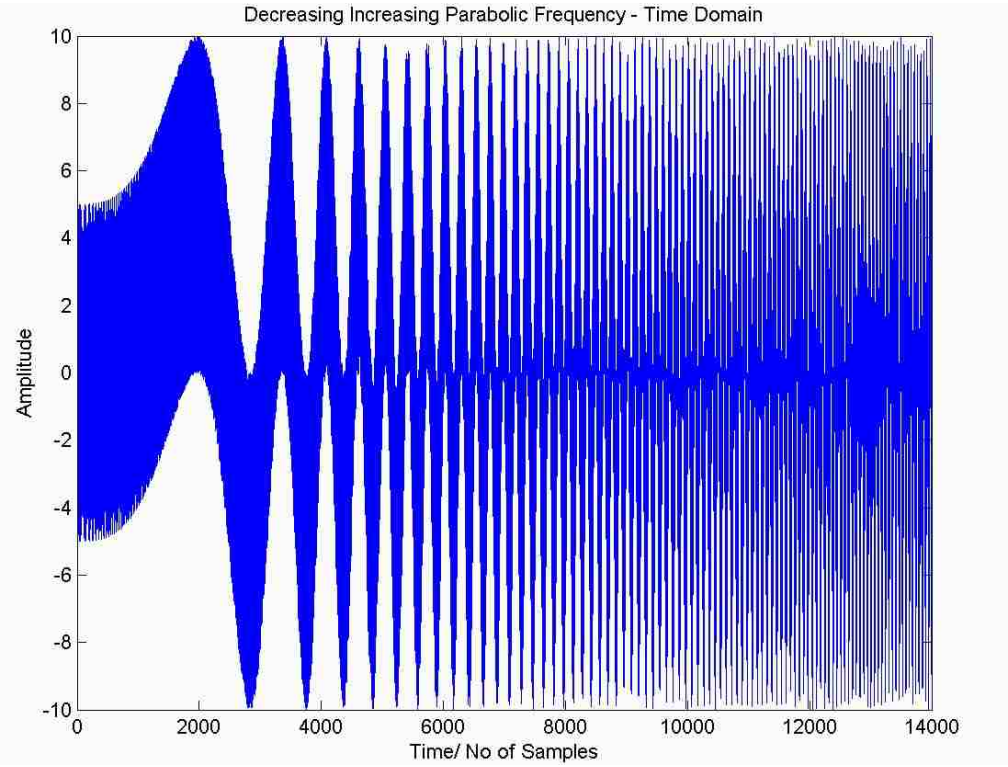


Figure 4.16. Signal having Increasing and Decreasing Parabolic Frequency

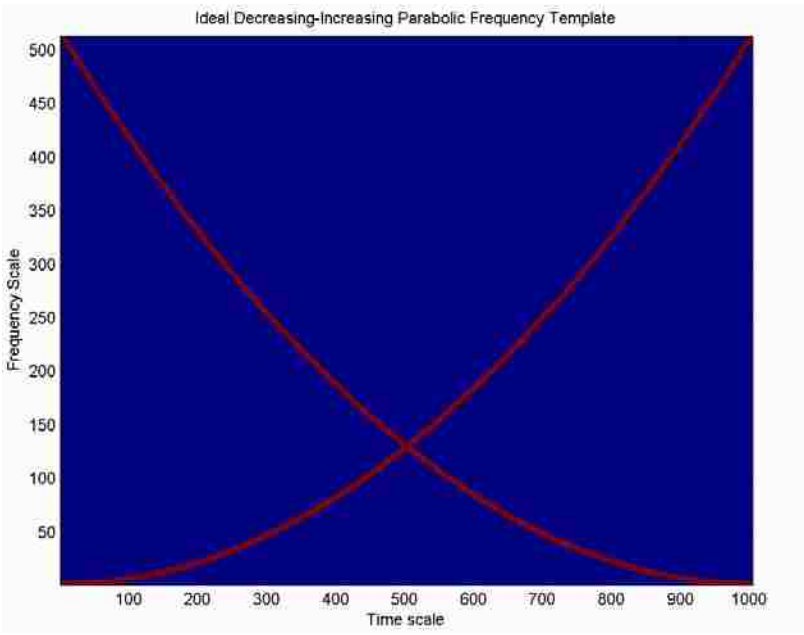


Figure 4.17. Ideal Template TFD of Increasing and Decreasing Parabolic Frequency

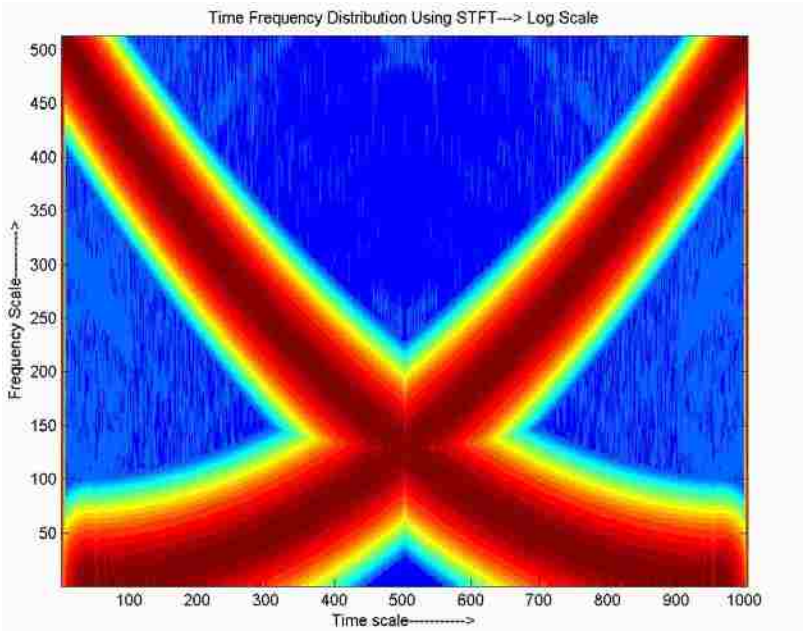


Figure 4.18. STFT of Increasing and Decreasing Parabolic Frequency

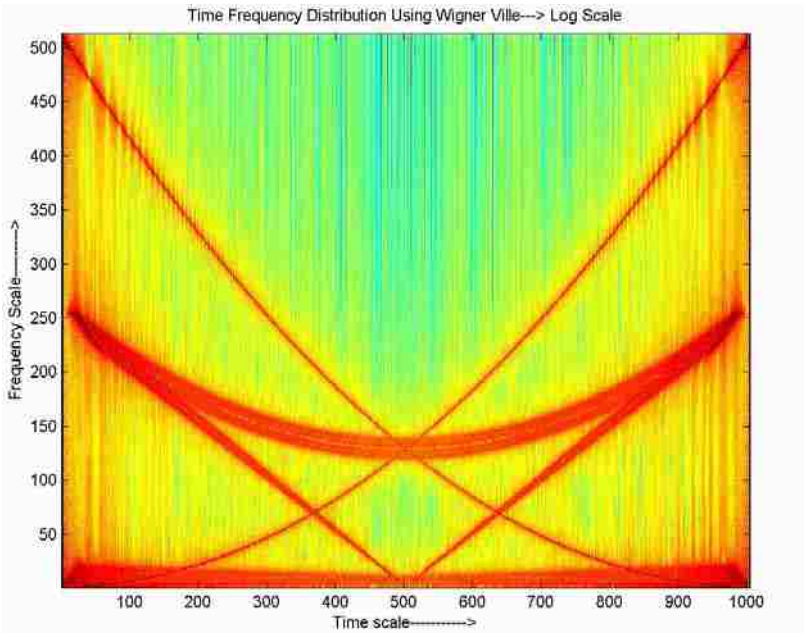


Figure 4.19. Wigner-Ville TFD of Increasing and Decreasing Parabolic Frequency

Figure 4.20. MWWV TFD of Increasing and Decreasing Parabolic Frequency

The power in the TFD from where the signal is present is computed from the template in each case and the remaining power in the TFD is summed up as noise or the artifact component. The SNR for all the cases are listed in Table 4.1. It can be concluded from the above test cases that the Multiple Window Wigner-Ville distribution retains the advantage offered by STFT, i.e., minimizing on the cross terms, and provides high resolution both in time and spectral domain.

Table 4.1. SNR for Different Signal Using STFT, Wigner-Ville and Multiple Window-Wigner-Ville

	Signal to Noise Ratio (SNR) in dB for different Signals		
	STFT	**Wigner-Ville**	**Multiple Window-Wigner-Ville**
Increasing Parabolic Signal	11.3053	17.0800	**17.6705**
Chirp Signal	10.8424	9.0912	**21.0331**
Dec and Inc parabolic	7.5860	7.2416	**12.9134**

5. DETECTION USING SINGULAR VALUE DECOMPOSITION

In the previous section, some methods for obtaining the time-frequency distribution have been studied. The time-frequency spectrum helps in the evaluation of the signal. So, it can be used for signal detection. The usage of the singular value decomposition (SVD) of the TFD which can be applied for signal detection is studied in this section. As shown in Figure 5.1, the input signal is band-pass filtered as a pre-processing step. Then its TFD is computed.

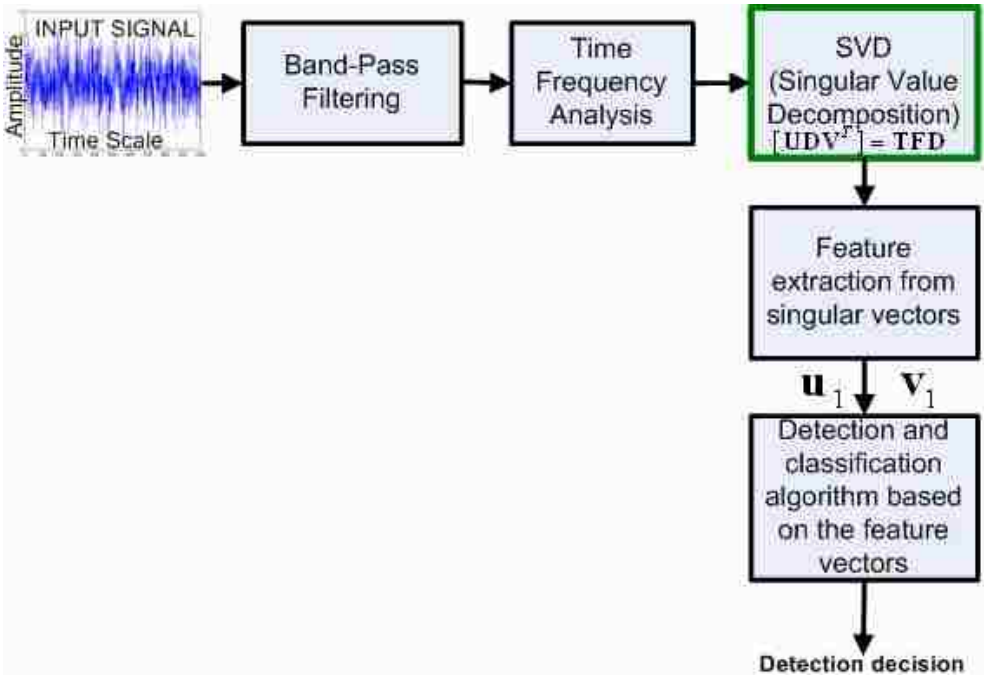


Figure 5.1. SVD Based Signal Detection

To make computations efficient, a method of dimensionality reduction is used since; TFD is large in its dimension space. SVD is one of the most popular methods used

for the decomposition of the TFD and so is used as a dimensionality reduction step. After dimensionality reduction, the most significant vectors shown ($\mathbf{u}_1$ and $\mathbf{v}_1$ from the left and right singular matrices) are used for the device detection because they represent the characteristics of the signal in time and frequency domains, respectively. Features of these vectors are, thus, used for the classification and detection purposes [10, 11, and 12]. SVD features are widely used for identification and classification in many applications including maternal and fetal EEG signals and brain signal analysis.

5.1. SVD BASED DECOMPOSITION OF TFD

The SVD decomposes the TFD into basis matrices, which are associated with certain weights. If the time-frequency distribution matrix is denoted by TFD, then it can be decomposed into basis matrices with corresponding weights [13].

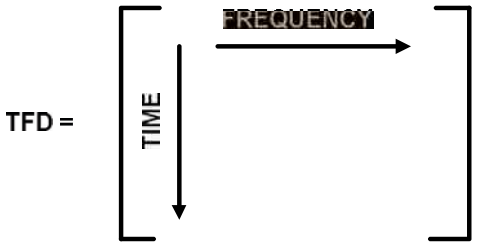
$$\mathbf{TFD} = \sum_{i=1}^{R} \sigma_i \mathbf{B}_i \tag{86}$$

where $\mathbf{B}_i$ is the basis matrix, and σ_i is the weight associate with $\mathbf{B}_i$. R is the rank of the matrix and $\mathbf{B}_i = \mathbf{u}_i \mathbf{v}_i^T$, where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the singular vectors associated left and right singular matrices which are expressed as follows:

$$\mathbf{TFD} = \begin{bmatrix} u_{11} & u_{21} & \cdots & u_{R1} \\ u_{12} & u_{22} & \cdots & u_{R2} \\ u_{13} & u_{23} & \cdots & u_{R3} \\ \vdots & \vdots & \cdots & \\ \vdots & \vdots & \cdots & \\ \vdots & \vdots & \cdots & \\ u_{1L-1} & u_{2L-1} & \cdots & u_{RL-1} \\ u_{1L} & u_{2L} & \cdots & u_{RL} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_R \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & v_{13} & \cdots & v_{1N-1} & v_{1N} \\ v_{21} & v_{22} & v_{23} & \cdots & v_{2N-1} & v_{2N} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ v_{R1} & v_{R2} & V_{R3} & \cdots & v_{RN-1} & v_{RN} \end{bmatrix}.$$

Here, L is the row dimension corresponding to the time scale and N is the column dimension corresponding to the frequency scale on the TFD matrix. This form of

decomposition of the TFD into left and right singular matrices helps to reduce the



dimensionality. Typically most of the energy of the signal is concentrated into a small set of the largest singular values. It is common to order the singular values from the largest to the smallest and the singular vectors correspondingly. It is expected that the first few singular vectors can be used for the TFD characterization, i.e., by using information from both time and frequency domains. The left singular vectors represent the significant characteristics of the signal in the time domain and the right singular vectors represent the significant characteristics of the signal in the frequency domain.

5.2. SVD FOR DEVICE DETECTION

Figure 5.2 shows the signal of a typical device before and after pre-processing. The TFD of the signal is shown in Figure 5.3. The first singular vector (that corresponds to the largest singular value) from the left singular matrix $\mathbf{u}_1$ is shown in Figure 5.4. The TFD of the signal is computed with a frame advance of 50, where "frame advance" is the number of samples between the first sample of each frame. In order to reduce the computational complexity a STFT has been used for obtaining the TFD. First order Hermite function with a length of 300 samples was used as a smoothing window. The distance between the peaks for this particular toy truck signal is approximately 200 samples. The randomness of the singular vector can be smoothened by an envelope detection method. To apply envelope detection to a signal *x*, the signal is squared and passed through a low pass filter, this reduces the high frequency content, in the form of spurious spikes. The output is shown in Figure 5.5.

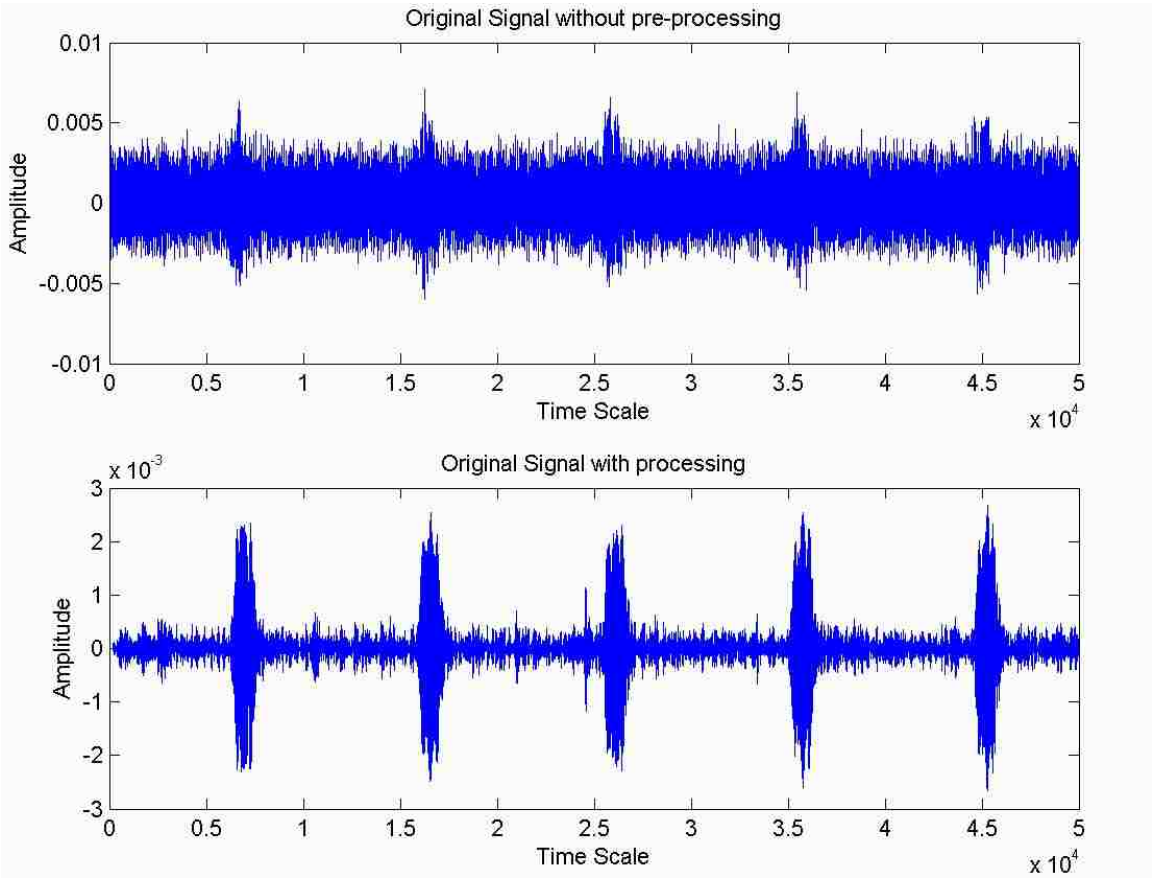


Figure 5.2. Signal Before and After Pre-processing

Figure 5.6, which shows the first right singular vector, $\mathbf{v}_1$, shows the peak frequency of the signal at a normalized scale of 50, for this signal. But, this frequency information alone cannot be used for device signal identification since the signal is bandpass filtered within the normalized frequency scale of approximately 33 to 66. Therefore, frequency domain information (the first right SV), along with the time domain information (the first left SV), is used for the signal detection. The frequency domain first right singular vector, after envelope detection is shown in Figure 5.7. Off-shift in peak from approximately 50 to 80 in normalized frequency scale is due to the group delay in the envelope detection operation when the SV is filtered by a 63 tap low pass filter.

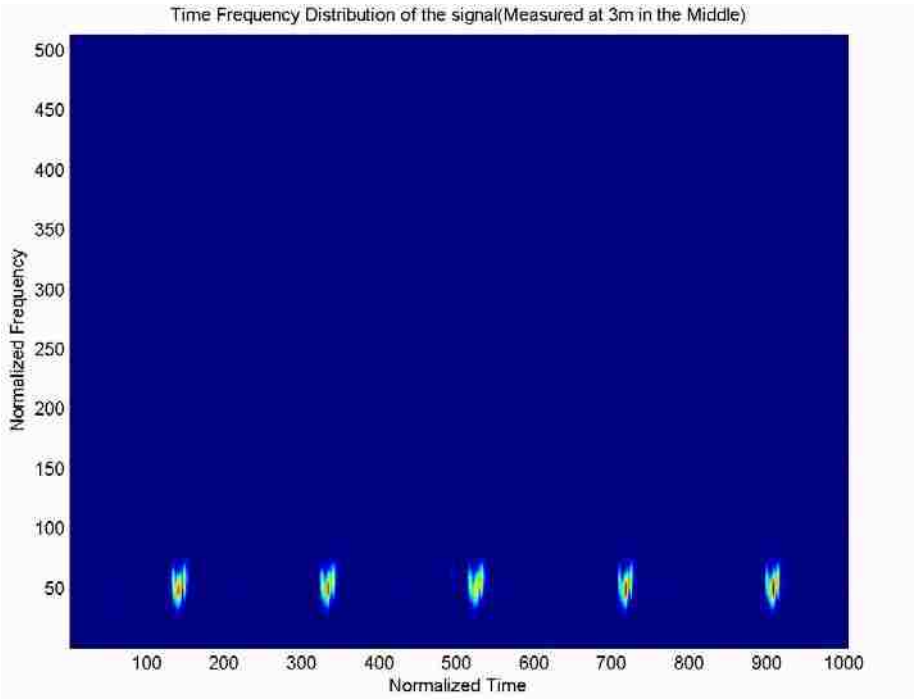


Figure 5.3 Time Frequency Distribution of the Device Signal

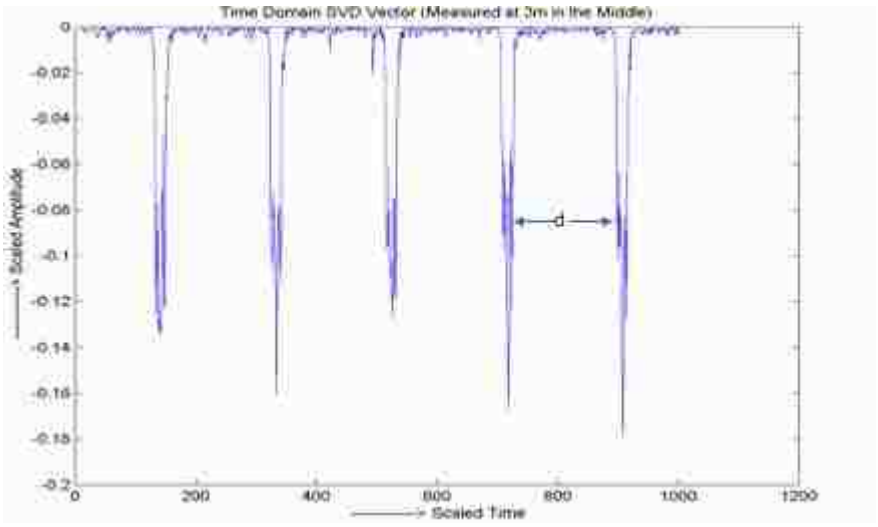


Figure 5.4. First SV of the Left Singular Matrix

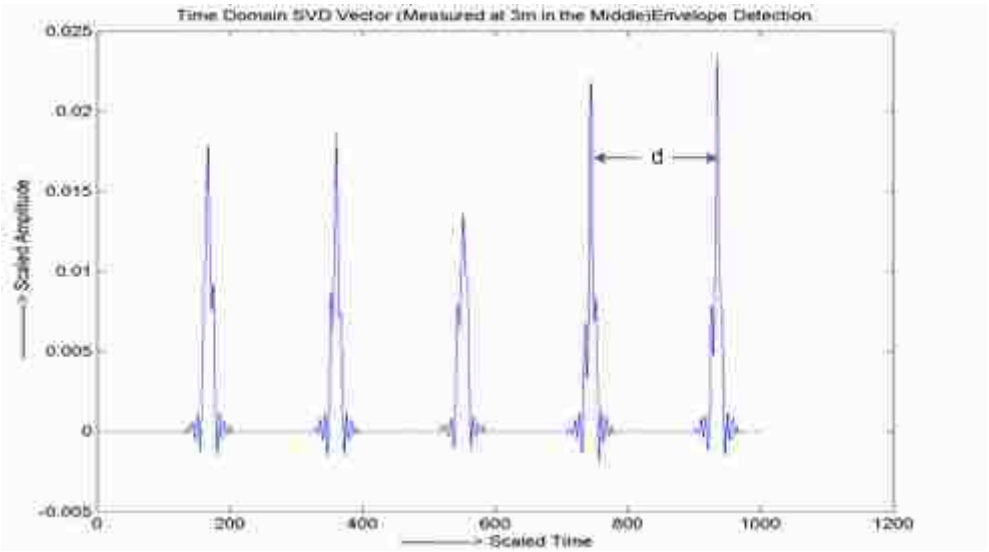


Figure 5.5. First SV of the Left Singular Matrix After Envelope Detection

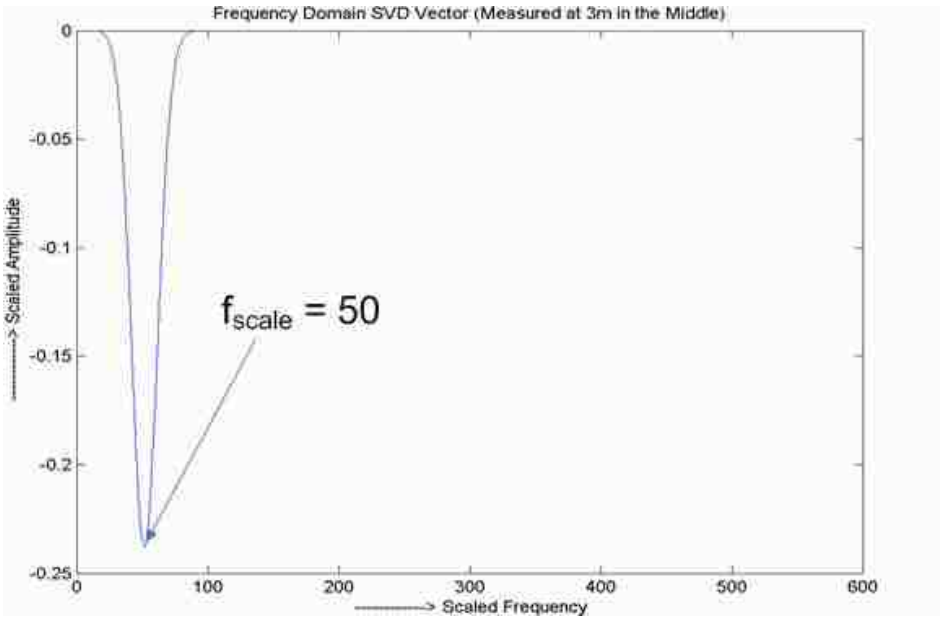


Figure 5.6. First SV of the Right Singular Matrix

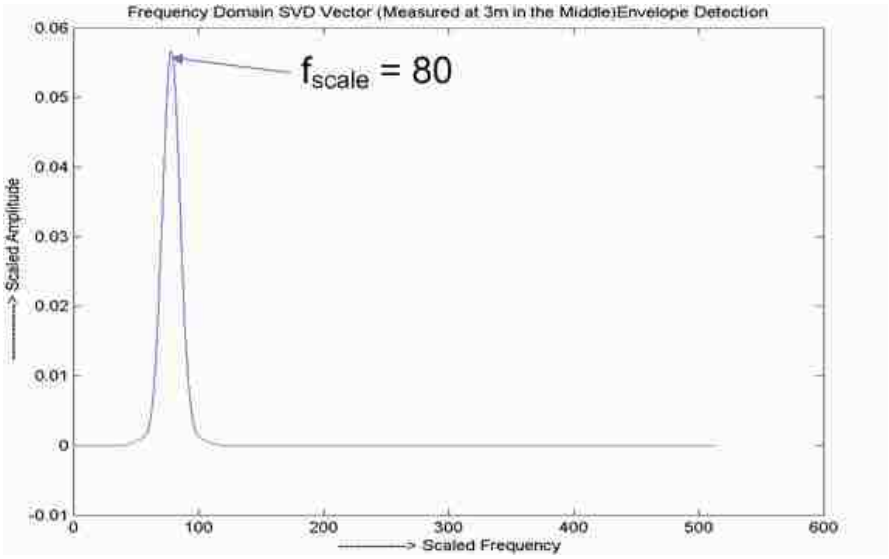


Figure 5.7. First SV of the Right Singular Matrix After Envelope Detection

The detection algorithm depends on 1) Peak detection in the first frequency SV, 2) Peak detection in the first time SV, and 3) the distance between the peaks in the first time SV. Figures 5.8 and 5.9 show first time and frequency SV's respectively. The following parameters, shown in the figures are defined as:

1) *d*, the distance between the peaks in the first time SV,

2) ς, the threshold for peak detection in the first time SV,

3) *D,* the time within over which the average peaks location in a local group are averaged,

4) Δf, the error on the frequency that can be tolerated. (For example, in this particular case the peak occurred at 50. Hence if $\Delta f = 5$, then peak of the frequency within a normalized frequency scale of 50 ± 5 units is tolerable and can be considered for detection.)

5) Δt, the error on the distance between the peaks in the time scale SV that can be tolerated. (If distance between the peaks is 200 and if $\Delta t = 5$, then distance between the peaks within 200 ± 5 samples is tolerable and can be considered for detection.)

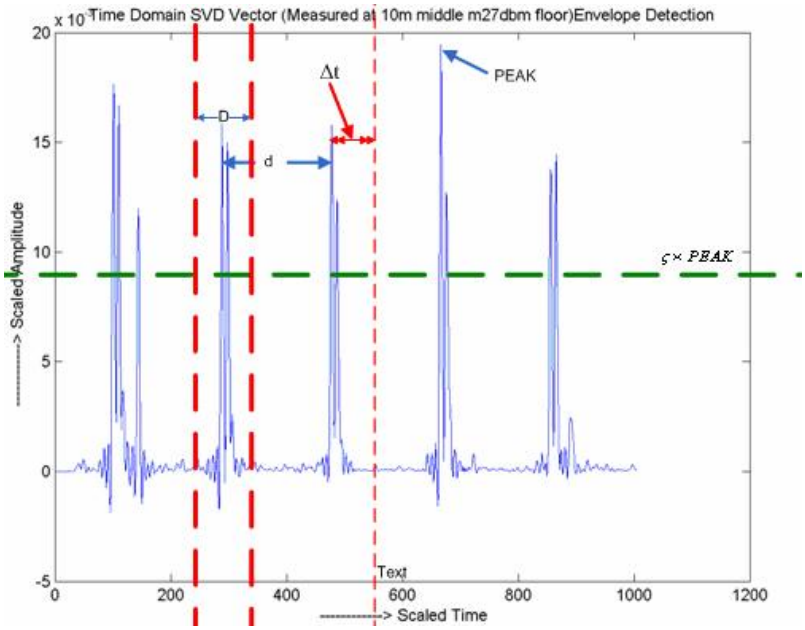


Figure 5.8. Variable Parameters Illustration Time Scale SV

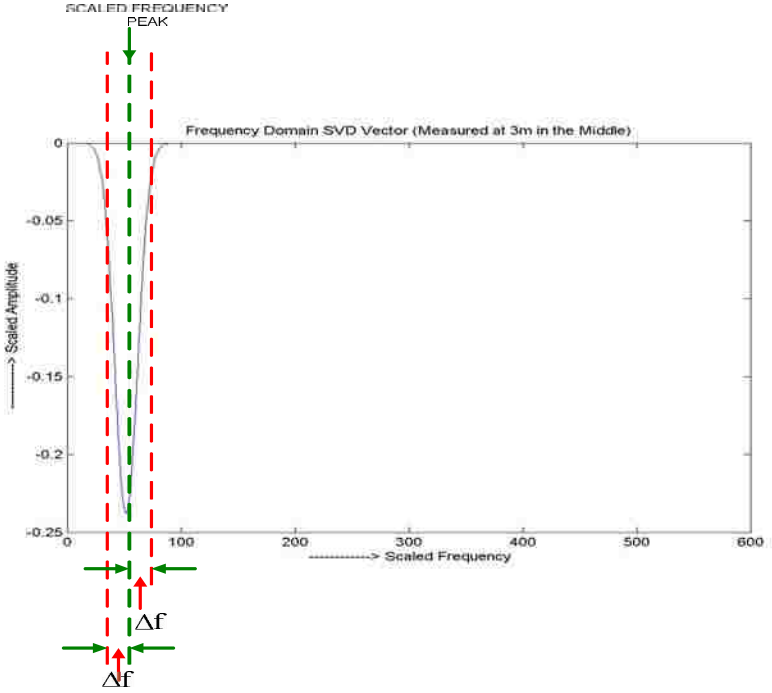


Figure 5.9. Variable Parameters Illustration Frequency Scale SV

The algorithm for the SVD based detection algorithm is shown in Figure 5.10. After the time-frequency analysis the first left and right SV's are obtained which are used for the device detection. Then, the average distance between the peaks in the first left SV (time scale) is computed. Then, the peak in the first right SV (frequency scale) is computed. The device detection has to satisfy two conditions:

1) The average distance between the peaks in the first left SV (time scale) should be within the tolerance given by Δt and

2) The peak in the first right SV (frequency scale) should be within the tolerance given by Δf.

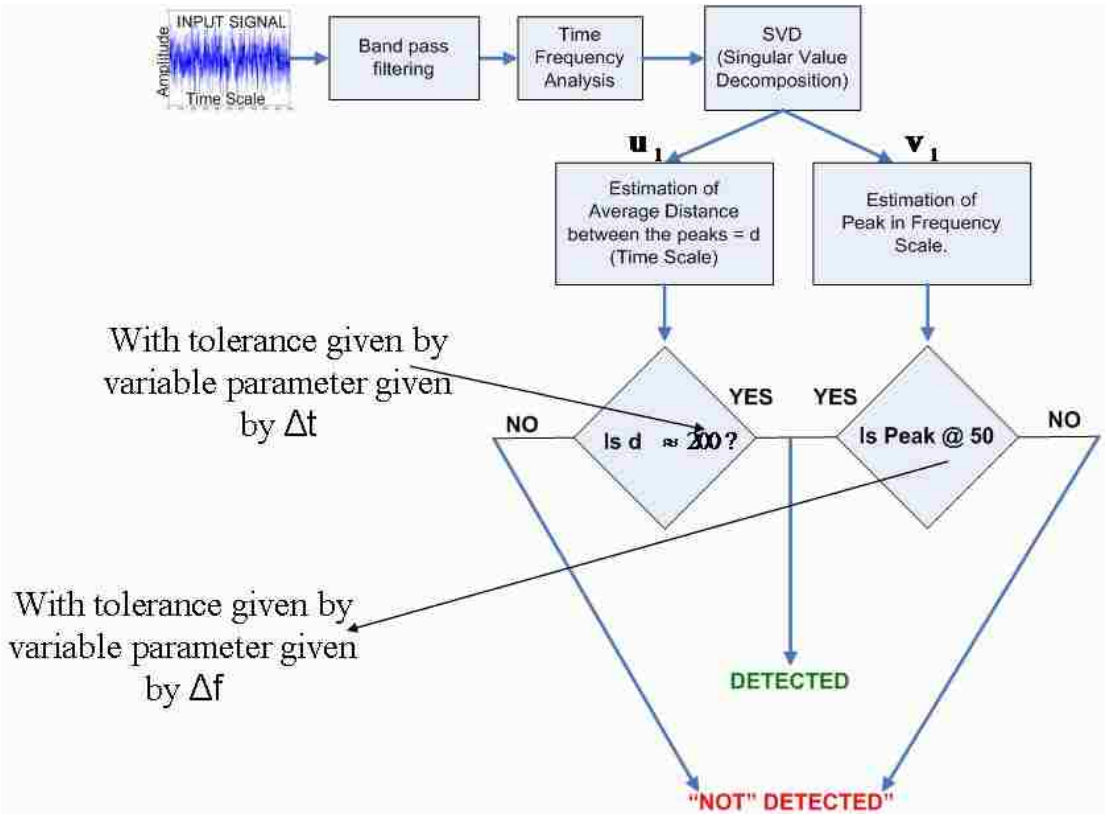


Figure 5.10. SVD Based Algorithm Flow

5.3. SVD BASED DETECTION RESULTS AND CONCLUSION

The ROC curve of SVD based detection curve is shown in Figure 5.11. The ROC curve has been computed by, varying the Δt from 15 to 30 and then estimating the "true detection" and the "false detection" percentages. The other parameters such as $\Delta f = 8, \varsigma = 0.27$, D = 120, d = 200, are kept constant. The best operating point obtained from the ROC curve is at $\Delta t = 18$, where the detection percentage is 92.1 and false alarm percentage is 11.07. Another operating point can also be chosen from the ROC curve which gives a detection percentage of 92.55 and false alarm percentage of 11.6. The SVD based method of detection is not as robust as the matched filtering based detection. Since, the number of parameters which needs to be tuned is high. The false detection percentage from the SVD based detection is considerably larger than the false detection percentage obtained from matched filtering based detection, while the detection percentage is comparable.

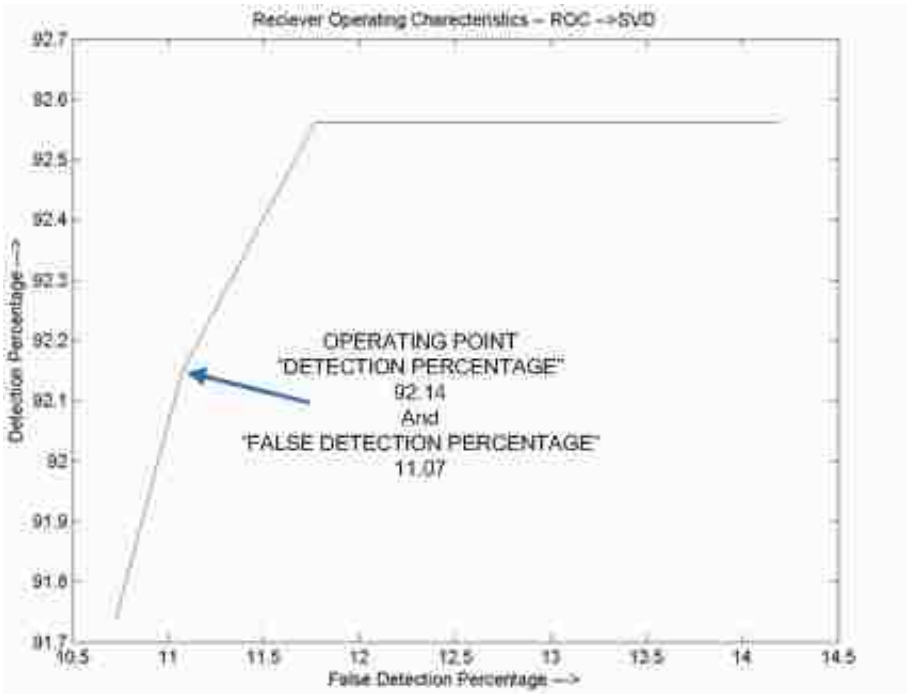


Figure 5.11. ROC Curve SVD Based Detection

6. ICA-R ON SVD FOR DEVICE DETECTION

As in the case of the MF, the SVD based detection cannot perform well if the signal power is low as compared to the noise. The TFD of such a signal is shown in Figure 6.1. The corresponding first three singular vectors from the left singular matrix are shown in Figure 6.2. None of the first three SV's signifies the device's presence. In a similar analysis of the sound signals in reference by Michael Anthony Casey [10], it is explained and will be proved further that the singular vectors are not statistically independent. And, because of this the distinct features that are seen for the device signal, its presence cannot be resolved [10]. The device signal, which has lower noise interference, the 1st SV can be used directly, because the device signal's characteristics are more prominent than the noise. In this section the methodology of using the ICA on the SV's in which the SV's are made statistically independent is explained. The idea is to then use the relevant independent component for the device detection. A similar idea has been used for the separation of the fetal ECG from a single mixture using the SVD-ICA method [14]. The basic block diagram is shown in Figure 6.3.

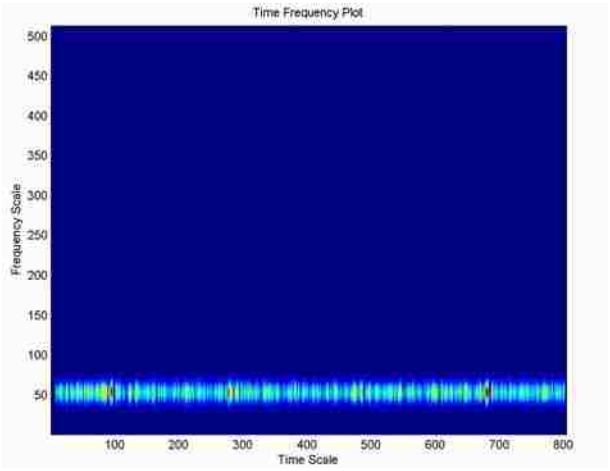


Figure 6.1. Time Frequency Distribution of Noisy Device Signal

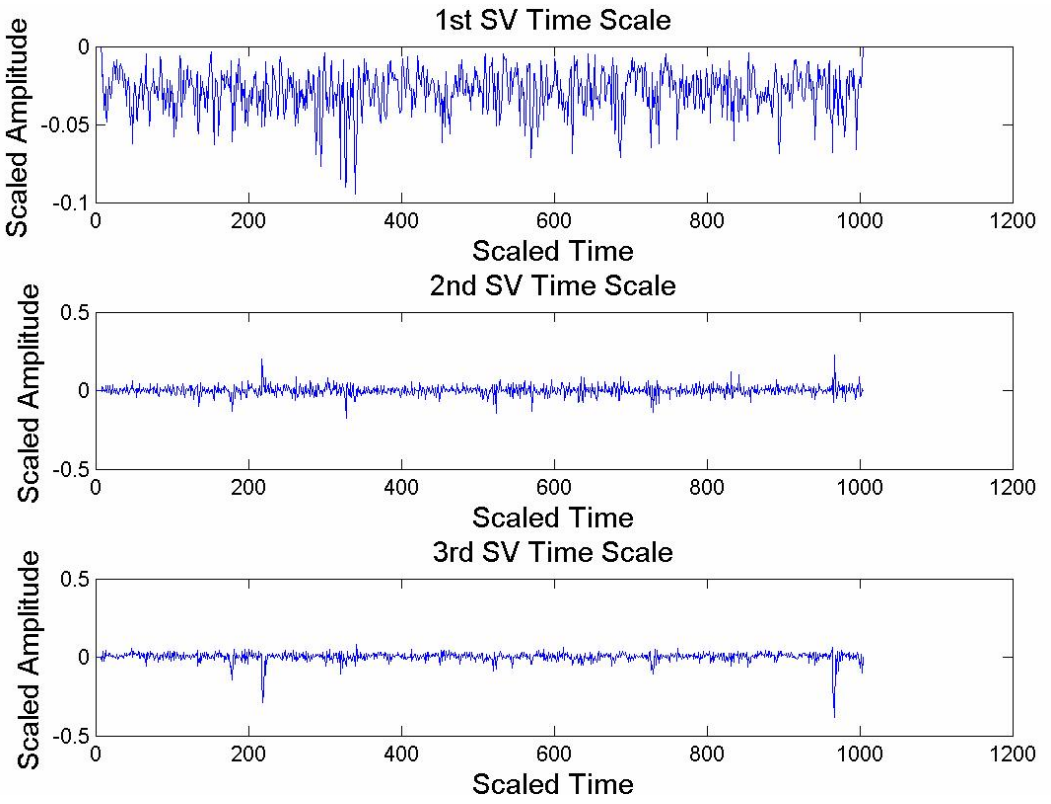


Figure 6.2. First 3 SV's from the Left Singular Matrix

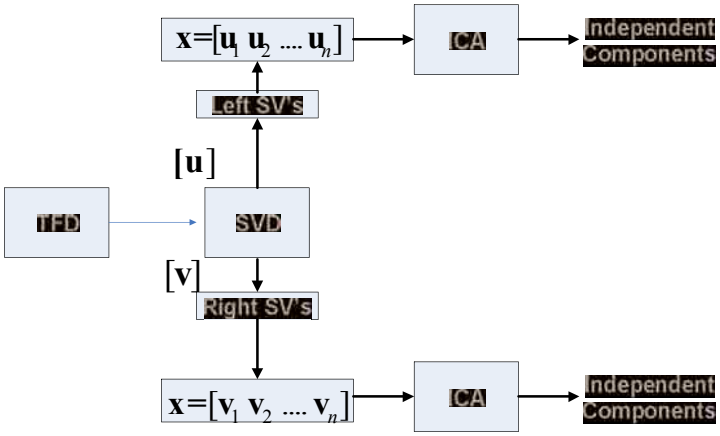


Figure 6.3. ICA on SV's

6.1. DEPENDENCE OF SV'S AND SEPARATION USING ICA

The example is as explained in [2]. Consider two stationary random processes $s_1(n)$ and $s_2(n)$. Both have a uniform probability distribution function (PDF), varying between 0 and 1, such that

$$P[s_1(n)] = p[s_2(n)] = 1/N;\ s_i \in \{0, \frac{1}{N}, \frac{2}{N}, ..., \frac{N-1}{N}\}. \tag{87}$$

Let

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}^T(0) \\ \mathbf{s}^T(1) \\ . \\ . \\ \mathbf{s}^T(L-1) \end{bmatrix} \tag{88}$$

where

$$\mathbf{s}(n) = \begin{bmatrix} s_1(n) \\ s_2(n) \end{bmatrix} \tag{89}$$

where $s_i(n)$ is the nth output out of random process s_i. From the scatter plot of **S**, as shown in Figure 6.4, it can be observed that variables $s_1(n)$ and $s_2(n)$ are independent of each other. The SVD of the input data is obtained as

$$\mathbf{S} = [\mathbf{U}\ \mathbf{D}\ \mathbf{V}^T]. \tag{90}$$

So,

$$\mathbf{U}=[\mathbf{u_1}\ \mathbf{u}_2] \tag{91}$$

corresponds to the transformation done by the SVD on the input data. Figure 6.5 shows the scatter plot of the SV's of **U** ($\mathbf{u}_1$ and $\mathbf{u}_2$) from which it can be inferred that the transformed data is no longer independent. So, with respect to the device detection problem, the signal SV component may be mixed with noise SV in such a manner that it could be rendered un-detectable in its mixed form. If the SV's are made independent, then the device characteristics could be present in one of the independent components. If the SV's are given as inputs to the ICA algorithm, as shown in Figure 6.3, the output data is independent, as shown in the scatter plot in Figure 6.6. So, the application of ICA on

the SV's is investigated to see whether or not it helps in the extraction of device characteristics.

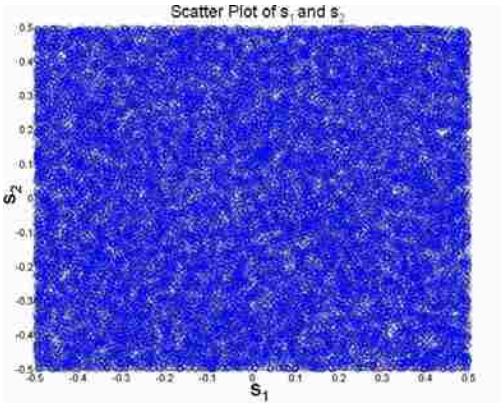


Figure 6.4. The Scatter Plot of Independent Random Variables

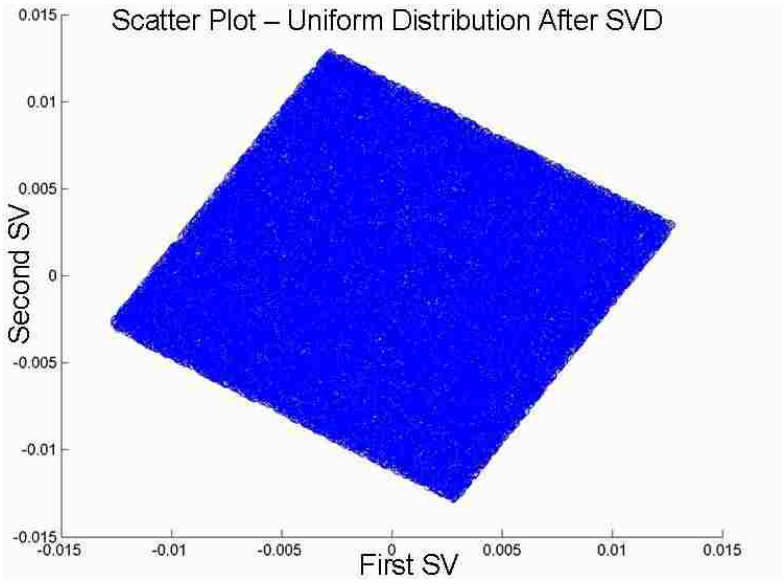


Figure 6.5. The Scatter Plot of SV's

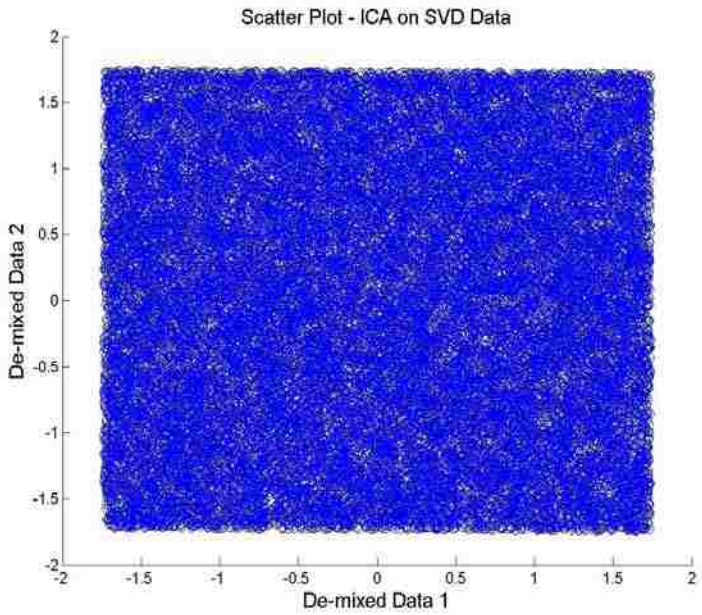


Figure 6.6. The Scatter Plot of De-mixed Data from ICA on SV's

Before the application of ICA on SVD is studied further, one of the simpler cases of detection in which the signal strength is much greater than the noise is explained. In this case, ICA on SVD is not required. Consider the signal measured at 3m distance, in the middle of the hallway. Its TFD using the Hermite window, with the corresponding 1st right singular vector is shown in the shown in Figure 6.7. The 1st right singular vector satisfies the criteria for detection based on the SVD based classification algorithm described in Section 5. The distance between the peaks is approximately 200 samples and the peak from the frequency domain singular vector is near a normalized scale of 50.

6.2. ICA-R ON SVD

The reference to the ICA-R algorithm for time-domain singular vector and frequency domain singular vector are shown in Figures 6.8 and 6.9 respectively. Figure 6.10 [10] shows the basic algorithm flow of the SVD, ICA-R algorithm [10].

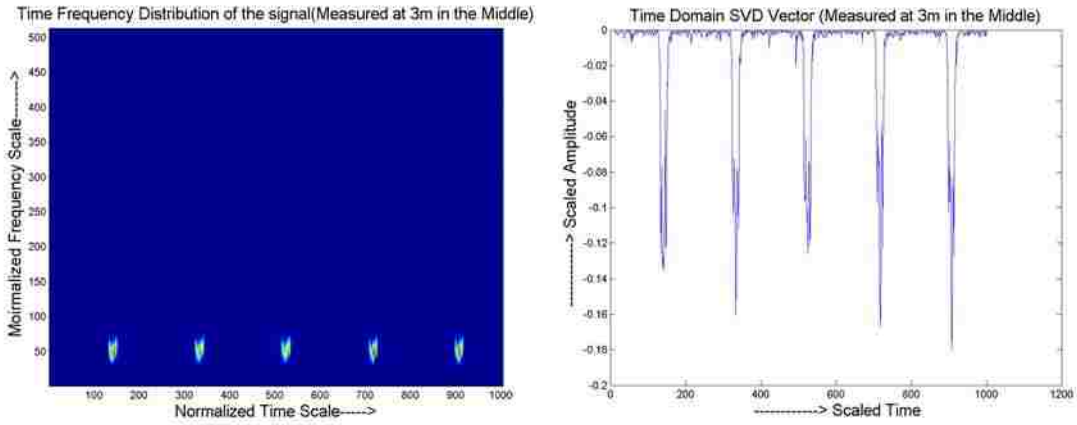


Figure 6.7. TDF & 1st right SV Using Hermite Window

1) As discussed, the SVD of the TFD is computed. The left and right singular vectors are computed. The singular values from the diagonal matrix **D** are used to calculate the number "n" of significant vectors. As illustrated "n" is obtained such that

$$\frac{\sum_{i=1}^{n}\mathbf{D}(i,i)}{\sum_{i=1}^{N}\mathbf{D}(i,i)} \geq 0.98 \text{ where dimension of } \mathbf{D} \text{ is } \mathrm{N}\times\mathrm{N}.$$

2) The input data to the ICA-R algorithm is now $\mathbf{x} = [\mathbf{u}_1\ \mathbf{u}_2\ \ldots.\ \mathbf{u}_n]$, which corresponds to the time domain characteristics [10].

3) The extracted component from the ICA-R is used along with the frequency domain SV by the SVD based detection algorithm for detecting the presence of the device signal. However, when the signal is bandpass filtered the peak of the first singular vector will occur at a normalized frequency scale of 50. Thus, the ICA-R on the SV's of the frequency domain can be skipped as shown in Figure 6.10.

Two cases are used to illustrate the application of ICA-R on SVD's for device detection: 1) ICA-R on SVD with signal pre-processing and 2) ICA-R on SVD without signal pre-processing.

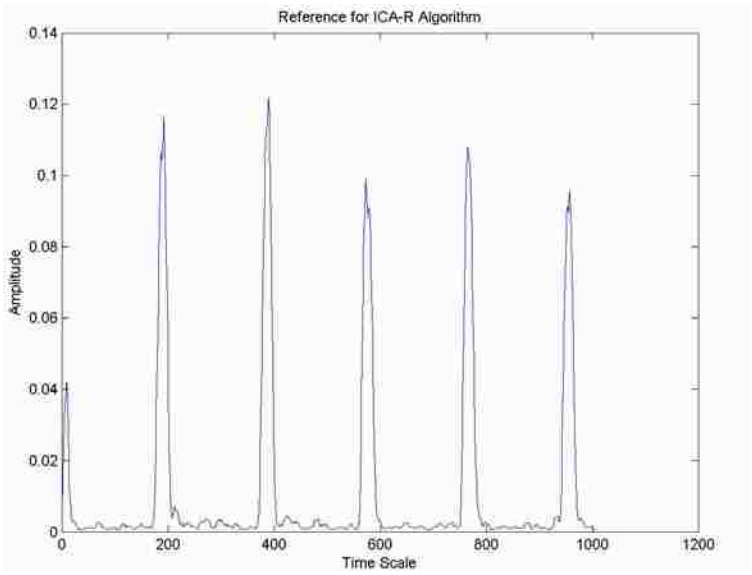


Figure 6.8. Reference ICA-R Algorithm Time Domain

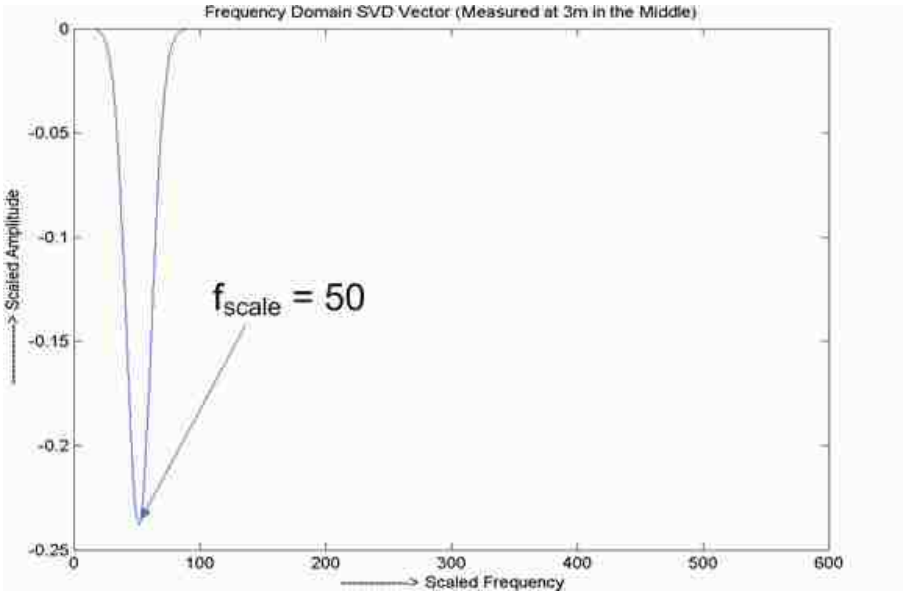


Figure 6.9. Reference ICA-R Algorithm Frequency Domain

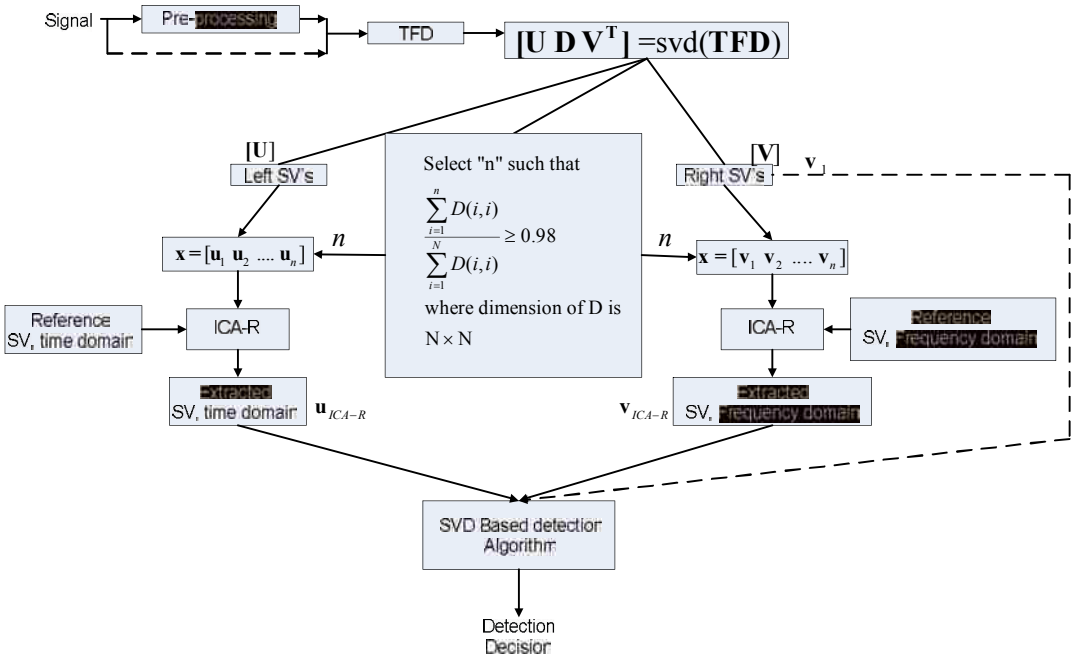


Figure 6.10. Algorithm Flow of ICA-R on SVD

6.2.1. ICA-R on SVD with Signal Pre-processing. In detection, the signal is pre-processed in order to improve the SNR and increase the possibility of detection if the device signal is present. The device signal is shown in Figure 6.11. Adding noise at different power levels to this device signal changes the SNR accordingly. Figure 6.12 shows the detection results from the 1st SV, of the TFD which is the scheme employed for the SVD based detection. Also, the detection results from the extracted component of the ICA-R are shown. From the graph it can be inferred that both schemes of detection are comparable.

One of the disadvantages of the SVD, ICA-R method is its computational complexity. The advantage of using the ICA-R with SVD based detection algorithm is better explained in the next section, in which the signal pre-processing is not conducted and can be generalized for the case when the signal-preprocessing method is not good enough for improving the SNR.

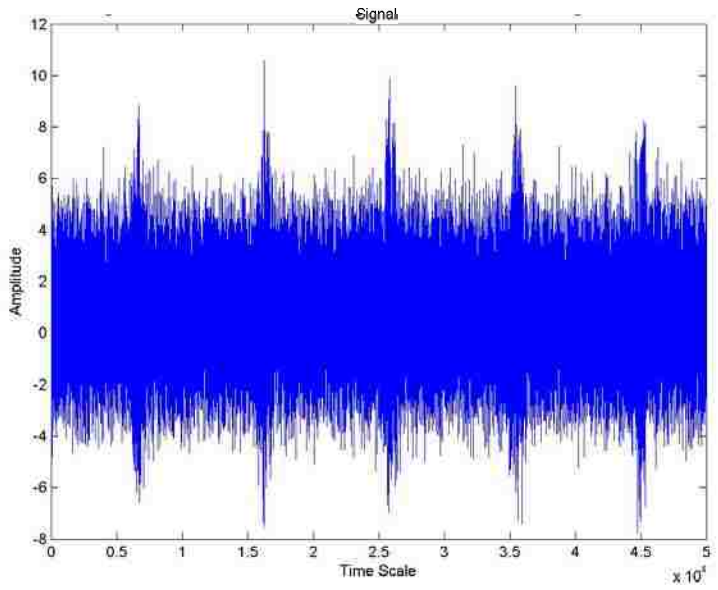


Figure 6.11. Device Signal Varied for Different SNR's for Testing.

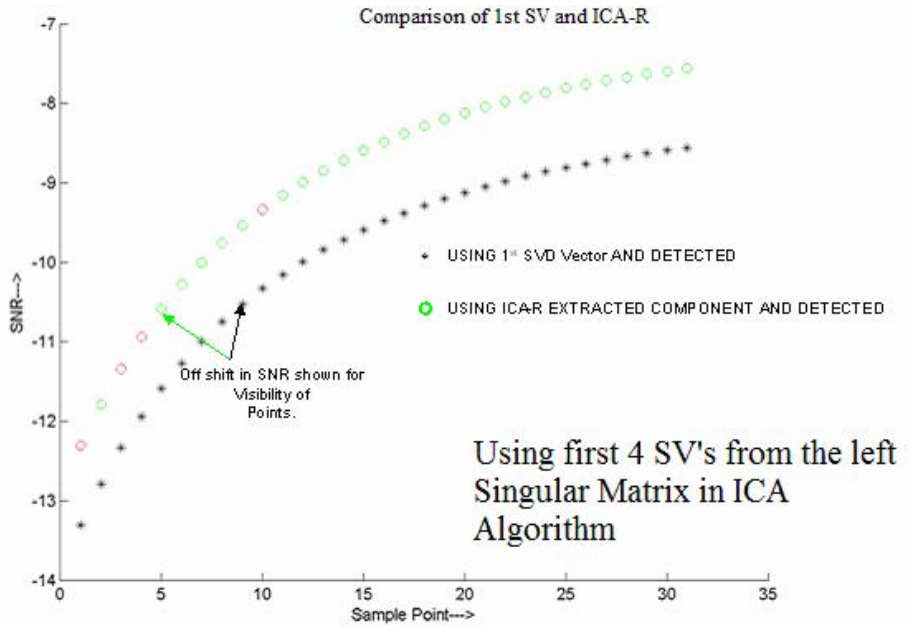


Figure 6.12. SVD Based Detection Using 1st SV, and the ICA-R Extracted Component

6.2.2. Failure Cases for ICA-R SVD with Signal Pre-Processing. It is observed for the high SNR that the ICA-R is not able to extract the component close to the reference device signal component from the SV's. The TFD of such a signal is shown in Figure 6.13. The reference and the extracted component are shown in Figure 6.14.

6.2.3. ICA-R on SVD Without Signal Pre-processing. In detection, pre-processing is done over the signals in order to improve the SNR and increase the possibility of detection if the device signal is present. In some of the detection problems, one may pre-process the signal or the signal pre-processing may not be good enough for improving the SNR. In that cases, the singular vectors obtained may be ordered differently, which implies that one cannot assume that the 1st singular vector obtained can be used for classification. This has been discussed in by Michael Anthony Casey [11] regarding the analysis of sound signals. In a device signal detection scenario, ICA-R over SVD's could prove useful. Consider the time domain signal shown in Figure 6.15. The Hermite window based TFD is shown in Figure 6.16. The TFD shows that the noise power is greater in the lower frequency range. The signal presence is also seen in the TFD, but at a much lower power compared to the noise. Hence, the first SVD vector obtained from the TFD corresponds to the noise presence in the lower frequency range of the signal. The second and third SV's could be used for device signal detection. Therefore, it is difficult to formulate a basis on which a particular singular vector can be chosen for signal detection. One solution to such a problem might be ICA-R. ICA-R helps in the extraction of the component which is close to the reference provided. So, without the knowledge of exactly which singular vector one needs to use for the device detection, ICA-R would extract that particular component close to the reference that one is looking for, which could be used further for detection. One such example is as follows: the 1st singular vector, and the component extracted by ICA-R is shown in Figure 6.17. The distinct peaks that occur in the extracted output signify the presence of the signal, while the 1st SV basically corresponds to the noise presence. Figure 6.18 shows the reference signal and the extracted component using first 4 SV's, i.e., $n = 4$.

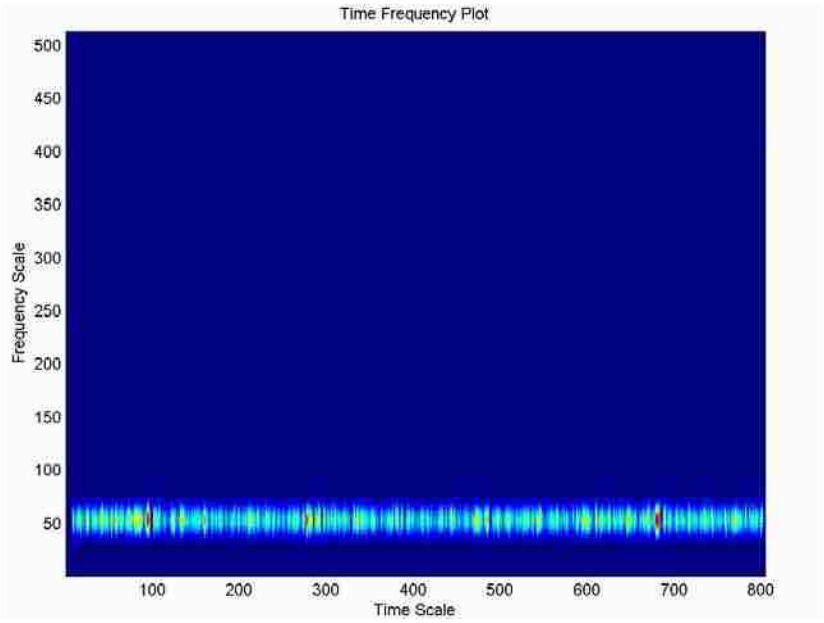


Figure 6.13. TFD of the Noisy Signal

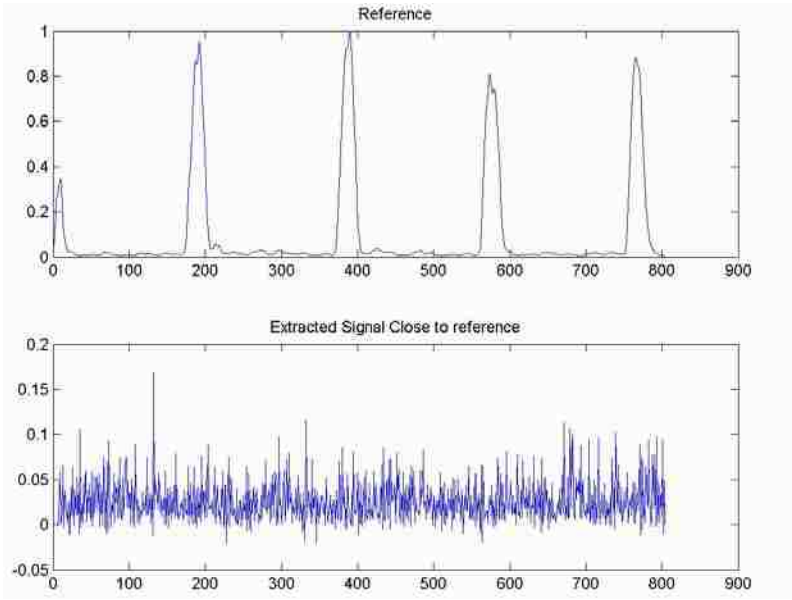


Figure 6.14. Reference and the Extracted Component from ICA-R, Noisy Signal

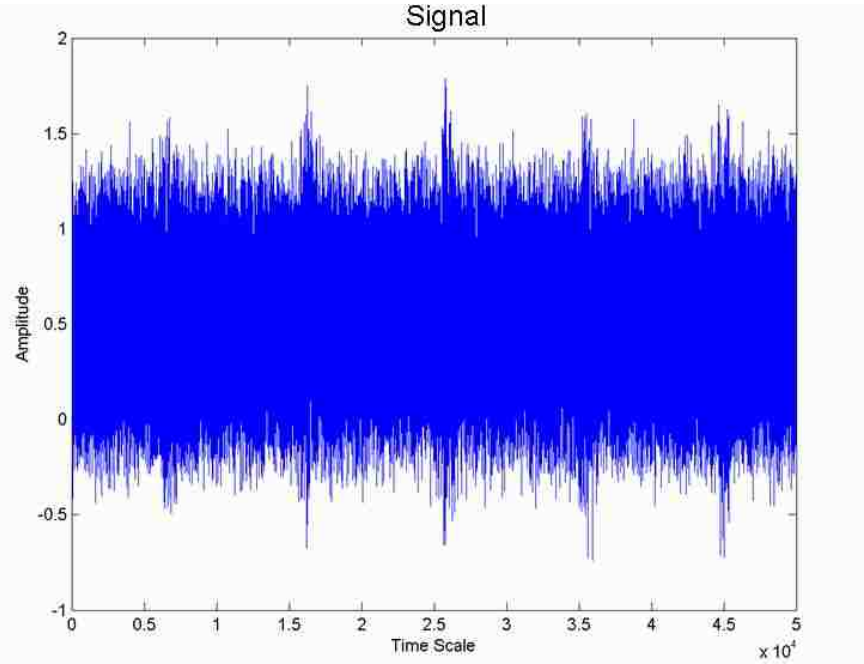


Figure 6.15. Device Signal with Noise for ICA-R SVD

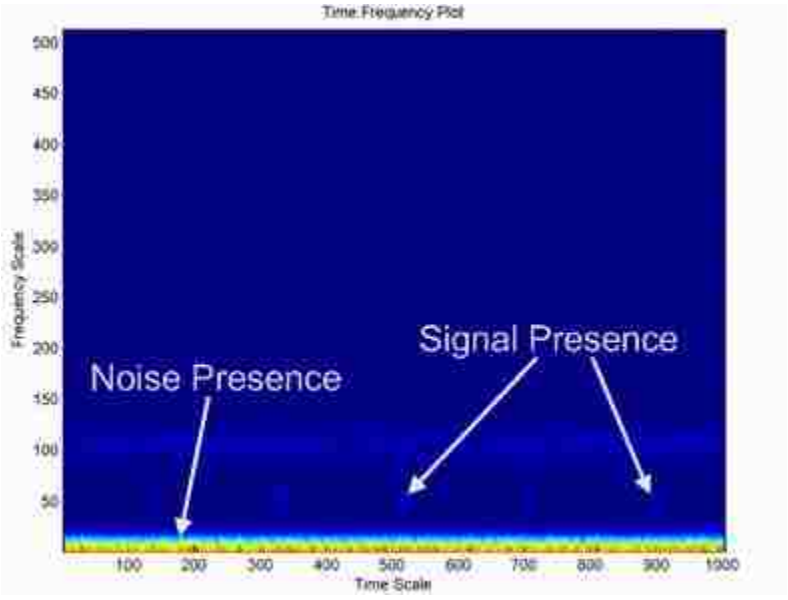


Figure 6.16. TFD of the Signal without Pre-Processing

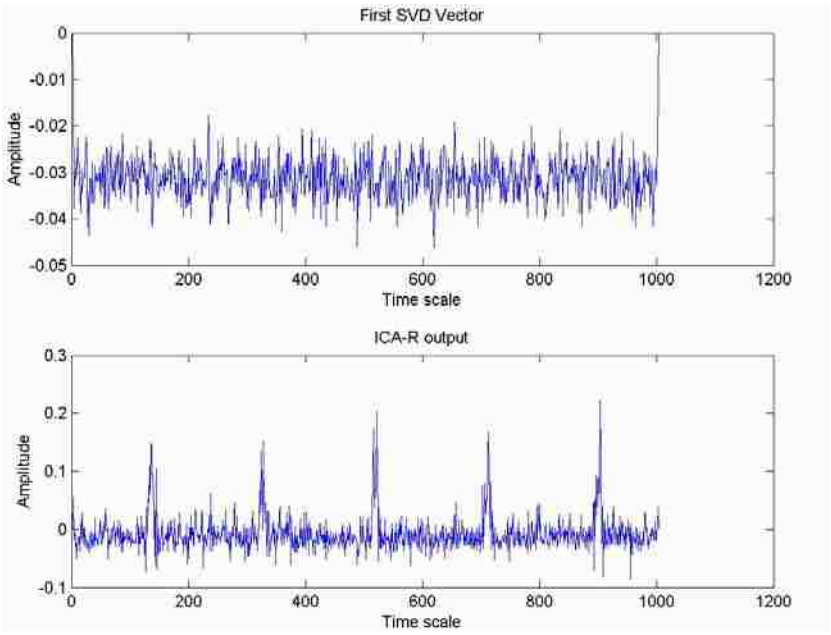


Figure 6.17. 1st SV and the ICA-R Extracted Component

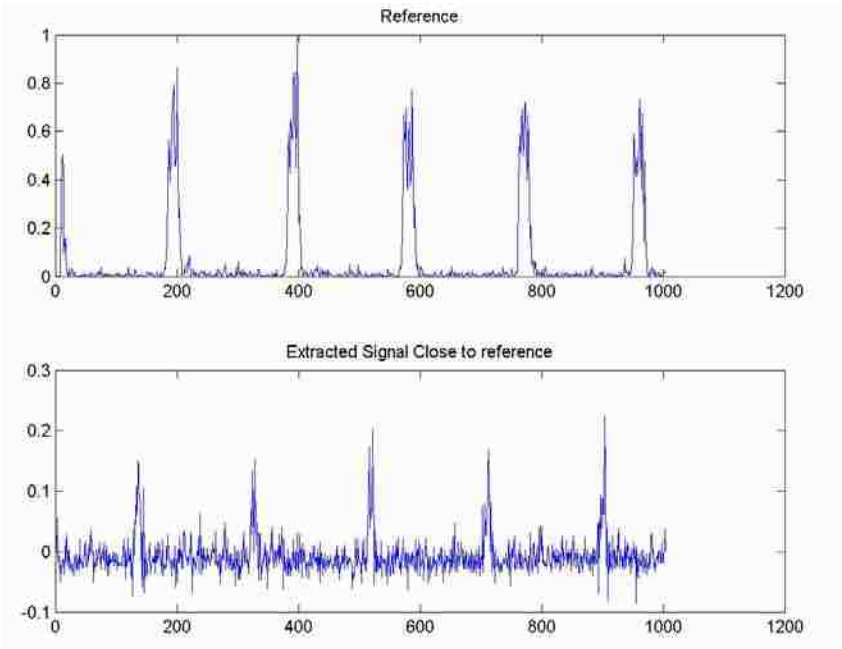


Figure 6.18. ICA-R Reference and the Extracted Component

7. SUMMARY OF RESULTS AND CONCLUSION

7.1. RESULTS AND CONCLUSION

The MF and the SVD detection algorithm were implemented for device signal detection. Since both the algorithms were based on peak detection and distance estimation, false detection probability can be high because of the spurious peaks that may occur. But, by suppression of false maxima and proper adjustments of the thresholds, false detection percentage could be controlled considerably.

With prior knowledge of the signal to be detected, the signal can be appropriately pre-processed to improve detection percentage and reduce the false alarm rate.

The percentage detection for the MF case was found to be 90.8% compared to 92.1% for SVD based detection. The false alarm percentage for SVD based detection was at 11.07%, as compared to a low false alarm rate from MF based detection which was around 3.7%. More tunable parameters were introduced in the case of SVD based detection than in the matched filtering based detection to reduce the false alarm percentage. Hence, it can be concluded that the matched filtering based detection is more reliable and robust than the SVD based detection

The Time-frequency distributions from the STFT, Wigner-Ville, and multiple Hermite windows were also studied. By creating suitable test bench signals was shown that the multiple Hermite window technique helped retain the resolution offered by the Wigner-Ville distribution as well as reduce the noise from cross terms.

One of the drawbacks of the MF was illustrated when the signal had a strong presence of noise or interference, which disallowed the detection of the signal. The concept of ICA and ICA-R was introduced with respect to the device signal detection. Simulation results consistently demonstrated that ICA-R extracted the signal component close to the reference provided, which further underwent MF for detection. ICA-R was also applied to the SV's for device detection. Two cases were studied. One case was when pre-processing of the device signal is done. With different SNR's the detection results offered by the simple SVD based detection were comparable to the ICA-R, SVD based detection. Hence, no depreciation of performance was observed. In another case there is no pre-processing on the signal. In this situation, the 1st SV which is used by the

SVD based detection algorithm does not characterize the signal. So, by using ICA-R on SV's, the device characteristics were extracted from an independent component. In section 6, ICA-R results were shown in which a suitable reference vector, the component from the SV's close to the reference, can be obtained, which could be used further for detection.

7.2. FUTURE WORK

The ICA-R with matched filtering can be used when the signal strength is very low. This can be tested in a real time scenario with two or more receivers. Along similar lines, the ICA-R processing of SVD could be tested for detection in a real time scenario. The SVD based detection algorithm can be tuned further to lower the false detection rate.

BIBLIOGRAPHY

[1] Harry L. Van Tress, "Detection, Estimation and Modulation Theory," 1st ed., John Wiley and Sons, Inc 2001.

[2] Aapo Hyvärinen and Erkki Oja, Tutorial on "Independent Component Analysis: Algorithm and applications," (http://www.cis.hut.fi/aapo/papers/IJCNN99_tutorialweb/), 2000.

[3] Aapo Hyvärinen, "New approximations in differential entropy for independent component analysis and projection pursuit," Advances in Neural Information Processing System. MIT Press, 1998, Pages 273-279.

[4] Wei Lu, Jagath C. Rajaphakse, "ICA with Reference," Neurocomputing, Volume 69, Issues 16-18, October 2006, Pages 2244-2257.

[5] Qiu-Hua, Lina,Yong-Rui, Zhenga,Fu-Liang Yin, Hualou Liang,Vince D.Calhounc "A fast algorithm for one-unit ICA-R," Information Sciences, Volume 177, Issue 5, 1 March 2007, Pages 1265-1275

[6] Mokhtar S. Bazaraa, Hanif D. Sherali, C. M. Shetty, "Nonlinear Programming: Theory and Algorithms", 2nd Edition, John Wiley and Sons, Inc 1993.

[7] L. Cohen, "Time-Frequency distributions-A review," Proceedings, IEEE, Volume 77, 1989, Pages 941-981.

[8] http://ce.caltech.edu/case/tfr/ Wigner-Ville Distributions (July 2006).

[9] Cakrak, F, Loughlin, P.L, "Multiple window non-linear time-varying spectral analysis," Acoustics, Speech, and Signal Processing, 1998. ICASSP '98. Proceedings of the 1998 IEEE International Conference, Volume 4, 12-15 May 1998, Pages 2409 – 2412.

[10] Michael Anthony Casey, PhD Dissertation, "Chapter III: Statistical Basis Decomposition of Time-Frequency Distributions," Massachusetts Institute of Technology Media Laboratory, February, 1998. (http://xenia.media.mit.edu/~mkc/thesis/)

[11] Sangkeun Lee, Hayes, M.H, "Properties of the Singular Value decomposition for efficient data clustering," Signal Processing Letters, IEEE, Nov. 2004,Volume 11, Issue 11, Pages 862- 866.

[12] Groutage D, Bennink D, "Feature sets for nonstationary signals derived from moments of the singular value decomposition of Cohen-Posch (positive Time-Frequency) distributions," Signal Processing, IEEE Transactions, Volume 48, Issue 5, Pages 1498-1503.

[13] Klema, V. Laub, A, "The Singular Value Decomposition: Its computation and some applications," Automatic Control, IEEE Transactions, Volume 25, Issue 2, Apr 1980, Pages 164- 176

[14] Ping Gao, Ee-Chien Chang, Wyse, L, "Blind separation of fetal ECG from single mixture using SVD and ICA," Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on Volume 3, Issue 15-18, Dec. 2003, Pages 1418 – 1422.

VITA

Narayana Ravirala Karthik Sadanandam was born in Mumbai, India, June 5th 1983. He earned his Bachelors in Engineering Degree from Mumbai University's Fr.Conceicao Rodriguez Institute of Technology in July of 2004, majoring in Electronics and Telecommunications. He received his Masters degree in Electrical Engineering from University of Missouri-Rolla in December of 2007. His interests generally include Digital Signal Processing and Digital Image Processing.