

LEAST SQUARES ESTIMATION OF MISSING SENSOR DATA FOR OVERSAMPLED ARRAYS

by

Thuykhanh Le  
A Thesis  
Submitted to the  
Graduate Faculty  
of  
George Mason University  
In Partial fulfillment of  
The Requirements for the Degree  
of  
Master of Science  
Electrical Engineering

Committee:

\_\_\_\_\_ Dr. Kathleen E. Wage, Thesis Director

\_\_\_\_\_ Dr. Jill K. Nelson, Committee Member

\_\_\_\_\_ Dr. Lloyd J. Griffiths, Committee Member

\_\_\_\_\_ Dr. Monson H. Hayes, Chairman, Department  
of Electrical and Computer Engineering

\_\_\_\_\_ Dr. Kenneth S. Ball, Dean for  
Volgenau School of Engineering

Date: \_\_\_\_\_ Summer Semester 2016  
George Mason University  
Fairfax, VA

Least Square Estimation of Missing Sensor Data for Oversampled Arrays

A thesis submitted in partial fulfillment of the requirements for the degree of  
Master of Science at George Mason University

By

Thuykhanh Le  
Bachelor of Science  
George Mason University, 2009

Director: Dr. Kathleen E. Wage, Professor  
Department of Electrical and Computer Engineering

Summer Semester 2016  
George Mason University  
Fairfax, VA

Copyright © 2016 by Thuykhanh Le  
All Rights Reserved

## Acknowledgments

I would like to give a special thank to my advisor, Dr. Kathleen E. Wage, who has guided my education and research.

# Table of Contents

	Page
List of Figures . . . . .	v
Abstract . . . . .	0
1 Introduction . . . . .	1
1.1 Motivation . . . . .	1
1.2 Ideal oversampled bandlimited signal and practical issues . . . . .	3
1.3 Goal . . . . .	4
1.4 Organization . . . . .	5
2 Background and Previous Work . . . . .	6
2.1 Problem Statement . . . . .	6
2.2 Previous work . . . . .	9
2.2.1 Frequency Approach . . . . .	9
2.2.2 Iterative Approach . . . . .	10
2.2.3 FIX Approach . . . . .	13
2.3 Summary . . . . .	14
3 The Least Squares Method and Simulation Results . . . . .	15
3.1 Least squares approach . . . . .	15
3.2 The simulation results . . . . .	17
3.2.1 Simulation Parameters . . . . .	17
3.2.2 Normalized Mean Square Error versus Signal Noise Ratio . . . . .	17
3.2.3 Normalized Mean Square Error versus Number of sensors . . . . .	24
3.3 Summary . . . . .	25
4 The results . . . . .	27
4.1 The SWellEx-96 experiment overview . . . . .	27
4.2 Spatial Spectra of signal in the SWellEx-96 experiment . . . . .	31
4.3 Comparing the performance between algorithms . . . . .	35
4.3.1 Comparison spatial Fourier transform between approaches . . . . .	35
4.3.2 Normalized MSE for some narrow band signals along 75 minutes recording . . . . .	41
4.4 Summary . . . . .	46
5 Conclusion . . . . .	47
Bibliography . . . . .	48

## List of Figures

Figure	Page
1.1 The beam pattern for 22-element uniformly linear arrays . . . . .	2
2.1 The Fourier transform of the truncated signal corrupted by additive white noise . . .	8
3.1 MSE vs. SNR comparison between Least squares approach and other approaches with $N = 22$ , and $r = 0.6$ . . . . .	19
3.2 MSE vs. SNR comparison between Least squares approach and other approaches with $N = 22$ , and $r = 0.5$ . . . . .	20
3.3 Normalized MSE vs. SNR comparison between Least squares approach and other approaches with $N = 22$ , and $r = 0.4$ . . . . .	21
3.4 Normalized MSE vs. SNR comparison between Least squares approach and other approaches with $N = 22$ , and $r = 0.3$ . . . . .	22
3.5 MSE vs. Number of Sensors on each side of a lost element . . . . .	25
4.1 Sound Speed Profile . . . . .	28
4.2 SWellEx-96 Event S5 Range to Sproul from each array. . . . .	29
4.3 Locations of sensors in Vertical Linear Array and Tilted Linear Array . . . . .	30
4.4 The Ideal Bandlimited Signal with a spatial cutoff radian frequency. . . . .	33
4.5 $k - \omega$ beamformer of data recorded in TLA of the narrowband signals having the temporal frequency less than $132Hz$ . . . . .	34
4.6 Comparing the spatial Fourier transform of the narrow band signal $f = 64$ Hz recorded by the Filled array with the missing element array, and recovered element array using the Frequency approach, the Iterative approach, the FIX approach, and the Least squares approach. . . . .	36
4.7 Comparing the spatial Fourier transform of the narrow band signal $f = 67$ Hz recorded by the Filled array with the missing element array, and recovered element array using the Frequency approach, the Iterative approach, the FIX approach, and the Least squares approach. . . . .	37
4.8 Comparing the spatial Fourier transform of the narrow band signal $f = 70$ Hz recorded by the Filled array with the missing element array, and recovered element array using the Frequency approach, the Iterative approach, the FIX approach, and the Least squares approach. . . . .	38

4.9	Comparing the spatial Fourier transform of the narrow band signal $f = 73$ Hz recorded by the Filled array with the missing element array, and recovered element array using the Frequency approach, the Iterative approach, the FIX approach, and the Least squares approach. . . . .	39
4.10	Comparing the spatial Fourier transform of the narrow band signal $f = 76$ Hz recorded by the Filled array with the missing element array, and recovered element array using the Frequency approach, the Iterative approach, the FIX approach, and the Least squares approach. . . . .	40
4.11	Normalized MSE for $f=64$ (Hz) along 75 minutes recording. . . . .	42
4.12	Normalized MSE for $f=67$ (Hz) along 75 minutes recording. . . . .	43
4.13	Normalized MSE for $f=70$ (Hz) along 75 minutes recording. . . . .	44
4.14	Normalized MSE for $f=73$ (Hz) along 75 minutes recording. . . . .	45
4.15	Normalized MSE for $f=76$ (Hz) along 75 minutes recording. . . . .	46

# Abstract

LEAST SQUARE ESTIMATION OF MISSING SENSOR DATA FOR OVERSAMPLED ARRAYS

Thuykhanh Le

George Mason University, 2016

Thesis Director: Dr. Kathleen E. Wage

The problem of recovering samples lost from time series or sensor data is important in signal processing. When the underlying signal is known to be bandlimited, and the sample rate is higher than the Nyquist rate, the samples are dependent. In this case a missing sample or samples can be recovered from the remaining samples. In the absence of noise, the accuracy of the sample estimates depends on the degree of oversampling and the total number of good samples available. In previous work, researchers often assumed that large numbers of high quality (high signal-to-noise ratio) samples were available. This assumption may not be valid in practice. In practice the number of samples is finite and the signal is corrupted by noise. The truncation and the noise will result in errors in the sample estimates. This thesis investigates a least squares solution to the problem, and uses the data from SwellEx-96 experiment to evaluate several approaches, including the least squares approach.

# Chapter 1: Introduction

## 1.1 Motivation

In array processing, it is common to lose data from one or more sensors. Deleting elements affects the performance of a linear array in several ways. The most significant effect is an increase in the sidelobe level of the array beampattern, which may mask important low level signals. The width of the mainlobe may also be increased if the array aperture is reduced by the loss of the sensors. This thesis focuses on the simplest problem of recovering a single missing sample from the remaining samples. Figure 1.1 shows the beampattern of linear array with 22 equally spaced elements and the beampattern of the same array with the 15<sup>th</sup> element missing. The beampattern of the array with the missing element array has significantly higher sidelobes at same angles compared to the sidelobe level of the filled array. In this case, the missing element should be recovered in order to improve performance.

SWellEx-96 is an example of an ocean experiment affected by sensor failure. This experiment included two arrays: the Vertical Linear Array (VLA) and the Tilted Linear Array (TLA). Both of these arrays had 22 equally spaced sensors spanning a 120 m aperture. The data was recorded from all the sensors of the TLA whereas the data recorded from the 15<sup>th</sup> sensor of the VLA was corrupted. Since the TLA data contains all 22 sensors, it provides a useful data set for testing algorithms for recovery of missing samples. In this experiment, the scenario for studying the performance of estimate of a lost sample can be set up as follows: the data from the desired location of the TLA is set to zero and the algorithm is applied to recover the deleted sample. The performance of the estimator will be analyzed by comparing the mean square error (MSE) and the conventional beamformer scanned response between using the zero-value at the desired location and using the estimated result.

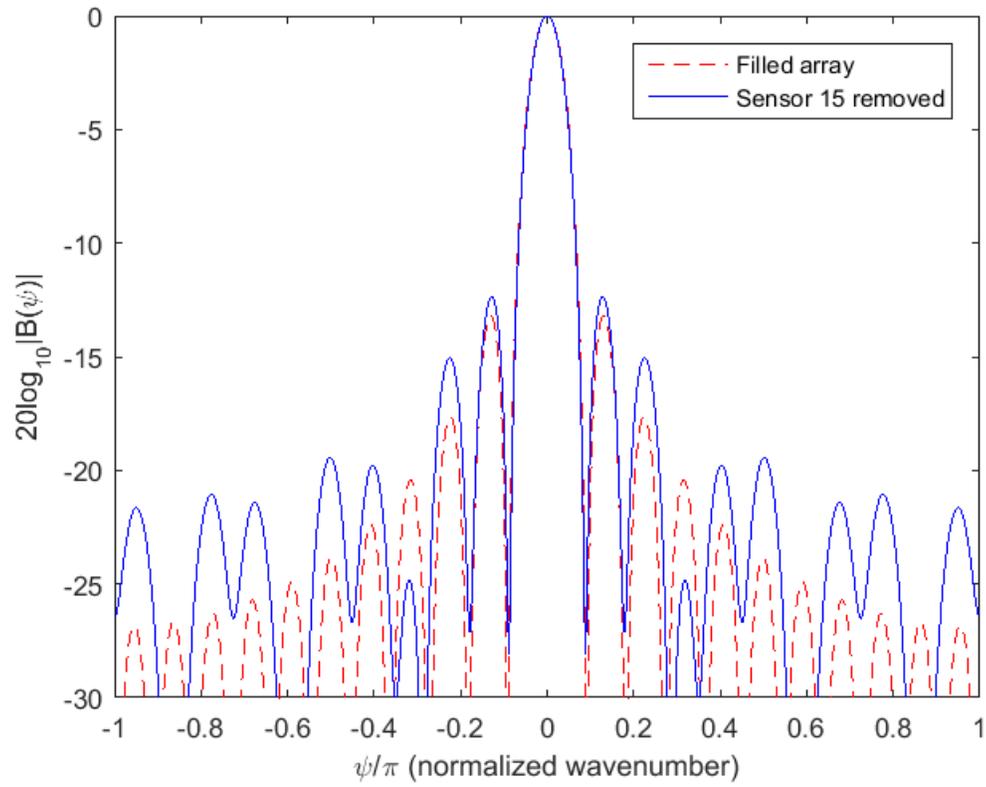


Figure 1.1: The dashed red line is the beam pattern for a uniform linear 22-element array with half wavelength spacing. The solid blue line is the beam pattern for uniform linear 22-element array with half wavelength spacing with the 15<sup>th</sup> sensor removed. A rectangular window is used for both arrays.

## 1.2 Ideal oversampled bandlimited signal and practical issues

A signal is called bandlimited if its Fourier transform is equal to zero above a specified frequency  $f_{max}$ . According to Papoulis [1], the samples are mathematically independent when the signal is sampled at its Nyquist rate which is twice the highest frequency  $f_{max}$  present in the signal. In contrast, when the sampling rate is higher than the Nyquist rate, the samples become mathematically dependent. In that case, a finite number of missing samples can be recovered from the remaining ones. It is well known that a finite number of missing samples can be recovered when the bandlimited signal is oversampled [2–9]. The ideal bandlimited signal requires an infinite number of samples to have Fourier transform (or power spectral density in the case of stochastic signal) equal to zero above a specified frequency. In practice, the number of samples or the number of sensors recorded the data are definitely finite, and the signal is always corrupted by noise. The truncation and the additive noise will result in error in estimating of the missing samples. The recovery of lost samples from an oversampled bandlimited signal has been approached in many ways. The paragraphs below describe three techniques for recovering lost samples.

The first approach is the Frequency approach. In [7] Marks restored a lost sample at the origin using the Fourier transform of the remaining samples at the specified frequency  $\psi = \pi$ . This method is used if only one sample is lost and the location of lost sample is known.

The second approach is an iterative approach. Marks [7] presented this approach to recover a finite number of lost samples rather than just a single missing sample. The first iteration of this approach consists of four steps: the first step is to compare the Fourier transform of the observed signal in which the defective samples are zeroed; the second step is apply lowpass filter with cut-off frequency  $\psi_c = \psi_{max}$  wherein  $\psi_{max}$  is the maximum radian frequency of the bandlimited signal; the third step is to take inverse Fourier transform; and the last step is to replace the values at missing locations by values from the inverse transform. Each following iteration of the approach is the same as the first iteration except the first step. The first step of each following iteration is Fourier transform of observed signal in which the values at defective locations come from the previous iteration. Several authors have used this method to reconstruct the bandlimited signal with a finite number of missing samples [2–8].

The third approach is denoted Fourier Interpolation of eXcitations( FIX). Stockhausen and Farrell [9] developed this method to reduce the performance degradation of a conventional beamformer due to missing elements in a uniform linear array. This approach is similar to the iterative approach described in the previous paragraph. In the iterative approach, a lowpass filter is used to eliminate any beam having frequency higher than  $\psi_{max}$ . In the FIX approach, the average power is used as a threshold to eliminate any beam having power lower than the average power. Stockhausen and Farrell note that the process can be repeated to reduce the variance of the beam power estimate.

### 1.3 Goal

In the frequency approach, Marks used only one value of the Fourier transform on the interval  $\psi_{max} < |\psi| < \pi$  to recover the missing sample at origin. That approach assumes that the Fourier transform of the original band limited signal at the designated frequency is equal to zero. In practice, the Fourier transform of the original bandlimited signal is not equal to zero at each  $\psi$  on the interval  $\psi_{max} < |\psi| < \pi$  for two main reasons. First, the signal is truncated due to the finite measurement aperture. A truncated bandlimited signal is no longer perfectly bandlimited. Second, most practical scenarios include additive measurement noise that is not bandlimited. Since the signal is not perfectly bandlimited, the frequency approach has higher errors. To reduce these errors, this thesis considers a least squares approach that uses multiple measurements outside the band of the signal to reduce the error variance. Let  $L$  be the number of multiple measurements available outside the the band of signal. The missing sample can be recovered by solving an overdetermined system of linear equations using least squares. This approach reduces the variance of the error by a factor of  $L$  compared to variance of error for the frequency approach.

The goal of the thesis is to investigate the least squares approach for estimating narrowband data for a missing sensor and comparing its performance to other standard approaches using simulated and experimental data.

## 1.4 Organization

The thesis is organized as follows. Chapter 2 presents the previous work on estimating data for a lost sensor. Chapter 3 presents the least squares approach and analyzes its performance using synthetic data. These results are compared to the three other approaches discussed above: the frequency approach, the iterative approach, and the FIX approach. Chapter 4 analyzes the performance of the least squares approach by comparing its performance with the performance of other three approaches using the real data in SwellEx-96. Chapter 5 concludes the thesis.

## Chapter 2: Background and Previous Work

This chapter defines the problem of recovering a missing sample from a truncated bandlimited signal in a noisy environment and reviews the previous work on the topic. Section 2.1 states the problem and Section 2.2 describes the previous approaches mentioned in Chapter 1. [10]

### 2.1 Problem Statement

Let  $x(t)$  be a bandlimited signal having a maximum frequency  $f_{max}$ .  $x[n]$  is samples of  $x(t)$  with the sampling rate  $f_s$ , which is greater than the Nyquist rate  $f_{NY} = 2f_{max}$ . The sampling rate parameter  $r$  is defined by

$$r \equiv \frac{f_{NY}}{f_s}. \quad (2.1)$$

The Fourier transform of the signal  $x[n]$  is  $X(e^{j\psi})$ , where  $\psi$  is radian frequency. For an ideal bandlimited signal that is oversampled that

$$X(e^{j\psi}) = 0; \quad \psi_{max} < |\psi| \leq \pi \quad (2.2)$$

The discrete-time radian frequency  $\psi_{max}$ , and sampling rate parameter  $r$  are related by

$$\psi_{max} = \frac{\Psi_{max}}{f_s} = \frac{2\pi f_{max}}{f_s} = r\pi \quad (2.3)$$

The equation 2.3 shows that the maximum discrete-time radian frequency  $\psi_{max}$  depends on the degree of oversampling  $r$ . As stated in Chapter 1, the finite missing samples can be recovered from the remaining samples in the oversampling case. To see how a single missing sample can be recovered from an ideal bandlimited signal, consider the following simple example. The sequence  $x[n]$  has one

sample at  $n = n_o$ , that is equal to  $A$ , i.e.,

$$x[n_o] = A\delta[n - n_o]. \quad (2.4)$$

Let  $\hat{x}[n]$  denote the signal with the missing sample at  $n_o$ , i.e.,

$$\hat{x}[n] = x[n] - A\delta[n - n_o]. \quad (2.5)$$

Taking the Fourier transform of both sides of Eq. 2.5 yields

$$\widehat{X}(e^{j\psi}) = X(e^{j\psi}) - Ae^{-j\psi n_o}. \quad (2.6)$$

Thus,

$$A = X(e^{j\psi})e^{j\psi n_o} - \widehat{X}(e^{j\psi})e^{j\psi n_o} \quad (2.7)$$

where  $X(e^{j\psi})$  is Fourier transform of the original signal and  $\widehat{X}(e^{j\psi})$  is Fourier transform of the original signal with the missing sample at  $n_o$ . Since  $X(e^{j\psi})$  is 0 on the interval  $\psi_{max} < |\psi| \leq \pi$ , we can determine  $A$  from

$$A = -\widehat{X}(e^{j\psi})e^{j\psi n_o} \quad \psi_{max} < |\psi| \leq \pi. \quad (2.8)$$

Therefore, the lost sample  $A$  at  $n_o$  can be completely recovered from the remaining samples using Equ. 2.8 with the assumption that  $X(e^{j\psi})$  is 0 on the interval  $\psi_{max} < |\psi| \leq \pi$ . This approach requires the original signal is an ideal bandlimited signal, which requires an infinite number of noise-free samples. These assumptions are rarely satisfied in practice. Any practical array has a finite number of sensors and noise is always present. The truncation and additive noise will result in error in estimating the missing samples.

Figure 2.1 shows the Fourier transform of the truncated signal imposed by additive white noise. The solid red line represents the ideal bandlimited signal, the dashed blue line represents the truncated bandlimited signal in which the signal is recorded by a finite number of sensors, and the dashdot black line represents the truncated bandlimited signal corrupted by additive noise. From

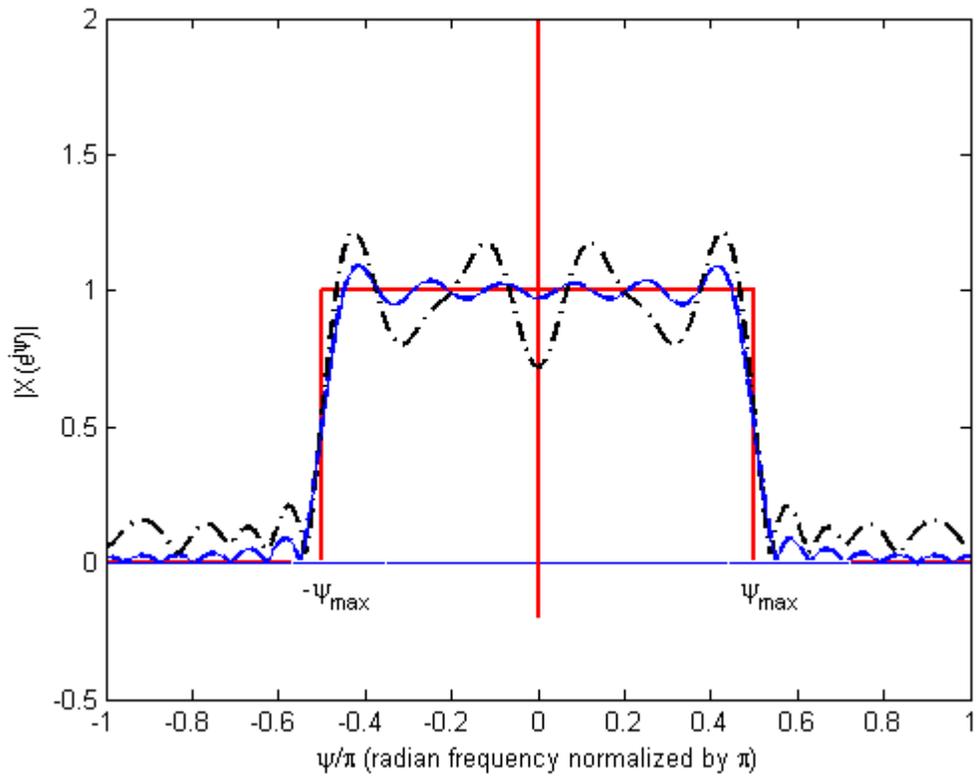


Figure 2.1: The solid red line is the Fourier transform of a ideal bandlimited signal with maximum frequency  $\psi_{max} = 0.5\pi$ . The dashed blue line is the Fourier transform of a bandlimited signal with maximum frequency  $\psi_{max} = 0.5\pi$  recorded by 22 sensors. The dashdot black line is the Fourier transform of a bandlimited signal with maximum frequency  $\psi_{max} = 0.5\pi$  recorded by 22 sensors and corrupted by additive noise.

the figure, one can see that the Fourier transform of the signal is no longer equal to zero outside the band of the signal. It leads to error in estimating the lost sample for the bandlimited signal. The main question is how the truncation and the noise affect the accuracy of the estimation. The following section presents previous approaches to missing sample estimation and analyzes the influence of truncation and noise on the accuracy of the estimate for each approach.

## 2.2 Previous work

The technique of recovering missing data from one or more sensors has been implemented by several of groups [2-9]. This section presents a brief overview of work that is relevant to this thesis.

### 2.2.1 Frequency Approach

The first method proposed for recovering one lost sample is based on the characteristic of the oversampled bandlimited signal in a frequency domain. Marks described this approach in detail [7]. Recall that the missing value  $A$  at the location  $n = n_o$  of the bandlimited signal  $x[n]$  can be completely recovered by implementing Eq. 2.8 at any value of  $\psi$  on the interval  $\psi_{max} < |\psi| \leq \pi$ , where  $\hat{X}(e^{j\psi})$  is Fourier transform of the remaining known samples, i.e.,

$$\hat{X}(e^{j\psi}) = \sum_{\substack{n=-\infty \\ n \neq n_o}}^{\infty} x[n]e^{-j\psi n}. \quad (2.9)$$

Marks [7] utilized this approach to recover the one lost sample at origin  $n_o = 0$  by claiming that an appropriate point  $\psi$  in the interval  $\psi_{max} < |\psi| \leq \pi$  is  $\psi = \pi$ . Hence,

$$x[0] = -\hat{X}(e^{j\psi})e^{j\psi n_o}|_{\psi=\pi, n_o=0}, \quad (2.10)$$

$$= -\sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} x[n]e^{-j\pi n}, \quad (2.11)$$

$$= -\sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} x[n](-1)^n \quad (2.12)$$

This approach is useful when only one sample of the bandlimited signal  $x[n]$  is missing.

## 2.2.2 Iterative Approach

The second method proposed is a iterative approach that can be used to recover a finite number of lost samples rather than just a single missing sample. This section describes steps required for this method.

Shannon [11] stated in Sampling Theorem that “if a function  $x(t)$  contains no frequencies higher than  $f_{max}$ , it is completely determined by giving it ordinates at a series of points spaced  $\frac{1}{2f_{max}}$  apart.”

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \text{sinc}(2f_{max}t - n), \quad (2.13)$$

where  $\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$ .

Using Eq. 2.13, Marks constructed the iterative approach as follows.

Define  $\mathcal{M}$  as the finite set of integers corresponding to the locations of the  $m$  lost samples, and  $x_M(t)$  is the result of the  $M^{th}$  iteration.

$$x_M(t) = \sum_{n=-\infty}^{\infty} x_M[n] \text{sinc}(2f_{max}t - n)$$

- The Fourier transform of  $x_M(t)$  is defined by

$$X_M(j\Psi) = \frac{1}{2\pi\Psi_s} \sum_{n=-\infty}^{\infty} x_M[n] e^{-j\psi n} \Pi\left(\frac{\Psi}{2\Psi_s}\right)$$

where

$$\Pi(\xi) = \begin{cases} 1 & \text{if } |\xi| < \frac{1}{2}, \\ 0 & \text{if } |\xi| > \frac{1}{2}, \\ \frac{1}{2} & \text{if } |\xi| = \frac{1}{2}. \end{cases}$$

Since  $x_M(t)$  has bandwidth  $f_{max}$ , the lowpass filter with cut-off frequency  $f_c = f_{max}$  is used.

This leads to step 2.

- Form the function

$$H_M(j\Psi) = X_M(j\Psi)\Pi\left(\frac{\Psi}{2\Psi_{max}}\right)$$

- The inverse Fourier transform of  $H_M(j\Psi)$  is determined by

$$h_M(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H_M(j\Psi)e^{-j\Psi t} dt.$$

The signal  $h_M(t)$  is sampled at a rate of  $f_s$ .

- The lost samples at  $\mathcal{M}$  locations are replaced by  $h_M[n]|_{n \in \mathcal{M}}$ .

$$x_{M+1} = \begin{cases} h_M[n]; & n \in \mathcal{M} \\ x_M[n]; & n \notin \mathcal{M} \end{cases}$$

Marks stated that in the limit

$$\lim_{M \rightarrow \infty} x_M[n] = x[n] \tag{2.14}$$

and the missing samples are regained. He proved that the final values of estimate by using the iterative approach is defined by

$$\mathbf{x} = [\mathbf{I} - \mathbf{S}]^{-1}\mathbf{k}, \tag{2.15}$$

where  $\mathbf{x}$  is an  $m \times 1$  column vector of unknown samples,  $\mathbf{I}$  and  $\mathbf{S}$  denote the identity matrix and Toeplitz matrix respectively determined by locations of lost samples and the sampling rate parameter  $r$ . The  $pq^{th}$  component of  $\mathbf{S}$  is defined as

$$S_{pq} = rsinc[r(p - q)]|_{(p,q) \in (\mathcal{M} \times \mathcal{M})}, \tag{2.16}$$

where  $\text{sinc}(x) = \sin(\pi x)/(\pi x)$ , and  $\mathbf{k}$  is an  $m \times 1$  column vector with elements given by

$$k_i = r \sum_{n \notin \mathcal{M}} x[n] \text{sinc}[r(i - n)]|_{i \in \mathcal{M}}, \quad (2.17)$$

where  $i$  is the location of the lost samples.

Ferreira [8] also proved Eq. 2.15 by implementing the generalized interpolation function

$$x(t) = r \sum_{n=-\infty}^{\infty} x[n] \text{sinc}(2f_{max}t - rn) \quad (2.18)$$

at known and unknown samples.

Implementing Eq. 2.15 for one lost sample at origin,

$$x[0] = \frac{r}{1-r} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} x[n] \text{sinc}(rn) \quad (2.19)$$

Marks asserted in [7] that convergence in Eq. 2.19 is better than in Eq. 2.12 due to the  $1/n$  factor from the sinc term. The FIX approach, on the other hand, does not require knowledge of  $r$ .

Implementing Eq. 2.15 for one lost sample at arbitrary location,

$$x[n_o] = \frac{r}{1-r} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} x[n] \text{sinc}[r(n_o - n)] \quad (2.20)$$

One can use Eq. 2.20 to recover one lost sample at  $n_o$ . Marks [7] considered the truncation effects in estimation by calculating the mean squares error for the signal having  $N$  samples from each side, the lost sample is at origin, the signal is corrupted by real additive zero mean wide sense stationary noise  $\zeta[n]$ , and the sampled noise  $\zeta[n]$  is uncorrelated with the sampled signal  $x[n]$ . Then the single lost sample  $x[0]$  is recovered by

$$x_{est}[0] = \frac{r}{1-r} \sum_{\substack{n=-N \\ n \neq 0}}^N (x[n] + \zeta[n]) \text{sinc}(rn) \quad (2.21)$$

The mean-square error of this estimate normalized by the noise variance is

$$\begin{aligned} \epsilon = & R_x(0) - \frac{4r}{1-r} \sum_{n=1}^N R_x(n) \text{sinc}(rn) \\ & + 2 \left( \frac{r}{1-r} \right)^2 \left[ \sum_{n=1}^N \text{sinc}^2(rn) \right. \\ & \left. + \sum_{n=1}^N \sum_{m=1}^N \left[ R_x(n-m) + R_x(n+m) \right] \text{sinc}(rn) \text{sinc}(rm) \right] \end{aligned} \quad (2.22)$$

Marks also proved that

$$\lim_{N \rightarrow \infty} \epsilon = \frac{r}{1-r} \quad (2.23)$$

Eq. 2.23 shows that as  $N \rightarrow \infty$  the normalized error depends on the sampling rate parameter  $r$  only.

### 2.2.3 FIX Approach

Stockhausen and Farrell developed the FIX (Fourier Interpolation of eXcitations) approach to recover the missing samples for a bandlimited signal. The FIX technique uses all information from the remaining samples. The authors used this approach to recover 5 missing elements at the 7<sup>th</sup>, 8<sup>th</sup>, 15<sup>th</sup>, 16<sup>th</sup>, and 25<sup>th</sup> sensors of a 30-element array. The field consisted of 4 components: one strong planewave signal near endfire, one small planewave signal 15° aft of broadside, three-dimensionally isotropic noise, and white noise. The small planewave signal and other two noises are 30dB, 20dB, and 50dB weaker than strong signal respectively. In the FIX approach, the defective elements are set to zero in the first iteration process, and the signal  $x[n]$  is windowed by a half cosine window

$$x_w[n] = x[n] \cdot w[n]. \quad (2.24)$$

A fast Fourier transform is applied for windowed signal. The normalization which follows the transform process has two factors. The first one is  $\frac{1}{N\Delta f\Delta t}$  which is required by the Fourier transform process. The second factor is defined by

$$\sqrt{\frac{\sum_{i=1}^N w_i^2}{\sum_{i=1}^N \varepsilon_i w_i^2}} \quad (2.25)$$

where  $\varepsilon_i$  is 1 if sensor is present and 0 if it is missing, and  $w_i$  is the weight from half cosine window applied to data. This second factor is designed to keep the power of obtained data unchanged. In the next step, each value of signal in frequency domain is converted to power and averaged. Any beam having power less than or equal to the averaged power is set to zero. The inverse transform is used for all remaining transform value, and inverse of the half cosine window is applied as well. The values at the missing locations is substituted by corresponding estimated values. The authors have stated that the entire process may be repeated for better estimation. The Frequency approach and the Iterative approach are utilized for oversampled signals. However, the FIX approach does not require an oversampled signal.

## 2.3 Summary

Previous results show that it is possible to recover a finite number of missing samples deleted from an oversampled bandlimited signal. The frequency approach uses the value of Fourier transform of the remaining samples at  $\psi = \pi$  to recover a single missing sample. The iterative approach uses all values of the Fourier transform of the remaining samples combined with a lowpass filter to recover a finite number of missing samples. The FIX approach is similar to the iterative approach, except the FIX approach uses an adaptive filter instead of a lowpass filter. For practical applications there are errors due to truncation and noise. When the noise is low enough, the truncation error dominates. At higher noise, the noise error dominates. The main question is how to further reduce the estimate error for oversampled arrays. Chapter 3 presents an alternative approach.

## Chapter 3: The Least Squares Method and Simulation Results

The main goals of this chapter are to present the least squares approach for recovering a missing sample and to compare the performance of the least squares approach with the previous approaches. The first section presents the mathematical method of the least squares approach. The second section uses synthetic data to illustrate its performance. The third section summarizes the chapter.

### 3.1 Least squares approach

Recall that the Fourier transform of the original ideal bandlimited signal  $X(e^{j\psi}) = 0$  on the interval  $\psi_{max} < |\psi| \leq \pi$ , so the lost sample  $x[n_o]$  will be recovered by Eq. 2.8. Suppose that Eq. 2.8 is implemented for  $L$  values of  $\psi$  between  $\psi_{max}$  and  $\pi$ , i.e.,

$$\begin{aligned}\hat{X}(e^{j\psi})|_{\psi=\psi_1} &= -Ae^{-j\psi_1 n_o} \\ \hat{X}(e^{j\psi})|_{\psi=\psi_2} &= -Ae^{-j\psi_2 n_o} \\ &\vdots \\ \hat{X}(e^{j\psi})|_{\psi=\psi_L} &= -Ae^{-j\psi_L n_o}\end{aligned}$$

These equations can be written more concisely in matrix form as

$$\mathbf{y} = \mathbf{h}A \tag{3.1}$$

where  $\mathbf{y}$  is an  $L \times 1$  column vector:

$$\mathbf{y} = \begin{bmatrix} \hat{X}(e^{j\psi_1}) \\ \hat{X}(e^{j\psi_2}) \\ \vdots \\ \hat{X}(e^{j\psi_L}) \end{bmatrix}.$$

$\mathbf{h}$  is also an  $L \times 1$  column vector:

$$\mathbf{h} = \begin{bmatrix} -e^{-j\psi_1 n_o} \\ -e^{-j\psi_2 n_o} \\ \vdots \\ -e^{-j\psi_L n_o} \end{bmatrix} = \begin{bmatrix} e^{-j(\psi_1 n_o + \pi)} \\ e^{-j(\psi_2 n_o + \pi)} \\ \vdots \\ e^{-j(\psi_L n_o + \pi)} \end{bmatrix}.$$

The goal is to estimate  $A$  by solving the overdetermined system of linear equations in (3.1). Let  $\hat{A}$  denote the estimated value of  $A$  and consider a value of  $\hat{A}$  that minimizes the squared differences

$$\min \|\mathbf{y} - \mathbf{h}\hat{A}\|^2. \quad (3.2)$$

This minimization problem has a unique solution

$$\hat{A} = \text{inv}(\mathbf{h}^H \mathbf{h}) \mathbf{h}^H \mathbf{y} = \frac{1}{L} \mathbf{h}^H \mathbf{y} \quad (3.3)$$

where  $\mathbf{h}^H$  denotes the Hermitian transpose of  $\mathbf{h}$ . In this least squares approach, we use  $L$  equations to search for one solution, each equation utilized one value Fourier transform of the original signal on the interval  $\psi_{max} < |\psi| \leq \pi$ . The variance of error should be reduced by a factor of  $L$  compared to the solution of Eq. 2.8 assuming these  $L$  equations are linearly independent. Jenkins et al. [12] stated that Discrete Fourier Transform (DFT) of  $N$ -samples signal at  $\psi_k = \frac{2\pi k}{Nfft}$  becomes uncorrelated if  $Nfft \leq N$  where  $Nfft$  is a number of FFT. In this thesis,  $Nfft$  is equal to the size of array. Truncation

of a bandlimited signal is considered as multiplying the bandlimited signal with a window. This leads to a transition band in the frequency domain. This transition band is considered in taking appropriate Fourier transform values to solve Eq. LSeStimation. Width of transition band relates to the width of the mainlobe of the window, and this mainlobe width is determined by length of window.

## 3.2 The simulation results

This section describes how the simulation data is generated for the examples in this thesis.

### 3.2.1 Simulation Parameters

This section uses simulations to compare the performance of the three algorithms from previous work with the performance of the least squares estimator. The synthetic bandlimited signal is generated as follows: white noise is run through a lowpass filter with a cutoff frequency  $\psi_c$ . The output is a bandlimited signal with the maximum radian frequency which is equal to the cutoff frequency of the lowpass filter  $\psi_{max} = \psi_c$ . The cutoff frequency in the lowpass filter governs the maximum radian frequency in the bandlimited signal. The length of the input is 100101, and the length of the filter is 101. Thus, the output is a bandlimited signal with the length is 100201. The bandlimited signal used to investigate the performance of estimation of the Frequency, Iterative, FIX, least squares approach is selected from center of synthesized bandlimited signal. The reason to select the samples at the center is to avoid transient samples in the synthesized bandlimited signal. The SNR is defined by

$$SNR = 10 \log_{10} \left( \frac{P_{sig}}{P_{noise}} \right) \quad (3.4)$$

where  $P_{sig}$  is power of the signal, and  $P_{noise}$  is power of noise.

### 3.2.2 Normalized Mean Square Error versus Signal Noise Ratio

This section analyzes how the estimation error of the least squares approach depends on Signal Noise ratio (SNR) at different sampling rate parameter  $r$ , and compares the performance of the least squares approach with other approaches. In this thesis, Mean Square Error (MSE) is defined

as

$$MSE = \sum_{i=1}^m \left| x^{(i)}[n] - x_{est}^{(i)}[n] \right|^2. \quad (3.5)$$

where  $x^{(i)}[n]$  is value of the signal at selected sensor, and  $x_{est}^{(i)}[n]$  is estimation of  $x^{(i)}[n]$  of the  $i^{th}$  trial. This MSE is normalized by the average signal power, i.e.,

$$\sum_{i=1}^m \left| x^{(i)}[n] \right|^2. \quad (3.6)$$

This leads to a definition for the normalized MSE

$$MSE_{Normalized} = \frac{\sum_{i=1}^m \left| x^{(i)}[n] - x_{est}^{(i)}[n] \right|^2}{\sum_{i=1}^m \left| x^{(i)}[n] \right|^2}. \quad (3.7)$$

In the results that follow an average value of 1000 Monte Carlo trials. The simulation parameters used in this section are number of sensors  $N = 22$  sensors and the location of the missing sensor is 15, which is the scenario of the TLA in SwellEx-96.

The following are the resulting of Normalized MSE of the estimation of 15<sup>th</sup> element in the 22 sensors array for different sampling rate parameters  $r$ .

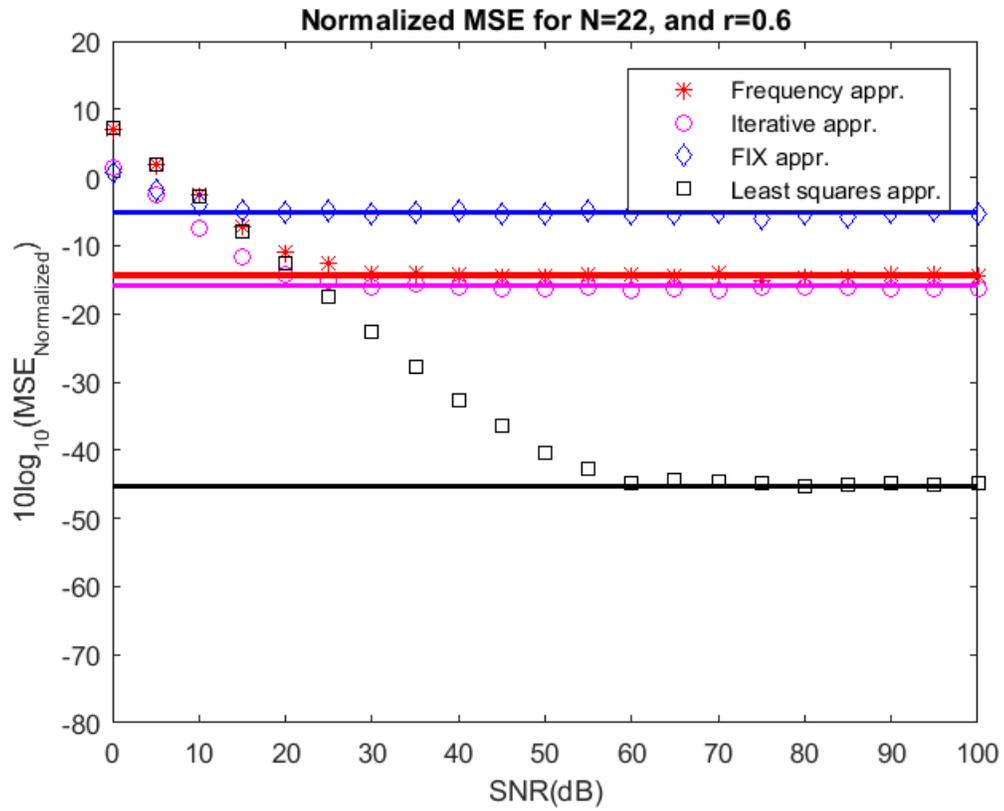


Figure 3.1: MSE vs. SNR comparison between Least squares approach and other approaches with  $N = 22$ , and  $r = 0.6$ .

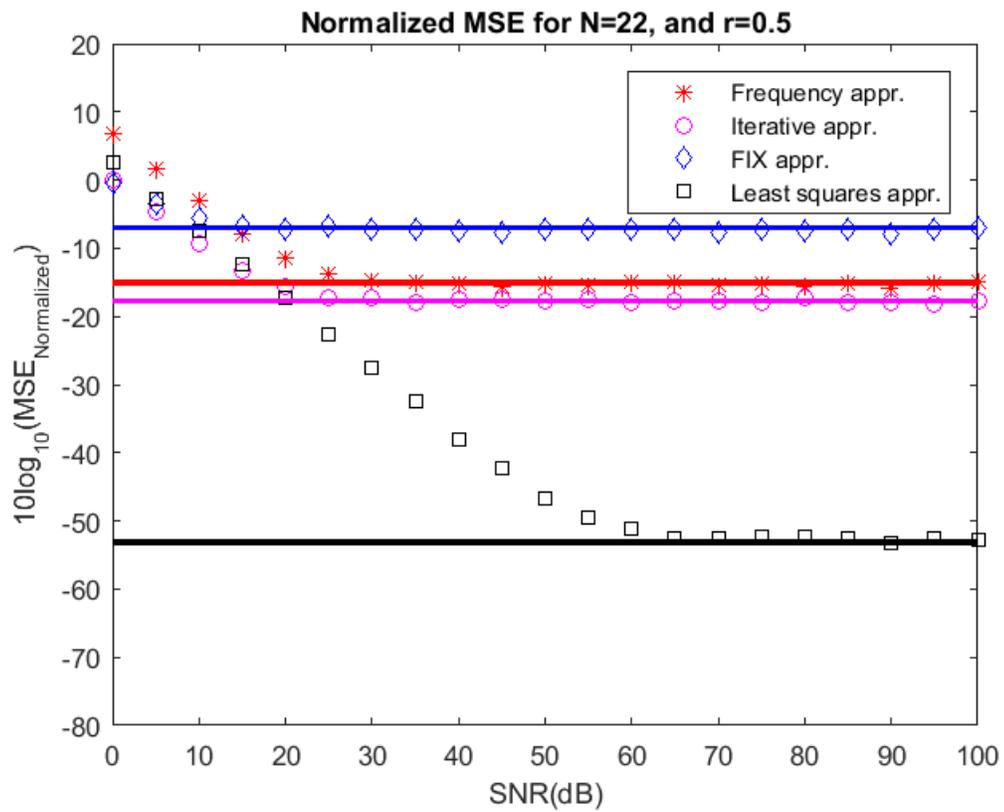


Figure 3.2: MSE vs. SNR comparison between Least squares approach and other approaches with  $N = 22$ , and  $r = 0.5$ .

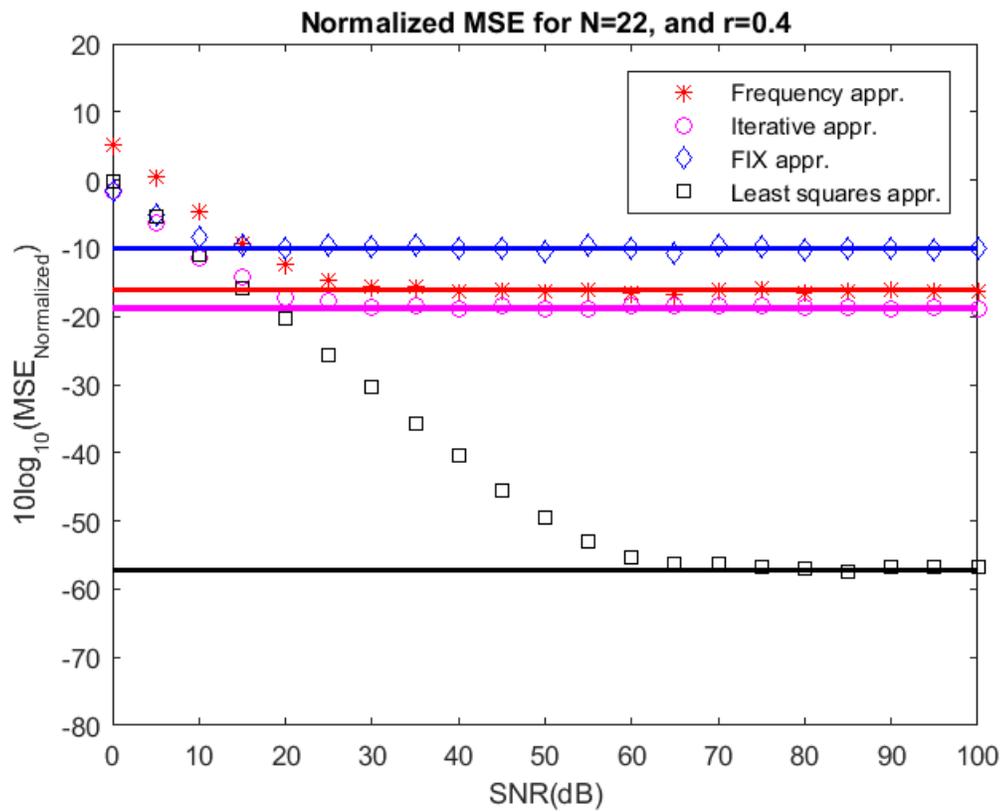


Figure 3.3: Normalized MSE vs. SNR comparison between Least squares approach and other approaches with  $N = 22$ , and  $r = 0.4$ .

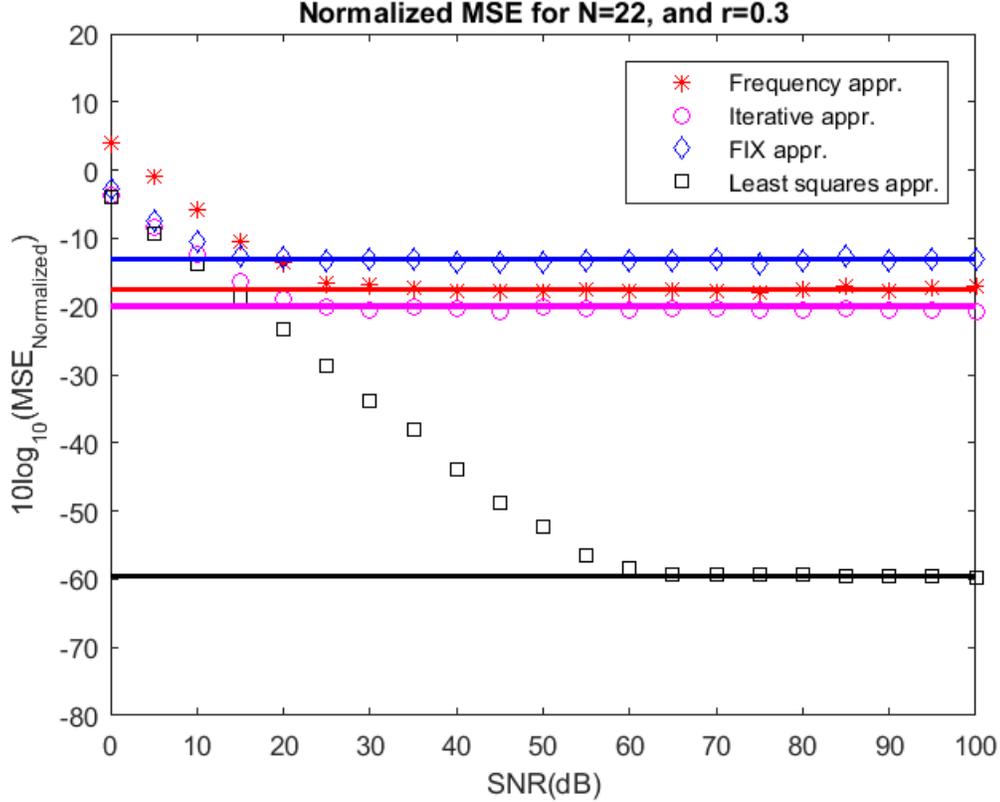


Figure 3.4: Normalized MSE vs. SNR comparison between Least squares approach and other approaches with  $N = 22$ , and  $r = 0.3$ .

In Fig. 3.1, the horizontal lines shows the results of the Normalized MSE of the estimation in all the approaches in the case signal having no noise. Fig. 3.1 shows that with no noise, Normalized MSE of estimation using the Frequency, Iterative, FIX, and least squares approach are  $-14.34$  dB,  $-15.92$  dB,  $-5.09$  dB, and  $-45.25$  dB respectively. The least squares approach has better estimation comparing with other approaches at SNR of 25 dB or higher. Fig. 3.1 also shows that the Least squares solution approaches the optimal result at SNR of around 60 dB.

Fig. 3.2 shows that Normalized MSE of the Frequency, Iterative, FIX, and Least squares approaches are  $-15.3$  dB,  $-15.03$  dB,  $-17.73$  dB,  $-7.03$  dB, and  $-53.16$  dB respectively in the case the signal has no noise. At SNR of 20 dB or higher, the Least squares approach has better estimation that other approaches. Fig. 3.2 also shows that the Least squares solution approaches the optimal result at SNR of around 60 dB.

The horizontal lines in Fig. 3.3 present the Normalized MSE of the approaches in the case signal

Table 3.1: Normalized Mean Square Error of approaches for no-noise case

	$r = 0.3$	$r = 0.4$	$r = 0.5$	$r = 0.6$
Frequency appr.	-16.68 dB	-16.10 dB	-15.03 dB	-14.34 dB
Iterative appr.	-19.92 dB	-18.75 dB	-17.73 dB	-15.92 dB
FIX appr.	-13.09 dB	-10.02 dB	-07.19 dB	-05.09 dB
Least squares appr.	-59.59 dB	-57.20 dB	-53.16 dB	-45.25 dB

has no noise. They are  $-16.19$  dB,  $-18.75$  dB,  $-10.02$  dB,  $-57.20$  dB for the Frequency, Iterative, FIX, and Least squares approaches respectively. The error in the Least squares approach is much lower than other approaches if signal has no noise. Fig. 3.3 shows that the Least squares approaches has better performance that other ones at SNR of 20 dB and the Least squares solution approaches the optimal result at SNR of around 60 dB.

The horizontal lines in Fig. 3.4 present the Normalized MSE of the approaches in the case signal has no noise. They are  $-16.68$  dB,  $-19.92$  dB,  $-13.03$  dB, and  $-59.59$  dB for the Frequency, Iterative, FIX, and Least squares approaches respectively. The error in the Least squares approach is much lower than other approaches if signal has no noise. Fig. 3.4 shows that the Least squares approaches has better performance that other ones at SNR of 15 dB and the Least squares solution approaches the optimal result at SNR of around 60 dB.

Overall, at a fixed number of sensor, the Least squares approach has much better estimation than others in the case signal has no noise. In the case signal has no noise, the number of sensors in the array controls the performance of the Least squares approach. In the case, the signal is embedded by noise, the Least squares has better estimation than others as the SNR reaches a threshold. For a 22-sensor array, the threshold levels are 15 dB, 20 dB, 20 dB, and 25 dB for the sampling rate  $r = 0.3$ ,  $r = 0.4$ ,  $r = 0.5$ , and  $r = 0.6$  respectively. These threshold depends on the sampling rate  $r$ . They decreases as the sampling rate decreases.

Eq. 3.3 indicates that the estimation  $\hat{A}$  is obtained by using  $L$  values of Fourier transform of the missing element signal in the interval between  $\psi_{max}$  and  $\pi$ . Eq. 2.3 in chapter 2 defined relationship of the radian frequency  $\psi_{max}$  and the sampling rate parameter  $r$  as

$$\psi_{max} = r\pi \tag{3.8}$$

It is clear that when the sampling rate parameter  $r$  is smaller, there are more values available to implement Eq. 3.3. This leads to better performance for the least square estimator. For the fixed number of sensors and the signal without noise embedded, the Normalized MSE of each of the approaches depends on sampling rate parameter  $r$ . Performance of the Iterative approach is a little better than the Frequency approach, Normalized MSE in the Iterative approach is less than Normalized MSE in the Frequency approach about from 2 dB to 3 dB for sampling rate parameter from  $r = 0.3$  to  $r = 0.6$ . This result is as Mark mentioned when comparing the performance between the Frequency approach and Iterative approach in [7]. Normalized MSE in the FIX approach is highest whereas Normalized MSE in the Least squares approach is lowest.

### **3.2.3 Normalized Mean Square Error versus Number of sensors**

This section investigates how the performance of the least squares approach depends on the number of sensors in the array for the no-noise case. The scenario in this section as follows: the missing sensor is at the origin and there are  $N$  sensors on either side of the origin.

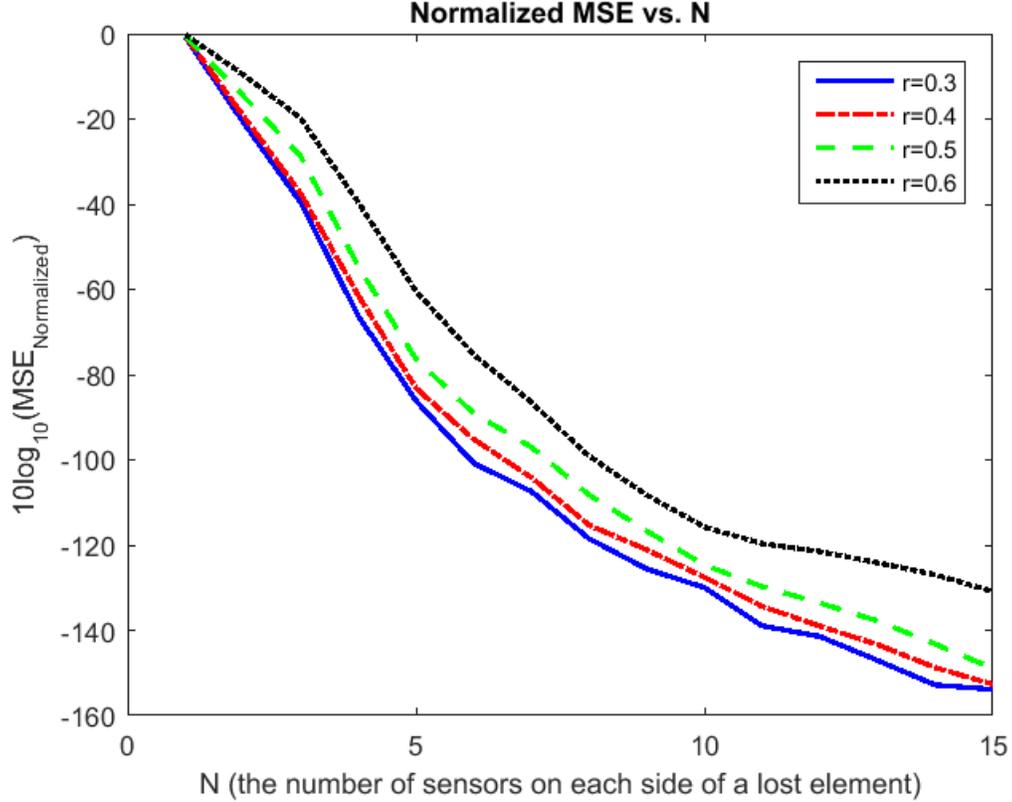


Figure 3.5: MSE vs. Number of Sensors on each side of a lost element.

Fig. 3.5 shows the Normalized MSE of estimation at different sampling rate parameters  $r$ , i.e.,  $r = 0.3$ ,  $r = 0.4$ ,  $r = 0.5$ , and  $r = 0.6$ . Fig. 3.5 shows that at a fixed sampling rate parameter, the Normalized MSE depends on the length of the array. The least squares approach has better performance with longer array. With longer array, there are more values available for implementing equation in the least squares approach. The performance of the Least squares approach also depends on the sampling rate. The sampling rate is smaller, the band outside of the signal is larger. As result, there are more values available for implementing equation in the least squares approach. This leads to the Least squares has better estimation.

### 3.3 Summary

In the case signal has no noise, the Normalized MSE in the Least squares approach depends on the length of array and the sampling rate  $r$ . The performance of the Least squares approach is

better with longer array and smaller sampling rate parameter  $r$ . In the case signal has no noise at a fixed sampling rate parameter  $r$ , the length of array controls performance of the Least squares approach. For a fixed number of sensors, the Normalized MSE of each of the approaches depends on SNR and sampling rate parameter  $r$ . The performance of each approach is better when the SNR is higher and the sampling rate parameter  $r$  is smaller. The least squares approach has better estimation comparing with other approaches once the SNR exceeds a threshold. For the sampling rate parameter from  $r = 0.3$  to  $r = 0.6$ , the threshold SNR is about 20 dB.

## Chapter 4: The experimental Results Using the SWellEx-96 Data Set

The previous chapter analyzed the performance of the Frequency approach, the Iterative approach, the FIX approach, and the Least squares approach for simulated data. This chapter discusses the results of estimating the missing data for the SWellEx-96 experiment. A brief overview of the SWellEx-96 experiment is given in Section 4.1. Section 4.2 presents the spatial spectra of signal in the experiment. Section 4.3 presents and analyzes the performance of the four approaches for the experiment data. Section 4.3 summarizes the chapter.

### 4.1 The SWellEx-96 experiment overview

The SWellEx-96 was deployed off the California coast from May 10, 1996 to May 18, 1996. The SWellEx-96 test site has been studied extensively. A conductivity, temperature, depth (CTD) survey was conducted during the experiment to provide water column sound speed data. The CTD data from the SWellEx-96 Experiment consisted of 51 CTD casts. Each CTD data file provides temperature, salinity, and sound speed as a function of depth. Figure 4.1 shows the sound speed profile recorded at station 12, which is closest to the position of the TLA.

There were two events of interest in the experiment: S5 event and S59 event. This chapter presents the result of processing data from the S5 event in which there were no loud interferers present. In this event, two sources were simultaneously towed by the R/V Sproul: a deep source at about 54 m depth and a shallow source at about 9 m depth. The shallow source transmitted 9 frequencies between 109 Hz and 385 Hz as shown in Table 4.1. The deep source transmitted 65 frequencies between 49 Hz and 400 Hz. These frequencies are divided into 5 sets and each set has 13 frequencies. The first set is projected at maximum level of 158 Hz, and the next four sets are transmitted at level of 132 dB, 128 dB, 124 dB, 120 dB, as shown in Table 4.2.

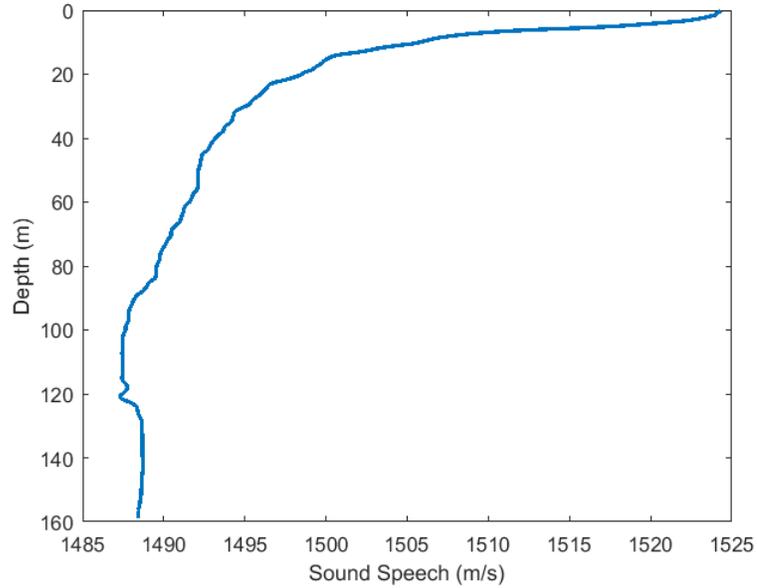


Figure 4.1: Sound Speed Profile.

Table 4.1: Shallow Source Tonal Set

Frequency (Hz)								
109	127	145	163	198	232	280	335	385

Table 4.2: Deep Source Tonal Set

Level	Frequency (Hz)												
158 dB	49	64	79	94	112	130	148	166	201	235	283	338	388
132 dB	52	67	82	97	115	133	151	169	204	238	286	341	391
128 dB	55	70	85	100	118	136	154	172	207	241	289	344	394
124 dB	58	73	88	103	121	139	157	175	210	244	292	347	397
120 dB	61	76	91	106	124	142	160	178	213	247	295	350	400

As described in Chapter 1, acoustic sensors deployed in the experiment included the VLA and the TLA. Both of these arrays consisted of 22 equally-spaced sensors. The data was recorded from all the sensors of the TLA whereas the time series recorded from the 15<sup>th</sup> sensor of the VLA was removed due to bad data. For this project, the data recorded on TLA is of particular interest because it offers an opportunity to investigate methods of estimating the missing data in bandlimited signal.

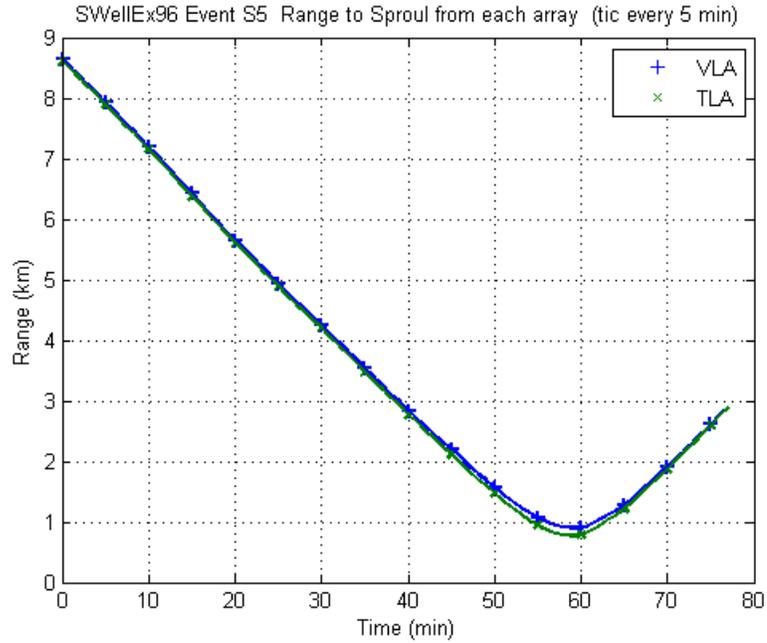


Figure 4.2: SWellEx-96 Event S5 Range to Sproul from TLA and VLA.

Figure 4.2 shows ranges from the towed source to VLA and TLA for each minute in the event S5. The closest point of approach is about 1 km and the furthest is almost 9 km. Figure 4.3 shows the locations of sensors in TLA. The TLA had a tilt of about 45 – 47 degrees. The first sensor of the TLA has a depth of 153 m, and the last sensor has a depth of 67.315 m.

The test of the missing sensor data estimation algorithms is set up as follows. First, the data from the 15<sup>th</sup> sensor of the TLA is set to zero and the algorithms as described in the previous chapters are applied to recover the deleted sample. Second, the performance of estimates from all of algorithms will be analyzed by comparing the spatial Fourier transform between using the zero-value at the 15<sup>th</sup> sensor of the TLA and using the estimated results.

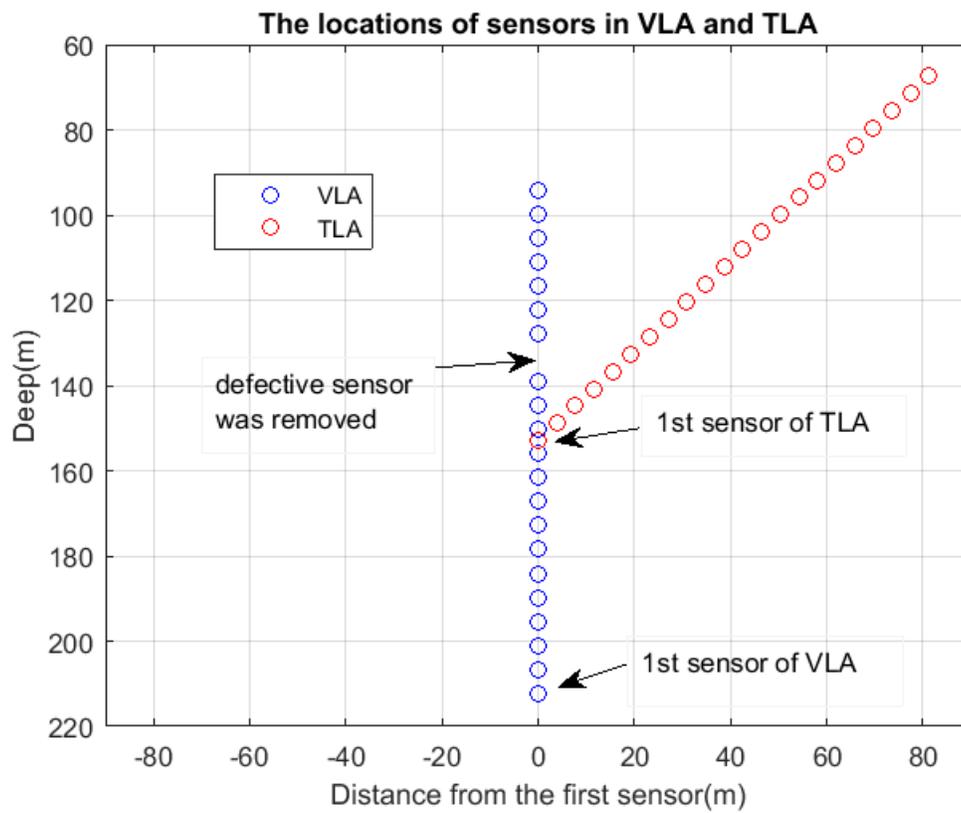


Figure 4.3: Locations of sensors in Vertical Linear Array and Tilted Linear Array.

## 4.2 Spatial Spectra of signal in the SWellEx-96 experiment

In this thesis, the  $k - \Omega$  beamforming technique is used to compute the frequency-wavenumber spectrum for the set of data recorded by the TLA.  $\Omega$  is the temporal radian frequency and  $k_z$  is the wavenumber. Two minute segments of data are used in this analysis. The first step of calculating the average  $k - \Omega$  spectrum is to split the data into set of overlapping time blocks. Each block consists of 4096 samples and an overlap of 50% is used between blocks. The second step is take the temporal Fourier transform of each of those windowed time blocks. The third step is take the spatial Fourier transform of the windowed narrowband array data. A Hanning window is used in both temporal and spatial Fourier transform [13]. Finally, the fourth step is to square the magnitude of the output and average over blocks to get power spectrum. Figure 4.5 shows the  $k - \Omega$  output for the TLA. This plot is derived from the two minutes data block starting at minute 58 of the recording, which is for the time segment where the source is closest to the array. Based on position of sensors in the TLA, the spacing  $d$  between two sensors in the TLA is about 5.625 m.

In a linear array, wavenumber  $k_z$  and arriving angle  $\theta$  is related by

$$k_z = \frac{\Omega}{c} \cos\theta$$

with limits on  $\theta$

$$0 \leq \theta \leq \pi$$

This leads to

$$-\frac{\Omega}{c} \leq k_z \leq \frac{\Omega}{c}$$

$$-\frac{2\pi}{\lambda} \leq k_z \leq \frac{2\pi}{\lambda}$$

The radian frequency  $\psi$  has value

$$-\frac{2\pi d}{\lambda} \leq \psi \leq \frac{2\pi d}{\lambda}$$

Consider normalizing the radian frequency  $\psi$  by  $\pi$ , i.e.,

$$-\frac{2d}{\lambda} \leq \frac{\psi}{\pi} \leq \frac{2d}{\lambda} \quad (4.1)$$

Eq. 4.1 defines visible region of the narrow band signal. Figure 4.4 illustrate the visible region of the narrow band signal. Figure 4.4 shows that the spatial maximum radian frequency  $\psi_{max}$  is defined by

$$\psi_{max} = \frac{2d}{\lambda}$$

This leads to

$$\psi_{max} = \frac{2df}{c} \quad (4.2)$$

Eq. 4.2 indicates that the spatial maximum radian frequency  $\psi_{max}$  depends on the temporal  $f$  frequency. It means that the temporal frequency controls the spatial maximum radian frequency. With smaller spatial maximum radian frequency, the Least squares approach has more available values to estimate the missing element.

For a linear array, a spatial Nyquist rate to avoid aliasing is defined by

$$d < \frac{\lambda}{2},$$

where

$$\lambda = \frac{c}{f}.$$

Solving for  $f$  leads to

$$f < \frac{c}{2d}.$$

For a velocity  $c = 1488$  m/s and spacing  $d = 5.625$  m (values for the SWellEx-96 experiment),  $f$  must be less than 132 Hz to avoid aliasing. It means that any narrowband signal having the temporal frequency less than 132 Hz has invisible region. Given this criteria, the thesis focuses on the SWellEx-96 narrowband signals with frequencies of 64 Hz, 67 Hz, 70 Hz, 73 Hz, and 76 Hz to

analyze the performance of the least squares estimator.

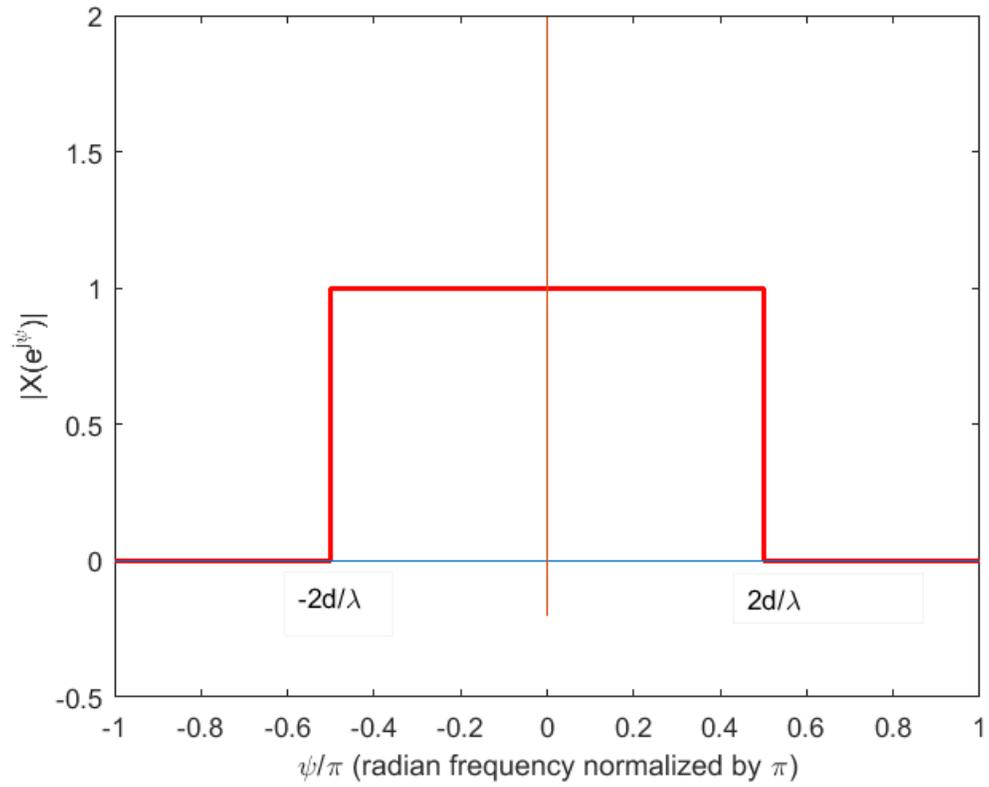


Figure 4.4: The Ideal Bandlimited Signal with a spatial cutoff radian frequency.

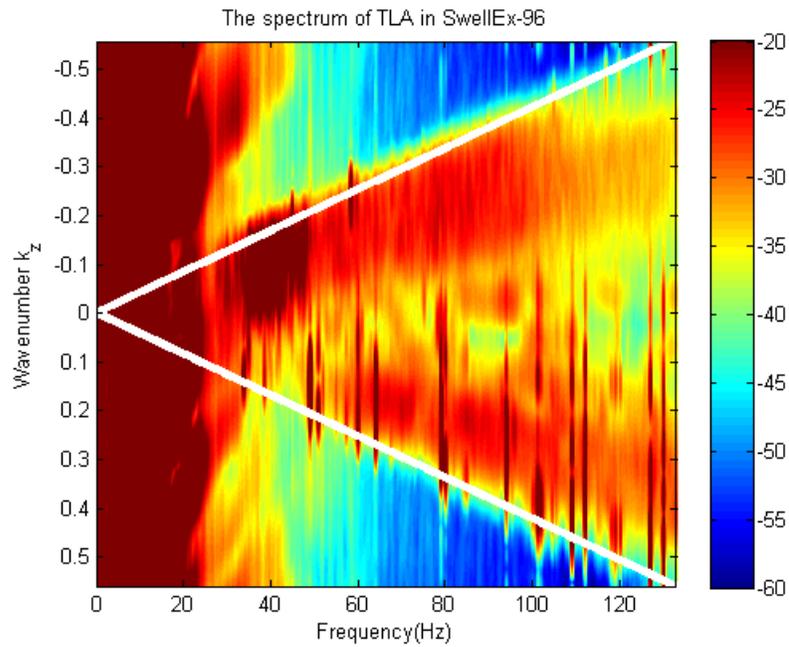


Figure 4.5:  $k-\Omega$  beamformer of data record in TLA. The two minutes length of data starting at 58<sup>th</sup> minute is used, a 4096 point Hanning window is applied to calculate the temporal Fourier transform for each of time block, Hanning window is also used to calculate the spatial Fourier transform. A overlap 50% is used between frames of data. The white lines are boundaries between the visible region and the invisible region.

### 4.3 Comparing the performance between algorithms

As stated in the previous section, any narrowband signals in TLA having the temporal frequency less than 132 Hz are spatially bandlimited and have samples in the invisible region that can be used by the least squares estimator. It implies that a missing sample in these signals can be recovered from the remaining samples to reduce the effects of missing element. For purpose of comparison, the spatial Fourier transform of the selected narrowed band signals is calculated for filled array, missing element array, and recovered element array using Frequency approach, Iterative approach, FIX approach, and Least squares approach. The Normalized MSE of estimation in each approach also calculated for comparison.

The following section presents results for some selected narrow band signals from TLA.

#### 4.3.1 Comparison spatial Fourier transform between approaches

Figures 4.6-4.10 show that the sidelobe of the spatial Fourier transform of the missing element array is higher than the sidelobe level of the spatial Fourier transform of the filled array as the data at 15<sup>th</sup> sensor of the TLA substituted by zero. In this section, the performance of all approaches are compared by observing the sidelobe level of the spatial Fourier transform of the signal in all approaches and the corresponding Normalized MSE. Figures 4.6 shows that the missing data at 15<sup>th</sup> sensor causes the sidelobe level of the Fourier transform of the narrow band signal  $f = 64$  Hz increase 12 dB from  $-45$  dB to  $-33$  dB. The Frequency approach, Iterative approach, FIX approach, and Least squares approach are applied to recover the missing element at 15<sup>th</sup> sensor. The second observation from Figure 4.6 is that the recovered missing element using all the approaches help lowering the sidelobe level of the spatial Fourier transform of the missing element array. The sidelobe level of the spatial Fourier transform of the signal with recovered missing element using the Frequency approach, Iterative approach, FIX approach, and Least square approach are about  $-42$  dB,  $-47$  dB,  $-42$  dB, and  $-49$  dB respectively. The results in Table 4.3 shows the corresponding Normalized MSE. The FIX approach has highest Normalized MSE, the Least square has lowest Normalized MSE. The Iterative approach has Normalized MSE smaller than the Frequency approach, this is consistent with depiction of Marks in [7].

Similarly, Figures 4.7 shows that the sidelobe level of the Fourier transform of the narrow band

signal  $f = 67$  Hz is increased from  $-52$  dB to  $-45$  dB because of affection of the missing element. Figures 4.7 shows that the spatial Fourier transform of the signal using the defined approaches has the similar characteristic in Figures 4.6, it means that the recovered missing element in all approaches help lower the sidelobe level of the missing element array. The Frequency, Iterative, FIX, and Least squares approach help lowering the sidelobe level of the missing element array from  $-44$  dB to  $-48$  dB,  $-51$  dB,  $-47$  dB, and  $-52$  dB respectively. The Normalized MSE in this case has order from lowest value to highest value as in the narrow band signal  $f = 64$  Hz.

The similar results are obtained from Figs. 4.8-4.10 for the narrow band signals  $f = 70$  Hz,  $f = 73$  Hz, and  $f = 76$  Hz.

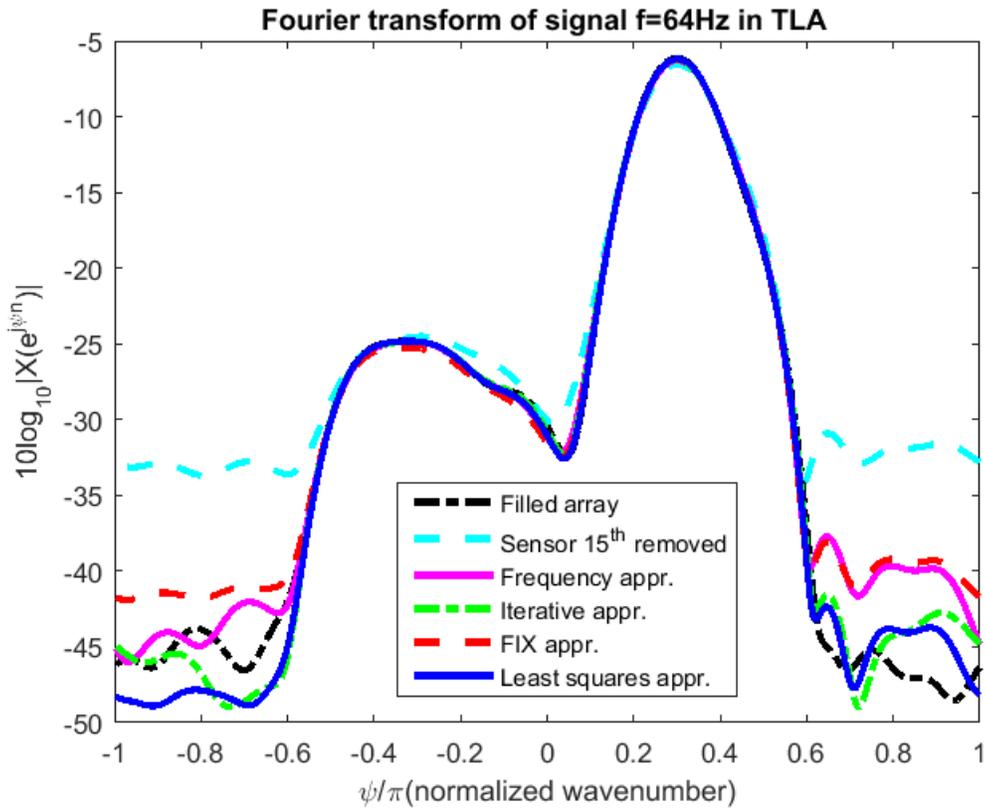


Figure 4.6: Comparing the spatial Fourier transform of the narrow band signal  $f = 64$  Hz recorded by the Filled array with the missing element array, and recovered element array using the Frequency approach, the Iterative approach, the FIX approach, and the Least squares approach.

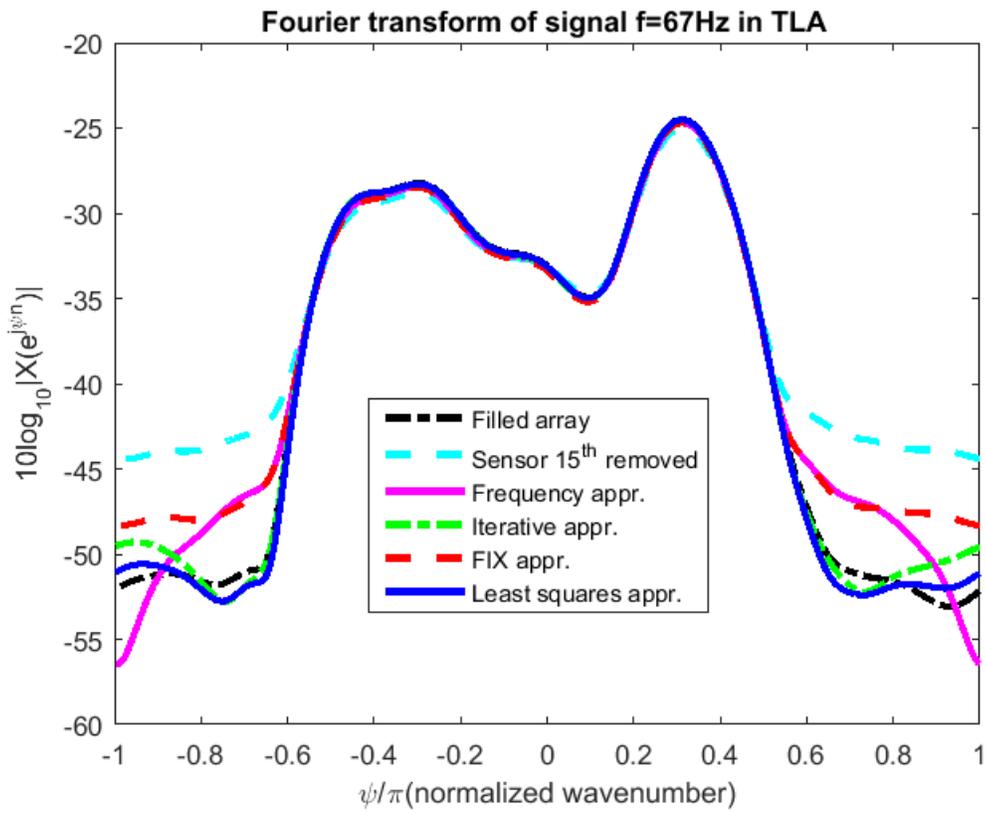


Figure 4.7: Comparing the spatial Fourier transform of the narrow band signal  $f = 67$  Hz recorded by the Filled array with the missing element array, and recovered element array using the Frequency approach, the Iterative approach, the FIX approach, and the Least squares approach.

Fig. 4.8 shows that the sidelobe level of Fourier transform of the original signal  $f = 70$  Hz is  $-52$  dB. This sidelobe level increases to  $-42$  dB as data at  $15^{\text{th}}$  sensor removed. The estimated data using the Frequency, Iterative, FIX, and Least squares approaches lower the sidelobe level from  $-42$  dB to  $-47$  dB,  $-51$  dB,  $-45$  dB, and  $-52$  dB respectively. The corresponding Normalized MSE are 0.3008, 0.2242, 0.5569, and 0.0380 as show in Table 4.3. In this case, the Fourier transform of the signal using estimated data of the Least squares approach is almost coincidence to the Fourier transform of the original signal.

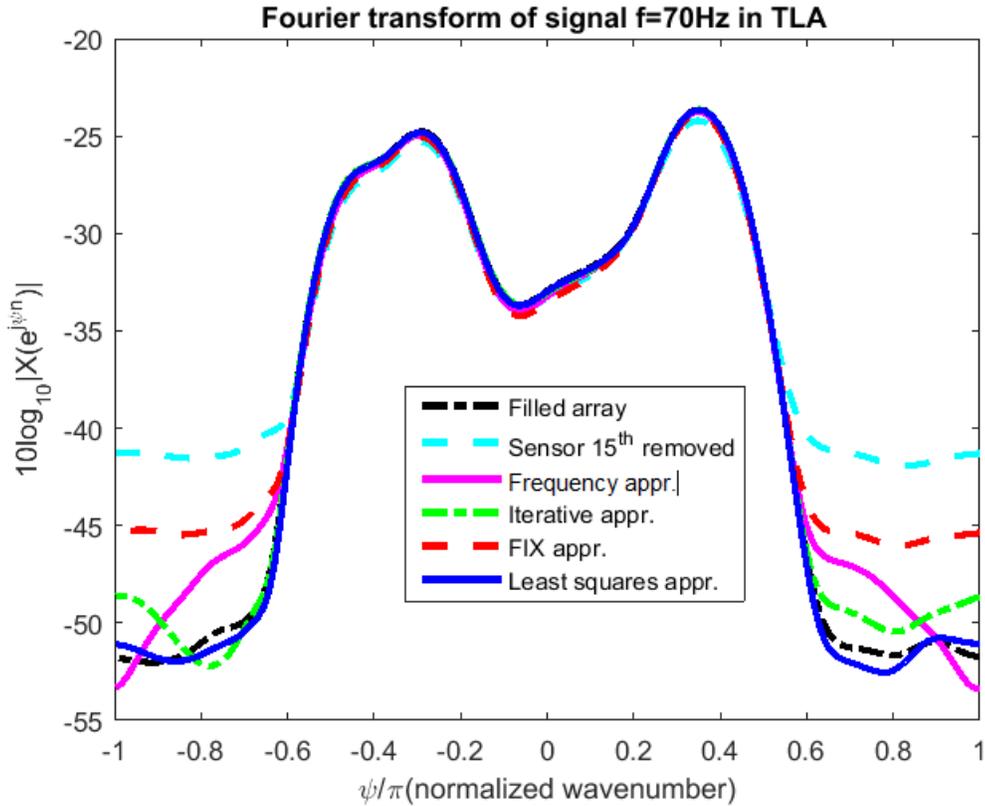


Figure 4.8: Comparing the spatial Fourier transform of the narrow band signal  $f = 70$  Hz recorded by the Filled array with the missing element array, and recovered element array using the Frequency approach, the Iterative approach, the FIX approach, and the Least squares approach.

Fig. 4.9 shows that the sidelobe level of Fourier transform of the original signal  $f = 73$  Hz is  $-52$  dB. This sidelobe level increases to  $-43$  dB as data at  $15^{th}$  sensor removed. The estimated data using the Frequency, Iterative, FIX, and Least squares approaches lower the sidelobe level from  $-43$  dB to  $-49$  dB,  $-51$  dB,  $-47$  dB, and  $-52$  dB respectively. The corresponding Normalized MSE are 1.9454, 2.0544, 7.1098, and 0.9349, as show in Table 4.3. In this case, the Fourier transform of the signal using estimated data of the Least squares approach is very close to the Fourier transform of the original signal.

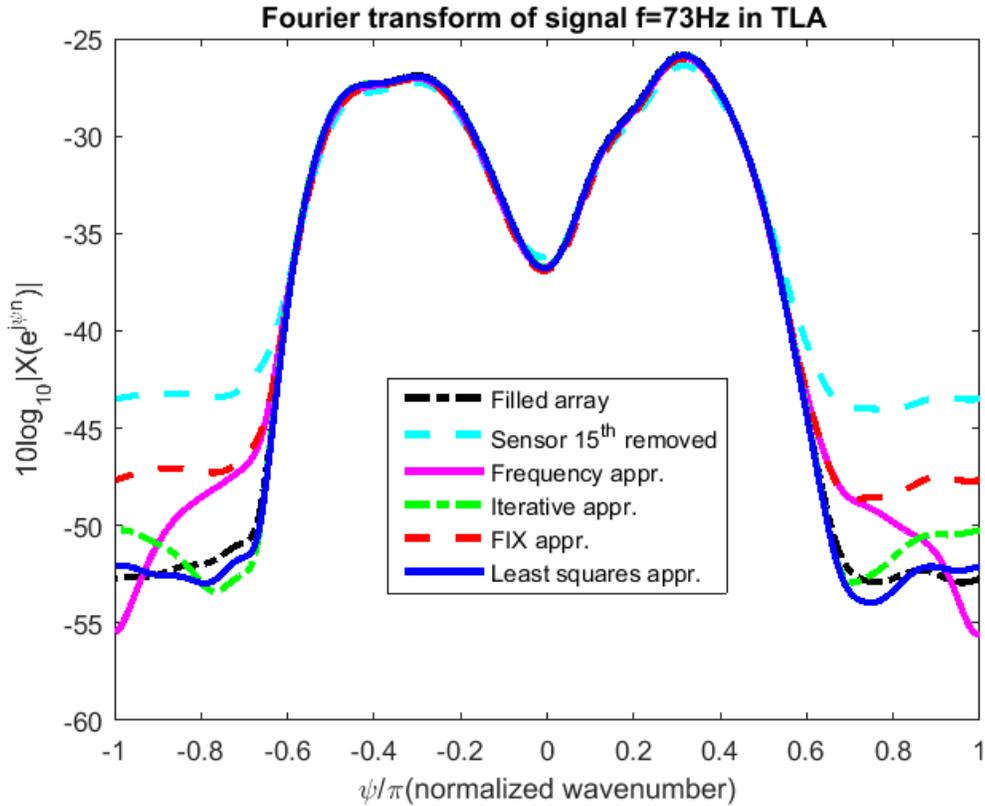


Figure 4.9: Comparing the spatial Fourier transform of the narrow band signal  $f = 73$  Hz recorded by the Filled array with the missing element array, and recovered element array using the Frequency approach, the Iterative approach, the FIX approach, and the Least squares approach.

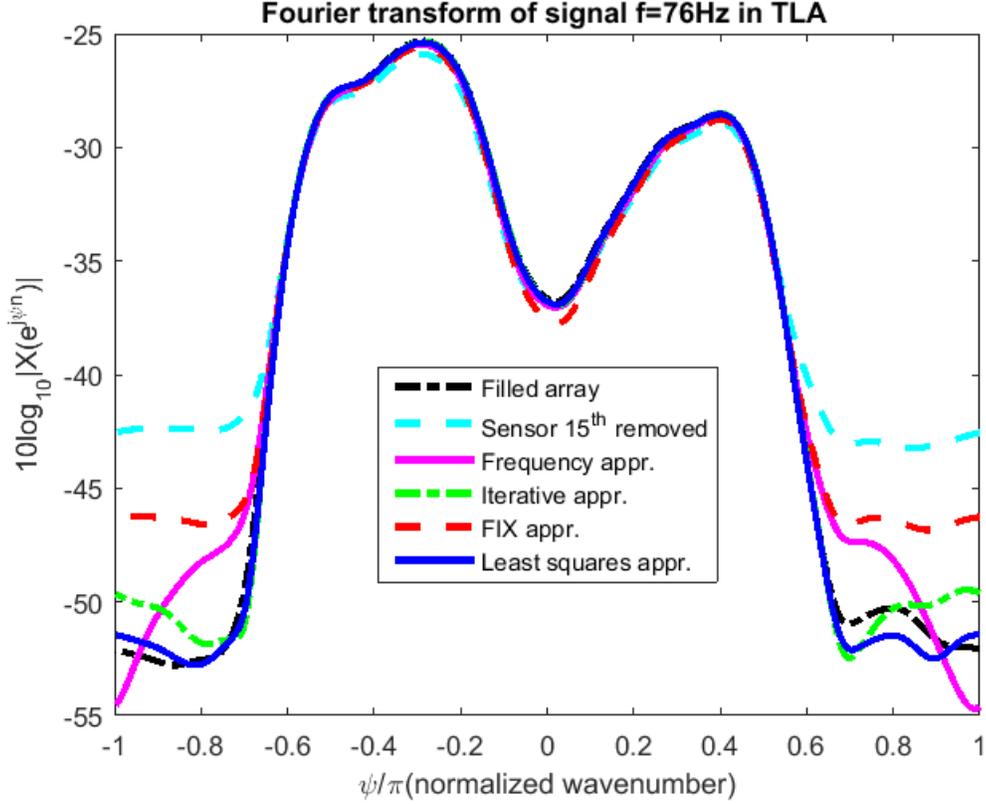


Figure 4.10: Comparing the spatial Fourier transform of the narrow band signal  $f = 76$  Hz recorded by the Filled array with the missing element array, and recovered element array using the Frequency approach, the Iterative approach, the FIX approach, and the Least squares approach.

Fig. 4.10 shows that the sidelobe level of Fourier transform of the original signal  $f = 76$  Hz is  $-52$  dB. This sidelobe level increases to  $-43$  dB as data at  $15^{th}$  sensor removed. The estimated data using the Frequency, Iterative, FIX, and Least squares approaches lower the sidelobe level from  $-43$  dB to  $-49$  dB,  $-51$  dB,  $-47$  dB, and  $-52$  dB respectively. The corresponding Normalized MSE are 0.5950, 0.1996, 0.5600, and 0.0747, as show in Table 4.3. In this case, the Fourier transform of the signal using estimated data of the Least squares approach is very close to the Fourier transform of the original signal.

Overall, the estimated data using the defined approaches reduced effect of the limited number of sensors in the array, and the embedded noise. For the narrow band signals  $f = 64$  Hz,  $f = 67$  Hz,  $f = 70$  Hz,  $f = 73$  Hz, and  $f = 76$  Hz in SWellEx-96 experiment, the Least squares approach have better performance than other approaches based on the comparison the Fourier transform and the corresponding Normalized MSE.

Table 4.3: Normalized Mean Square Error

	Frequency appr.	Iterative appr.	FIX appr.	Least Squares appr.
64Hz	0.1925	0.1079	0.2937	0.0523
67Hz	0.3019	0.2422	0.4795	0.0558
70Hz	0.3008	0.2242	0.5569	0.0380
73Hz	1.9454	2.0544	7.1098	0.9349
76Hz	0.5950	0.1996	0.5600	0.0747

### 4.3.2 Normalized MSE for some narrow band signals along 75 minutes recording

In this section, the performance of the approaches are analyzed by comparing Normalized MSE along 75 minutes recording for all approaches. Normalized MSE is calculated for each segment of every two minutes recording, and overlap 50% is used between each segment.

Fig. 4.11 shows that the Normalized MSE of the estimation using Least squares approach is smallest comparing with the Normalized MSE of other approaches along 75 minutes recording. Along 75 minutes recording, all Normalized MSE is below 0.5 except Normalized MSE at segment 24<sup>th</sup>. The Normalized MSE for this segment is 0.845, which is smaller than Normalized MSE of other approaches. The Normalized MSE for the Frequency, Iterative, and FIX approaches are 1.202, 3.341, and 3.912 respectively.

The first observation from Fig. 4.12 is the Normalized MSE of the Least squares approach has the highest value at the 10<sup>th</sup> segment, which is 0.1804 at along 75 minutes recording. The second observation is the Normalized MSE of the Least squares approach at each segment is smaller than Normalized MSE of other approaches.

Fig. 4.13 shows that Normalized MSE of the Least squares approach has the highest value at the 52<sup>th</sup> segment comparing with other segments along 75 minutes recordings. At this segment, the Normalized MSE of the Frequency, Iterative, FIX, and Least squares approach are 4.500, 3.312, 2.108, and 1.571 respectively.

Fig. 4.14 shows that Normalized MSE of the Least squares approach is below 0.2 at each segment except Normalized MSE at 58<sup>th</sup> segment. At this segment, the Normalized MSE of the Frequency, Iterative, FIX, and Least squared approaches are 1.9454, 2.0544, 3.100, and 1.202 respectively.

Fig. 4.14 shows that Normalized MSE of the Least squares approach is below 0.18 at each segment

except Normalized MSE at 68<sup>th</sup> and 70<sup>th</sup> segments.

Overall, the Least squares approach achieves lower than other approaches. For the signals  $f = 64$  Hz,  $f = 67$  Hz,  $f = 70$  Hz,  $f = 73$  Hz, and  $f = 76$ , the Normalized MSE of the Least squares approach is below 0.5 at almost segments along the recording.

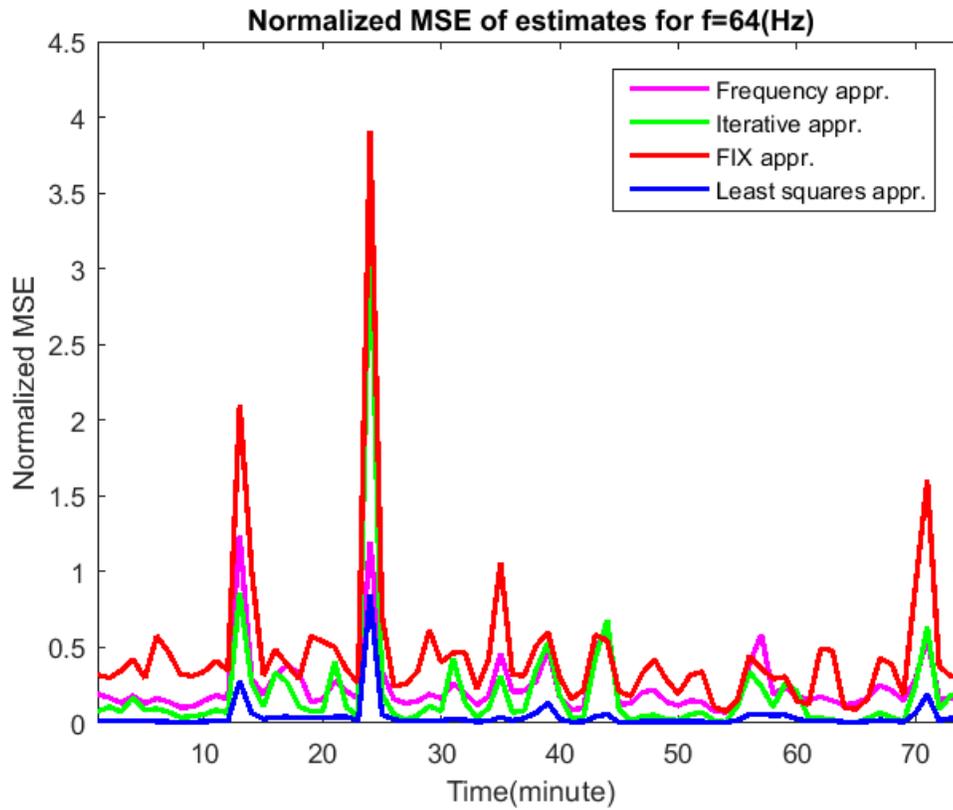


Figure 4.11: Normalized MSE for  $f=64$ (Hz) along 75 minutes recording.

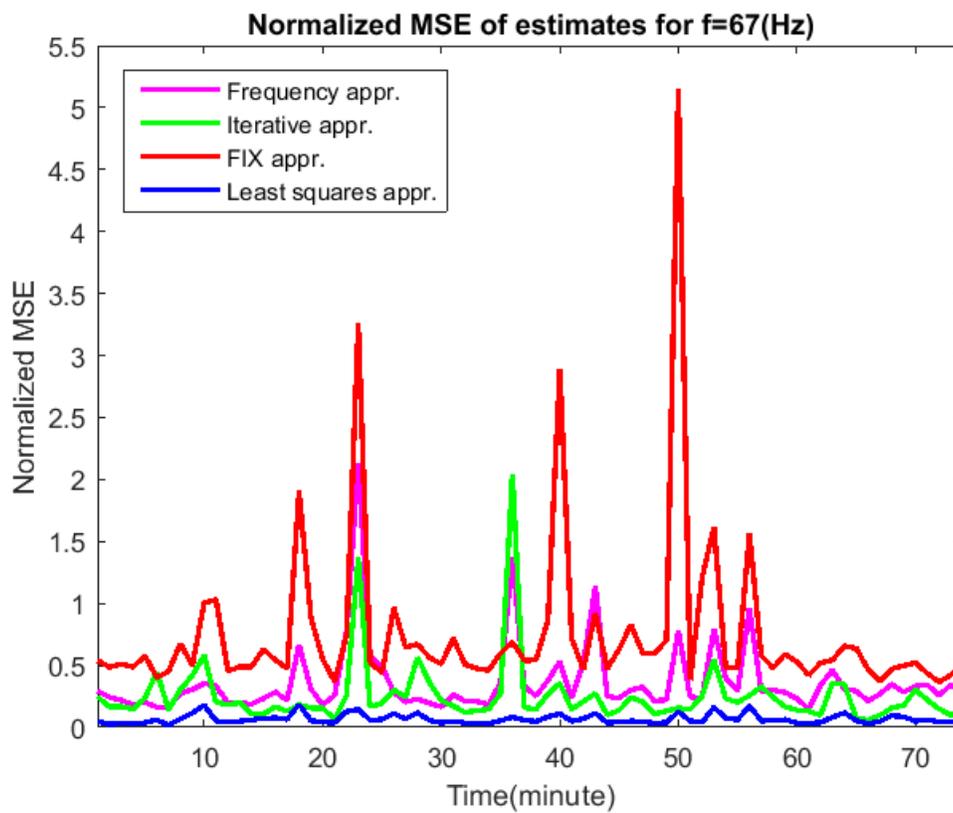


Figure 4.12: Normalized MSE for  $f=67(\text{Hz})$  along 75 minutes recording.

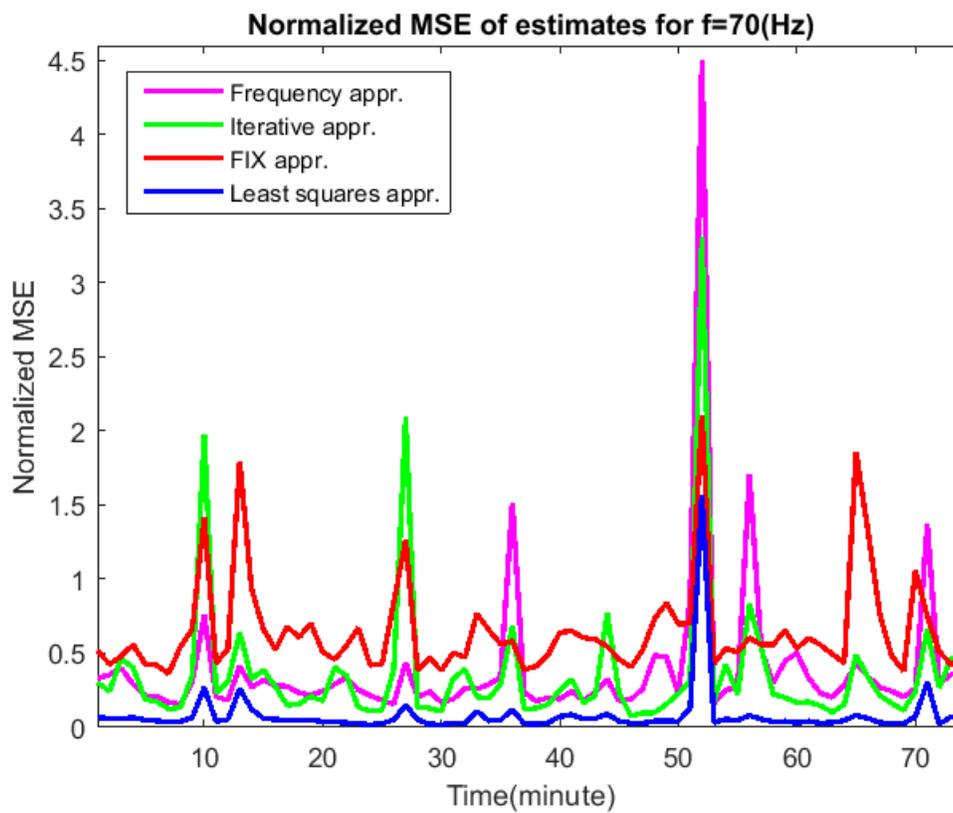


Figure 4.13: Normalized MSE for  $f=70(\text{Hz})$  along 75 minutes recording.

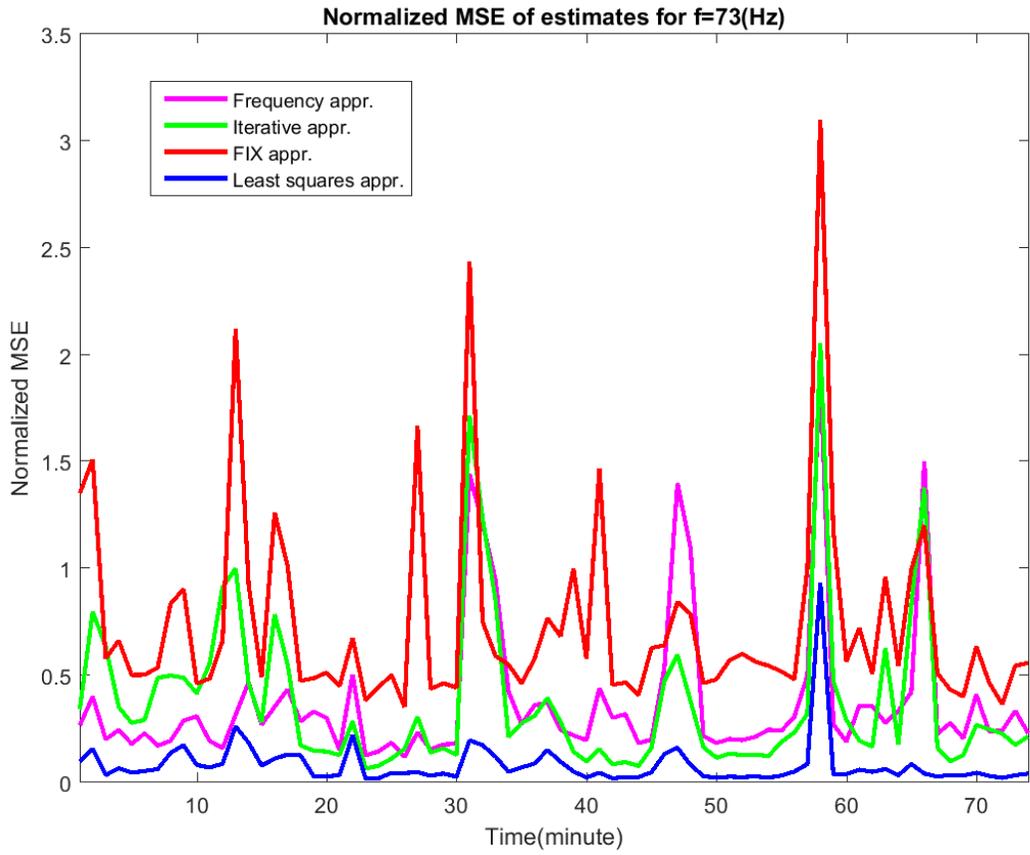


Figure 4.14: Normalized MSE for  $f = 73$  Hz along 75 minutes recording.

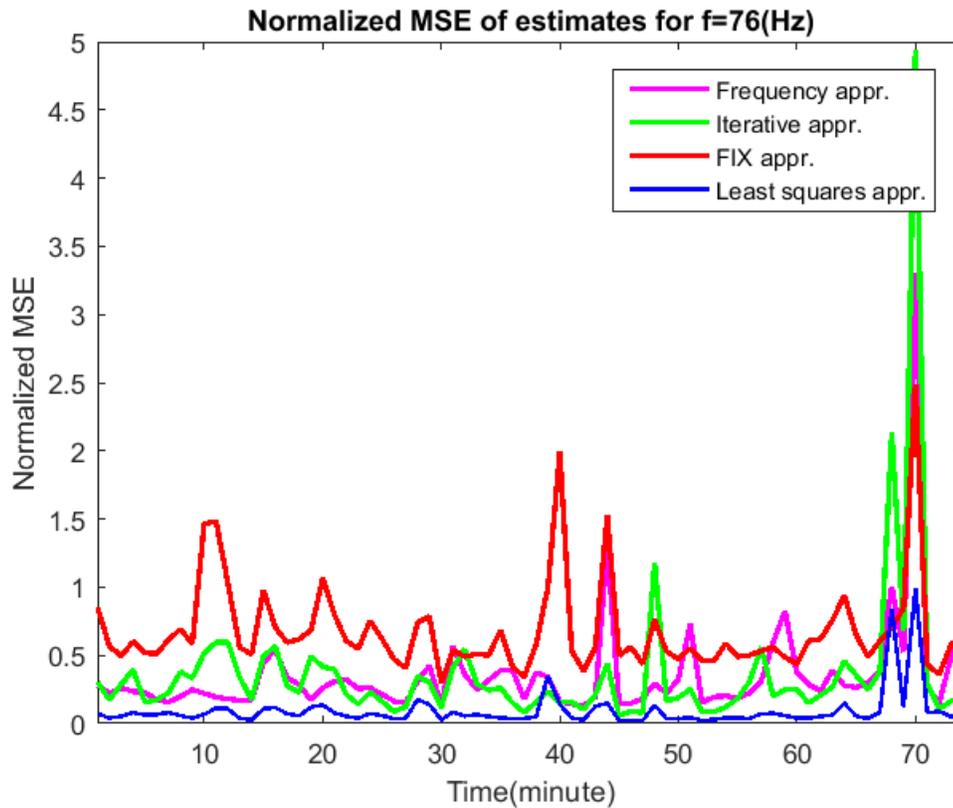


Figure 4.15: Normalized MSE for  $f = 76$  Hz along 75 minutes recording.

## 4.4 Summary

This chapter presented the performance of the Frequency, Iterative, FIX, and Least squares approaches using data in the SWellEx-96 experiment. By comparing the spatial Fourier transform and the corresponding Normalized MSE, the results show that the Least squares approach has better performance than other approaches.

## Chapter 5: Conclusion

This thesis investigated the least squares solution to the problem of estimating data for a missing sensor, assuming a oversampled bandlimited signal. Several practical issues were considered, including the effect of noise and a finite aperture. First chapter 3 of the thesis analyzed the performance of the least squares approach by comparing the mean square estimation error for the least squares approach with the mean square error of estimation used the Frequency, Iterative, and FIX approaches. The comparison assumed a fixed number of sensors and different SNR levels. The thesis shows that the least squares approach perform better than the other approaches if SNR of the signal exceeds a threshold. For a fixed number of sensors, this threshold SNR depends on the sampling rate parameter. A higher sampling rate parameter requires higher SNR level. Second chapter 3 analyzed the performance of the least squares approach as a function of array size for different sampling rate parameters. The thesis shows that all approaches perform better with longer array. The number of sensors in the array controls the number of values available to solve the equation in the least squares approach. As the number of sensors increases, the number of values available increases. The thesis also shows that the number of available values depends on the sampling rate parameter  $r$ . The least squares approach has better estimation with smaller sampling rate parameter. Chapter 4 used the data from SWellEx-96 experiment to evaluate least squares and other approaches. The thesis shows that the least squares approach is useful in estimating a missing value for some narrow band frequency signals in which those narrow band signals have a number of values and level of noise reaching requirement for this approach.

This thesis suggests several avenues for future work. First, it would be interesting to derive an analytical prediction of MSE as a function of number of sensors at a specific level of SNR. Second, the effects of the location of missing element also needs to be considered in analyzing the performance of the least squares approach.

## Bibliography

- [1] A. Papoulis, *Signal Analysis*. New York Inc: New York: McGraw-Hill, 1977.
- [2] R. J. Marks II, "Restoring lost samples from an oversampled band-limited signal," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 31, no. 3, pp. 752 – 755, Jun. 1983.
- [3] R. J. Marks II and D. Radbel, "Error of linear estimation of lost samples in an oversampled band-limited signal," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 3, pp. 648 – 654, Jun. 1984.
- [4] R. Wiley, "Recovery of bandlimited signals from unequally spaced samples," *Communications, IEEE Transactions on*, vol. 26, no. 1, pp. 135 – 137, Jan. 1978.
- [5] F. Marvasti and M. Analoui, "Recovery of signals from nonuniform samples using iterative methods," in *IEEE International Symposium on Circuits and Systems*, May 1989, pp. 1021 –1024 vol.2.
- [6] F. Marvasti, P. Clarkson, M. Dokic, U. Goenchanart, and C. Liu, "Reconstruction of speech signals with lost samples," *IEEE Transactions on Signal Processing*, vol. 40, no. 12, pp. 2897 –2903, Dec. 1992.
- [7] R. J. Marks II, *Introduction to Shannon Sampling and Interpolation Theory*, 1st ed. Springer-Verlag, 1991.
- [8] P. Ferreira, "Incomplete sampling series and the recovery of missing samples from oversampled band-limited signals," *IEEE Transactions on Signal Processing*, vol. 40, no. 1, pp. 225 –227, Jan. 1992.
- [9] J. B. Stockhausen and J. B. Farrell, "Fix- a beamforming technique for line arrays with missing elements," *Defence Research Establishment Atlantic*, vol. 84/0, Aug. 1984.

- [10] C. Shannon, "Communication in the presence of noise," *Proceedings of the IEEE*, vol. 72, no. 9, pp. 1192 – 1201, Sept. 1984.
- [11] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IEEE*, vol. 86, no. 2, pp. 447–457, Feb 1998.
- [12] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*. San Francisco: Holden-Day, 1968.
- [13] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, Jan 1978.

## Curriculum Vitae

Thuykhanh Le was born in Angiang, Vietnam. She has been living in the United States since 2000. She finished her bachelor's degree at George Mason University (GMU) in the fall of 2009, and started her graduate studies, also at GMU, in the spring of 2010. She worked on signal processing, her favorite area, with Dr. Kathleen Wage as a research assistant from the spring of 2009 to the fall of 2012. From the spring of 2013, she has started working at the Patent and Trade Office as an examiner in speech processing.