NOVEL SIGNAL PROCESSING TECHNIQUES FOR FIBER OPTIC DISTRIBUTED ACOUSTIC SENSING

by

İbrahim Ölçer

B.Sc., Electronics and Comm. Eng., İstanbul Technical University, 1992

M.Sc., Electronics and Comm. Eng., İstanbul Technical University, 1995

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Electrical & Electronics Engineering

Boğaziçi University

2019

NOVEL SIGNAL PROCESSING TECHNIQUES FOR FIBER OPTIC DISTRIBUTED ACOUSTIC SENSING

APPROVED BY:

Assoc. Prof. Ahmet ÖNCÜ (Thesis Supervisor)	.....
Prof. Ayşın Baytan ERTÜZÜN	.....
Assist. Prof. Heba YÜKSEL	.....
Assist. Prof. Onur FERHANOĞLU	.....
Prof. Ali SERPENGÜZEL	.....

DATE OF APPROVAL: 13.06.2019

To my departed father, Nejat Ölçer.

ACKNOWLEDGEMENTS

First of all, I would like to thank Prof. Avni Morgül and Prof. Selim Şeker for inviting me and encouraging me to study for the doctoral degree in Boğaziçi University.

My sincere thanks are due to my thesis advisor Assoc. Prof. Ahmet Öncü for permitting to study the research topic. He always supported my personal vision and attempts to develop a cooperative project with industry where I spent significant effort beside my academic objectives.

Assoc. Prof. Aydın Yeniay deserves special thanks for introducing to me the phase-sensitive optical time domain reflectometry concept. It was his demonstration in the TÜBİTAK facilities that triggered my intention and motivation to adapt my signal processing background to fiber optic sensing field.

Thanks go to Bülent Candan for helping me in assembling the optical hardware components of our first experimental system and setup in the laboratory.

I am very grateful to the directorate of the National Research Institute of Electronics and Cryptology for supporting my research work. Especially, I would like to express my gratitude to our former institute director Assist. Prof. Fatih Üstüner for supporting my ideas and project proposals relevant to this thesis research.

I also express my gratitude to Erdal Bayram, the current institute director for maintaining the same supporting attitude towards my research.

I owe my gratitude to Prof. Hacı Ali Mantar, the Chairman of Informatics and Information Security Research Center for his support and granting the funding via FOTAS project, and understanding for my executive efforts to develop nationwide awareness of distributed fiber optic sensing technology.

My colleague Dr. Umut Uludağ has a special place in my thesis study. He had contributed to my research by executing most of the field tests according to my scheduled test plans. Most of the field tests of this research would not have been possible without his passionate support.

I am grateful to Mustafa Akşit, the general manager of SAMM Teknoloji A.Ş. for supporting our field tests, and providing fiber optic cables for our field tests. Without any charge and personal interest, he provided the technical staff of their company in our campus to speed up our experimental research. It is a big pleasure for me to have met him and his agile company members.

Special thanks are reserved for my dissertation committee members Assist. Prof. Heba Yüksel and Assist. Prof. Onur Ferhanoğlu for reviewing the progress reports and the comments provided during my research; for Prof. Ayşın Baytan Ertüzün and Prof. Ali Serpengüzel for the critical reading of the dissertation.

I am deeply thankful to my children Zeynep and Emre for the enjoyment that they brought to my life. They deserve special recognition for their patience during my tough times.

I thank my beloved wife Piraye for her efforts to cheer me up when I was down.

And finally, I thank God for all the wisdom and the opportunities He gifted through my life.

ABSTRACT

NOVEL SIGNAL PROCESSING TECHNIQUES FOR FIBER OPTIC DISTRIBUTED ACOUSTIC SENSING

In this thesis, to reduce the effect of noise in phase-sensitive optical time domain based (ϕ-OTDR) distributed acoustic sensing (DAS) systems, two novel approaches are proposed and a real experimental ϕ-OTDR system is developed for validation. The first approach is the temporal adaptive processing of ϕ-OTDR signals which is based on maximizing the signal-to-noise ratio (SNR) at the output of an adaptive linear filter. When the vibration frequency of interest is known a priori, it is called the adaptive matched filter (AMF). The second approach is based on the largest eigenvalue computation of the optical covariance matrix which does not require any prior information about the vibration frequencies. Both methods utilize the correlation properties of the measured data. In the first method, the noise covariance matrix is estimated to compute an adaptive weight vector for optimum linear filtering. In the second method, the eigenvalues of the covariance matrix are computed and the maximum eigenvalue is used as the test statistic for detecting the vibrations along the fiber optic cable route. This so called maximum eigenvalue detection (MED) technique is assisted by the random matrix theory (RMT) to establish the binary detection threshold.

First, the efficacy of the proposed methods was demonstrated with Monte Carlo simulations. In the second phase, a ϕ-OTDR system was developed in the laboratory to gather real data and to verify the AMF and MED techniques with indoor experiments. In the last phase, extensive field tests, with both buried fibers and fibers on fences, were carried out to validate the proposed techniques in real-world conditions. The results show that more than 20 dB of SNR values can be achieved without any reduction in the system bandwidth and using any optical amplifier stage in the hardware.

ÖZET

DAĞITIK FİBER OPTİK AKUSTİK ALGILAYICILAR İÇİN YENİ SİNYAL İŞLEME TEKNİKLERİ

Bu tez çalışmasında, fiber optik dağıtık akustik sensör (DAS) sistemlerindeki gürültünün etkilerini azaltmak için iki yeni yaklaşım önerilmiş; önerilen yöntemlerin de geçerliliğini kanıtlamak için gerçek bir faz-duyarlı OTDR sistemi geliştirilmiştir. Birinci yaklaşım, faz-OTDR işaretlerinin zamansal uyarlamalı işlenmesine dayanmaktadır. Bu yöntem uyarlamalı doğrusal bir süzgeç çıkışındaki işaret-gürültü oranının en iyileştirilmesini temel alır ve ilgili titreşim frekansı önceden bilindiğinde uyarlamalı uyumlu süzgeç (AMF) olarak anılır. Tezde sunulan ikinci yaklaşım ise titreşim frekanslarının önceden bilinmesini gerektirmez ve özilinti matrisinin en büyük özdeğerinin hesaplanmasına dayalıdır. Her iki yöntem de ölçümlerin özilişki özelliklerinden yararlanmaktadır. Birincisinde, gürültünün en uygun şekilde ve doğrusal olarak süzülmesi için gereken uyarlamalı bir ağırlık vektörünün hesaplanması için gürültü ilinti matrisi kestirilir. İkincisinde, ilinti matrisinin özdeğerleri hesaplanır ve en büyük özdeğer fiber güzergâhı üzerindeki titreşimlerin algılanması için test istatistiği olarak kullanılır. En büyük özdeğer algılama (MED) yöntemi olarak bilinen bu yöntem, ikili sezim eşiğinin tespiti amacı ile rastlantı matrisi kuramı ile desteklenmektedir.

İlk olarak önerilen yöntemlerin etkinliği Monte Carlo benzetimleri ile gösterilmiştir. İkinci aşamada, gerçek veri toplamak, AMF ve MED yöntemlerinin laboratuvar testleri ile kullanılabilirliğini doğrulamak amacı ile gerçek bir faz-OTDR sistemi geliştirilmiştir. Son olarak, önerilen yöntemlerin gerçek koşullarda geçerliliğini test etmek amacı ile, hem zemin altında gömülü hem de çit üzerindeki fiber optik kablolar kullanılarak kapsamlı saha testleri gerçekleştirilmiştir. Test sonuçları, herhangi bir optik kuvvetlendirici donanımı kullanmadan ve sistemin frekans yanıtını bozmadan, 20 dB'den yüksek SNR değerleri elde edilebildiğini göstermektedir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv
ABSTRACT vi
ÖZET vii
LIST OF FIGURES xi
LIST OF TABLES xxi
LIST OF SYMBOLS xxii
LIST OF ACRONYMS/ABBREVIATIONS xxvi
1. INTRODUCTION 1
1.1. What is Fiber Optic Distributed Acoustic Sensing? 1
1.2. Distributed Acoustic Sensing: Historical Perspective and Applications . 2
1.2.1. Intrusion Sensing and Infrastructure Security 4
1.2.2. Seismic Imaging and Detection of Earthquakes 6
1.2.3. Structural Health and Condition Monitoring 7
1.2.4. Transportation Monitoring 8
1.3. Motivation and Objectives 8
1.4. Organization of the Thesis 9
1.5. Contributions 11
2. PHASE-SENSITIVE OPTICAL TIME-DOMAIN REFLECTOMETRY (ϕ-OTDR) FOR DISTRIBUTED ACOUSTIC SENSING 13
2.1. Introduction 13
2.2. Operational Principles of OTDR 14
2.3. Literature Review and State of the Art 19
2.4. Theoretical Model for the Rayleigh Backscatter 26
2.5. Numerical Implementation of Rayleigh Backscatter 28
2.6. The ϕ-OTDR Signal Model 33
2.7. Statement of the Problem 37
3. OPTIMAL FILTERING FOR NOISE REDUCTION IN ϕ-OTDR 38
3.1. Introduction 38
3.2. Optimum Temporal Filtering of ϕ-OTDR Signals 39

3.2.1. Maximum SNR Approach 41
3.2.2. The Adaptive Matched Filter (AMF) Algorithm 43
3.3. Applications of AMF for Vibration Detection 45
3.3.1. Single-Strong Vibration Case 45
3.3.2. Single-Weak Vibration Case 46
3.3.3. Multiple Vibrations 48
3.4. Conclusions 53
4. EIGENVALUE BASED ACOUSTIC VIBRATION SENSING 55
4.1. Introduction 55
4.2. Covariance Structure of the ϕ-OTDR Data 56
4.3. Maximum Eigenvalue of the Sample Covariance Matrix 57
4.4. Eigenvalue Statistics of the Sample Covariance and the Random Matrix Theory 58
4.4.1. Bulk Statistics of the SCM: The Marcenko-Pastur Law 60
4.4.2. Edge Statistics of the SCM: The Tracy-Widom Law 62
4.5. Maximum Eigenvalue Detection (MED) 65
4.6. Applications of MED for Vibration Detection 69
4.6.1. Single-Strong Vibration Case 70
4.6.2. Single-Weak Vibration Case 73
4.6.3. Multiple Vibrations 75
4.7. Empiricial Eigenvalue Distributions 78
4.8. Conclusions 82
5. LABORATORY EVALUATIONS WITH REAL ϕ-OTDR DATA 83
5.1. Introduction 83
5.2. Real ϕ-OTDR System Development 83
5.3. Indoor Experiments 85
5.3.1. AMF Test results 88
5.3.2. MED Test results 103
5.3.3. Empiricial Eigenvalue Distributions of the Indoor Data 113
5.3.4. Conclusions 117
6. FIELD TEST EVALUATIONS 119

6.1. Introduction 119
6.2. Outdoor Test Sites 119
6.2.1. Field Test Site-1: Buried Fiber Optic Cables 119
6.2.2. Field Test Site-2: Fiber Optic Cables on Fences 121
6.3. Applications at Test Site-1 124
6.3.1. AMF Test Results 127
6.3.2. MED Test Results 135
6.4. Applications at Test Site-2 138
6.5. Conclusions 147
7. CONCLUSION AND FUTURE WORK 150
7.1. Sample Covariance of the Optical Backscatter 150
7.2. Random Matrix Theory 151
7.3. Real DAS System Development 152
7.4. Future Prospects 154
7.4.1. Spatial Beamforming 154
7.4.2. Frequency Diversity Techniques 155
7.4.3. Machine Learning 155
REFERENCES 157
APPENDIX A: DERIVATION OF THE MAXIMUM SNR EQUATION . . . 168

LIST OF FIGURES

Figure 1.1. Distributed acoustic sensing concept. 3

Figure 2.1. Schematic of the direct and heterodyne detection schemes for OTDR. 14

Figure 2.2. Intensity profile of real $\phi-$OTDR data for a 28 km long fiber cable. 17

Figure 2.3. Relative strength of different noise sources in a ϕ-OTDR system according the distance. 18

Figure 2.4. LTI system representation of ϕ-OTDR system. 26

Figure 2.5. Discrete model for Rayleigh backscattering in the fiber. 28

Figure 2.6. A typical single ϕ-OTDR trace of the Monte Carlo simulation. . . 32

Figure 2.7. Impact of averaging on ϕ-OTDR traces. 32

Figure 2.8. Spectrogram of the simulated 19.2 seconds long ϕ-OTDR data with a synthetic vibration of 833.33 Hz at the 84. range bin. 33

Figure 2.9. K virtual acoustic channels synthesized on the sensing fiber 34

Figure 2.10. A-scope presentation of three consecutive ϕ-OTDR traces: Samples denoted by red dashed lines are collected on the slow-time axis. 35

Figure 2.11. Signal parameters related to the slow-time and fast-time axes. . . 36

Figure 2.12. 3-D demonstration of six consecutive real ϕ-OTDR traces. 36

Figure 3.1. AMF results for the 1. data set for different values of N. 46

Figure 3.2. AMF results for the 2. data set for different values of N. 47

Figure 3.3. AMF results with steering vector tuned to non-vibrating frequencies. 47

Figure 3.4. AMF test results for the multiple vibrations at the range bins 84 and 244 with identical frequencies with $\rho = 10$ and $L_s = \{1, 2, 3\}$. 49

Figure 3.5. AMF test results for the multiple vibrations at the range bins 84 and 244. FIR is tuned to 833.33 Hz with with $\rho = 10$ and $L_s = \{1, 2, ..., 80\}$ 49

Figure 3.6. AMF test results for the multiple vibrations at the range bins 84 and 244. FIR is tuned to 833.33 Hz with $\rho = 1$ and $L_s = \{1, 2, ..., 80\}$ 50

Figure 3.7. AMF test results for the multiple vibrations at the range bins 84 and 244. FIR is tuned to 833.33 Hz with $\rho = 0.1$ and the secondary data including all the range bins exluding the 84. and 244. 51

Figure 3.8. AMF test results for the multiple vibrations of 833.3 Hz at the 84. range bin and 625 Hz at the 244. range bin. FIR is tuned to 833.3 Hz with $\rho = 1$ and the secondary data including all the range bins. 52

Figure 3.9. AMF test results for the multiple vibrations of 833.3 Hz at the 84. range bin and 625 Hz at the 244. range bin. FIR is tuned to 625 Hz with $\rho = 1$ and the secondary data including all the range bins. 52

Figure 3.10. AMF test results for the multiple vibrations for two different DL levels. The FIR is tuned to 625 Hz. 53

Figure 4.1. Empirical and analytical eigenvalue distributions of a Wishart random matrix for 4 different values of κ after 500 trials with size 50×50. 61

Figure 4.2. TW distributions representing the distribution of the largest eigenvalue of the SCM for 3 different (real, complex and quaternion valued) SCMs. 64

Figure 4.3. Segmentation of the input data for the MED technique. 65

Figure 4.4. Probability of false alarm versus threshold: (a) for fixed TPI and with various κ values; (b) for square data matrices ($\kappa = 1$) for various TPI durations. 68

Figure 4.5. MED(M, L) processing results of the 1. data set with $M = 5$, $L = \{10, 20, 40, 100\}$ and $P_{FA} = 0.01$. 70

Figure 4.6. MED(M, L) processing results of the 1. data set with $M = 10$, $L = \{10, 20, 40, 100\}$ and $P_{FA} = 0.01$. 71

Figure 4.7. MED(M, L) processing results of the 1. data set with $M = 15$ $L = \{10, 20, 40, 100\}$ and two P_{FA} vaues:$P_{FA} = 0.01$ and $P_{FA} = 0.001$. . 71

Figure 4.8. MED test results of the 1. data set for 4 different square sized data matrix ($M = L$) selections. $L = \{10, 20, 30, 40\}$ with $P_{FA} = 0.01$. 72

Figure 4.9. MED test results of the second data set with matrix $M = 5$, $L = \{10, 20, 40, 100\}$ and $P_{FA} = 0.01$ 73

Figure 4.10. MED(M, L) processing results of the 2. data set with $M = 10$, $L = \{10, 20, 40, 100\}$ and $P_{FA} = 0.01$. 74

Figure 4.11. MED(M, L) processing results of the 2. data set with $M = 15$, $L = \{10, 20, 40, 100\}$ and two P_{FA} vaues:$P_{FA} = 0.01$ and $P_{FA} = 0.001$. 74

Figure 4.12. MED test results of the second data set for 4 different square sized data matrix ($M = L$) selections. $L = \{10, 20, 30, 40\}$ with $P_{FA} = 0.01$ 75

Figure 4.13. MED test results of the 4. data set with $M = 15$, $L = \{20,30,40,50, 100,200\}$ and $P_{FA} = 0.01$ 76

Figure 4.14. MED(M,L) processing results of the 4. data set with $M = 20$, $L = \{20, 30, 40, 50, 100, 200\}$ and $P_{FA} = 0.01$ 77

Figure 4.15. MED test results of the 4. data set with $M = 25$, $L = \{20,30,40,50, 100,200\}$ and $P_{FA} = 0.01$ 78

Figure 4.16. The MP bounds and the empirical eigenvalue distributions of the 1. data set computed from 50 consecutively TPI blocks at the 80. range bin (H_0 case) and the 84. range bin (H_1 case). 79

Figure 4.17. The MP bounds and the empirical eigenvalue distributions of the 2. dataset computed from 50 consecutively TPI blocks at the 80. range bin (H_0 case) and the 84. range bin (H_1 case). 80

Figure 4.18. The outliers of the empirical eigenvalue distribution at the vibration range bin for the 4th dataset: 200 consecutive TPI's were processed with six different data matrix sizes with $M = 20$ and $L = \{20, 25, 30, 40, 50, 75\}$. 81

Figure 5.1. The general indoor setup for gathering real ϕ-OTDR data. 84

Figure 5.2. The FOTAS system: the 1st prototype developed for real data gathering. 85

Figure 5.3. Typical ϕ-OTDR trace recorded indoor by the FOTAS system. . . 87

Figure 5.4. AMF test results for f_{p_1}= 833.33 Hz. 89

Figure 5.5. AMF test results for f_{p_2}= 1225 Hz. 90

Figure 5.6. Indoor test results with the steering vector tuned to non-vibrating frequencies (i.e., $f \neq f_{p_i}, i = 1, 2$). 91

Figure 5.7. The slow-time power spectrum densities of the simulated and measured vibrations at the target range bins with 833.33 Hz vibration. 92

Figure 5.8. SNR vs N plots of the simulated and measured ϕ-OTDR data. . . 92

Figure 5.9. Power spectrum density plots for the last two data sets at the target range bin. The upper row is the time-frequency plot for the whole recording while the graphs in the lower designate the periodogram plots. 93

Figure 5.10. SNR vs N curve for the strong and weak vibration ϕ-OTDR data with $f_p = 833.33$ Hz. 94

Figure 5.11. AMF test results for the weak vibration of 833.33 Hz without the acoustic isolation. 95

Figure 5.12. Normalized intensity at the AMF output for the vibration of 1225 Hz at the 104. range bin with the acoustic isolation. 96

Figure 5.13. Normalized intensity at the AMF output for the vibration of 841 Hz at the 307. range bin with the acoustic isolation. 96

Figure 5.14. Normalized intensity outputs for the weak vibration of 1225 Hz with the acoustic isolation. 97

Figure 5.15. Normalized intensity at the AMF output for the vibration of 841 Hz with the acoustic isolation for two DL level selections. 98

Figure 5.16. Normalized intensity at the AMF output for the vibration of 1225 Hz with the acoustic isolation for two DL levels and secondary data regions. 98

Figure 5.17. Normalized intensity at the AMF output for the vibration of 841 Hz with the acoustic isolation for two DL levels and secondary data regions. 99

Figure 5.18. AMF filter output versus ρ values for the 1. test. 101

Figure 5.19. AMF filter output versus ρ values for the 2. test. 102

Figure 5.20. Sample test results of MED(M, L) for the strong 833 Hz vibration with $M = 10$ and $L = \{15, 20, 25, 30, 40, 50\}$. 103

Figure 5.21. Sample test results of MED(M, L) for the strong 1225 Hz vibration with $M = 10$ and $L = \{15, 20, 25, 30, 40, 50\}$. 104

Figure 5.22. Sample test results of MED$(10, L)$ for the weaker 833 Hz vibration with $L = \{15, 20, 25, 30, 40, 50\}$. 105

Figure 5.23. Sample test results of MED$(15, L)$ for the weaker 833 Hz vibration with $L = \{15, 20, 25, 30, 40, 50\}$. 106

Figure 5.24. Sample test results of MED$(15, L)$ for the weaker 833 Hz vibration with $L = \{15, 20, 25, 30, 40, 50\}$ after high-pass filtering with a cut-off frequency of 500 Hz. 107

Figure 5.25. Sample test results of MED$(15, L)$ for the weaker 833 Hz vibration with $L = \{15, 20, 25, 30, 40, 50\}$ after high-pass filtering and whitening. 108

Figure 5.26. MED$(10, L)$ test results for the multiple vibrations with $L =$\{10,20, 50,75,100,150 \}. 109

Figure 5.27. MED test results for the multiple vibration data with various square data matrix choices. 110

Figure 5.28. MED test results for the multiple vibration data with various square data matrix choices. 111

Figure 5.29. P_D versus P_{FA} for two different vibration strengths. 112

Figure 5.30. The bulk eigenvalue statistics at the vibration range bin for the 833 Hz vibration and their comparison with the relevant MP bounds. . 114

Figure 5.31. The empirical largest eigenvalue distribution of the noise-only SCMs computed for the 1. data set for $\beta = 1$ and various (M, L) choices. 115

Figure 5.32. The empirical largest eigenvalue distribution of the noise-only SCMs computed for the 3. data set for $\beta = 1$ and various (M, L) choices. 115

Figure 5.33. The empirical largest eigenvalue distribution of the noise-only SCMs computed with I-Q recorded data for $\beta = 2$ and various (M, L) choices. 116

Figure 6.1. Generic schematic of the outdoor test points for the 1. test site. . 120

Figure 6.2. The route of the buried FO intranet line at 1. test site. 120

Figure 6.3. Typical raw ϕ-OTDR trace recorded by FOTAS at the 1. test site. 121

Figure 6.4. Generic schematic of the outdoor test points for the 2. test site. . 122

Figure 6.5. The route of the FO cables for the 2. test site. 122

Figure 6.6. Aerial fiber installation near the railway. 123

Figure 6.7. Typical raw ϕ-OTDR trace recorded by FOTAS at the 2. test site. 123

Figure 6.8. Two sample ϕ-OTDR intensity data recorded near TP1 and TP2. 125

Figure 6.9. Recorded intensities at different range bins for TP1 and TP2. . . . 126

Figure 6.10. AMF outputs for TP1 and TP2 with $N = 500$ and $N = 1000$. . . . 128

Figure 6.11. AMF(1000) outputs for TP1 and TP2 signals with for varying DL levels. 129

Figure 6.12. Raw and enhanced 147. channel data for TP3 and TP4. 130

Figure 6.13. Acoustic vibration profiles along the FUT axis after AMF processing with three different N values for the 8. dataset. 132

Figure 6.14. Acoustic vibration profiles along the FUT axis after AMF processing with three different N values for the 9. dataset. 133

Figure 6.15. SAM(100,50) processing results for the five hammer hit durations of the 9. dataset. 134

Figure 6.16. AMF(5000) processing results for the five hammer hit durations of the 9. dataset when the FIR is tuned to 50 Hz. 134

Figure 6.17. The MED test results at a time segment including the 1. hammer hit at 20 m distance from the FUT. 135

Figure 6.18. The acoustic vibration profiles detected by the MED processing at TP3 for the time segment including the 2. hammer hit. 137

Figure 6.19. The normalized FOTAS range profile of the second test site with 100 OTDR frames averaged. The aerial background is significantly different from the buried section. 139

Figure 6.20. AMF and MED processing results for the train pass at 10. and 40. seconds of the record. 141

Figure 6.21. Waterfall presentation of AMF and MED processing of the whole data record for the train pass of the 10. data set. 142

Figure 6.22. AMF and MED processing results for the vehicle passes at 50. and 120. seconds of the record. 143

Figure 6.23. Waterfall presentation of AMF and MED processing of the whole data record for the vehicle passes of the 11. data set. 144

Figure 6.24. AMF and MED processing results for the vehicle passes at 0.2 and 4.8 seconds of the record. 146

Figure 6.25. Waterfall presentation of AMF and MED processing of the whole data record for the digging activities of the 12. data set. 148

LIST OF TABLES

Table 2.1. Review of the state-of-the-art DSP methods used for DAS. 25

Table 2.2. DAS Simulation parameters. 31

Table 4.1. Some values of $F_1(\lambda)$. 64

Table 5.1. General test parameters used during indoor experiments. 86

Table 5.2. Indoor data used during testing. 87

Table 5.3. ρ_{opt} and the largest eigenvalue comparison for the 1. test. 102

Table 5.4. ρ_{opt} and the largest eigenvalue comparison for the 2. test. 102

Table 6.1. Modified DAS system parameters for field tests. 124

Table 6.2. Test signals used at test site-1. 125

Table 6.3. SNR results with $f_p = 50$ Hz at TP3. 132

Table 6.4. SNR results with $f_p = 50$ Hz at TP4. 133

Table 6.5. SNR [dB] comparison for MED and AMF methods with varying TPIs. 136

Table 6.6. Summary of the test signals used at test site-2. 138

Table 6.7. Test signals and associated FOTAS system parameters used. . . . 140

LIST OF SYMBOLS

$\{.\}$	Set representation used for range bin indices
$\#\{\cdot\}$	The number of elements of the set $\{\cdot\}$
$\mathbf{1}_{(,)}$	Indicator function
$\lfloor \cdot \rfloor$	Floor function
$[.]^T$	Transpose operator
$[.]^H$	Hermitian transpose operator
$a(t_m)$	Amplitude of the m-th scatterer
$\mathbf{A}^N$	Hermitian random matrix of size N
B	System bandwidth
c	Speed of light in vacuum
$\boldsymbol{e}_i$	i-th eigenvector of the optical covariance matrix
$e(t)$	Excitation signal
$E\{\cdot\}$	Expectation operator
E_0	Launched electric field amplitude at the beginning of the fiber
E_{LO}	The local oscillator electric field amplitude
E_{RBS}	The Rayleigh backscatter electric field amplitude
E_s	Magnitude of the optical Rayleigh backscatter
f	Frequency axis
f_{p_i}	Vibration frequency of the i-th perturbation
f_s	Sampling frequency of the analog-to-digital converter
$f_{\boldsymbol{A}_N}(\cdot)$	Probability density function of the eigenvalues of $\boldsymbol{A}_N$
$F^{\boldsymbol{A}_N}(\cdot)$	Cumulative distribution function of the eigenvalues of $\boldsymbol{A}_N$
$F_\beta(\cdot)$	Tracy-Widom cumulative distribution function of order β
$h(t)$	Impulse response of the fiber optic cable
H_0	Null (vibration absence) hypothesis
H_1	Alternative (vibration presence) hypothesis
$I(t)$	The optical intensity denoted after photo-detection
$\boldsymbol{I}_N$	$N \times N$ identity matrix

k	Range bin index
K	Number of virtual acoustic channels (total range bins)
l	Training cell bin index
l_j	Location of the j-th perturbation
L_s	Set of the range indices used for secondary (training) data
L_{FUT}	Length of the fiber under test
M	Segment size of of the MED technique
M_s	Number of the reflective mirrors per spatial resolution
n	OTDR pulse index
n_{eff}	Refractive index of the FO cable
$\boldsymbol{n}_{rbs}$	Rayleigh backscatter component of the noise vector
$\boldsymbol{n}_w$	Receiver noise term of the noise vector
N	Size of the measurement vector or the FIR filter
p_i	Range bin index of the i-th perturbation (vibration)
P_D	Probability of detection
P_{FA}	Probability of false-alarm
P_0	Power of the acoustic vibration signal
P_{LO}	Power of the optical local oscillator
P_m	Random polarization of the m-th reflecting partition
P_{RBS}	Power of the Rayleigh backscatter
P_t	Laser source power
$q(y)$	Solution to Painleve II differential equation
r_m	Random reflectivity of the m-th reflecting partition
$\boldsymbol{R}_{k-dl}$	Diagonally loaded covariance matrix
$\boldsymbol{R}_s$	Signal covariance matrix
$\boldsymbol{R}_{rbs}$	Covariance matrix of the Rayleigh backscatter
$\boldsymbol{R}_x$	True covariance matrix
$\hat{\boldsymbol{R}}_x$	Sample (estimated) covariance matrix
$\boldsymbol{s}(\omega)$	Temporal steering vector
$SNR_{(k)}$	The fast-time signal-to-noise ratio for the k-th channel
t	Time axis

t_m	Relative delay of the m-th reflecting partition
$T(\cdot)$	Pulse repetition interval
T_p	Pulse repetition interval
TW_β	Tracy-Widom distribution of order β
$u(t)$	Modulating pulse of the optical interrogator
v_g	Group velocity of the fiber
$\boldsymbol{v}(\omega)$	Perturbation(vibration) signal term of the backscatter
w_i	Complex valued tap weight for the i-th pulse
$\boldsymbol{w}$	Weight vector
$\boldsymbol{W}$	Wishart matrix
$x(t)$	Optical Rayleigh backscatter (response of the fiber to $e(t)$)
$x_i(t)$	In-phase of the optical Rayleigh backscatter
$x_q(t)$	Quadrature component of the optical Rayleigh backscatter
$x_{n,k}$	Electrical baseband backscatter signal component at the n-th frame of the k-th range bin
$\boldsymbol{x}_k$	Generic temporal snapshot vector for the k-th range bin
$\boldsymbol{x}_{k/H_0}$	Vibration absent $\boldsymbol{x}_k$
$\boldsymbol{x}_{k/H_1}$	Vibration present $\boldsymbol{x}_k$
$\boldsymbol{X}$	Matrix ensemble of temporal snapshots
$\boldsymbol{X}^T$	Transpose of $\boldsymbol{X}$
z_k	Scalar output of the FIR filter at the k-th range bin
$\boldsymbol{Z}$	Data matrix after the whitening process
α	Rayleigh attenuation coefficient of the fiber optic cable
β	Order of the Tracy-Widom distribution
γ	Binary detection threshold value
$\delta(\cdot)$	Dirac delta function
Δ_f	Frequency linewidth of the laser source
Δz	Spatial resolution of the measurement system
$\Delta\phi$	Relative phase between the OLO and the RBS
η	A scalar constant used during SNR optimization
θ	Relative polarization angle between the OLO and RBS

κ	The ratio of the rows to the columns of the data matrix
λ_1	Maximum (1st) eigenvalue of the covariance matrix
λ_N	Minimum (N-th) eigenvalue of the covariance matrix
λ_{max}	Maximum eigenvalue bound
λ_{min}	Minimum eigenvalue bound
$\lambda_{max}(\cdot)$	Maximum eigenvalue of $(\cdot)$
λ_t	Wavelength of the laser source
$\mu_{ML,\beta}$	Parameter for shifting the center of $f_{\boldsymbol{A}_N}()$
π	Pi number
ρ	Diagonal loading factor
$\hat{\rho}_{opt}$	Optimum diagonal loading factor
$\sigma_{ML,\beta}$	Parameter for scaling the variance of $f_{\boldsymbol{A}_N}()$
σ_p^2	Vibration signal power
σ_w^2	White noise power
σ_η^2	Average total noise power
τ	Laser pulse width
ϕ_p	Phase of the perturbation
ω_s	Laser source frequency
ω_{IF}	Intermediate angular frequency (of the AOM)
ω	Angular frequency
$\overline{\omega}$	Normalized sampling frequency (f/PRF)

LIST OF ACRONYMS/ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
a.u.	arbitrary units
A/D	Analog-to-Digital
ADC	Analog-to-Digital Converter
AMF	Adaptive Matched Filter
AOM	Acousto Optic Modulator
APD	Avalanche Photo Diode
ASE	Amplified Spontaneous Emission
AWG	Arbitrary Waveform Generator
BİLGEM	(*tr*). Bilişim ve Bilgi Güvenliği İleri Teknolojiler Araş. Mer.
BLF	Bilateral Filter
BPD	Balanced Photo Dedector
CD	Coherent Detection
CDF	Cumulative Distribution Function
CFAR	Constant False Alarm Rate
CW	Continuous Wave
CWT	Continuous Wavelet Transform
DAQ	Data Acquisiton Card
DAS	Distributed Acoustic Sensing
DD	Direct Detection
DL	Diagonal Loading
DSP	Digital Signal Processing
DVS	Distributed Vibration Sensing
DWT	Discrete Wavelet Transform
EDFA	Erbium Doped Fiber Amplifier
EMD	Empirical Mode Decomposition
EOM	Electro Optic Modulator

FAR	False Alarm Rate
FBG	Fiber Bragg Grating
FIR	Finite Impulse Response
FFT	Fast Fourier Transform
FPGA	Field Programmable Gate Array
FM	Frequency Modulation
FO	Fiber Optic
FOTAS	(*tr.*) Fiber Optik Tabanlı Akustik Sensör
FUT	Fiber Under Test
GPU	Graphical Processing Unit
GOE	Gaussian Orthogonal Ensemble
GUE	Gaussian Unitary Ensemble
GUI	Graphical User Interface
HHT	Hilbert-Huang Transform
HMI	Human Machine Interface
LFM	Linear Frequency Modulation
LO	Local Oscillator
LOE	Laguerre Orthogoal Ensemble
LTI	Linear Time Invariant
LUE	Laguerre Unitary Ensemble
IF	Intermediate Frequency
IL	Insertion Loss
I/Q	In-phase/Quadrature
IIR	Infinite Impulse Response
MAD	Moving Average and Diferential
MED	Maximum Eigenvalue Detection
MSE	Mean Square Error
MMSE	Minimum Mean Square Error
MP	Marcenko Pastur
MR	Maximum Range
MSNR	Maximum Signal-to-Noise Ratio

MSPS	Mega Samples per Second
MWD	Multiscale Wavelet Decomposition
N/A	Not Available (or Not Applicable)
NEP	Noise Equivalent Power
NLL	Narrow Linewidth Laser
OA	Optical Amplifier
OID	Optical Interrogator Device
OLO	Optical Local Oscillator
OMP	Optimum Matching Pursuit
OTDR	Optical Time Domain Reflectometry
ϕ-OTDR	Phase-sensitive Optical Time Domain Reflectometry
PD	Probability of Dedection
PDF	Probability Density Function
PFA	Probability of False Alarm
PIDS	Physical Intrusion Detection System
PM	Polarization Maintaining
POTDR	Polarization Optical Time Domain Reflectometry
PRF	Pulse Repetition Frequency
PZT	Lead Zirconate Titanate (Piezoelectric Transducer)
RBS	Rayleigh Backscatter
RF	Radio Frequency
RIN	Relative Intensity Noise
RMT	Random Matrix Theory
ROC	Receiver Operating Characteristics
SAK	Spatial Average Kurtosis
SAM	Separate Averaging Method
SAVTAG	(*tr.*) Defence & Security Technologies Research Grant Group
SBS	Stimulated Brillouin Scattering
SCM	Sample Covariance Matrix
SDR	Software Defined Receiver
SHM	Structural Health Monitoring

SMF	Single Mode Fiber
SNR	Signal-to-Noise Ratio
SOP	State of Polarization
SR	Spatial Resolution
STAP	Space-Time Adaptive Processing
STFT	Short-time Fourier Transform
TAD	Trace Averaging and Differentiating
TED	Two-dimensional Edge Detection
TP	Test Point
TPI	Temporal Processing Interval
TÜBİTAK	(*tr.*) Türkiye Bilimsel ve Teknolojik Araştırma Kurumu
TW	Tracy-Widom
YHT	(*tr.*) Yüksek Hızlı Tren

1. INTRODUCTION

1.1. What is Fiber Optic Distributed Acoustic Sensing?

It has been more than 30 years that the fiber optic (FO) cables have found usage as sensor elements in industrial applications. The revolutionary technological advances in fiber optics have resulted in the production of various sensor devices for measuring temperature, pressure and vibration, and development of chemical probes [1]. These fiber optic sensors can be divided into two main types as "extrinsic" and "intrinsic" sensors according to the location of the sensing mechanism. In the former case, the transduction mechanism for the sensing application takes place in a remote location and the FO cable serves as the signal conveying element from the sensor to the processing electronics. This type of measurement can be also considered as a single point sensor and have many uses depending on the application. Single point sensors can be multiplexed along the length of a fiber by using different wavelengths of light for each sensor, or by sensing the time delay as the light wave propagates along the fiber through each sensor. In the latter case, the FO cable itself is being used as the sensing element and owing to the capability of taking measurements at every point along the FO cable; this type of sensing is called as "distributed sensing" [1, 2].

The concept of distributed sensing is based on measuring the scattered light along the FO cable that depends on the physical properties of the fiber. There exist three types of scattering mechanisms according to light intensity irradiating the fiber material. These scattering phenomena are known as Rayleigh, Brillouin and Raman scattering which are all suited for monitoring slow-changing parameters along the FO cable. Rayleigh scattering is directly correlated with the wavelength of the laser source used which is called elastic scattering. Brillouin and Raman scattering are called inelastic scattering which shows a frequency shift in the scattered light. In fused silica fibers, Brillouin and Raman scattering are weaker than Rayleigh scattering unless stimulated by high input powers [1, 3].

Among these three scattering mechanisms, in this thesis research, we focus on the Rayleigh scattering which is the most dominant type of scattering. Rayleigh scattering is caused by density and composition fluctuations created in the material during the manufacturing process. It occurs due to random microscopic variations in the index of refraction of the fiber core. When a narrow pulse of light is launched into a fiber, the variation in Rayleigh backscatter can help to determine the approximate spatial location of these variations. Although being relatively insensitive to temperature, Rayleigh backscattering can still be utilized for developing a distributed sensing technique for temperature and strain; hence, it is often regarded as a promising and emerging technology [1].

1.2. Distributed Acoustic Sensing: Historical Perspective and Applications

Since their initial development, both intrinsic and extrinsic sensors have been developed as commercially available products in a wide range of industries. Among the distributed sensing technology, distributed temperature sensing (DTS) seems to be the most common application in the commercial sense which has wide application in diverse areas [1, 2]. Several other uses of distributed sensing technology range from leak detection in oil and gas pipes, monitoring high-voltage electrical distribution lines, complementary solutions for securing the borders of critical infrastructures, onboard diagnostic tools for condition monitoring in avionics and concrete structures, vehicle and railway traffic monitoring and seismic imaging of large scale events. These application areas all exploit the sensitivity of the fiber optic cable to mechanical vibrations that induce strain changes on the fiber medium. Mechanical vibration is a periodic displacement of the medium in which it takes place and distributed vibration sensors respond to these mechanical vibrations as they are converted to dynamic strain changes on the sensing fiber. Apart from temperature sensing, it is quite difficult to tackle the detailed dynamics of distributed vibration sensing (DVS) - which is sometimes used interchangeably with DAS - and the transduction mechanism how the incoming acoustic or seismic waves impinge on the fiber gauge length.

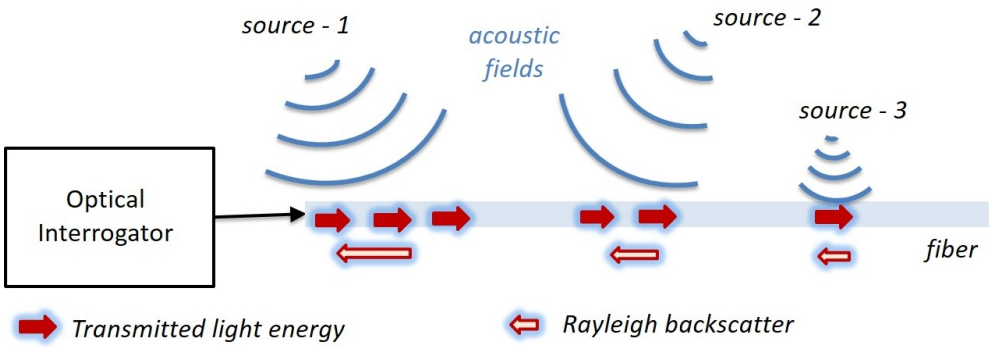


Figure 1.1. Distributed acoustic sensing concept.

As illustrated very basically in Figure 1.1, a DAS system consists of an interrogator device which sends optical pulses to a fiber optic cable and analyzes the weak Rayleigh backscatter. If there is an acoustic field in the surrounding medium, it causes changes in the refractive index of the fiber which in turn modulates the phase of the signal. The sources of the acoustic waves can be of any type, ranging from natural hazards to manmade events such as vehicle motion, and pedestrian walking. The modulations due to these acoustic perturbations can be analyzed to detect and classify such events at any point along the fiber. In this regard, the whole fiber optic cable can be regarded as an array of microphones listening to the environment. The spacing of these microphones is dependent on the interrogation parameters.

There are several governing factors influencing the dynamics of the acoustic waves and their interaction with the fiber gauge length. The phase modulation of the Rayleigh backscattering, hence the response of the DAS/DVS systems are often not related to a direct and absolute measurement of a measurand, but rather concerned with a detection of an "event". One step further on this topic is the "classification of the detected events" that requires advanced signal processing tasks and is perhaps the most important issue which adds "value" to DAS/DVS systems when real-world and commercial use is of concern. The event classification is out of our scope in this thesis research. The most common application areas of DAS/DVS systems are briefly explained below.

1.2.1. Intrusion Sensing and Infrastructure Security

Intrusion detection at the perimeters of critical facilities seems to be the first application of DAS technology according to our survey in the literature and the commercial products in the market. The first application dates back to 1993 of Taylor who introduced the use of phase-sensitive optical time-domain reflectometry (ϕ-OTDR) technology for detecting intruders for perimeter security applications by using a coherent laser source and detecting the changes in the backscattered Rayleigh energy [4]. Taylor's patent later triggered attention on similar applications and several research outcomes have been published since then. Perimeter intrusion detection systems (PIDS) can be deployed either with buried fibers $0.2 - 0.4$ meters deep or fibers included with fences surrounding the critical infrastructure. Footsteps, digging activities, a vehicle approaching or passing by, or similar events can be detected at varying proximities to the fiber installation depending on the strength of the vibration [5–7]. The reliability of PIDS systems is related to their success in correctly deciding in favor of an acoustic event whether it is a "threat" or "benign" event. Reliable detection is closely related to the signal-to-noise ratio (SNR): Setting a high threshold can result in missing very weak responses which may be due to a threatening situation. Conversely, setting a lower threshold will probably increase the false alarm rate (FAR) of the PIDS. These classical problems of signal processing can be alleviated by improving the SNR of the system. Balancing the FAR and probability of miss is not only related to the SNR of the system, but also the discrimination capability of the system between natural events and the events predefined as threats. As an example, for fence installations, the PIDS will raise an alarm when someone is climbing the fence. This alarm is likely to be raised due to fence movements caused by strong winds, animals, or other natural phenomena. Shortly speaking, the credibility of the PIDS systems is highly dependent on its digital signal processing (DSP) power in discriminating the threats from benign events. In this regard, not only detection issues, but feature extraction and pattern recognition methods should be taken into account in the DSP domain for successful PIDS applications.

It is evident that the capability of synthesizing hundreds of virtual acoustic sensors along the full length of FO cable makes this technology an extremely efficient method for monitoring relatively longer distances for homeland security. Other than the PIDS applications which require relatively shorter installations, the borders of a country or the critical pipelines in the oil and gas industries are always big demanders of safety and security monitoring systems. Due to the longer routes ranging from several hundreds of km's to a thousand km long applications, the installation and maintenance cost of the systems required for protecting these assets become very critical. In this sense, the FO based DAS technology again provides a very cost-effective and complementary solution for protecting the pipelines, monitoring wells and reservoirs in the oil and gas industry [8,9]. The biggest challenge in long distance monitoring is the handling of data depending on the number of acoustic channels synthesized. For a 100 km long distance, the interrogation rate setting would be 1 kHz and if 5 meter separation is desired between the channels, then 20,000 channels have to be monitored in real-time for a single installation. When several hundreds of km's long pipeline protection is considered, the situation about data handling becomes more important with repetitively installed several systems. If no event is detected, then the data has to be discarded. Concerning the huge number of channels in such long perimeter or pipeline security applications, there is always merit in temporarily storing the data, and considering backward off-line processing for event classification.

The placement and the protection of the sensing fiber cable itself is also a critical issue in the pipeline applications. It must be buried sufficiently deep or with special protective installation to protect it from unwanted events such as burrowing of animals, water running, etc. For most cases, the fiber is installed on the pipe – either above or below- if leakage or corrosion detection is the primary objective. For most PIDS applications, the fiber is preferably used under the ground in a shallow manner, typically buried 30 cm below the surface. For both short perimeter and long pipeline applications, the relevant system is expected to discriminate the human-based threats from natural events such as animal activities in security related applications of DAS/DVS systems.

1.2.2. Seismic Imaging and Detection of Earthquakes

It is not difficult to foresee the opportunities of DAS systems to be used as alternative candidates for seismometers after the applications mentioned above, since the fiber cables are not only sensitive to external acoustic fields above the ground, they are also sensitive to acoustic fields which can impinge on the fiber from any direction. Hence, it is obvious that the existing optical network infrastructure established in almost every country have a great potential to be used for research and development of the detection of relatively large scale seismic events such as earthquakes [10–13]. Preinstalled fiber optic cables along pipelines, roads or telecommunication fibers can provide a dense sampling of passing seismic wave fields and single-mode fiber (SMF) cables up to several tens of kilometers can provide thousands of acoustic channels with spatial sampling less than 10 m. Such fiber optic cable configurations may not be as sensitive as long-period seismometers, but given a large number of sensor points along the FO route, and depending on the array size and the distance of seismic source, even small and medium- size earthquakes can be detected with good accuracy. One such example is reported for the 480 km long pipeline installation in Turkey, where the DAS sensor array along the pipeline detected earthquakes in the vicinity with magnitudes ranging from 3.6 to 7.2 centered near Van, Turkey [14]. Due to the success of observing such teleseismic events, several research studies and field experiments have been carried out recently to utilize the capability of DAS for monitoring the earthquake wave-fields. As fiber optic infrastructure is widespread, studies on this topic are receiving growing interest and evolving with exemplifying field tests. It was demonstrated that DAS records showed a high degree of correlation with collocated seismometers and indicated that the DAS sensitivity is very similar at 0.8-1.6 Hz. DAS technology also enables to gather additional information such as the direction of seismic energy with special array configurations. Although the application of DAS in earthquake seismology is not mature yet, it seems that it will play an important role in passive seismology and extend the domain of seismological measurements into many un-instrumented fields. DAS technology has great potential to provide complementary solutions for present early warning systems and hazard prevention.

DAS can be considered as a low-cost acquisition system for seismic profiling in wells that already contain fiber optic cables [15]. When the dense sampling capability of FO based DAS is compared with conventional geophones, vertical seismic profiling with fibers is more affordable than renting and locating several geophones [16]. Several experimental field tests have been reported in the literature for comparison of DAS capability with conventional geophones.The current sensitivity of the DAS technology have been reported insufficient to detect P-waves during the conducted experiments, since the fiber optic cables are sensitive to strain along the cable length but not to the strain perpendicular to the fiber. In addition, the SNR values observed were 40-50 dB below to the ones achieved with geophones. Nevertheless, the DAS sensitivity could be improved with proper SNR enhancement techniques which have never been adapted in this application area before.

1.2.3. Structural Health and Condition Monitoring

Long term reliability and ease of installation of DAS/DVS technology to real-life structures has also triggered applications in monitoring of cracks in concrete structures. Distributed fiber optic sensing allows potential monitoring of different types of structures where it is possible to detect and locate not only the premature emergence of cracks and stresses, but also their evolution and behavior [17]. Several reports have been published regarding the potential use of DAS for structural health monitoring (SHM) applications [18].

Integration of DAS/DVS with civil infrastructure is an active research area [19]. Several demonstrations have been carried out for monitoring buildings, piles, and bridges. The dominating sensor type in these applications is based on fiber bragg grating (FBG) and array processing of many FBGs which is considered as a quasi-distributed sensing. Nevertheless, “fully distributed” sensing is a promising technology for monitoring large areas in diverse areas of SHM.

1.2.4. Transportation Monitoring

ϕ-OTDR based DAS/DVS systems have also been introduced as a way of vehicle traffic monitoring in [20]. Since then, although there has never been an implemented commercial application for monitoring traffic on highways, there exists efforts and ongoing interest to evaluate this technology for tracking high-speed trains for railway safety and monitoring applications [21, 22]. Since trains are relatively massive compared to vehicles on the roadways, the vibrations induced are strong enough to be detected by ϕ-OTDR systems with buried fibers parallel to the railway. Without any complicated signal processing, and by only analyzing the received energy at every acoustic channel, the train motion can be tracked real-time, and its speed can be easily estimated.

1.3. Motivation and Objectives

In this thesis, our primary objective is to seek robust and efficient signal detection algorithms for DAS/DVS applications in terms of adaptivity to real-world conditions and sensitivity to weak vibrations. The core motivation behind this objective is due to the limited DSP related research published for DAS applications. Although numerous experimental research studies have been carried out, and publications for DAS are available, according to our literature survey, most of the research is based on optical means of signal processing rather than DSP methods. Efforts are mostly dedicated to methods for increasing the sensing range of ϕ-OTDR systems and SNR enhancement.

In the first phase of our study, we focused on adapting robust signal detection algorithms from radar theory to DAS systems. Since the system operation principles and the surrounding environmental conditions in DAS applications resemble to a large extent the operational principle and the clutter and interference conditions of radar systems, we considered in many aspects, the adaptation of noise mitigation techniques which have not been tested for DAS applications before. Thus, we started from adaptive techniques which have – surprisingly never – been used for such sensor systems. The adaptive filtering technique reported in this thesis study requires the estimation of

the noise statistics from "signal-free" assumed noise samples. In addition, the spectral response of the target vibration is also required for optimum detection. When this prior information is known, a matched filter structure is employed after a whitening stage which is inherent in the proposed processing scheme. Hence, one of the main disadvantages of the AMF is requirement of the vibration frequency information for optimum adaptive processing. In the absence of this spectral information, a brute force search methodology can be implemented in DAS applications at the expense of increasing complexity and costs.

In the second phase, we adapted another technique which is based on computing the maximum eigenvalue of the covariance of the optical backscatter. Eigenvalue based detection techniques have found applications in cognitive radio and radar applications. As a second objective in this research, we pragmatically aimed to adapt these techniques in the temporal domain for DAS applications.

Last but not least, one of the core objectives of this thesis research was based on experimental work with real data. Although we started to test our above mentioned approaches with computer simulations, a special emphasis was given to the development of a real-world DAS system. Significant effort was dedicated for setting up different ϕ-OTDR system configurations in the laboratory and conducting numerous tests both indoor and outdoor test facilities. Since TÜBİTAK (The Scientific and Technological Research Council of Turkey) BİLGEM (Informatics and Information Security Research Center) has a well-established infrastructure with several relevant R&D laboratories, the second core motivation in this research was to develop a real system, and use it to validate our proposed signal processing techniques.

1.4. Organization of the Thesis

In Chapter 2, the basic detection schemes employed in Rayleigh scattering based ϕ-OTDR systems are summarized. Brief information about the shortcomings of conventional DSP techniques is included, and state of the art is reviewed with a summary of the literature survey on the recent DSP methods. The theoretical model for

the Rayleigh backscattering is included with numerical implementation information. The problem of vibration detection is described with the ϕ-OTDR signal model at the end of the chapter.

Chapter 3 presents our first approach for DAS systems: the AMF technique. The performance metric used in this thesis is introduced and the AMF algorithm is summarized. The performance of the AMF is tested with computer simulated data and sample results are included at the end with remarks and discussion.

Chapter 4 presents the MED technique and a constant false-alarm rate (CFAR) threshold selection scheme. Special emphasis was given to RMT and the two main theorems are described: Marchenko-Pastur and Tracy-Widom. These two theorems are of utmost importance in the context of eigenvalue based detection techniques to understand how the eigenvalues of a sample covariance matrix can be evaluated for vibration detection in DAS systems. The threshold level for the binary decision rule is based on the Tracy-Widom distribution which is briefly described. Similarly, the MED technique is tested again with computer simulated data, and results are discussed.

Chapter 5 is dedicated to the introduction of the real ϕ-OTDR system developed, and the first experiments which were carried out indoor for the verification of the proposed algorithms. The system was developed within the scope of an internally funded project by BİLGEM, called FOTAS[1] which was commenced by the initiatives of the author of this thesis. The indoor built ϕ-OTDR configuration was a very lean system; it did not include any optical amplification neither at the transmit stage nor at the receiver stage. The indoor experiments were carried out with the DAS system connected to long fiber spools, and synthetic vibrations were induced with a piezoelectric transducer (PZT) based phase shifter connected in between these fiber cables. Both AMF and MED techniques were tested with real data, and results are summarized.

[1]Fiber Optik Tabanlı Akustik Sensör (in Turkish)

Chapter 6 includes sample results gathered from the field tests of the real DAS system prototype. Two different outdoor test sites were configured in the TÜBİTAK Gebze campus for the validation of the AMF and the MED techniques. The first outdoor test configuration was based on an approximately 315 m long buried intranet line, and the second test site was configured with 2.4 km long fibers installed on fences. This second line with longer routing enabled more realistic tests as the number of channels was increased. It was also routed in parallel to a railway line so that the DSP methods were also tested to detect and track high-speed trains which can be considered as a transport monitoring application.

And finally, Chapter 7 provides our concluding remarks and future prospects for DAS systems and for our proposed signal processing approaches.

1.5. Contributions

The main contributions of this thesis work can be summarized as follows:

- The research reported here is a part of ongoing experimental work mainly focused on the investigation of adaptive signal detection methods for DAS systems. Due to the similarity of the operational concepts of the conventional ϕ-OTDR with the radar principles, the classical binary detection problem was visited with emphasis on space-time adaptive processing (STAP) concept. STAP is an innovative tool used in phased-array deployed radar systems to cancel out the clutter and interference terms [23]. In STAP applications, due to the antenna diversity and multiple receiver channels, the received signal is a function of both space and time and clutter rejection techniques are employed in both Doppler and spatial domains [23]. In our DAS application, a similar approach is applied by considering the optical interrogator as a single receiver channel and adaptive weight computations are done only in the temporal domain.

- Inspired by the successful application of eigenvalue based detection techniques in cognitive radio, the second contribution of our research is the application of maximum eigenvalue detection for acoustic vibration detection. The eigen-structure of the correlation of the optically measured ϕ-OTDR data is exploited and the MED technique is used for detecting acoustic vibrations. The main advantage of the MED is that it does not require information about the frequency content of the vibration and a CFAR can be achieved with recent results from the RMT [24]. RMT is another popular research area which studies the spectral distribution of random matrices and has gained interest in diverse fields of science including statistical inference, applied physics, finance, cognitive radio, and radar. In the absence of vibration for any DAS application, the received samples are pure noise; hence, the data and the correlation matrices are pure random matrices. In this research, the noise only assumed observations were exploited for the binary detection problem, and we opened the door for efficient use of RMT for DAS/DVS applications.

Briefly, the two detection approaches mentioned above and the application of RMT, to the best of our knowledge, have been applied for ϕ-OTDR based DAS systems for the first time.

2. PHASE-SENSITIVE OPTICAL TIME-DOMAIN REFLECTOMETRY (ϕ-OTDR) FOR DISTRIBUTED ACOUSTIC SENSING

2.1. Introduction

One of the most popular approaches employed in DAS applications is based on the analysis of Rayleigh backscattering (RBS) via ϕ-OTDR which is often regarded as a promising and emerging technology [1]. In this chapter, we briefly present this common method by describing the general detection schemes. The mathematical model for the RBS is also visited to carry out computer simulations and to evaluate the performance of the proposed DSP methods with simulated data.

The conventional RBS based OTDR was invented in early 1970s which is an effective technique for investigating characteristics of fiber optic cables [25]. The ϕ-OTDR technology that we focus in our research works very similarly to a conventional OTDR system. In the conventional OTDR, a broadband (incoherent) light source is used while in the ϕ-OTDR case narrow linewidth laser (NLL) sources are required. In both cases, light energy is transmitted to a FO cable and the optical RBS is analyzed in the time-domain. If a continuous wave (CW) laser is used with pulse modulation, which is the most common method employed in ϕ-OTDR, the operation resembles the classical pulsed radar in many aspects. The backscattered signal is easily associated with a position along the FO cable route with the time-of-flight information between the sent pulses and detected backscattered signals.

In analogy to the pulsed radar case, to increase the maximum range one must also decrease pulse repetition frequency (PRF). Besides, "the coherence length of laser source" used in ϕ-OTDR is the most critical factor governing the maximum detectable range. Coherence length of a laser defines the maximum length that waves of different path lengths in the medium sum up coherently, to produce interference. This is of

great importance in order to achieve phase sensitive detection. The theoretical bound for setting the "maximum detectable range" for the fiber under test (FUT) given by

$$L_{FUT(max)} = c/\left(2n_{eff}\Delta_f\right) \tag{2.1}$$

where n_{eff} is the refractive index of the FO cable, Δ_f is the frequency linewidth of the laser source and c is the speed of light in the vacuum. If a DAS system is to be designed for monitoring a fiber optic line of 50 km, given that $n_{eff} = 1.467$, the linewidth of the laser source must satisfy $\Delta_f \leq 3.10^8[m/s]/2\times1.467\times5.10^4[m]$. Hence, a laser source with a linewidth of at most 2 kHz should be selected.

2.2. Operational Principles of OTDR

The generic OTDR system is represented in Figure 2.1 below. The detection techniques encountered in OTDR systems generally fall into two main classes that employ similar architecture and hardware: direct detection and coherent detection (heterodyning) [1, 2]. In the direct detection OTDR (DD-OTDR), the RBS signal power is directly detected by a photo-detector, usually by an avalanche photo-diode (APD), as depicted by dashed lines.

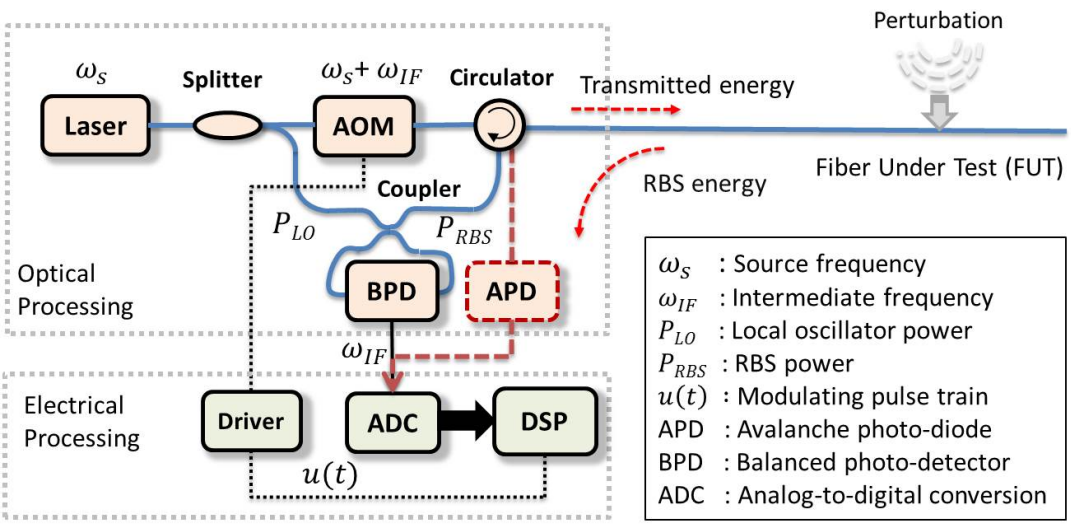


Figure 2.1. Schematic of the direct and heterodyne detection schemes for OTDR.

Currently, due to its simplicity and insensitivity to the spectral characteristics of the optical source, DD-OTDR is used by most commercial OTDR systems in fault inspections. The other detection scheme is the above mentioned heterodyne detection which is also called coherent detection OTDR (CD-OTDR). In CD-OTDR based systems, some of the source power is used as the local oscillator (LO) for coherent mixing with the backscattered optical signal.

If the width of the laser pulse propagating along the fiber is τ seconds, then the minimum distance required for distinguishing two closely spaced events along the fiber length can be computed by $\Delta z = c\tau/2n_{eff}$. The resolving capability of the OTDR system is called the "spatial resolution (SR)". As the laser pulses are repetitively transmitted along the fiber axis, the repetition rate is bounded by the monitored fiber length L_{FUT} which is given by $PRF = c/(2L_{FUT}.n_{eff})$ [Hz]. The number of "virtual sensors" K, synthesized along the sensing fiber is then equal to the sensing fiber length divided by the spatial resolution, i.e., $K = \lfloor (L_{FUT}/\Delta z) \rfloor$ where $\lfloor . \rfloor$ denotes the floor function.

When the FUT length to be monitored is not very long (e.g. 20 km), the strength of the RBS is not so weak, and DD-OTDR systems are often preferred as an effective solution due to their simplicity and low complexity compared to CD-OTDR systems.

For monitoring long distances, the signals to be detected are significantly weak and CD-OTDR becomes an attractive solution to improve the sensitivity of the DAS systems. The underlying principles of CD-OTDR systems can be better understood by analyzing the system depicted in Figure 2.1. A NLL CW laser is used as the light source. The output of the laser is split into two branches by a fused-fiber directional coupler. One signal path is routed for transmission along the FO cable, while the other branch is used as the optical local oscillator (OLO). The transmitted CW signal is pulse modulated by either an acousto-optic modulator (AOM) with predetermined pulse width and PRF values appropriate the required SR and maximum range (MR) requirements. An electro-optic modulator (EOM) can be used instead of an AOM according to the desired application requirements (e.g. very narrow pulse).

The AOM also introduces an intermediate frequency (IF) shift ω_{IF} (typically of a few hundred MHz) to the optically transmitted signal. Booster amplifiers at the transmit stage or pre-amplifiers at the receiver stages might be of concern to enhance the SNR and to increase the dynamic range of the system [26] for both detection types. Although it is not shown in Figure 2.1, the AOM may be either preceded or followed by a specially designed booster Erbium Doped Fiber Amplifier (EDFA), when necessary. A low noise pre-amplifier EDFA followed by an optical band-pass filter might be of concern at the receiver stage to improve the signal quality, especially for DD-OTDR systems.

In the CD-OTDR systems, the backscattered signal is mixed with the reference OLO signal by using a 2-by-2 coupler before the PD stage. After the PD circuitry, the optical intensity denoted by $I(t)$ will be proportional to the interference intensity of the LO signal E_{LO} with the backscattered signal E_{RBS} which can be expressed as

$$I(t) \propto E_{LO}^2 + E_{RBS}^2 + 2E_{LO}E_{RBS}\cos(\theta(t))\cos(\omega_{IF} + \Delta\phi(t)) \tag{2.2}$$

where, $\theta(t)$ and $\Delta\phi(t)$ are the relative polarization angle and phase between the OLO signal and the backscattered optical signal, respectively. With the assumption that the 180 degree phase shift is preserved between the arms of the coupler, a balanced photo-detector (BPD) with perfectly matched arms will reject the direct detection component (DC) and common mode terms. Following this further SNR improvement at optical-electrical conversion stage, the electrical signal with the intermediate frequency ω_{IF} is down-converted to baseband with an electrical LO and appropriate filtering. This stage is usually an RF in-phase & quadrature (I/Q) direct down-conversion stage followed by an analog-to-digital conversion (ADC). The complete chain of electrical signal processing blocks can form an ensemble of a single electrical subsystem as depicted in Figure 2.1. The resultant baseband signal can be given by

$$I(t) \propto 2E_{LO}E_{RBS}\cos^2(\theta(t) - \Delta\phi(t)) \tag{2.3}$$

The differences between the direct and the coherent detection systems can be easily deduced from (2.3). The main advantage of the heterodyning comes from the fact that the intensity of signal of interest ($\propto E_{LO}E_{RBS}$) will be higher than the intensity of the signal using direct detection ($\propto E_{RBS}$). Among all conventional detection types used in optical systems such as the lidar and the conventional OTDR systems, the average SNR gain in coherent detection is over 20 dB higher than that of the direct-detection at 1 MHz detection bandwidth [26–28]. By using appropriate OLO power, the SNR of coherent detection can be made shot noise limited for weak signals. Therefore, coherent detection is generally dependent on changes of the state of polarization (SOP) along the optical paths of E_{LO} and E_{RBS} since the amplitude of the interference between the LO and backscattered signal strongly depends on their relative SOP as can be seen from (2.3). This amplitude can even be zero for orthogonal polarizations. Thus, the coherent detection setups are always very sensitive to SOP changes, which is the main drawback of this heterodyne based detection scheme.

The final backscattered vibration signal in ϕ-OTDR systems is usually very weak and has to be carefully processed. A typical ϕ-OTDR data received from a CD-OTDR system with 28 km FUT length and an adjusted measurement range of 40 km is shown in Figure 2.2.

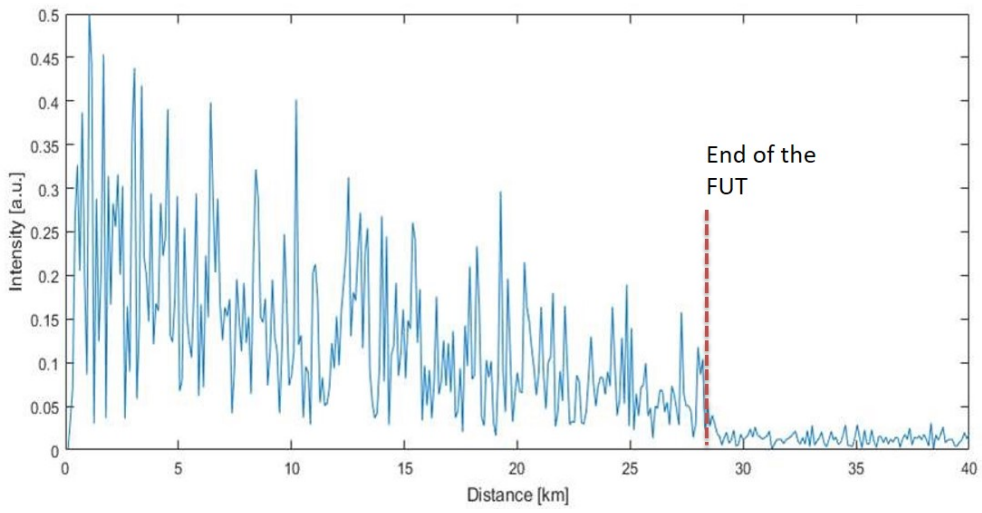


Figure 2.2. Intensity profile of real $\phi-$OTDR data for a 28 km long fiber cable.

CD-OTDR frames exhibit high amplitude fluctuations due to the stochastic nature of the RBS and the strength of this backscattered power decays exponentially due to the inherent loss factor of the fiber. The backscattered power at the beginning of the FUT is relatively much higher than that of the backscatter at the far end, thus making the RBS based DAS system very sensitive to environmental acoustic sources while this background impact becomes much lower at longer distances.

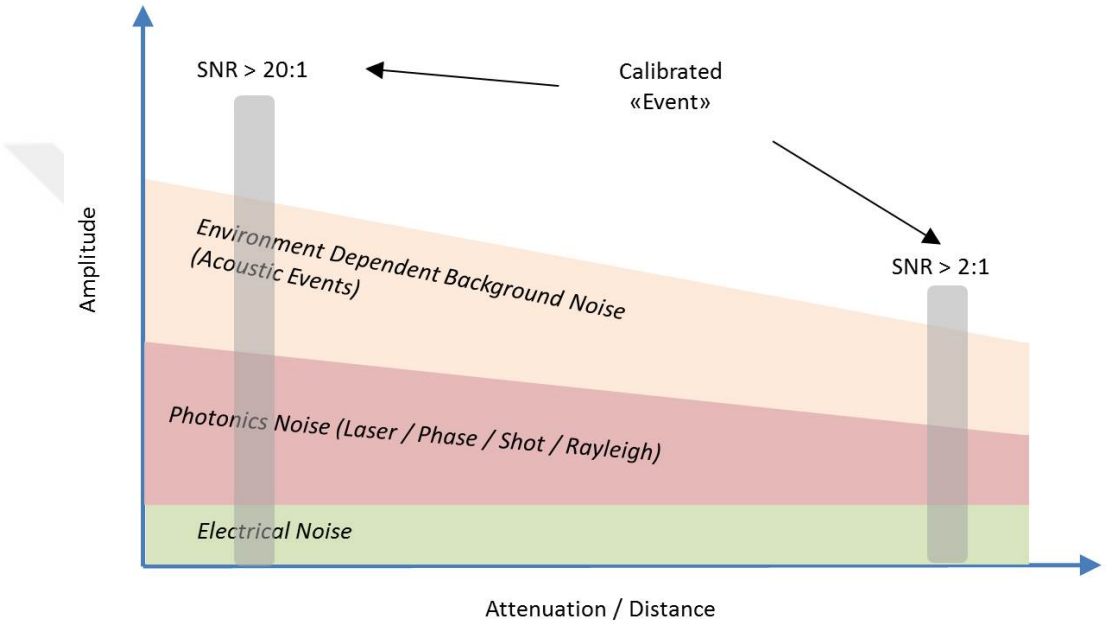


Figure 2.3. Relative strength of different noise sources in a ϕ-OTDR system according the distance.

ϕ-OTDR based systems experience a degradation of signal with distance, due to the attenuation of the fiber optic cable, which can impact sensitivity. Systems should be tested at their maximum reach to verify worst case detection capability. Noise in a fiber optic sensing system includes electrical noise, a number of photonics-related noise sources (RIN: relative intensity noise/phase/shot/Rayleigh noise), and can also include environmental or background activity noise. It is recommended by experienced manufacturers of DAS systems that the worst case SNR must be greater than 2 for meaningful detection and classification. For short distances typical SNR is > 20. At 40-50 km typical SNRs are $\simeq 2$ to 5 as depicted in Figure 2.3.

2.3. Literature Review and State of the Art

Several techniques have been studied and reported in the literature to enhance the performance of the ϕ-OTDR based DAS systems in terms of maximum sensing range, spatial resolution and SNR. As mentioned previously, in the conventional architecture with pulse modulation, the maximum acoustic bandwidth, or the maximum detection bandwith (MDF) is inversely proportional to the maximum range. Due to this trade-off, some of the research efforts reflect a competitive challenge in achieving better performance for both parameters at the same time. This could be possible by using dual wavelengths and interleaving the pulses. As there is no accepted generic architecture for DAS systems, hybrid methods that incorporate different technologies are also of interest to increase acoustic bandwidth while maintaining the maximum sensing range. One novel method that can achieve high range detection while maintaining the maximum detectable frequency, or vice versa, are based on using hybrid sensing technologies, such as adding an interferometric subsystem to the ϕ-OTDR detection scheme [29]. Diversity techniques in the modulation frequency of the laser source can also be used to improve the MDF without sacrificing the target range [30]. A rule of thumb to increase the maximum sensing range is to use optical amplification for higher SNR levels. However, due to the non-linear scattering mechanisms inherent to the fiber optic physics, the maximum optical power launched at the FO cable is limited. Stimulated Brillouin Scattering (SBS) is one such nonlinear effect that limits the optical power when transmission over fibers is considered [31]. Above some critical SBS threshold, the phase noise dramatically increases which seriously deteriorates the signal and decreases the sensitivity of the system. Techniques based on phase modulating the laser source and linewidth broadening can be employed to overcome these challenges [32].

Some part of the laser source power can be used at the receiver stage to enhance the weak scatter response as mentioned previously in Section 2.2 [26, 33]. Although appropriate mixing the source power with the weak RBS signal enhances the detection in the heterodyning, the change of SOP between the interfering optical fields is one of the main problems encountered for this detection scheme. The randomly varying

phase difference between the source and backscattered energy may even result in a zero current at the photodetector output when the difference between the phase of the OLO and the phase of the RBS signal is an integer multiple of π. Hence, due to these factors, the amplitude fluctuations can mask the phase modulation of the perturbation. The instability of coherent detection based ϕ-OTDR systems can be overcome by maintaining the SOP constant. This can be accomplished by using all polarization maintaining (PM) components or by using polarization diversity techniques in the setup including the sensing fiber [33–35]. On the other hand, using PM components significantly increases systems costs as well as adding new ADC channels to the system to process the extra polarization information.

Raman amplification is one of the popular methods when an increase in the sensing range is the primary objective. It can be used to assist the ϕ-OTDR pulse power to maintain its power high enough along the whole fiber length, thus extending the maximum sensing range. Several research studies have been reported which uses first-order or second-order Raman amplifications to increase the sensing range over 125 km with a spatial resolution of 8 m and vibration measurements up to 380 Hz [36–38].

A very recent study reported a maximum sensing range of 125 km without the use of any optical amplification [39]. In this study, a suitable SNR value is obtained by combination of different fiber types with different attenuation coefficients. This novel use of different types of fibers to optimize the far end of the FUT seems to have achieved the highest maximum sensing range currently known without any inline amplification.

It is clear that the above mentioned techniques are related to the improvements or novelties in the optical path, and during our literature survey, it was observed that most of the research efforts were dedicated to signal processing by optical means rather than the digital processing. A concise summary of the above mentioned techniques and the relevant literature for DAS/DVS applications are given in [40], and a tabulated review of the applied techniques for ϕ-OTDR are given in chronological order in [38].

Apart from the methods employed in the optical path, several alternative DSP techniques exist for improving the SNR. Trace averaging and differentiating (TAD) techniques can be used with proper data treatment to improve the SNR by $\sqrt{N}$ where is N the number of averages [4, 33]. The basic TAD technique is simply based on coherently adding the sequentially arriving ϕ-OTDR traces, taking the average of them and differentiating to reveal the acoustic vibrations. This separate averaging method (SAM) significantly reduces the maximum detection bandwidth. Using a moving average instead of seperate windows can reduce this effect but this method still removes higher frequencies, therefore bandwidth reduction still occurs [41]. Both SAM and the moving average and differential (MAD) methods suffer from the inherent low bandwidth of the averaging technique. According to the Nyquist theorem, the MDF is half of the sampling frequency (i.e. PRF/2 Hz). When using SAM with an averaging size of M, the MDF is reduced to PRF/$2M$, while MDF is only increased at most twice this value with the MAD method (i.e. MDF=PRF/M).

Image processing techniques can be applied to DAS systems by considering the ϕ-OTDR traces in 2-D format. One method based on 2-D edge detection (TED) and 2-D processing of backscattered optical traces was introdeced [42] which can increase the SNR to 8.5 dB compared to these conventional methods. In this method, a 2-D image composed of Rayleigh traces is resolved into a series of small matrices and the gradient of the images are calculated at each point by convolving with Sobel operator. The SR value can also be enhanced with some degree with this method, but due to this gradient operation, the frequency response of the system deviates from its original response. Another 2-D technique based on 2-D bilateral filtering (BLF) can be used to enhance the SNR of the ϕ-OTDR based DAS systems better than the TED method presented in [42] in the expense of increased processing power [43]. The BLF method recovers the original vibration signal by a weighted averaging method which results in a denoising performance of 13 dB higher than that of the TED approach. Another TED method based on Prewitt operator was visited for DAS applications in [44] to identify the temporal and spatial distribution of vibrations in the Rayleigh traces, which allows SNR values higher than 30 dB.

Besides the conventional time-domain techniques mentioned above, transform domain methods can be used as alternatives for vibration detection. The classical fast Fourier transform (FFT) based spectrum analysis techniques were used to cope with the SOP changes in DVS systems [45] and multiple simultaneous vibrations can be observed by taking the FFT of every acoustic channel of the DVS system. SNR values as high as 20 dB are reported in [46], but it is clear that this is directly related to the power of the original vibration. If the frequency of vibrations is very low (e.g. < 1 Hz), which is the case in seismic waves in geophysics applications, then the combination of TAD with FFT can be very efficiently used in real-time. Similar to the 2-D processing, by simply Fourier transforming the RBS data matrix, 9.5 dB of SNR was achieved with a spatial resolution of 3.7 m in [46]. This SNR was better than the one obtained with 3×3 Sobel operator of the TED approach presented in [43].

The concept of using the popular wavelet based techniques for distributed fiber sensors was mentioned for the first time in [47], and experimental results of wavelet transform based denoising method over a 14 km long fiber for PIDS applications were reported in [48]. Qin et al. used a wavelet technique for noise reduction with resulting increases in both highest detectable frequency and spatial resolution of 8 kHz and 0.5 meters, respectively [49]. In the wavelet method used in [49], after computation of the discrete wavelet transform (DWT) for the raw ϕ-OTDR traces, an appropriate selection of the threshold is critical to disregard the wavelet coefficients associated with the noise components. After this partial removal of the wavelet coefficient for the noise-only assumed bands, the stationary vibration signal is recovered by the inverse DWT computation. The authors extended the use of wavelet based techniques in another study by using the continuous wavelet transform (CWT) to analyze the non-stationary vibrations [50]. After low-pass filtering of the RBS to eliminate the high-frequency noise and wavelet denoising these filtered traces, the CWT method can be applied to reveal the timely-evolved features of acoustic vibrations. Recently, the popular wavelet denoising was effectively used to assist ultra-long PIDS applications; over a whole 131.5 km fiber with 8 m resolution [51], and over a 175 km long fiber with 25 m resolution [10]. Empirical mode decomposition (EMD) combined with the wavelet denoising is introduced in [52] again for monitoring long distances. The most

detailed wavelet analysis for ϕ-OTDR based DAS applications is given in [53] where the multiscale-wavelet decomposition (MWD) was shown to be an an effective tool for detecting the human disturbances embedded in the natural interference. Although high SNR improvements as high as 35 dB were reported, the degree of the SNR improvement with wavelet based techniques will depend on the application, and the signal type to be detected. Hence, the thresholding required for rejecting the transform domain coefficients is a challenge for the wavelet based denoising. Additionally, the level of decomposition should guarantee the real-time application.

Non-stationarity or transient behavior of vibrations in the environment, will require the analysis of signals both in time and in frequency domains. Besides the popularity of short-time Fourier transform (STFT) and DWT for such applications, Hilbert–Huang transform (HHT) is an attractive choice for time-frequency analysis [54]. When compared to DWT, one of the most attractive feature of HHT is that, it is data driven and there is no need to make prior adjustments as it is done for the selection of a mother wavelet in the DWT case. Experimental studies also report that HHT is superior to DWT in the accuracy of instantenous frequency measurements [55]. Hence, HHT was introduced to ϕ-OTDR based DAS technology in [56] for analyzing multiple vibrations and the experimental results carried out with synthetic PZT driven vibrations verify that HHT is a promising tool. In the common sense, although HHT is known to suffer from computational load, the authors of [56] report that their HHT implementation was much faster than that of the STFT and CWT. Nevertheless, in the general sense, the Hilbert spectral analysis seemed to be a powerful to provide high frequency resolution for both stationary and nonstationary signals, but requires more experimental research to be mature for DAS field applications.

Another DSP method was recently introduced which is based on calculating the kurtosis of acoustic signals at each position of the fiber in a short time period [57]. In this method, after the moving average on the spatial dimension, the spatial average kurtosis (SAK) is computed and the peak is compared with a certain threshold value to accurately locate the acoustic vibration segments along the fiber. The SAK method was found to perform well to discriminate the instantaneous perturbations

from the system noise and certain ambient environmental interferences. The experimental results presented in [57] show that, when compared with the previous methods (MAD, TED, DWT), the SAK method can locate the vibrations with a higher SNR. Pencil-break events (which are used to simulate the cracks in concrete structures in the laboratory environment for SHM applications), and digging activities were detected with the SAK method with improved SNR values of 16.6 dB and 17.3 dB, respectively. On the other hand, the SAK method is considered to be effective for short-duration events, and long-term vibrations need more features to be identified for real PIDS applications.

More recently, a signal processing method based on compressive sensing was presented for DAS applications in [58] which allows for compressing and denoising the raw data simultaneously. The denoising method is based on orthogonal matching pursuit (OMP) algorithm which recovers the signal under the sparsity determined by a threshold rule, and its denoising performance is compared with DWT and the MAD. Indoor experimental results carried out for the detection of 100 Hz vibration event over a fiber of 3 km long show that, SNR values above 30 dB are achievable with a compression ratio of 18.9; while the DWT and MAD techniques attained SNR values of 25.6 and 18.2, respectively.

As a conclusion, the research regarding the DSP related techniques in ϕ-OTDR systems is very limited, and there is not any published research nor any evidence for a DAS application based on adaptive filtering or eigen-analyis of the RBS up to now. Besides, all system configurations presented in these references employ at least one extra stage (e.g. EDFA) for optical amplification to increase the SNR of the ϕ-OTDR system significantly before the DSP post processing. Hence, none of the above-mentioned DSP methods seem to be convenient to detect weak vibrations that are very close to the noise level. Additionally, almost all experimental studies were carried out with the indoor data and very limited emphasis was on real-world conditions. In this thesis research, as will be discussed and presented in the following chapters, the proposed AMF [59] and MED techniques were firstly tested with Monte Carlo simulations, the results were verified and validated with real ϕ-OTDR and field data.

The performance of the above mentioned DSP techniques applied for ϕ-OTDR systems are summarized in chronological order in Table 2.1 below. The last column includes the abbreviation of the main DSP method(s) used in the related research, while the 8. column gives the number of optical amplification (OA) stages used in the system, since the strength of the RBS is very dependent on the type of the amplifications in the system. For that reason, it is worth emphasizing that these results are not comparable directly with each other, and the Table 2.1 should only be considered as a snapshot of the state-of-the-art DSP methods.

Table 2.1. Review of the state-of-the-art DSP methods used for DAS.

Date	Ref.	MR [km]	SR [m]	MDF [kHz]	Detection Scheme	MSNR [dB]	OA Stages	DSP Method(s)
Jun 05	[7]	12	1000	3,33	DD	20	2	TAD
Apr 08	[48]	14	50	10	DD	18	1	DWT
Jul 08	[45]	1	10	N/A	DD	20	-	FFT
Nov 10	[41]	1,2	5	10	DD	20	1	TAD
Aug 11	[34]	10	5	5	CD	7,6	1	TAD
Apr 12	[49]	1	0,5	25	CD	20	2	DWT
Aug 12	[50]	10	1	5	CD	20	1	DWT+CWT
Sep 13	[42]	1	3	5	DD	8,4	1	TED
Dec 14	[56]	50	50	1	CD	30	1	HHT
Apr 15	[46]	2,7	3,7	18	CD	9,5	1	FFT
Aug 15	[53]	23,7	N/A	1	CD	35	1	DWT
Sep 15	[52]	60	6,8	0,5	CD	14	2	DWT+EMD
Mar 17	[44]	12,5	24	1	CD	39	1	TED
Jun 17	[43]	27,6	10	N/A	CD	20	2	BLF
Jun 17	[59]	28,3	25	1,25	CD	39	-	AMF
Aug 18	[57]	5,1	10	0,5	CD	23,4	1	MAD+SAK
Oct 18	[58]	3	3	3	CD	34,4	2	CS+OMP
Jun 19	N/A[2]	28,3	25	1,25	CD	25,6	-	MED

2.4. Theoretical Model for the Rayleigh Backscatter

Rayleigh scattering in the fiber occurs due to random microscopic variations in the index of refraction of the fiber core, which are caused by the imperfections during the manufacturing process. It is due to the integrated effect of large number of signals from these randomly spaced microscopic variations, which are also considered as isotropic scatterers. The analytical models used for predicting the behaviour of Rayleigh backscattering in ϕ-OTDR systems treat these random scatterers as a large number of reflectors or mirrors [60].

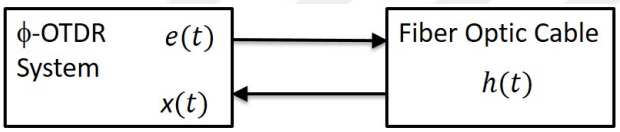


Figure 2.4. LTI system representation of ϕ-OTDR system.

Let us consider the ϕ-OTDR based DAS system as a linear time-invariant (LTI) system as depicted in Figure 2.4, where $e(t)$ and $x(t)$ are the excitation and backscatter signals, respectively, and $h(t)$ is the FO backscatter impulse response which can be given by

$$h(t) = \sum_{m=1}^{M_s} e^{-\alpha v_g t_m} a(t_m) \delta(t - t_m) \tag{2.4}$$

where $a(t_m)$ is the amplitude of the m-th scatterer, t_m is the relative delay of the m-th scatterer, M_s is the total number of scatterers per SR in the fiber medium, α is the Rayleigh attenuation coefficient, v_g is the group velocity in the fiber and $\delta(t)$ is the Dirac delta function (i.e., $\delta(t - t_m) = 1$ when $t - t_m = 0$, zero otherwise). The exponential term is due to the attenuation in the fiber. Polarization effects of each scatterer have been neglected in (2.4) and the fiber attenuation α is the same in each direction.

[2]Research results not published yet, but reported in the thesis.

The RBS response $x(t)$ of the system is simply the convolution of the $h(t)$ with the excitation $e(t)$, i.e. $x(t) = h(t) * e(t)$. In the conventional ϕ-OTDR system, the excitation signal is a rectangular pulse of τ seconds, of quasi monochromatic optical carrier with angular frequency ω_s. The source coherence time is assumed to be large compared to the duration of τ of the pulse. Then $x(t)$ becomes

$$x(t) = \sum_{m=1}^{M_s} e^{-\alpha v_g t_m} a(t_m) rect\left(\frac{t - t_m}{\tau}\right) \cos\left[\omega_s(t - t_m)\right] \tag{2.5}$$

The correlation properties of $x(t)$ have been studied in the literature [61, 62]. It can be shown that the RBS is a nonstationary process, and the maximum correlation distance of $x(t)$ is τ when the excitation signal is a short duration pulse of τ seconds. We can recognize that $x(t)$ is a narrowband process by rewriting (2.5) as

$$x(t) = x_i(t)\cos(\omega_s t) + x_q(t)\sin(\omega_s t) \tag{2.6}$$

where

$$x_i(t) = \sum_{m=1}^{M_s} e^{-\alpha v_g t_m} a(t_m) rect\left(\frac{t - t_m}{\tau}\right) \cos\omega_s(t_m) \tag{2.7a}$$

$$x_q(t) = \sum_{m=1}^{M_s} e^{-\alpha v_g t_m} a(t_m) rect\left(\frac{t - t_m}{\tau}\right) \sin\omega_s(t_m) \tag{2.7b}$$

Both $x_i(t)$ and $x_q(t)$ are slowly varying compared to $\cos(\omega_s t)$ and $\sin(\omega_s t)$. Since the sums in the equations above are over a large number of statistically independent, identically distributed random variables, adhering to the central limit theorem, both $x_i(t)$ and $x_q(t)$ are now Gaussian distributed with zero mean and equal variance. Now we can write the bandpass process

$$x(t) = E_s(t)\cos(\omega_s t - \phi(t)) = Re\{E_s(t)e^{-j\phi(t)}e^{j\omega_s t}\} \tag{2.8}$$

where

$$E_s(t) = \sqrt{x_i^2(t) + x_q^2(t)} \tag{2.9a}$$

$$\phi(t) = \tan^{-1}\left[\frac{x_q(t)}{x_i(t)}\right] \tag{2.9b}$$

The sum of two Gaussian distributed random variables can be shown to be Rayleigh distributed. Hence, the envelope of the backscatter response $E_s(t)$ is Rayleigh distributed, while the phase of the backscatter is uniformly distibuted over the interval $(-\pi, \pi)$ [61].

2.5. Numerical Implementation of Rayleigh Backscatter

In order to simulate the RBS based ϕ-OTDR system, as the fiber sensor is spatially sampled with Δz, we can consider the discrete model of Rayleigh backscattering shown in Figure 2.5 below.

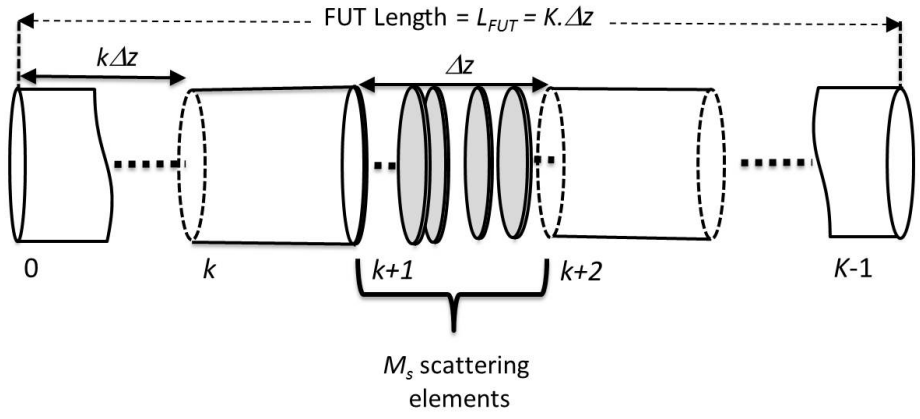


Figure 2.5. Discrete model for Rayleigh backscattering in the fiber.

In this discrete model, the fiber is divided into K equispaced individual scattering partitions. Each partition has a width of Δz where the backscattering process is described by a set of randomly spaced low-reflectance mirrors. The electric field emitted backward from each scattering object within Δz has independent random phase and amplitude. The behavior of a complex sum of scattering centers in Δz can be considered as a random walk in the complex plane. Omitting the carrier frequency term after demodulation and down-conversion, the total baseband Rayleigh backscatter term at the k-th partition is the superposition of back-scattered fields from different scattering centers which can be written as

$$E_{RBS}(k\Delta z) = E_0 e^{-\alpha k \Delta z} \sum_{m=k}^{k+M_s-1} P_m r_m e^{j\phi_m} \tag{2.10}$$

where r_m , ϕ_m and P_m are the random reflectivity, phase, and polarization of the k-th reflecting partition, and E_0 is the field amplitude of incident light (i.e. the launched electric field at the beginning of fiber).The m-th scattering center has a reflectivity $r_m \in [0, 1]$ (usually $r_m \ll 1$). The probability density functions of the reflectivity of a reflector, r and the phase ϕ are Rayleigh and uniform distributed as stated previously:

$$p(r) = \left(\frac{r}{\sigma^2}\right) exp\left(r^2/2\sigma^2\right) \qquad (r > 0) \tag{2.11a}$$

$$p(\phi) = 1/2\pi \qquad (-\pi < \phi < \pi) \tag{2.11b}$$

where σ^2 is approximately equal to 10^{-7} [63]. If a perturbation exists on the same k-th partition, the electric field term can be written as

$$E_{RBS}(k\Delta z) = E_0 e^{-\alpha k \Delta z} \sum_{m=k}^{k+M_s-1} P_m r_m e^{j(\phi_m+\theta_p)} \tag{2.12}$$

where θ_p is perturbation phase factor. Thus, in the case of perturbation, a phase shift will be introduced which can be written in mathematical form as:

$$\theta_p = \begin{cases} 0, & k < p \\ 2\phi_p, & k \geq p \end{cases} \tag{2.13}$$

Here, p is the point where the perturbation occurs, and ϕ_p is the phase of the perturbation. The factor of 2 multiplying ϕ_p results from the fact that the light waves pass through the perturbation region twice. To simulate the above analytical model in (2.12), we assume that the resultant electric field at the photodetector output, after demodulation and digitization, is represented by one sample during the digitization process. In other words, no oversampling is employed and every Δz represents a single acoustic channel or range bin. Additionally, we neglect the random polarization effects in microscopic scale (i.e., we assume $P_m = 1$ for all m).

Computer simulations were performed with the above assumptions and the analytical model. The random locations of the scattering centers were simulated with the Monte Carlo method and the simulations were carried out in MATLAB^{TM}. It was assumed that the light source is monochromatic at a wavelength of 1.559 μm which is modulated by a pulse of width τ. The number of scattering centers M_s was assumed to be the same for every acoustic channel and assumed to be homogeneously distributed along the FUT. Selection of M_s was determined in a trivial manner by observing the amplitude fluctuations dependency on the selected simulation parameters and by comparing the resultant intensity traces with conventional OTDR observations with the same interrogation parameters in the laboratory. As the backscattering intensity is directly proportional to the pulse energy, in order to develop a realistic simulation model, the average number of reflectors was coupled to the pulse width via $M_s = 2000.round(\tau/10^{-6})$ Hence, a one microsecond pulse-width selection will result in 2000 scattering centers per channel. The simulation parameters were chosen almost the same with the parameters of the laboratory setup that was developed to gather real-world data. This was in purposely done to evaluate the performance of the computer codes and to make the simulation results directly comparable with the real data. The simulation parameters are summarized in Table 2.2.

Table 2.2. DAS Simulation parameters.

System Parameter	**Symbol**	**Value**
Laser Power	P_t	22 mW
Laser Pulse Width	τ	860 ns
FO Refractive Index	n_{eff}	1.467
Pulse Repetition Frequency	PRF	2.5 kHz
Length of the FUT	L_{FUT}	28321 m
Total Range Bins	K	456
Range Resolution	Δz	87.67 m
Perturbation Location	l_p	7364 m
Perturbation Range Bin Number	p	84
Photo-Detector Noise Equivalent Power	NEP	6 pW/$\sqrt{Hz}$
Perturbation Frequency	f_p	833.33 Hz
Number of Scattering Centers per range bin	M_s	2000
Total System (Insertion) Loss	IL	6 dB
System Bandwidth	B	1 MHz

Vibration presence cases were simulated by injecting a perturbation signal for a selected range bin. The strength of the vibrations, hence the SNR values were adjusted by varying the reflectivity values of the mirrors in a trivial manner. After every modification in the reflectivity, the strength of the simulated vibration was verified by observing the frequency spectrum of the range bins. Several datasets with varying perturbation and record durations were obtained with computer simulations throughout the thesis study. A single ϕ-OTDR trace simulation result is shown in Figure 2.6 below as a sample in arbitrary units (a.u.), where 833.33 Hz vibration is injected at 7364 m corresponding to the 84. range bin. Since it is a raw data, the vibration is not observable on the fast-time without any proper signal processing. The end of the FUT is not clear either. Coherently adding the traces pulse-to-pulse will naturally enhance the SNR to reveal out any non-stochastic behavior along the axis as can be seen in Figure 2.7. When the averaged traces is 100, the peak expected at the 84. range bin is not clear; while it is observable when the averaging size is increased to 1000.

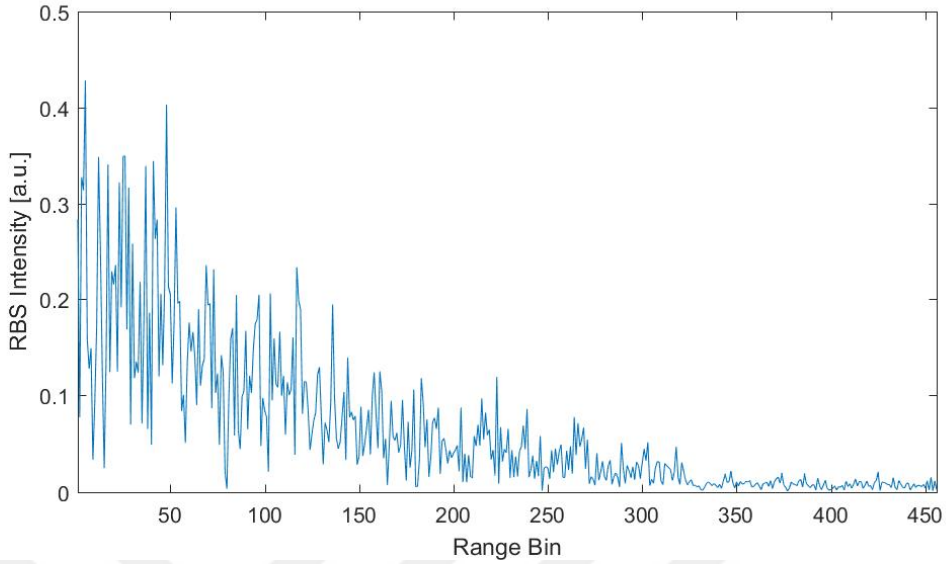


Figure 2.6. A typical single ϕ-OTDR trace of the Monte Carlo simulation.

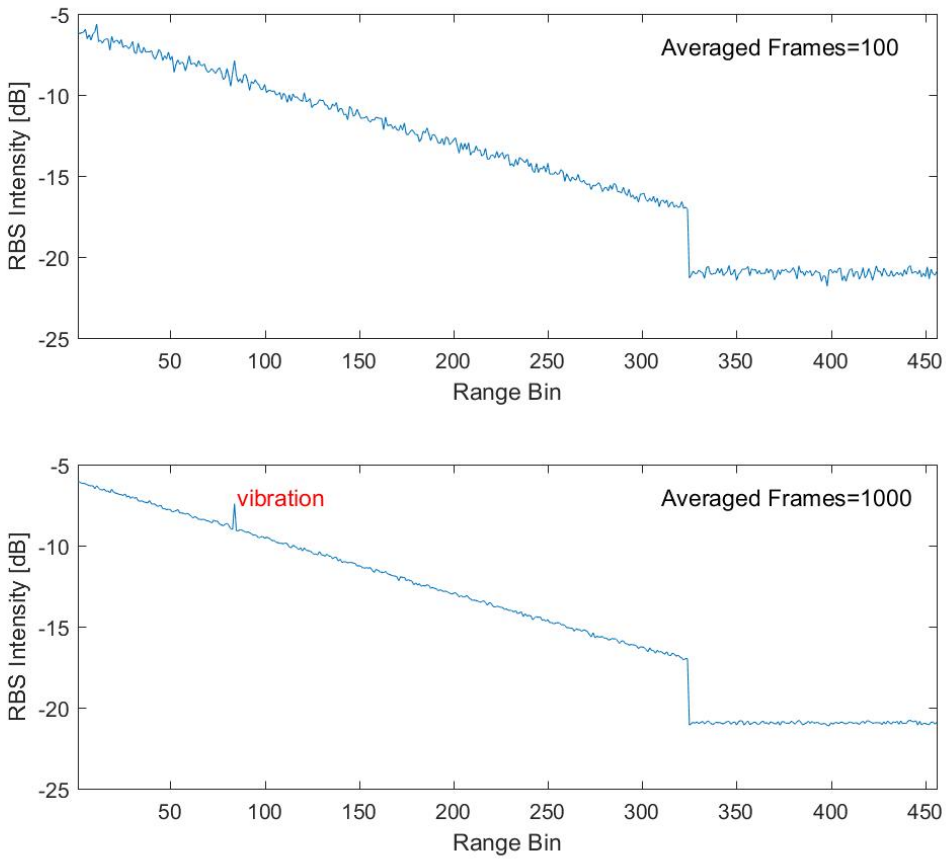


Figure 2.7. Impact of averaging on ϕ-OTDR traces.

Averaging N traces enhances the SNR of the data with $\sqrt{N}$, but the MDF is reduced by N. Hence, this basic enhancement method is only appropriate for detecting very low frequency vibrations. The vibration signal is clearly observable when the 2-D frequency spectrum of the raw OTDR data is computed as shown in Figure 2.8.

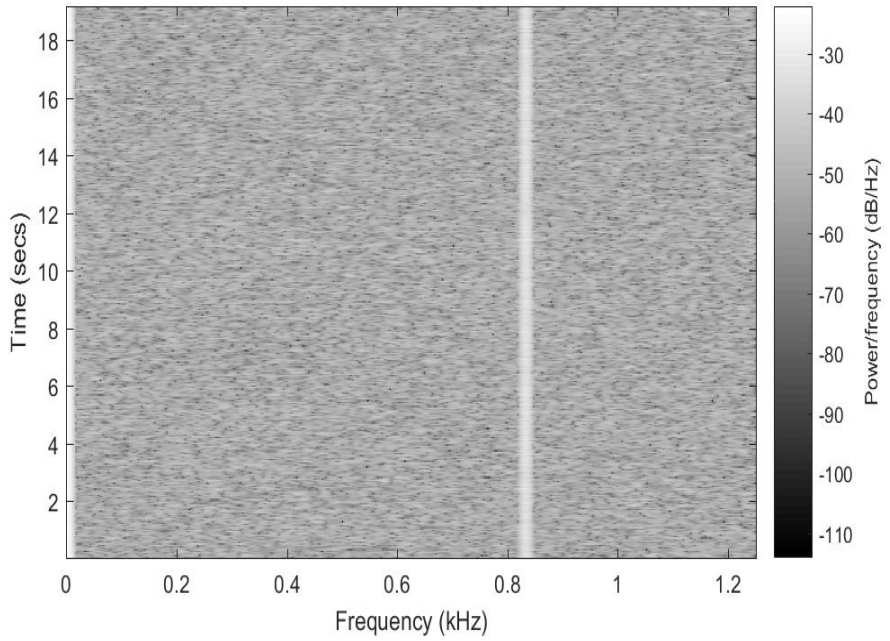


Figure 2.8. Spectrogram of the simulated 19.2 seconds long ϕ-OTDR data with a synthetic vibration of 833.33 Hz at the 84. range bin.

2.6. The ϕ-OTDR Signal Model

When compared with the classical radar detection theory, the weak signal detection problem encountered in the DAS system resembles the radar detection in many aspects. The DAS system emits a burst of laser pulses and records the returns between successive pulses. As sketched in Figure 2.9 to present important DAS signal parameters, the ϕ-OTDR system interrogates the FUT and spatially samples the RBS at different locations along its whole length. The location of the sampled regions, which are illustrated as spaced dots in Figure 2.9, and the distance between them are defined through DAS system configuration. The numbering system is defined such that the first spatial sample starts with "0", and successive samples are numbered with positive

integers. These sample locations are usually known as "acoustic channels" in the DAS literature, but will be referred to as "range bins" in this context. The pyhsical separation between these range bins defines the spatial sampling interval which should not be confused with the SR of the DAS system. In general, the SR may consist of more than one spatial sample according to the sampling rate of the ADC of the system, which determines the distance between these range bins that are collected on the "fast-time axis".

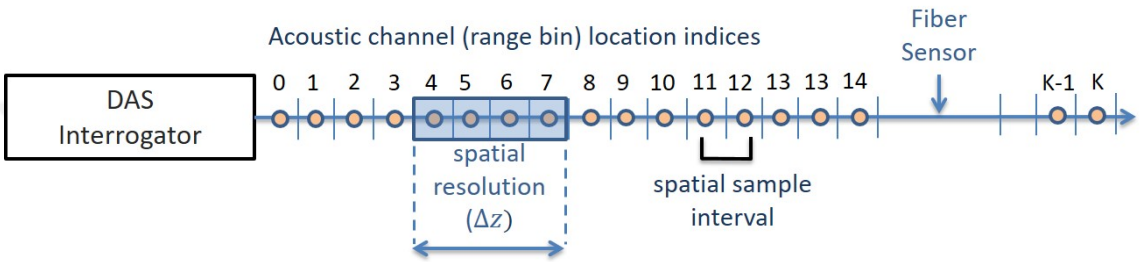


Figure 2.9. K virtual acoustic channels synthesized on the sensing fiber

The ϕ-OTDR based DAS operation generally involves the collection and processing of these spatial samples for every channel separately according to the application. One single shot of a total K samples is usually referred to as a "single ϕ-OTDR trace". The continuous interrogation of the FUT is equivalent to sampling every range bin with a sampling rate equal to the PRF. This sampling on the "slow-time axis" is illustrated in Figure 2.10 by with three consecutive ϕ-OTDR traces. The returns from N consecutive ϕ-OTDR traces for a given k-th range bin forms a temporal snapshot of dimension $N \times 1$:

$$\boldsymbol{x}_k = \begin{bmatrix} x_{0,k} & x_{1,k} & x_{3,k} & ..... & x_{(N-2),k} & x_{(N-1),k} \end{bmatrix}^H \tag{2.14}$$

where $\boldsymbol{x}_k$ is the vector of collected baseband RBS samples, which is subject to processing in this research study. Here, $x_{n,k}$ is the baseband signal detected for the n-th pulse at the k-th range bin, for $n = 0, 1, ...N-1$ and $k = 0, 1, ..., K-1$. The letter H on the upper right side of the equation denotes the conjugate transpose operation on the collected samples which is a row vector and thus $\boldsymbol{x}_k$ is a column vector.

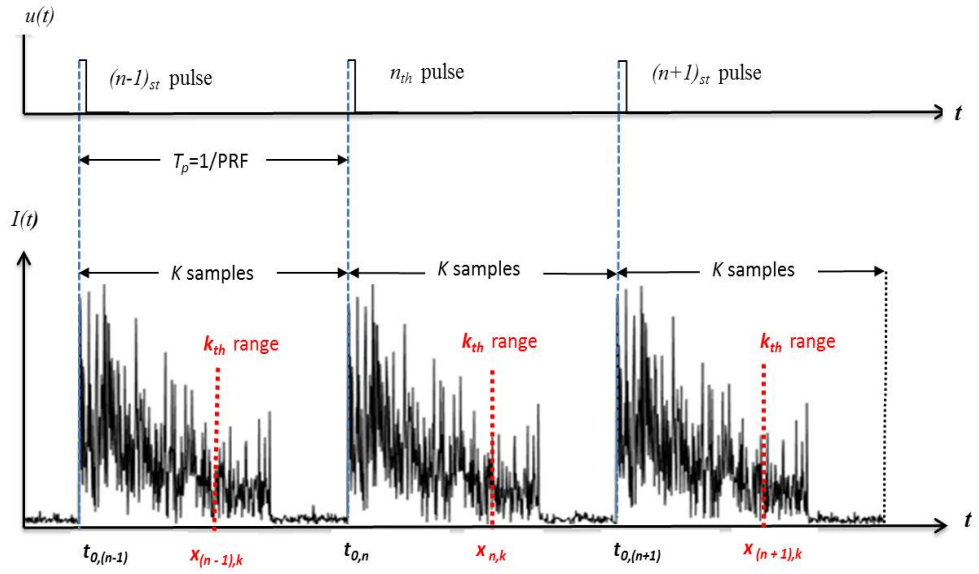


Figure 2.10. A-scope presentation of three consecutive ϕ-OTDR traces: Samples denoted by red dashed lines are collected on the slow-time axis.

The slow-time samples as depicted by dashed red lines in Figure 2.10 should not be confused with the ϕ-OTDR traces which are plots of returned optical power versus time. The time interval over which the N pulse returns are collected is referred to as the temporal processing interval (TPI), as depicted in Figure 2.11. Thus, the signal processor's TPI length is N/PRF seconds. Slow-time and fast-time axes are shown together in Figure 2.11, and a 3-D visualization of both axes with real ϕ-OTDR traces is shown in Figure 2.12.

A pulse length of τ seconds and bandwidth B Hertz is assumed. For each received TPI frame, the all electrical subsystem executes the following chains: the electrical IF signal detected at the photo-detector output is down-converted to baseband, low-pass filtered and digitized for subsequent DSP routines. The sampling frequency during digitization is taken at least twice the bandwidth, (i.e. $f_s \geq 2B$ Hertz). With appropriate decimation, the number of baseband samples per range bin can be reduced to one. Thus, each 2-D TPI frame consists of exactly KN samples.

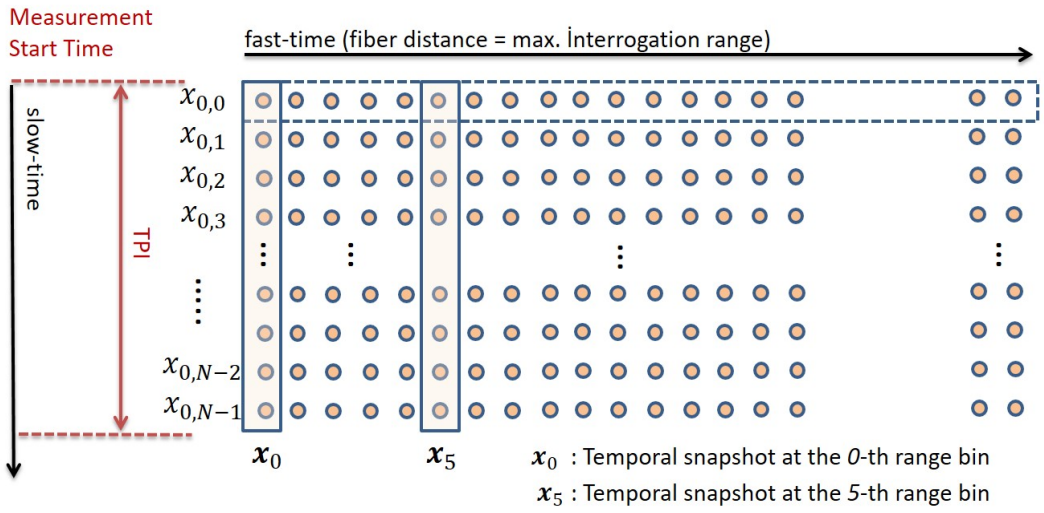


Figure 2.11. Signal parameters related to the slow-time and fast-time axes.

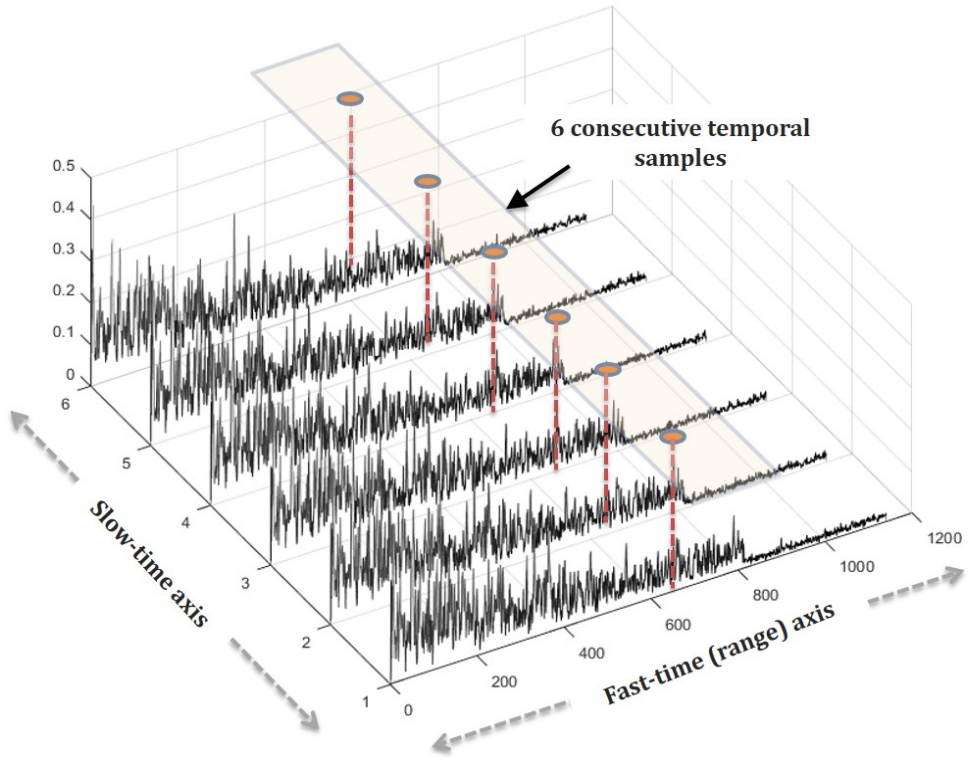


Figure 2.12. 3-D demonstration of six consecutive real ϕ-OTDR traces.

2.7. Statement of the Problem

In this research, we work primarily with data from a single range gate to be processed sequentially along the FUT axis and our goal is to make a decision whether there is a vibration or not at a specific location. In general, along the FUT axis, the detected signal consists of two main components:

$$H_0 : \qquad \boldsymbol{x}_{k/H_0} = \boldsymbol{n}_{rbs,k} + \boldsymbol{n}_{w,k} \tag{2.15}$$

where the condition H_0 is the null hypothesis denoting the vibration (or perturbation) absence, and $\boldsymbol{x}_{k/H_0}$ is the temporal snapshot vector for the noise only case. The first term on the right side of the (2.15) refers to the highly fluctuating noise term due to the Rayleigh backscattering and the second term is due to the receiver thermal noise. Suppose that a sinusoidal vibration occurs at a specific range point, then the received signal term can be written as:

$$H_1 : \qquad \boldsymbol{x}_{k/H_1} = \boldsymbol{v}(\omega_p) + \boldsymbol{n}_{w,k} \tag{2.16}$$

where H_1 denotes the alternative hypothesis condition, $\boldsymbol{x}_{k/H_1}$ is the temporal snapshot vector in the vibration presence case, and $\boldsymbol{v}(\omega_p)$ is the response of the system to a hypothetical perturbation with a perturbation frequency of $\omega_p/2\pi$ [Hz].

For PIDS applications, the frequency of interest is relatively low and averaging techniques may suffice for all times to detect such low-frequency vibrations. For detection of seismic waves in geophysical applications, the vibration frequencies of interest are even much lower which may be as low as 0.1-1Hz. However, for some particular application areas in SHM such as crack detection of materials and anomaly detection of engines, the frequency range of the perturbation events could be as high as several kHz to MHz. In these cases, it is desired to have DAS systems with high-frequency response and large distance from the anomaly point to sensing fiber. Here we focus on improving the SNR of the DAS system without sacrificing the frequency bandwidth.

3. OPTIMAL FILTERING FOR NOISE REDUCTION IN ϕ-OTDR

3.1. Introduction

Weak signal detection in severely noisy environments has always been a cumbersome problem in diverse areas of signal processing. Noise reduction based on adaptive filtering is a heavily used signal processing technique to detect specifically weak signals under harsh conditions. The adaptive filter theory has found numerous applications in radar, sonar and speech processing fields and it is well documented in the literature [64, 65].

Motivated by the success of the application of adaptive signal processing algorithms for noise reduction in many diverse fields, in this chapter, we propose to adapt the classical linear filtering algorithms for DAS systems. Similar to basic optimal filtering techniques employed in radar [66] and speech processing [67] , we seek a similar approach for the distributed fiber sensors. The term "adaptive" used above refers to the fact that, as will be described in the following sections, the linear filter coefficients will be based on observing the statistics of the environmental noise which needs to be updated regularly for real-world applications. There exists different optimality criteria and nomenclature in designing the filter coefficients. The celebrated Wiener filtering approach is one method which is based on minimizing the mean squared error (MSE) between the estimated and the desired vibration signal [66, 67]. This minimum MSE (MMSE) criterion leads to the optimum filter coefficients under Gaussian observations and white noise assumptions. While the MMSE approach seems to be popular in statistical inference and communications signal processing, the maximum SNR (MSNR) criterion is the most popular method in radar and general detection theory. It has been proven theoretically that the MMSE and MSNR estimators are in fact equivalent even when the noisy observations are non-Gaussian [68] and it can be shown that these two filters are equivalent up to a scaling factor.

Based on such previous analytical proofs [50] and for the sake of reality, the optimum noise filtering approach considered in this thesis is based on the MSNR approach. The maximization procedure will be of concern for every range bin of the DAS system, hence the weight computations are done for every range bin individually.

Our main performance measure in this section is based on how much background noise and amplitude fluctuations due to RBS are suppressed. Thus, in order to highlight the visibility of a particular vibrating range bin along the FUT axis (i.e. fast time axis), the following fast-time SNR definition is used in this research study:

$$SNR_{(fast-time)} = 10\log_{10}\left(\frac{\textit{the signal energy at vibrating range bin(s)}}{\textit{mean value of the total noise energy at other range bins}}\right) \tag{3.1}$$

The result of (3.1) is expressed in dB and will be used as the main performance metric throughout this research study.

3.2. Optimum Temporal Filtering of ϕ-OTDR Signals

In order to suppress the noise and to recover the acoustic vibration and its spatial location along the fiber route, we propose to design a finite impulse response (FIR) filter in this chapter. In digital signal processing applications, either FIR or infinite impulse response (IIR) type of filters are utilized according to application requirements. Since FIR filters are inherently stable [69], it would be a better choice than an IIR filter for a dynamically changing environment where the filter is to be updated in the run-time. Another distinct advantage of the FIR over the IIR filter is their phase linearity which is critical for processing signals that convey information in their phase, and simultaneous processing of multiple signals that will be later combined [69]. Since the phase of the optical backscatter is altered according to the environment along multiple channels, the true phase information is critically important for real-time implementation for DAS applications.

Considering the fact that DAS systems are deployed in dynamically changing environments and the optimality conditions will probably require updating important system parameters in the run-time, we only consider the FIR type of filter to guarantee the stability of the proposed algorithm in this study. In this approach, to temporal snapshot vector $\boldsymbol{x}_k$ is subject to an FIR filter and the intensity at the output is computed by

$$z_k = \sum_{n=0}^{N-1} w_{k,n}^{*} x_{n,k} = \boldsymbol{w}_k^H \boldsymbol{x}_k \tag{3.2}$$

where z_k is the FIR output, $w_{k,n}$ is the complex weighting coefficient computed for the n-th pulse at the k-th range bin, and $\boldsymbol{w}_k$ is a column vector of size $N \times 1$ containing these coefficients for the k-th range bin (acoustic channel):

$$\boldsymbol{w}_k = \begin{bmatrix} w_{k,0} & w_{k,1} & \dots & w_{(k,N-2)} & w_{(k,N-1)} \end{bmatrix}^H \tag{3.3}$$

Hence, the received ϕ-OTDR samples at the k-th range bin given by (3.2) are linearly weighted and a scalar output value is produced for an observation length of N samples. This operation can be implemented in a block-by-block fashion as new samples arrive at the receiver. Each data block of size N can be consecutively input to the FIR filter with overlapping segment sizes equal to $N-1$. In the vibration presence case, the temporsal signal vector $\boldsymbol{v}(\overline{\omega})$ (i.e. vibration signal) of H_1 can be written as:

$$\boldsymbol{v}(\overline{\omega}) = \sigma_p \boldsymbol{s}(\overline{\omega}) \tag{3.4a}$$

$$\boldsymbol{s}(\overline{\omega}) = \begin{bmatrix} 1 & e^{j2\pi\overline{\omega}} & e^{j4\pi\overline{\omega}} \dots & e^{j(N-2)\pi\overline{\omega}} & e^{j(N-1)\pi\overline{\omega}} \end{bmatrix}^H \tag{3.4b}$$

where σ_p denotes the random complex voltage of the perturbation, $\overline{\omega} = f/PRF$ is the normalized sampling frequency and $\boldsymbol{s}(\overline{\omega})$ is the temporal steering vector.

3.2.1. Maximum SNR Approach

One method in optimum filtering is to design such an FIR filter which maximizes the output SNR, so called as MNSR approach [48][49]. For this objective, we now define the "slow-time" SNR which is given as the ratio of the power at the output of the FIR filter due to the desired signal (P_{signal}) to the power at the output due to noise only (P_{noise}):

$$SNR_{(slow-time)} = \frac{P_{signal}}{P_{noise}} = \frac{E\{\boldsymbol{w}_k^H \boldsymbol{v}(\overline{\omega})\boldsymbol{v}^H(\overline{\omega})\boldsymbol{w}_k\}}{E\{z_{k/H_0} z_{k/H_0}^H\}} \tag{3.5}$$

where $E\{\cdot\}$ is the expectation operator and z_{k/H_0} is the FIR output for the noise-only (H_0) observations. By substituting (3.2) and (3.4a) into (3.5) we can write

$$SNR_{(slow-time)} = E\{|\sigma_p^2|\}\frac{\boldsymbol{w}_k^H \boldsymbol{s}(\overline{\omega})\boldsymbol{s}^H(\overline{\omega})\boldsymbol{w}_k}{\boldsymbol{w}_k^H E\{\boldsymbol{x}_{k/H_0}\boldsymbol{x}_{k/H_0}^H\}\boldsymbol{w}_k} = E\{|\sigma_p^2|\}\frac{|\boldsymbol{w}_k^H \boldsymbol{s}(\overline{\omega})|^2}{\boldsymbol{w}_k^H \boldsymbol{R}_{\boldsymbol{x},k/H_0}\boldsymbol{w}_k} \tag{3.6}$$

where $\boldsymbol{x}_{k/H_0}$ is the noise only snapshot vector defined by (2.15) previously, and $\boldsymbol{R}_{\boldsymbol{x},k/H_0} = E\{\boldsymbol{x}_{k/H_0}\boldsymbol{x}_{k/H_0}^H\}$ is the noise-only covariance matrix of the optical RBS. By simple mathematical manipulations, the optimum weight vector or the optimal matched filter that maximizes the above equation can be written as

$$\boldsymbol{w}_{k,opt} = \eta \boldsymbol{R}_{\boldsymbol{x},k/H_0}^{-1}\boldsymbol{s}(\overline{\omega}) \tag{3.7}$$

where η is an arbitrary real scaling factor. The derivation of the above equation can be found in Appendix A. Since $\boldsymbol{R}_{\boldsymbol{x},k/H_0}^{-1/2}$ is a whitening filter, the filter acts as a correlation canceler (i.e. the whitening filter de-correlates the colored noise input, hence making it look like white noise). In white noise case, $\boldsymbol{s}(\overline{\omega})$ is the matched filter, it has a band-pass response and maximizes SNR. We may interpret the optimal weight vector as [23] :

$$\boldsymbol{w}_{k,opt} = \eta \underbrace{(\boldsymbol{R}_{\boldsymbol{x},k/H_0}^{-1/2})}_{\substack{\text{Whitening} \\ \text{Filter}}} \underbrace{(\boldsymbol{R}_{\boldsymbol{x},k/H_0}^{-1/2}\boldsymbol{s}(\overline{\omega}))}_{\text{Matched Filter}} \tag{3.8}$$

where the process in the right-bracket above accommodates the linear transformation applied to the temporal signal vector during the whitening stage. Equations (3.7) and (3.8) state that the weights chosen to maximize the SNR are determined via the statistics of the background noise. This background statistics includes both the speckle-like profile of Rayleigh backscattered data and receiver noise contributions. The constant term of the MSNR calculation η is usually set to unity or taken as

$$\eta = \frac{1}{\sqrt{\boldsymbol{s}^H(\overline{\omega})\boldsymbol{R}^{-1}_{\boldsymbol{x},k/H_0}\boldsymbol{s}(\overline{\omega})}} \tag{3.9}$$

For the steering vector computation, if there is no a priori information about the frequency region of interest, the processor has to sweep across the frequency axis to detect the vibration frequencies. Let p and f_p denote the range bin index of the perturbation and the frequency of the perturbation, respectively. When $k = p$ and the steering vector is tuned to the actual vibration frequency of the perturbation (i.e., $f = f_p$), the output scalar value of the filter is expected to be maximized. This procedure is repeated for every range bin of the ϕ-OTDR trace to search for all possible vibrations (i.e., along the FUT axis).

For covariance matrix estimation, the situation is quite more complex. For every range bin under interest, the noise covariance is to be estimated from the received samples and the noise statistics are estimated by the following:

$$\hat{\boldsymbol{R}}_{k/H_0} = \frac{1}{\#\{L_s\}} \sum_{\forall k \in L_s} \boldsymbol{x}_{k/H_0}\boldsymbol{x}^H_{k/H_0} \tag{3.10}$$

where L_s is the set of range bin indices used in the secondary data set and $\#\{L_s\}$ is the total number of range bins used, where $\#\{\cdot\}$ denotes the number of the elements in a set given in the brackets. The temporal snapshot vector $\boldsymbol{x}_{k/H_0}$ is also called "the secondary data" . The above given estimated covariance matrix is commonly called as the "sample covariance matrix (SCM)". The range bin under test is considered as the one containing the primary data while secondary data inputs are assumed to contain only noise that are collected from the neighboring range bins.

Thus, the weights are computed via (3.7) with this estimated value of the covariance matrix. The matrix inversion in (3.7) is sensitive to the training set size and the matrix is not invertible when the size of the training set is small which is unavoidable in most practical situations. The traditional way to overcome this ill-conditioning is to add some amount of white noise on the diagonal elements of the SCM:

$$\hat{\boldsymbol{R}}_{k-dl} = \hat{\boldsymbol{R}}_{k/H_0} + \rho \boldsymbol{I}_N \tag{3.11}$$

where ρ is the scalar loading factor and $\boldsymbol{I}_N$ is the identity matrix of size N. This is so called the "diagonal loading (DL)" method which is popularly employed especially in beamforming applications. Besides, there is no definite rule for selecting the optimum value for the loading factor. In array processing and spatial beamforming applications, it is admitted that 5-10 dB loading above the receiver noise floor can be employed as a rule of thumb [70].

3.2.2. The Adaptive Matched Filter (AMF) Algorithm

Under the assumptions that the statistical covariance and the spectral information about the vibration are known, the optimum filter weight vector which maximizes the output SNR is computed by (3.8). In practice, the actual statistics cannot be known but estimated from a limited number of observations as given by (3.11). The other requirement associated with the steering vector selection requires information about the frequency content of the vibration. Lack of this prior information may require seeking all possible vibration frequencies up to the MDF of the system. This computational burden can be easily alleviated by today's high-performance processors or parallel computing architectures. Under this assumption and for the sake of the borrowed methodology from radar theory, we will nominate this technique as the "adaptive matched filter (AMF)" which can be summarized as follows:

(i) The ϕ-OTDR data is collected and TPI duration (value of N) is determined.

(ii) A training strategy is determined and the training region along the sensing fiber axis with size L_s is selected. Using this training data set and attaining a DL level, the covariance matrix of background noise is estimated by (3.11).

(iii) An adaptive weight vector is computed by (3.8) with the steering vector tuned to the specific vibration frequency of interest.

(iv) Using this weight vector the complex scalar output z_k for the k-th range is computed along the sensing fiber by (3.2), i.e. for all values of $k = 0, 1, 2, ..., K-1$ and the results are plotted.

The above steps are repeated for all pre-selected or possible vibrations in the frequency axis as mentioned previously. The training and FIR filter size selections are critical factors that have an impact on the noise suppression performance of the AMF but not the only consideration.

In real-world applications, the noise impact of the operating environment is unknown a priori. For relatively short-range applications such as SHM, the environmental noise may be stationary and a fixed training region may be used along the whole FUT. But for long-range DAS applications such as pipeline or border security, the background noise is most likely to change with time. Similar to the clutter ridge problems encountered in radar signal processing in heterogenous environments , the background heterogeneity in DAS applications will impact the performance of the system. In a dynamically changing environment, the estimated covariance, hence the filter coefficient vector needs to be updated on a regular basis according to the application requirements.

3.3. Applications of AMF for Vibration Detection

The performance of the AMF algorithm was evaluated for both strong and weak vibration conditions. Four different data sets with varying SNR values have been simulated as described in Section 2.5. In the first two, only one range bin included a vibration with two different slow-time SNR values. In the last two datasets, weak and strong perturbations were simulated at two range bins simultaneously. After all simulations were completed, the slow-time SNR values were computed in the frequency domain by estimating the frequency spectrum of the slow-time samples of the target range bin with the Welch method in MATLAB [71, 72].

3.3.1. Single-Strong Vibration Case

In the first dataset, a strong vibration of 833.33 Hz with a slow-time SNR value of 30 dB is synthetically included at 7364 m of a 28321 m long fiber which corresponds to the 84. range bin. A typical raw ϕ-OTDR data trace is shown in Section 2.5, where the high amplitude fluctuations conceal the vibration signal at 84. range bin. The results of AMF method for this dataset with six different FIR sizes are shown together with the computed fast-time SNR values in Figure 3.1. When the FIR size is 5, the noise reduction is not sufficient to clearly reveal the vibration at the target bin although the computed fast-time SNR is 12 dB, since other peaks levels are comparable with the output at the 84. bin.

The vibration becomes visible when N is greater than 50 and significantly higher SNR values are achieved when N is greater than 200 reaching up to 30 dB. Increasing the FIR size, hence the TPI duration improves the noise suppression of the AMF method. Diagonal loading level and the size of the secondary data were fixed as 10 dB and 10, respectively. A total of 11 range bins from $k = 10$ to $k = 20$ were selected to estimate the noise only SCM.

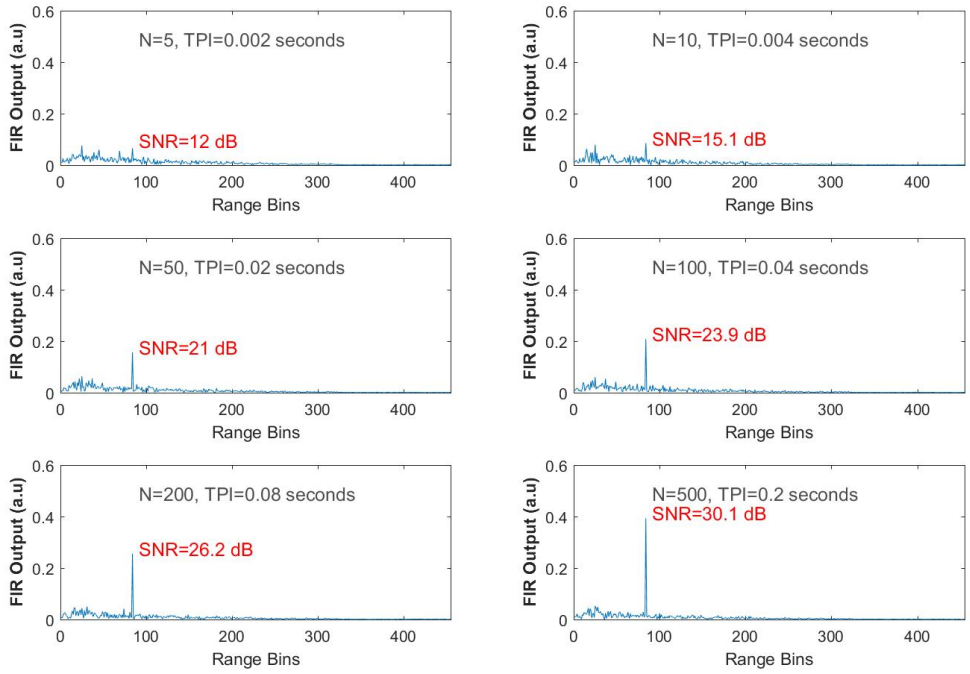


Figure 3.1. AMF results for the 1. data set for different values of N.

3.3.2. Single-Weak Vibration Case

In the second dataset, a relatively weak vibration of 833.33 Hz with a slow-time SNR value of 10 dB is synthetically included at the same range bin. This data with weaker vibration was subject to AMF method with the same DL and secondary data selections of the previous one. The results with six different FIR sizes are shown together with the computed fast-time SNR values in Figure 3.2. Fast-time SNR was computed by taking all range bins other than the target bin as noise cells. It can be observed from the plots that the vibration at the 84. range bin is not solely observable unless the FIR size is above several hundreds. As expected, the weaker the signal the longer the required TPI to achieve the same amount of fast-time SNR or noise suppression. In this case, a FIR size of $N = 500$ was required to achieve a fast-time SNR of 20 dB while it was only $N = 50$ in the strong vibration case.

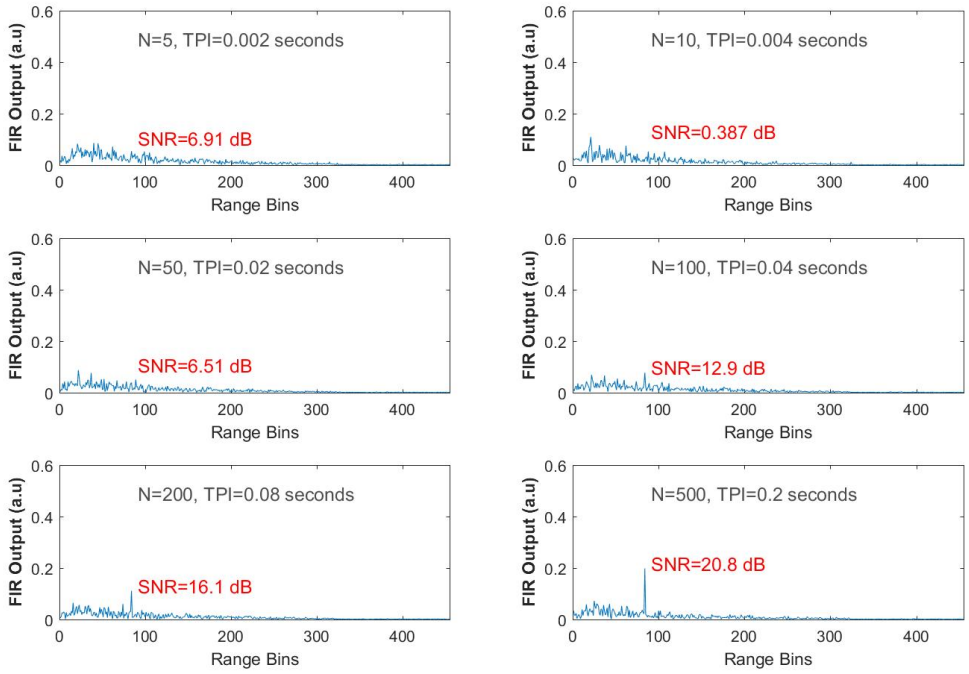


Figure 3.2. AMF results for the 2. data set for different values of N.

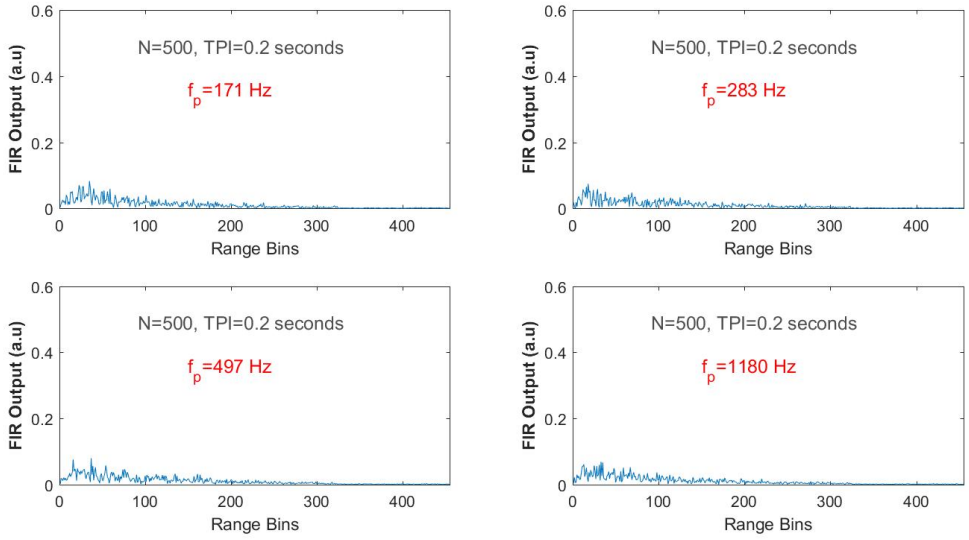


Figure 3.3. AMF results with steering vector tuned to non-vibrating frequencies.

For the second dataset, the filter outputs were tested for some arbitrary frequencies that are not equal to the vibration frequency such as 171 Hz, 283 Hz, 497 Hz and 1180 Hz. It can be seen in Figure 3.3 , that none of these tuned frequencies yielded spikes at the output of the FIR filter. As the algorithm seeks vibrations over all range bins, the filter output is only maximized when the test frequency matches the real vibration. In all results, the first 3 range bins were selected as the secondary data for estimating the noise only SCM and a linear diagonal loading level of 10 was used. (i.e. $\rho = 10$ and $L_s = \{1, 2, 3\}$).

3.3.3. Multiple Vibrations

A second synthetic vibration was induced at the 244. range bin to analyze the multiple vibration case. In the first case, the vibration frequency was selected to be the same as the vibration at the 84. range bin. The slow-time SNR values of the vibrations at 84. and 244. range bins were 18 dB and 15 dB, respectively. The results of AMF are shown in Figure 3.4 for various FIR sizes with $\rho = 10$ and $L_s = \{1, 2, 3\}$.

For this multiple vibration case, the fast-time SNR for $k = 84$ was computed by selecting all the range bins up to 200. cell excluding the 84. cell; and for $k = 244$, the fast-time value was computed by selecting the range bins between 201 and 324 by excluding the 244. one. It is clear from Figure 3.4 that the speckle profile of the FUT is not suppressed significantly unless $N = 200$ and the second vibration at the far end of the FUT is solely observable for $N = 500$.

In Figure 3.5, the AMF processing results are shown for the cases where the secondary data region is extended close the 1. vibration. When the first 80 range bins are used for the noise only SCM estimate, there is a slight increase in the SNR_{84} value. In other words, when almost all the range bins up to 84. cell is used for training and the size of the secondary data is increased, the noise suppression is improved by 2.2 dB, as can be seen by the new $SNR_{84} = 23.2$ dB.

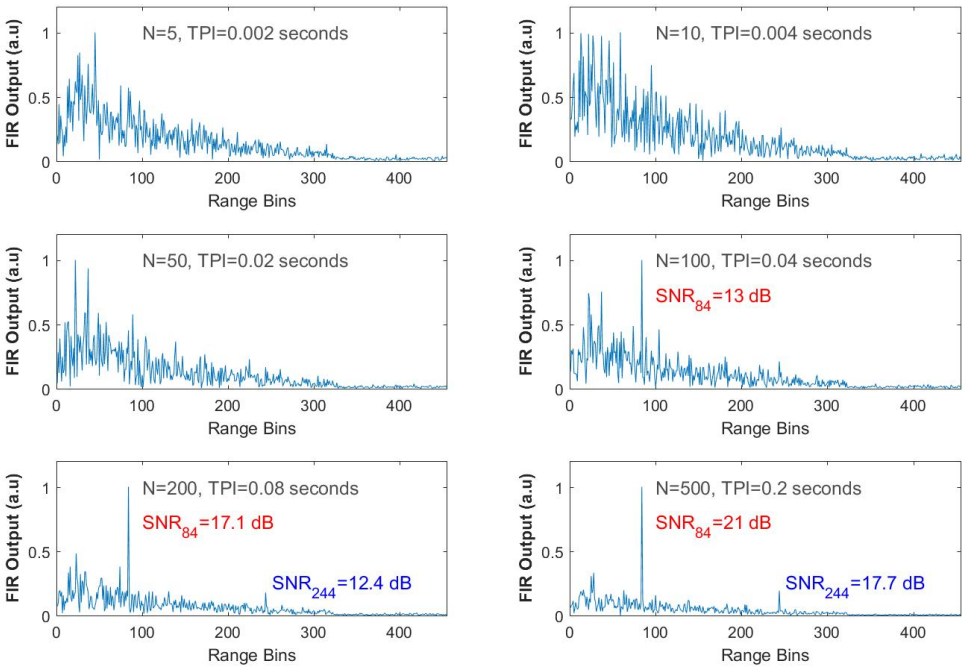


Figure 3.4. AMF test results for the multiple vibrations at the range bins 84 and 244 with identical frequencies with $\rho = 10$ and $L_s = \{1, 2, 3\}$

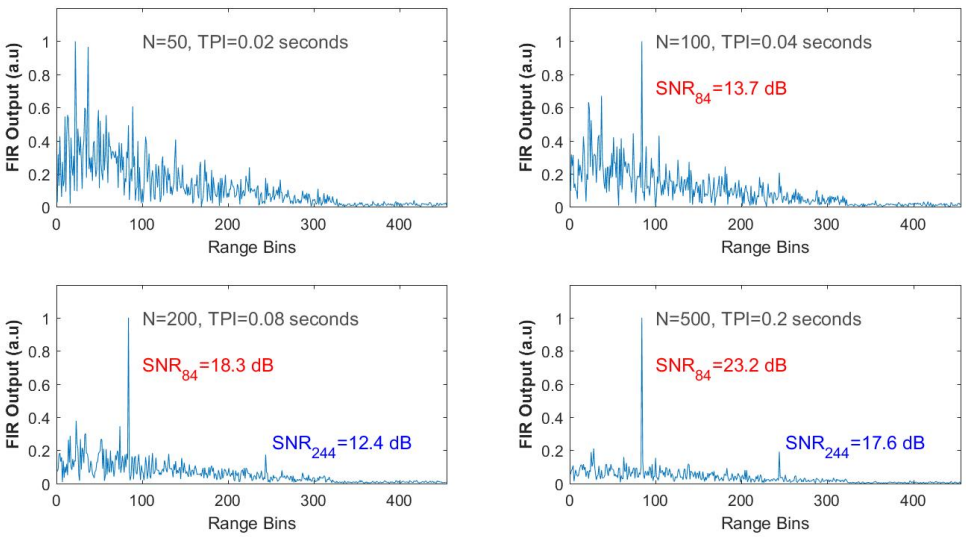


Figure 3.5. AMF test results for the multiple vibrations at the range bins 84 and 244. FIR is tuned to 833.33 Hz with with $\rho = 10$ and $L_s = \{1, 2, ..., 80\}$

The impact of another parameter modification is shown in Figure 3.6 by reducing the DL level to 1 (i.e., 0 dB) by keeping the same secondary data. It is quite interesting to see that when the DL level is reduced 10 dB, the noise suppression is significantly improved. The suppressed level at the beginning of the fiber is obvious.

Another interesting result is shown in Figure 3.7, where the training set is extended to include all the cells along the FUT route except the 84. and 244. bins (i.e.,$\#\{L_s\} = 323$). In this case, both vibrations are better visualized and SNR_{244} is improved significantly.

Numerous tests can be executed with varying L_s and DL level selections. These sample results all show that both the secondary data and DL level selections are critical factors influencing the AMF outcomes.

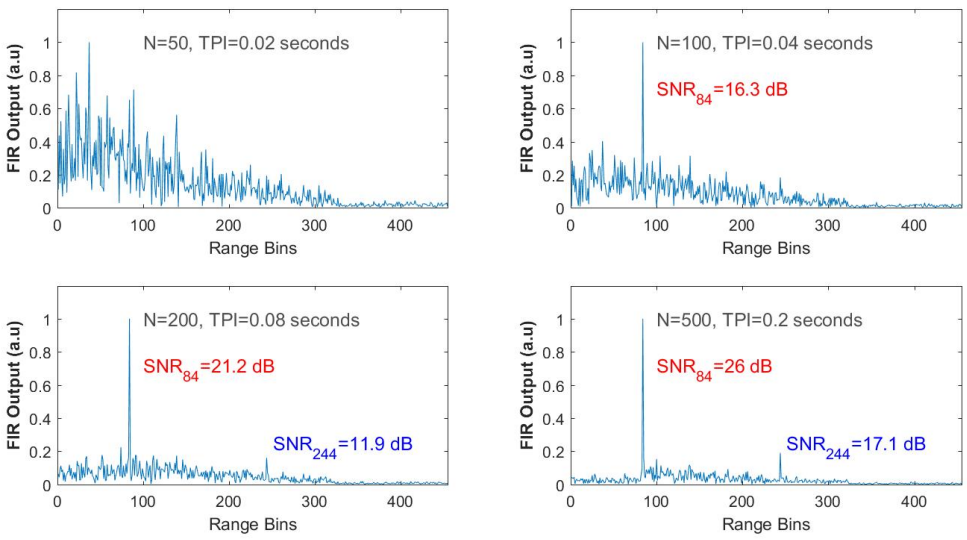


Figure 3.6. AMF test results for the multiple vibrations at the range bins 84 and 244. FIR is tuned to 833.33 Hz with $\rho = 1$ and $L_s = \{1, 2, ..., 80\}$

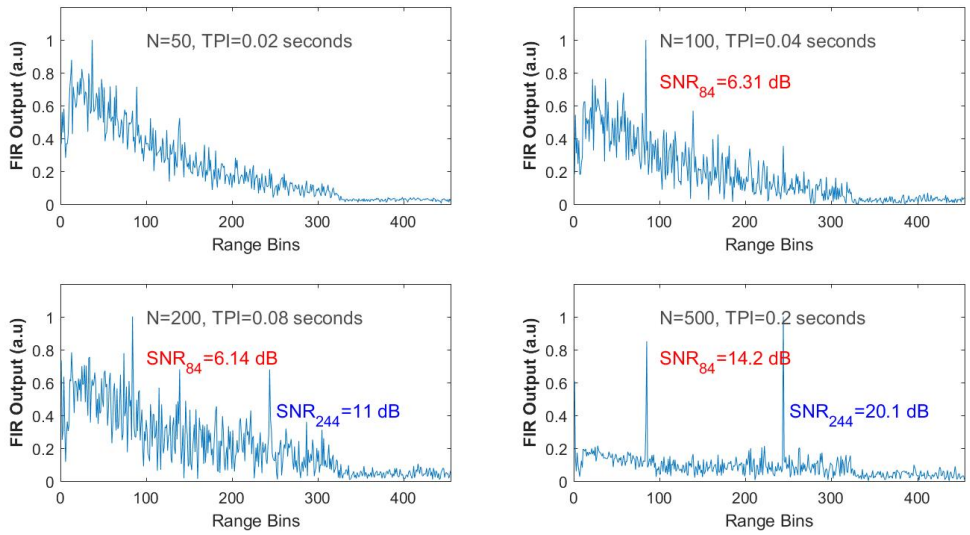


Figure 3.7. AMF test results for the multiple vibrations at the range bins 84 and 244. FIR is tuned to 833.33 Hz with $\rho = 0.1$ and the secondary data including all the range bins exluding the 84. and 244.

For the last case, multiple vibrations with different frequencies were subjected to AMF processing in a similar fashion given above. The frequency of the second vibration at the 244. range is 625 Hz with an SNR of 11 dB. All of the range bins along the FUT were selected as the secondary data, and $\rho = 1$ (0 dB). When the AMF is tuned to 833.33 Hz, only the relevant vibration at the 84. bin is revealed as can be seen Figure 3.8. More than 10 dB SNR is achieved when $N \geq 200$. When the temporal steering vector is tuned to 625 Hz, the vibration at the far end with the same frequency is extracted out of the suppressed background noise which is shown in Figure 3.9. These results are in good agreement with the results shown in Figure 3.3 as the AMF output is maximized if and only if there is a vibration at the tuned frequency. And the impact of the DL level selection is once more emphasized in Figure 3.10 where two different extreme DL level values are tested with $\rho = 100$ (20 dB) and $\rho = 0.01$ (-20 dB) for the last dataset.

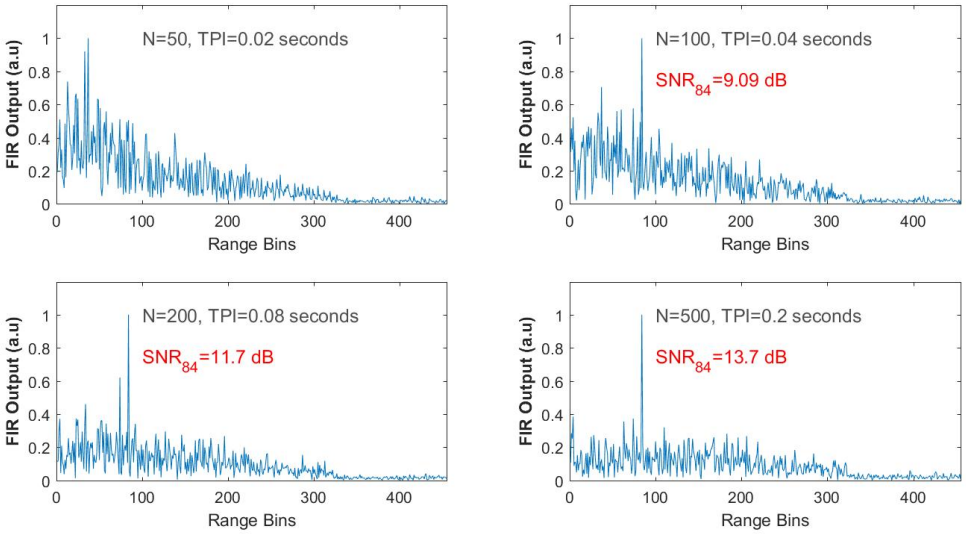


Figure 3.8. AMF test results for the multiple vibrations of 833.3 Hz at the 84. range bin and 625 Hz at the 244. range bin. FIR is tuned to 833.3 Hz with $\rho = 1$ and the secondary data including all the range bins.

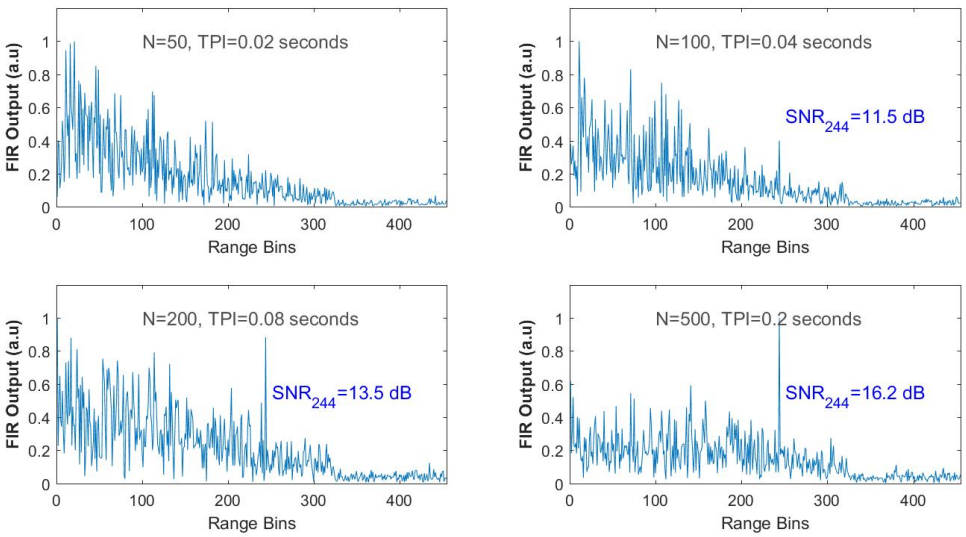


Figure 3.9. AMF test results for the multiple vibrations of 833.3 Hz at the 84. range bin and 625 Hz at the 244. range bin. FIR is tuned to 625 Hz with $\rho = 1$ and the secondary data including all the range bins.

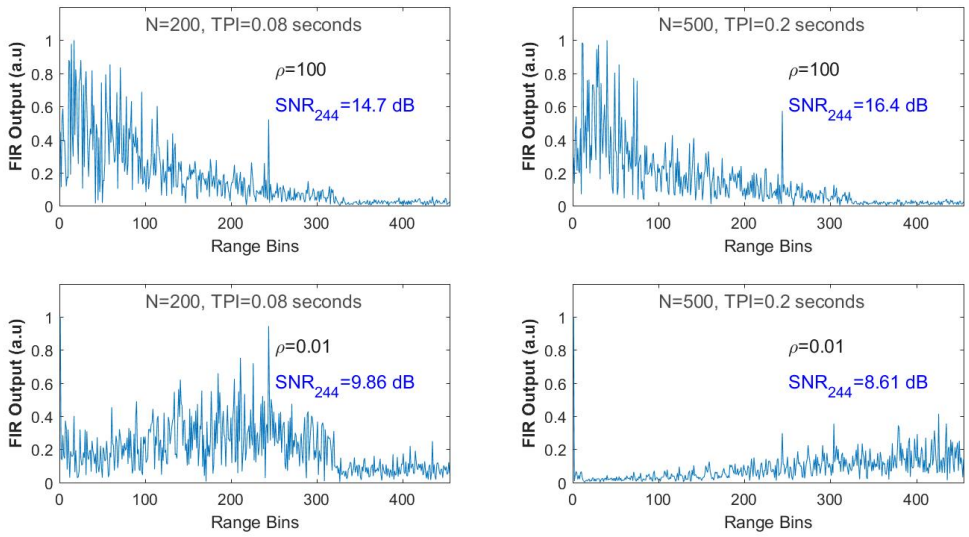


Figure 3.10. AMF test results for the multiple vibrations for two different DL levels. The FIR is tuned to 625 Hz.

3.4. Conclusions

The proposed adaptive filtering based detection algorithms were applied to Monte Carlo simulated ϕ-OTDR data. Based on the MSNR approach which were proven to be equivalent to MMSE [68], FIR filtering of ϕ-OTDR samples was found successful in reduction of the highly fluctuating amplitudes of Rayleigh based backscattering. Thus, perturbations embedded in the speckle like profile of the ϕ-OTDR traces were successfully detected without any reduction in frequency response of the detection system which is one of the drawbacks of conventional frame averaging methods. The AMF de-correlates the input data stream first, and then applies the warped matched filter to the whitened data to maximize the tuned frequency response.

Besides the adaptive structure of the AMF to dynamic environments, one of its attractive features is that it is inherently a CFAR detector [73]. Controlling the FAR is critical for most demanding DAS applications, and the CFAR property of the AMF makes the direct application of the conventional and popular CFAR techniques [74] possible for ϕ-OTDR posterior to the FIR filtering.

On the other hand, it was observed that the performance of the adaptive FIR filter is mainly dependent on the two parameters: the secondary data size and the DL level. It is clear that the training region selections during AMF processing with simulated data did not reflect real-world situations. In real DAS applications, the background along the fiber route will be heterogeneous most of the time; hence the impact of secondary data selection strategy is expected to be much more severe for real applications [75]. Nevertheless, the results presented herein clearly identify the main features of the AMF technique which can be efficiently implemented with appropriate CFAR techniques.

Selection of the DL value is another critical task. Since there is no definite rule for selecting this value, the general rule of thumb used in radar and array signal processing can be borrowed for DAS applications. The optimum level for DL will be visited again in the following chapters.

4. EIGENVALUE BASED ACOUSTIC VIBRATION SENSING

4.1. Introduction

In the previous chapter, we demonstrated the efficacy of AMF approach with simulated ϕ-OTDR signals and obtained promising results for the detection of weak acoustic perturbations. The term adaptive refers here to the structure of the filter that the optimum weight coefficients can be adapted to the environmental changes as it requires the covariance measurements of the input data. Additionally, when the steering vector is tuned to the actual vibration spectrum, the output SNR is expected to be maximized. Since the AMF method relies on the prior information about the vibration spectra, a brute force method to seek the entire frequency spectrum seems to be crucial for most practical applications. Depending on the resolution of this scanning process, from nearly DC to maximum detectable frequency which is equal to half of the PRF of the ϕ-OTDR system may not be an appropriate method. Using multicore processing power or utilizing parallel programming with appropriate hardware for this task can be easily accomplished, but it naturally introduces additional complexity and costs to the DAS hardware. In this regard, it is of importance to seek robust techniques to either improve the AMF method or to find more efficient methods.

In this chapter, we introduce a new approach for DAS/DVS systems which is based on the eigenspectra of the measured covariance. Similar to the AMF processing, our main concern is again the SCM of the vibration-free case. The eigenvalue distribution is exploited for vibration detection and a CFAR detection technique is developed by adapting the recent results from the RMT. Eigenvalue-based detection techniques are successively applied especially in cognitive radio applications where the efficient use of the frequency is of the main concern.

Adhering to the similarity of the signal model used in cognitive radio [76–78] and the effectiveness of the eigenvalue based techniques in detection primary users in the spectrum, the covariance structure of the ϕ-OTDR system is reviewed and the detection methodology is developed for DAS/DVS systems in this chapter.

4.2. Covariance Structure of the ϕ-OTDR Data

The hypothesis testing problem for vibration detection in a ϕ-OTDR system can be interpreted in terms of the correlation properties of the measured data. Since the noise and signal contributions are independent and uncorrelated, the covariance matrices of the two hypotheses can be given as

$$\boldsymbol{R}_x = \begin{cases} \boldsymbol{R}_{rbs} + \sigma_w^2 \boldsymbol{I}_N & , \quad H_0 \\ \boldsymbol{R}_s + \boldsymbol{R}_{rbs} + \sigma_w^2 \boldsymbol{I}_N & , \quad H_1 \end{cases} \tag{4.1}$$

where $\boldsymbol{R}_{rbs}$ and $\sigma_w^2 \boldsymbol{I}_N$ are the covariance matrices of RBS and receiver noise, respectively; $\boldsymbol{R}_s = P_0 \boldsymbol{s}(\overline{\omega}) \boldsymbol{s}(\overline{\omega})^T$ is the rank one signal covariance matrix with $P_0 = |\sigma_p|^2$ being the signal power.The RBS term is inherently correlated on the fast-time axis but it can be assumed uncorrelated if samples are collected on the same range bin (i.e. the slow-time axis). In most practical cases, the real ϕ-OTDR samples will exhibit some degree of correlations due to the band-pass structure of the data acquisition systems. In general, these samples can be decorrelated by pre-whitening techniques as it was inherently embedded in the AMF structure. Nevertheless, in this chapter the collected samples for each acoustic channel will be considered white and for simplicity, we will assume $\boldsymbol{R}_{rbs} = \sigma_{rbs}^2 \boldsymbol{I}_N$ with σ_{rbs}^2 being the average power of the RBS. In this case, the SCM for the noise only case can be approximated by $\boldsymbol{R}_{x/H_0} = (\sigma_{rbs}^2 + \sigma_w^2) \boldsymbol{I}_N = \sigma_\eta^2 \boldsymbol{I}_N$ where σ_η^2 is the average noise power. Finally, in the general case, in the case of a pure sinusoidal vibration presence, the SCM will be the sum of two independent components:

$$\boldsymbol{R}_{x/H_1} = P_0 \boldsymbol{s}(\overline{\omega}) \boldsymbol{s}(\overline{\omega})^T + \sigma_\eta^2 \boldsymbol{I}_N \tag{4.2}$$

4.3. Maximum Eigenvalue of the Sample Covariance Matrix

It can be easily seen that the signal vector is one of the eigenvectors of $\boldsymbol{R}_x$ if we multiply (4.2) at both sides with $\boldsymbol{s}(\overline{\omega})$

$$\boldsymbol{R}_x\boldsymbol{s} = (P_0\boldsymbol{s}\boldsymbol{s}^T + \sigma_\eta^2\boldsymbol{I}_N)\boldsymbol{s} = P_0\boldsymbol{s}\boldsymbol{s}^T\boldsymbol{s} + \sigma_\eta^2\boldsymbol{s} = (NP_0 + \sigma_\eta^2)\boldsymbol{s} \tag{4.3}$$

where in the last step $\boldsymbol{s}\boldsymbol{s}^T = ||\boldsymbol{s}||^2 = N$ and we took $\boldsymbol{s} = \boldsymbol{s}(\overline{\omega})$ for simplicity in notation. Thus, the corresponding eigenvalue – which is also the maximum eigenvalue of the SCM – is equal to $(NP_0 + \sigma_\eta^2)$. Since $\boldsymbol{R}_x$ has N orthonormal eigenvectors, all of the other $(N-1)$ eigenvectors will be orthogonal to $\boldsymbol{s}$ (i.e. $\boldsymbol{s}^T\boldsymbol{e}_n = 0$ where $\boldsymbol{e}_n$ is the n-th eigenvector).

$$\boldsymbol{R}_x\boldsymbol{e}_1 = P_0\boldsymbol{s}\boldsymbol{s}^T\boldsymbol{e}_1 + \sigma_\eta^2\boldsymbol{e}_1 = \sigma_\eta^2\boldsymbol{e}_1 \tag{4.4}$$

The equation (4.4) states that all of the other eigenvalues of the SCM are equal to σ_η^2. From above, it can be easily deduced that we can use the largest eigenvalue of the SCM as a test statistic for our binary decision problem. In this case, the test statistic can be heuristically transformed to

$$\lambda_{max}(\boldsymbol{R}_x) > \gamma\sigma_\eta^2 \tag{4.5}$$

where γ is a threshold value greater than 1. Principally, the vibration frequency is also determined from the largest eigenvector as the $\boldsymbol{e}_1$ is a complex exponential with a vibration frequency ω_1.

4.4. Eigenvalue Statistics of the Sample Covariance and the Random Matrix Theory

As mentioned above, the eigenvalues of the RBS covariance will contribute to the noise space in the absence of vibration. For this case, since the samples of the data are all noise which is a purely stochastic process; the matrix is simply called as a "random matrix"; which in turn results in pure randomness of the RBS covariance. Since our vibration detection methods proposed here will be based on extreme eigenvalues of the noise only SCM, it is worth to review the eigenvalue distributions of random matrices which exhibit very interesting properties.

The spectra of random matrices and its spectral properties are studied by RMT which has gained big attention of researchers and engineers recently. The RMT is being exploited in numerous research fields ranging from nuclear spectra, atomic physics [79], multidimensional data analysis [80, 81], machine learning [82], and wireless communications [78]. To handle the classical problem of separating noise and signal components from received samples, the RMT related detection methodologies are based on the eigenvalue statistics of the received data which are successfully tailored for signal detection purposes in cognitive radio and radar applications [80] .

Let the sorted real eigenvalues for a $N \times N$ random and Hermitian matrix $\boldsymbol{A}_N$ be denoted by $\lambda_1, \lambda_2, ..., \lambda_N$. Central to the large random matrices is the empirical eigenvalue distribution function of $\boldsymbol{A}_N$ with N real eigenvalues (counted with multiplicity) which is defined as

$$F^{\boldsymbol{A}_N}(\lambda) = \frac{1}{N}\#\{n : \lambda_n \leq \lambda\} \tag{4.6}$$

where #{ } denotes the number of elements of the set { }. The function $F^{\boldsymbol{A}_N}(\lambda)$ is also called the cumulative distribution function (CDF) of the discrete uniform random variable which can be written in a more mathematical form as

$$F^{\boldsymbol{A}_N}(\lambda) = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{(-\infty,\lambda)}(\lambda_n) \tag{4.7}$$

where $\mathbf{1}_{(a,b)}(\lambda) = 1$ for $a \leq \lambda \leq b$ and is zero otherwise. The empirical eigenvalue probability density function (PDF) $f_{\boldsymbol{A}_N}(\lambda)$ can be defined as the derivative of $F^{\boldsymbol{A}_N}(\lambda)$:

$$dF^{\boldsymbol{A}_N}(\lambda) = f_{\boldsymbol{A}_N} d\lambda = \frac{1}{N}\sum_{n=1}^{N} \delta(\lambda - \lambda_n) d\lambda \tag{4.8}$$

where $\delta(\lambda)$ is a Dirac delta function. For a broad class of random matrices, in the limit case $N \to \infty$, $f_{\boldsymbol{A}_N}(\lambda)$ converges to a non-random limiting eigenvalue density function $f_{\boldsymbol{A}}(\lambda)$.

In general, there are two types of eigenvalue distributions of interest: bulk statistics and edge statistics. The former refers to the full set of eigenvalues a random matrix while the latter deals with the fluctuations of the largest and smallest eigenvalues. There exist three theorems which can be counted as the main theorems of the RMT: Wigner's semi-circle, Marcenko-Pastur and Tracy-Widom. Wigner's semi-circle law is attributed to symmetric random matrices while the other two theorems concern Wishart type random matrices.

The extreme eigenvalues of a pure random matrix are of special interest most of the time as they are used to describe important physical mechanisms in diverse applications. Since the SCM in our application is a Wishart type random matrix, we are heuristically interested in using the two theorems of RMT for DAS/DVS applications as described below.

4.4.1. Bulk Statistics of the SCM: The Marcenko-Pastur Law

The fundamental tool to analyze the spectrum of large sample covariance matrices is the Marcenko-Pastur(MP) [83] theorem which states the following:

Theorem 4.1. *Let $\boldsymbol{X}$ be a sequence of random $N \times M$ matrices $(N = 1, 2, ...)$, with $M > N$, satisfying the following:*

- *The elements x_{ij} of each $\boldsymbol{X}$ are independent random variables with zero mean and unit variance.*
- *M depends on N in such a way that the ratio of rows to columns $\kappa = N/M < 1$ as $N \to \infty$.*

Then the distribution of the eigenvalues of the Wishart matrix $\boldsymbol{W} = (1/M)\boldsymbol{X}\boldsymbol{X}^T$ asymptotically approaches the following:

$$f_{\boldsymbol{W}}(\lambda) = \left(1 - \frac{1}{\kappa}\right)\delta(\lambda) + \frac{\sqrt{(\lambda - \lambda_{min})(\lambda_{max} - \lambda)}}{2\pi\lambda\kappa}\mathbf{1}_{(\lambda_{min},\lambda_{max})}(\lambda) \tag{4.9}$$

where $\lambda_{min} = (1 - \sqrt{\kappa})^2$ and $\lambda_{max} = (1 + \sqrt{\kappa})^2$ for eigenvalue λ within the bound $\lambda_{min} < \lambda < \lambda_{max}$.

Hence, the minimum eigenvalue λ_N is bounded by λ_{min} and the maximum eigenvalue λ_1 is bounded by λ_{max}. Figure 4.1 shows the analytical and empirical eigenvalue distributions for four different values of κ. The empirical distributions were computed in MATLAB by simply running the $\boldsymbol{X} = randn(m, n)$ command and computing the eigenvalues of the realized Wishart matrices for 500 trials. It can be easily seen that histograms (yellow) are in good agreement with the analytical computation (blue). During the computations N was fixed to 50 and M was computed for $\kappa \in \{0.02, 0.05, 0.1, 0.2\}$ with M=round(N/κ). The realized Wishart type random matrices were all real valued random matrices which are called Laguerre Orthogonal Ensembles (LOE). The MP Law is also valid when the matrix elements are complex valued which is called the Laguerre Unitary Ensemble (LUE) in the literature.

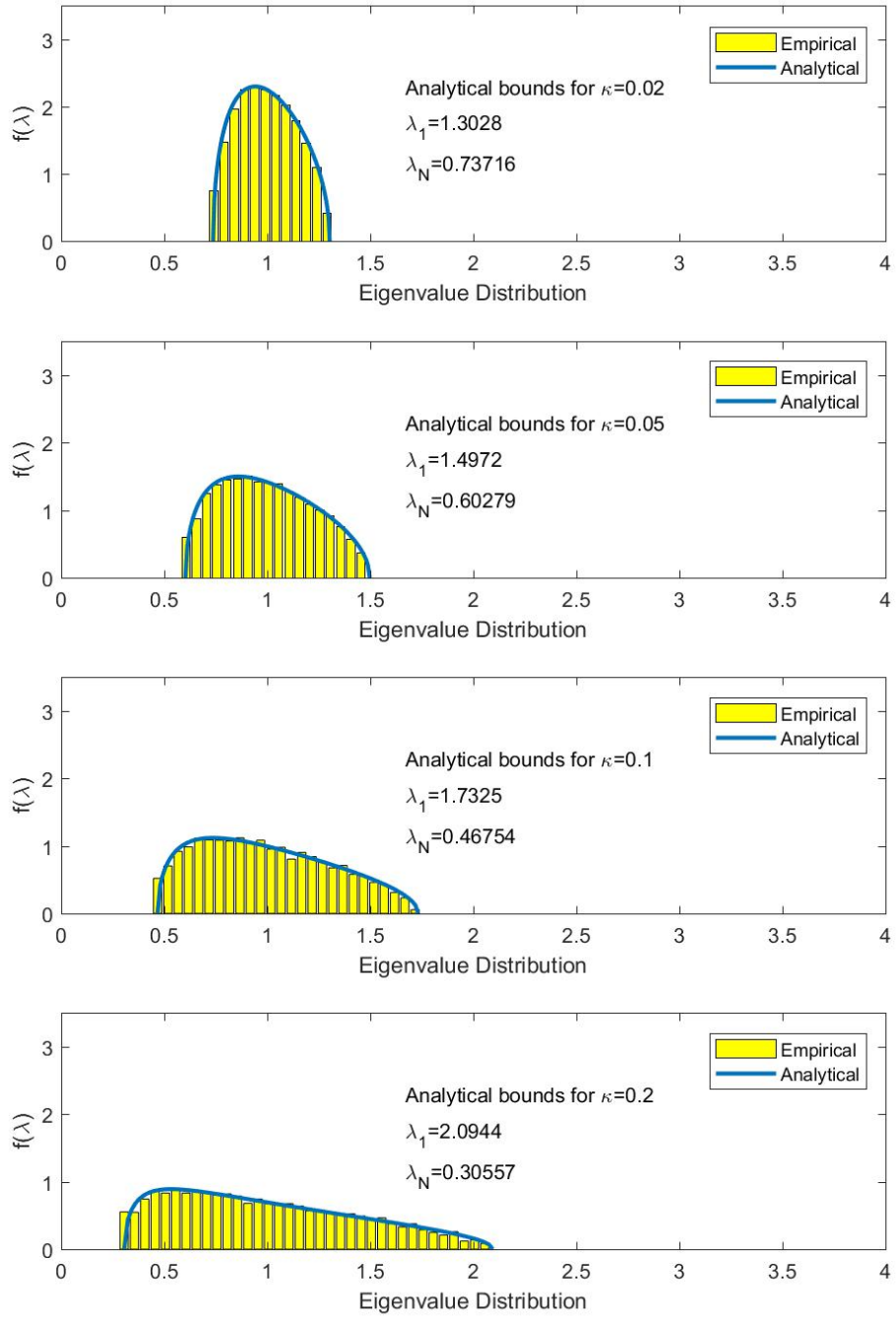


Figure 4.1. Empirical and analytical eigenvalue distributions of a Wishart random matrix for 4 different values of κ after 500 trials with size 50×50.

It is easy to see that the "maximum" and the "minimum" eigenvalues which are denoted by λ_{max} and λ_{min} can be computed (estimated) without the eigen-decomposition of SCM, by just knowing the "observation dimension" N and "the number of observations" M. The lower the κ value, the more compact the eigenvalue distribution of the covariance matrix. Although (4.9) is only valid for $M \to \infty$ and $N \to \infty$, even finite values of M and N, the distribution of the eigenvalues of the random matrix fits the distribution function given by (4.9). Any outlier out of these bounds should indicate the presence of a coherency in the data. In other words, the bounds λ_{max} and λ_{min} can be used to track the statistical distribution of the SCM and to separate noise and signal contributions.

4.4.2. Edge Statistics of the SCM: The Tracy-Widom Law

Although the MP Law gives the minimum and the maximum eigenvalues of a random matrix in the limit, it does not tell anything about the fluctuations of these extreme eigenvalues. The SCM defined in Section 3.2.2 and given by (3.19) is called a real (complex) Wishart matrix when the received samples are real (complex). In the case when κ (the ratio of number of observations to the observation dimension), tends to a constant, the limiting distribution of largest eigenvalue the Wishart type matrices with identity covariance is described again by the Tracy-Widom law [84] as proved by Johnstone [85]. It has been also reported in [85] that the Tracy-Widom approximation is satisfactory for N and M as small as 10. The result of Johnstone can be formulated as follows:

Theorem 4.2. *Let $\boldsymbol{W}$ be a white Wishart matrix and λ_{max} be its largest eigenvalue. Then the distribution*

$$\frac{\lambda_{max}(\boldsymbol{W}) - \mu_{ML,\beta}}{\sigma_{ML,\beta}} \tag{4.10}$$

approaches to the Tracy-Widom distribution of order β denoted by F_β with $\beta = 1$ if $\boldsymbol{W}$ is real and $\beta = 2$ if $\boldsymbol{W}$ is complex.

The center and scaling parameters are given by

$$\mu_{ML,1} = (\sqrt{M-1} + \sqrt{L})^2 \tag{4.11a}$$

$$\sigma_{ML,1} = \sqrt{\mu_{ML,1}}(\frac{1}{\sqrt{M-1}} + \frac{1}{\sqrt{L}}) \tag{4.11b}$$

for $\beta = 1$ *and*

$$\mu_{ML,2} = (\sqrt{M} + \sqrt{L})^2 \tag{4.12a}$$

$$\sigma_{ML,2} = \sqrt{\mu_{ML,2}}(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{L}}) \tag{4.12b}$$

for $\beta = 2$ *respectively.*

F_1 refers to a particular distribution function from a family of cumulative distribution functions (CDF) F_β and related probability distribution functions (PDF) f_β for β=1; 2; 4. F_β appear as the limiting distributions for the largest eigenvalues in the Gaussian Orthogonal (GOE), Gaussian Unitary (GUE) and Gaussian Symplectic ensembles (GSE), correspondingly. The three CDFs are given as

$$F_2(\lambda) = exp\left(-\int_\lambda^\infty (y-\lambda)q^2(y)dy\right) \tag{4.13a}$$

$$F_1(\lambda) = exp\left(-\frac{1}{2}\int_\lambda^\infty q(y)dy\right)[F_2(\lambda)]^{1/2} \tag{4.13b}$$

$$F_4(2^{-2/3}\lambda) = cosh\left(-\frac{1}{2}\int_\lambda^\infty q(y)dy\right)[F_2(\lambda)]^{1/2} \tag{4.13c}$$

Here $q(y)$ is the unique solution to the Painleve II equation

$$q(\lambda)'' = \lambda q(\lambda) + q(\lambda)^3 \tag{4.14}$$

satisfying the boundary condition

$$q(\lambda) \sim Ai(y), \quad y \to +\infty \tag{4.15}$$

where $Ai(\lambda)$ denotes the Airy function. The equations given by (4.13) are difficult to evaluate and the random matrix model for the GSE has not been studied much in the literature. On the other hand, the matrix models for GOE and GUE are often visited and there are already calculated tables for $\beta = 1$ and $\beta = 2$ and MATLAB codes to compute the thresholds [86, 87]. In this research study, we focus on the intensity data (i.e.real valued data) at the photo-detection output, hence we are concerned with the GOE case. Some values of F_1 are given in Table 4.1.

Table 4.1. Some values of $F_1(\lambda)$.

λ	-3.90	-3.18	-2.78	-1.91	-1.27	-0.59	0.45	0.98	2.02
$F_1(\lambda)$	0.01	0.05	0.10	0.30	0.50	0.70	0.90	0.95	0.99

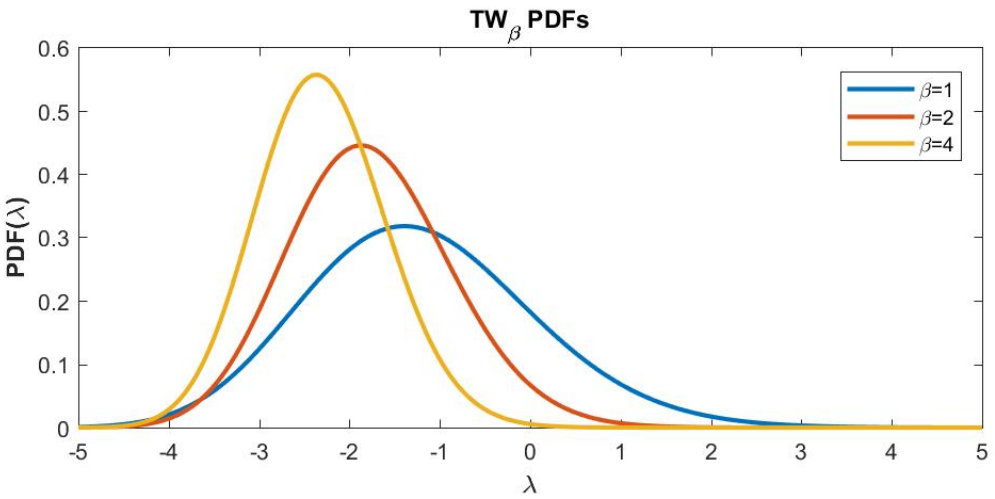


Figure 4.2. TW distributions representing the distribution of the largest eigenvalue of the SCM for 3 different (real, complex and quaternion valued) SCMs.

4.5. Maximum Eigenvalue Detection (MED)

Our objective is to evaluate the largest eigenvalue of the SCM for vibration detection. To proceed with our objective, we start by segmenting the input ϕ-OTDR data stream of size N into L equal length consecutive blocks. In other words, the TPI will be interpreted as an $N = ML$ long sequence constructed from slow-time samples as illustrated in Figure 4.3.

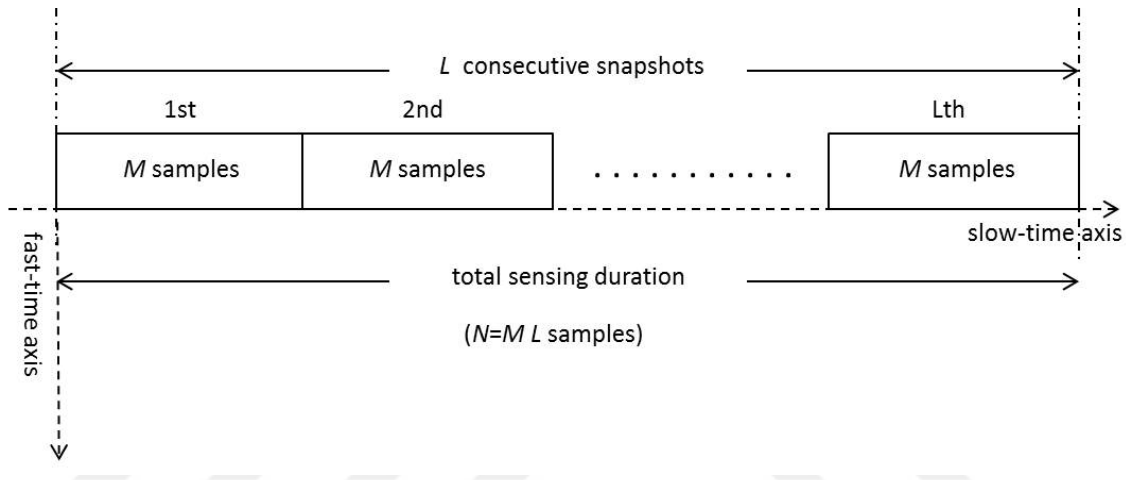


Figure 4.3. Segmentation of the input data for the MED technique.

For each k-th channel of the DAS system, by considering the each segment as a column vector and concatenating them in order, we now organize the ϕ-OTDR RBS data matrix of the form:

$$\boldsymbol{X}_k = [\boldsymbol{x}_{k,1} \quad \boldsymbol{x}_{k,2} \quad ... \quad \boldsymbol{x}_{k,L}] = \begin{bmatrix} x_{k,1}(1) & x_{k,2}(1) & ... & x_{k,L}(1) \\ x_{k,1}(2) & x_{k,2}(2) & ... & x_{k,L}(2) \\ . & . & ... & . \\ . & . & ... & . \\ x_{k,1}(M) & x_{k,2}(M) & ... & x_{k,L}(M) \end{bmatrix} \tag{4.16}$$

As we are interested in the correlation properties of the received samples, we need the statistical covariance matrix of the data. Due to the inherent finiteness of the measured data, instead of the real statistics, we have the SCM which can be estimated by

$$\hat{\boldsymbol{R}}_{x,k} = \frac{1}{L}\boldsymbol{X}_k\boldsymbol{X}_k^T = \frac{1}{L}\sum_{l=1}^{L}\boldsymbol{x}_{k,l}\boldsymbol{x}_{k,l}^T \tag{4.17}$$

where L is the number of segments as shown in Figure 4.3, and $l = 1, 2, ..., L$. Equation (4.17) is no different from the SCM estimated in the AMF approach given by (3.9), where the noise only assumed vectors were taken into account for noise reduction processing. Here, the situation is only different that the spectrum of the SCM is of interest. By remembering the two hypotheses written in terms of the SCMs in (4.1), the largest eigenvalue is evaluated for our binary detection problem. In this case, for the k-th acoustic channel, the test statistic denoted by $T(\boldsymbol{x}_k)$ can be heuristically transformed to

$$T(\boldsymbol{x}_k) = \begin{cases} H_0 & , \quad \text{if } \lambda_{max}(\hat{\boldsymbol{R}}_{x,k}) < \gamma\sigma_\eta^2 \\ H_1 & , \quad \text{if } \lambda_{max}(\hat{\boldsymbol{R}}_{x,k}) \geq \gamma\sigma_\eta^2 \end{cases} \tag{4.18}$$

where γ is the predetermined threshold as stated before. Setting this threshold in binary detection problem is critical because a too low threshold will result in an increase in the FAR while a too high threshold will lead to missing weak vibrations. In that case, the probability distribution of the largest eigenvalue of the SCM becomes the key issue in setting the appropriate threshold. Since it is impossible to have a priori knowledge about the vibration of interest in most cases, we can't know the probability distribution function of the SCM for the H_1 case. For this reason, we will rely on the noise-only assumed samples and follow the well-known Neyman-Pearson criterion [88, 89] to set a threshold for a predetermined fixed FAR.

The probability of false alarm (P_{FA}) of the largest eigenvalue based detection technique proposed here can be computed as follows [90]:

$$\begin{aligned} P_{FA} &= P(\lambda_{max}(\hat{\boldsymbol{R}}_x) > \gamma\sigma_\eta^2) = P\left(\frac{\sigma_\eta^2}{L}\lambda_{max}(\boldsymbol{W}) > \gamma\sigma_\eta^2\right) \\ &= P(\lambda_{max}(\boldsymbol{W}) > \gamma L) = P\left(\frac{\lambda_{max}(\mathbf{W}) - \mu_{ML,\beta}}{\sigma_{ML,\beta}} > \frac{\gamma L - \mu_{ML,\beta}}{\sigma_{ML,\beta}}\right) \\ &\approx 1 - F_\beta\left(\frac{\gamma L - \mu_{ML,\beta}}{\sigma_{ML,\beta}}\right) \end{aligned} \tag{4.19}$$

Remember that M is the number of noise-only snapshots used in estimating the SCM (i.e. the column size of the data matrix). (4.19) can be written as

$$F_\beta\left(\frac{\gamma L - \mu_{ML,\beta}}{\sigma_{ML,\beta}}\right) = 1 - P_{FA} \tag{4.20}$$

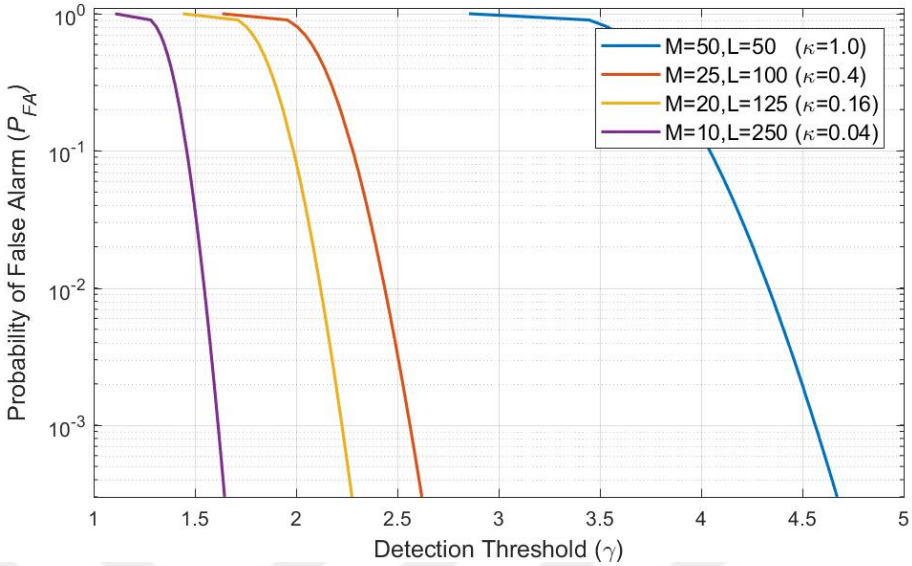
Thus, we can write

$$\frac{\gamma L - \mu_{ML,\beta}}{\sigma_{ML,\beta}} = F_\beta^{-1}(1 - P_{FA}) \tag{4.21}$$

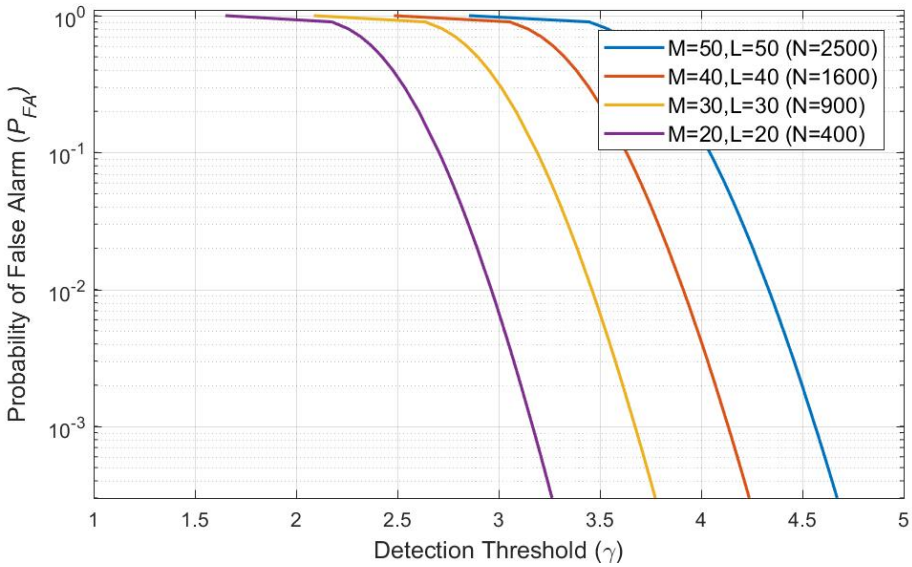
By using the definitions of $\mu_{ML,\beta}$ and $\sigma_{ML,\beta}$ given above the threshold can be computed as:

$$\gamma = \frac{\sigma_{ML,\beta} \cdot F_\beta^{-1}(1 - P_{FA})}{L} + \frac{\mu_{ML,\beta}}{L} \tag{4.22}$$

It can be seen by (4.22) that the threshold value is dependent on the data matrix size and the preselected FAR value. P_{FA} as a function of κ was computed for some κ and TPI values and the results are shown in Figure 4.4.



(a)



(b)

Figure 4.4. Probability of false alarm versus threshold: (a) for fixed TPI and with various κ values; (b) for square data matrices ($\kappa = 1$) for various TPI durations.

By considering all the above analytical derivations, the following steps can be followed for a maximum eigenvalue based detection algorithm which will be called shortly as the MED algorithm from now on:

(i) A segmentation strategy is determined for the block size (M) and number of consecutive blocks (L) to process at a time and the ϕ-OTDR data is collected.

(ii) The data and the associated SCM are constructed by via (4.16) and (4.17). The columns of the covariance matrix are properly scaled to have zero mean and unit variance.

(iii) The threshold κ is computed via (4.22) for a preselected FAR rate.

(iv) The eigenvalues of the SCM, hence the largest eigenvalue $\lambda_{max}(\hat{\boldsymbol{R}}_{x,k})$ is computed and compared with the threshold. The decision is made in favor of an acoustic vibration if $\lambda_{max}(\hat{\boldsymbol{R}}_{x,k}) \geq \kappa$, or vice versa. The above steps are to be repeated for every acoustic channel in a similar manner done before for the AMF algorithm in Chapter 3.

4.6. Applications of MED for Vibration Detection

The performance of the MED algorithm was evaluated for both strong and weak vibration conditions as we did in the previous chapter and the same four data sets with varying TPI durations were used. As there are numerous choices for the selection of the M and L; limited numbers of test results are given for the MED with TPI durations comparable of those tested with the AMF approach. Additionally, the empirical eigenvalue distributions were also computed and compared with the analytical equations of Section 4.4. The application of the MED algorithm for various (M, L) pairs will be denoted by MED(M, L) throughout the text. The associated TW threshold values computed for the selected (M, L) pair will be shown by dashed red lines on every plot. Since the simulated data was based on intensity detection and real-valued, the TW thresholds were computed for $\beta = 1$.

4.6.1. Single-Strong Vibration Case

As done before in the previous chapter, we start with the ϕ-OTDR data with a clearly visible vibration having a slow-time $SNR = 21$ dB at the 84. range bin. The results for this 1. dataset with selections of MED(5,10), MED(5,20), MED(5,40) and MED$(5, 100)$ with a selected FAR rate of $P_{FA} = 0.01$ are shown in Figure 4.5.

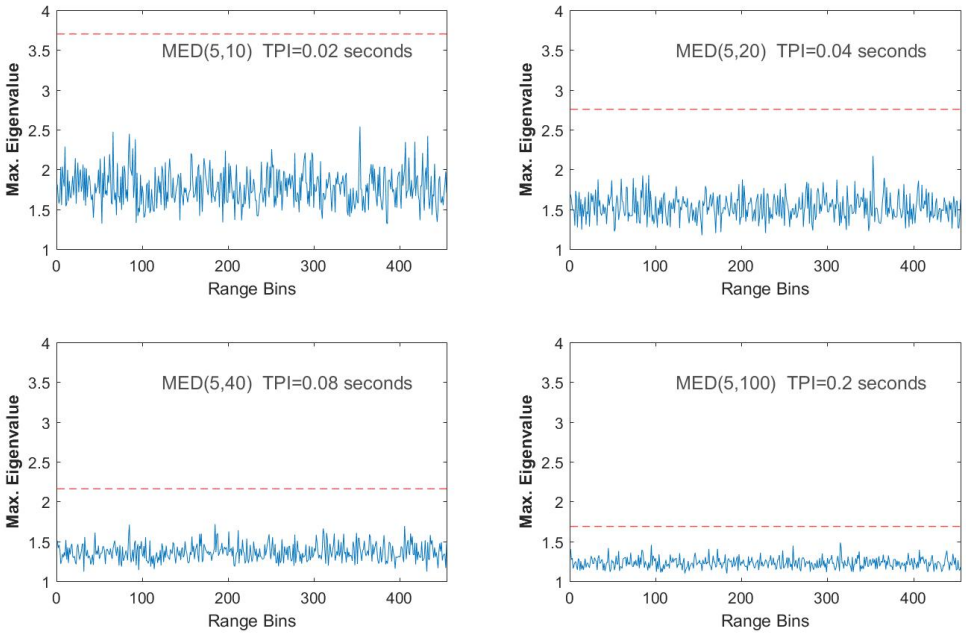


Figure 4.5. MED(M, L) processing results of the 1. data set with $M = 5$, $L = \{10, 20, 40, 100\}$ and $P_{FA} = 0.01$.

Here, the segment size was fixed at $M = 5$ and the TPI durations were increased from $N = 50$ to $N = 500$ by increasing the number of consecutively observed segments. Only four descriptive samples were presented here and none of the (M, L) selections yielded a peak at the expected range bin after MED processing as shown in Figure 4. 5; hence the target is missed. When M is increased to 10 with the same increasing order of L in Figure 4.5, the target vibration is still missed except for MED(10,50) where the vibration at 84. range bin revealed out but hardly detected as can be seen in the bottom left plot of Figure 4.6. The relevant TW threshold for $P_{FA} = 0.01$ with MED(10,20) and MED(20,40) is above the λ_1 of the SCM.

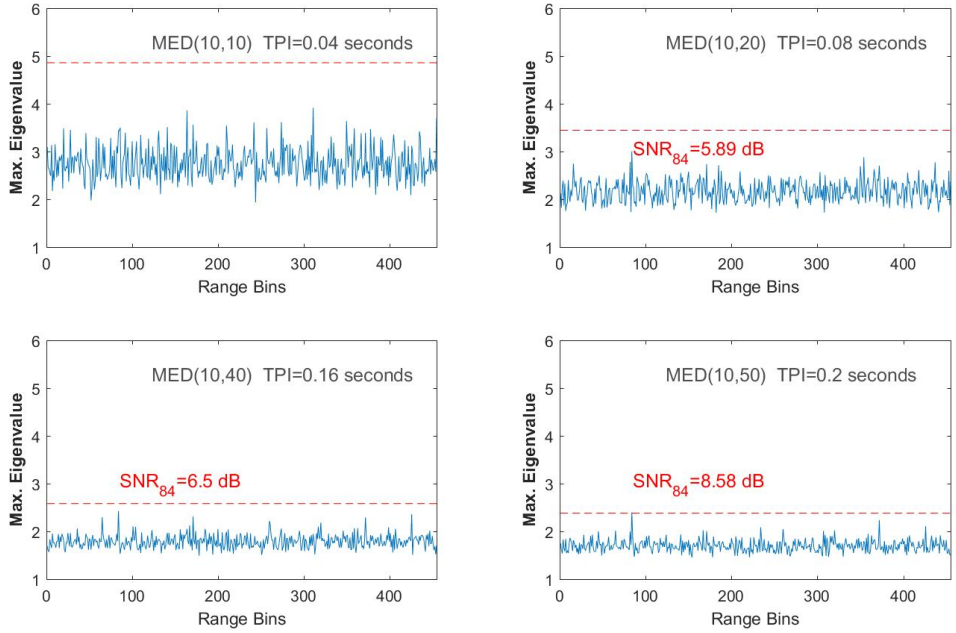


Figure 4.6. MED(M, L) processing results of the 1. data set with $M = 10$, $L = \{10, 20, 40, 100\}$ and $P_{FA} = 0.01$.

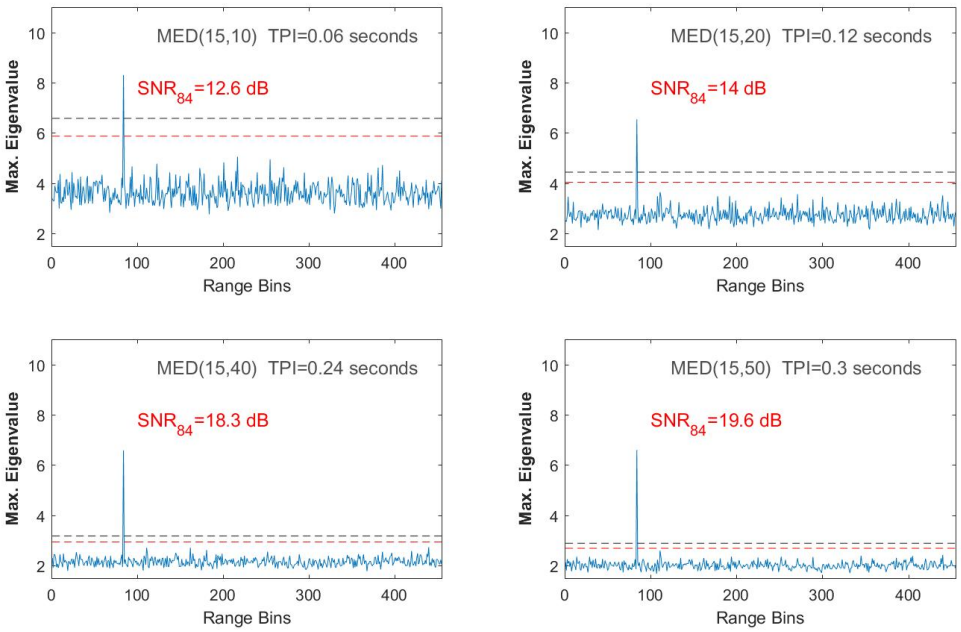


Figure 4.7. MED(M, L) processing results of the 1. data set with $M = 15$ $L = \{10, 20, 40, 100\}$ and two P_{FA} vaues:$P_{FA} = 0.01$ and $P_{FA} = 0.001$.

The results of MED processing for $M = 15$ are shown in Figure 4.7. The thresholds were computed for two different FAR rates: $P_{FA} = 0.01$ (dashed red line) and $P_{FA} = 0.001$ (dashed gray line). It is clear that for all TPI selections the vibration is solely observed and successfully detected with the selected FAR rate of 0.01. Since all the observed fast-time SNR values are high, the vibration will be still detectable even with a more stringent FAR rate of $P_{FA} = 0.001$ where all the computed relevant thresholds for (M, L) selections are lower than the largest eigenvalue of the SCM.

The performance of MED processing with square data matrix (i.e., $\kappa = 1$) selections was also analyzed as shown in Figure 4.8 for 4 different (M, L) pairs. Except for $M = 10$, the vibration was detected with $P_{FA} = 0.01$ for all other cases. A critical situation of missing the vibration exists where FAR is reduced to 0.001 since the fast-time SNR computed for MED(20,20) is at a critical level and just below the threshold. An interesting result was observed for MED(40,40) where the fast-time SNR was below the fast-time SNR observed for MED(30,30) although the TPI duration is longer with (40,40) than that of the TPI with (30,30).

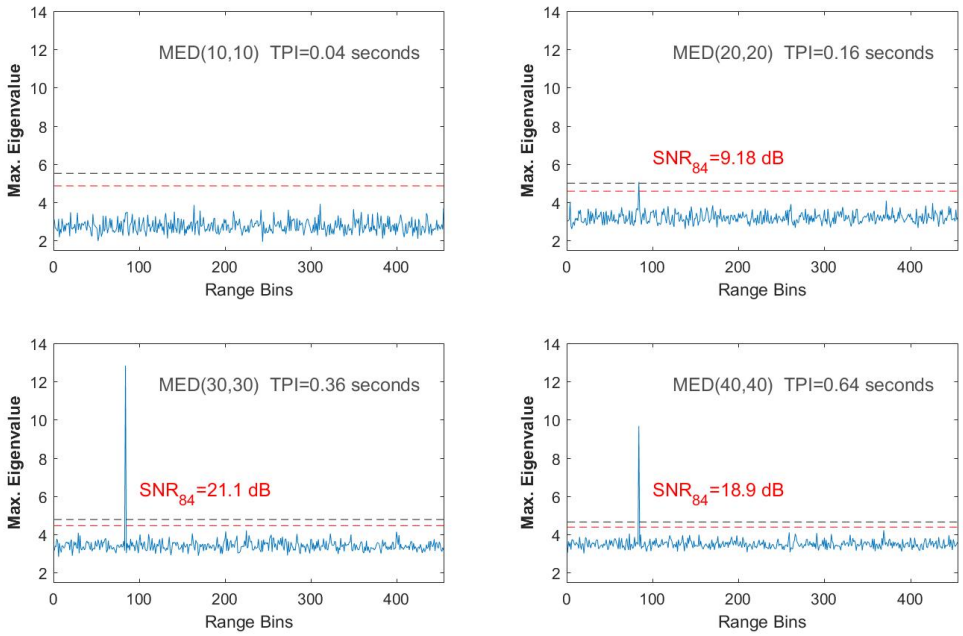


Figure 4.8. MED test results of the 1. data set for 4 different square sized data matrix $(M = L)$ selections. $L = \{10, 20, 30, 40\}$ with $P_{FA} = 0.01$.

4.6.2. Single-Weak Vibration Case

The MED processing of the previous section was repeated for the relatively weaker vibration data with the same (M, L) selections. First sample results, with a fixed selection of $M = 5$ and TPI increases with $L = \{10, 20, 40, 100\}$ are shown in Figure 4.9. None of the MED processing yielded a peak at the 84. range bin and the vibration is missed.

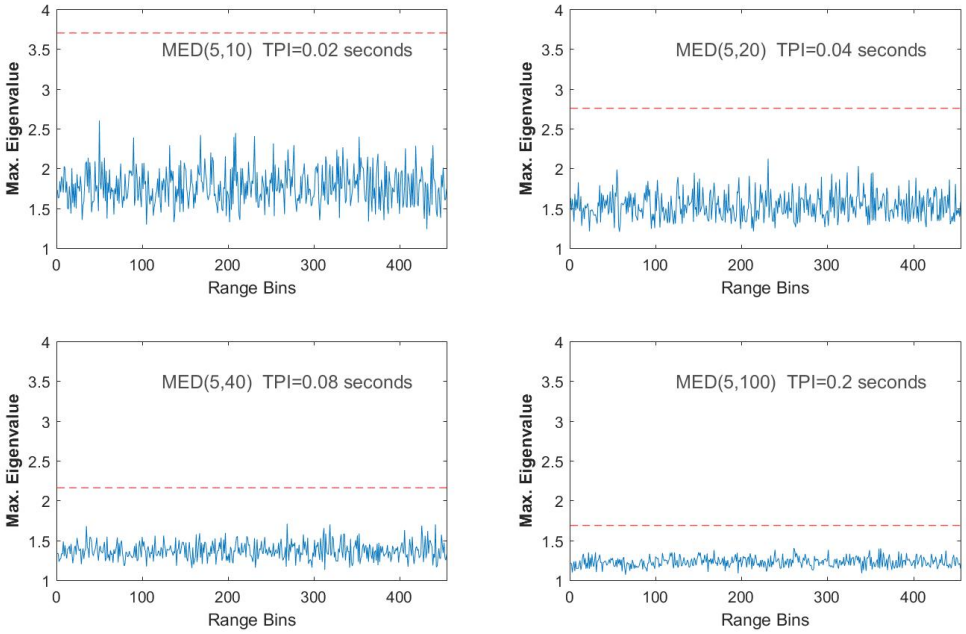


Figure 4.9. MED test results of the second data set with matrix $M = 5$, $L = \{10, 20, 40, 100\}$ and $P_{FA} = 0.01$.

When the segment size is increased to $M = 10$, the situation is the same and the vibration is not detected for the identical number of observation choices as depicted in Figure 4.10. The TPI is further increased by $M = 15$ and the vibration becomes detectable when the number of consecutive segments are 40 or above, as shown in Figure 4.11.

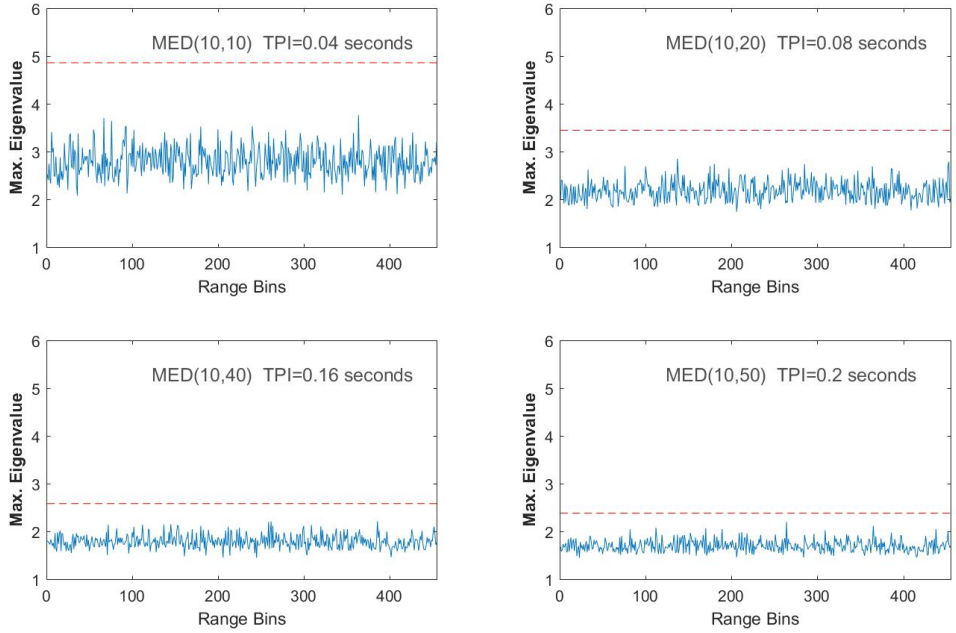


Figure 4.10. MED(M, L) processing results of the 2. data set with $M = 10$, $L = \{10, 20, 40, 100\}$ and $P_{FA} = 0.01$.

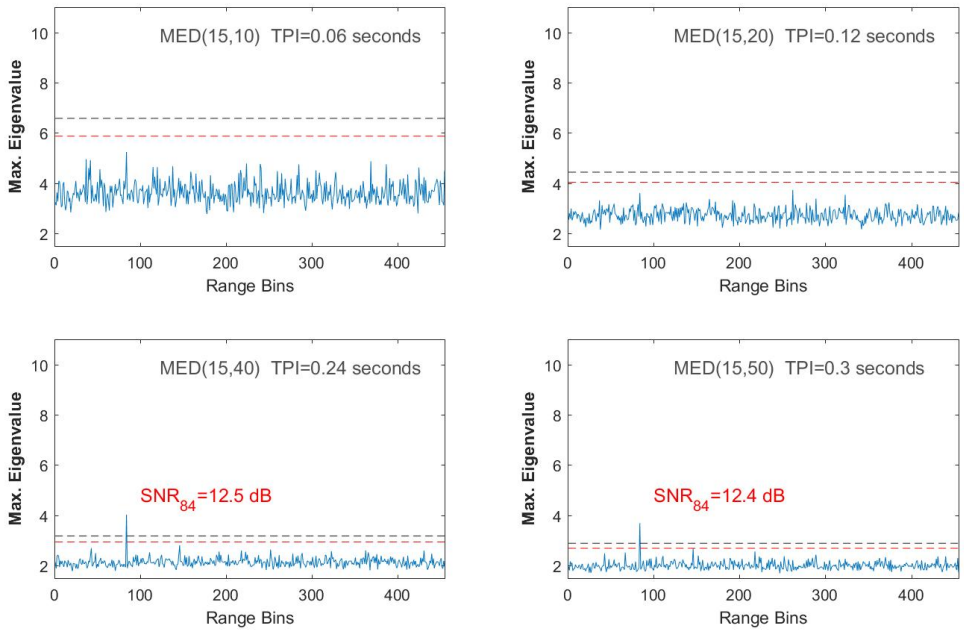


Figure 4.11. MED(M, L) processing results of the 2. data set with $M = 15$, $L = \{10, 20, 40, 100\}$ and two P_{FA} vaues:$P_{FA} = 0.01$ and $P_{FA} = 0.001$.

The MED tests were repeated with square data matrix selections for $M = \{10, 20, 30, 40\}$ and the results are presented in Figure 4.12. As expected, the fast-time SNR values were lower than the SNR values obtained with the 1. data set. The result of having a higher fast-time SNR with MED(30,30) than MED(40,40) processing repeated again in this data, as shown in Figure 4.12.

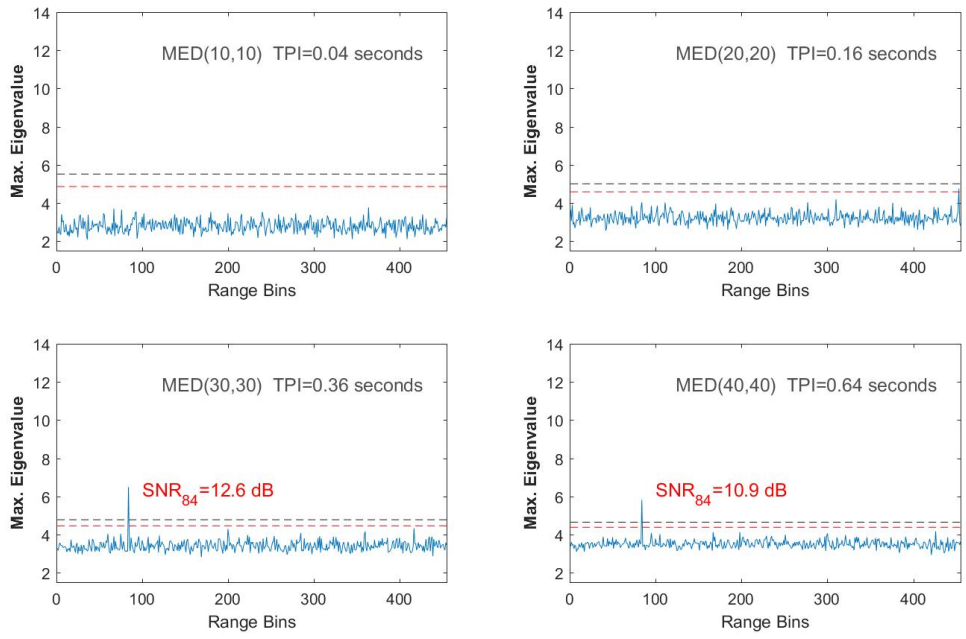


Figure 4.12. MED test results of the second data set for 4 different square sized data matrix ($M = L$) selections. $L = \{10, 20, 30, 40\}$ with $P_{FA} = 0.01$.

4.6.3. Multiple Vibrations

To avoid the redundancy, multiple vibrations with the identical frequency were not analyzed in this section. The case where two vibrations at two different range bins with different vibration frequencies was subjected to testing with the MED approach. The results are shown in Figure 4.13 below for six different processing durations with for $M = 15$ and $P_{FA} = 0.01$. Only the vibration at the 84. range bin is detected while the vibration close to the fiber end is missed. False alarms were observed at the 252. range bin for MED(15,40), at the 154. range bin for MED(15,100) and MED(15,200).

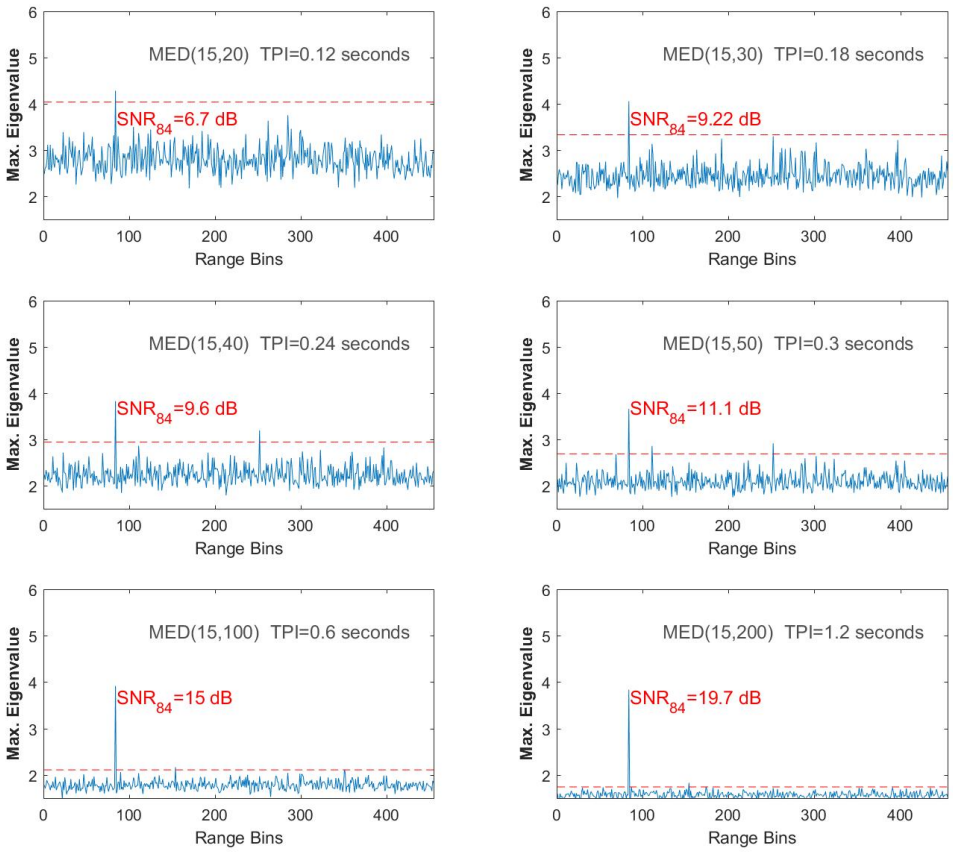


Figure 4.13. MED test results of the 4. data set with $M = 15$, $L = \{20,30,40,50,100,200\}$ and $P_{FA} = 0.01$.

When the segment size was increased to $M = 20$ but the number of observations was kept the same as before, interesting results was observed as neither of the vibrations were detected for MED(20,20) and MED(20,30) selections. When the number of consecutive observations is increased to $L = 30$, only the further vibration was revealed and the 1. vibration is still missed as shown in the second row of Figure 4.14. Both vibrations become discernable when the TPI is increased with MED(20,50) and MED(20,100). The fast-time SNR of both vibrations exceed 15 dB when the TPI duration is over 1 second.

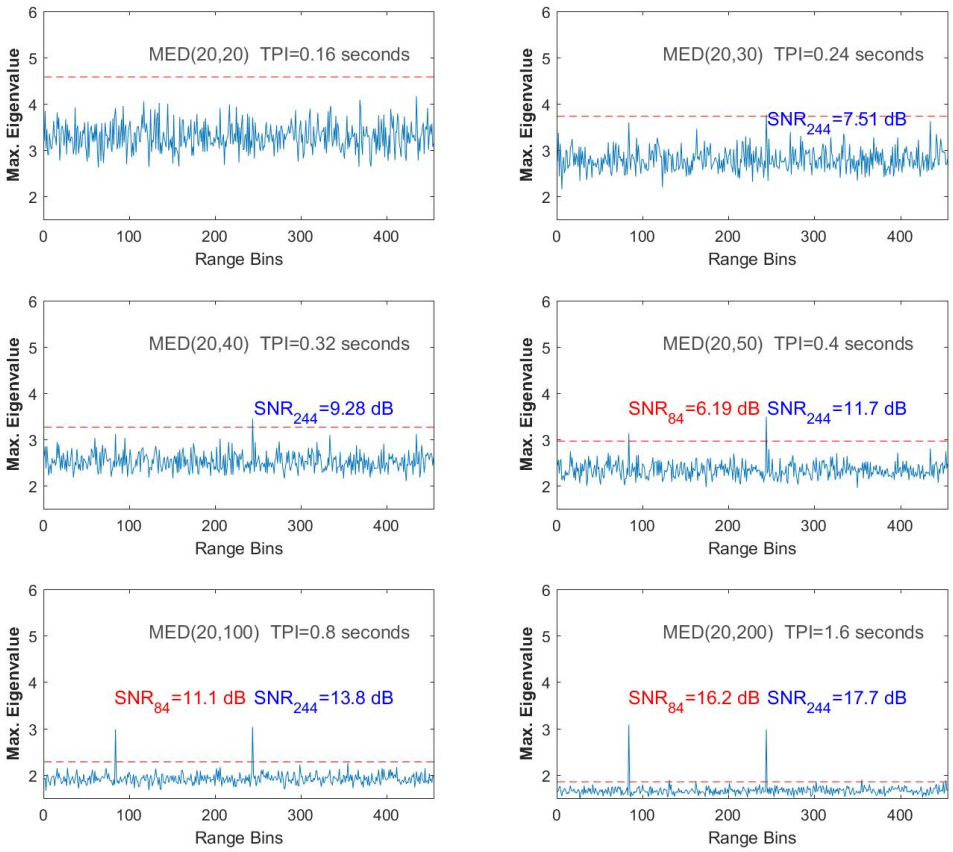


Figure 4.14. MED(M,L) processing results of the 4. data set with $M = 20$, $L = \{20, 30, 40, 50, 100, 200\}$ and $P_{FA} = 0.01$.

When the segmentation is modified by increasing the segment size to 25, the vibration at the 84. range bin is much more emphasized and the 2. vibration of 625 Hz at the 244. range bin became undiscernible unless $L = 200$ as can be seen clearly in Figure 4.15. This is quite interesting, because previous attempts with $M = 20$ and with shorter processing durations such as MED(20,40) and MED(20,50) successfully detected the 2. vibration while longer processing times with $M = 25$ did not reveal the 2. vibration; it even did not yield any peaks close to the threshold as can be seen in the 2. and 3. rows of Figure 4.15.

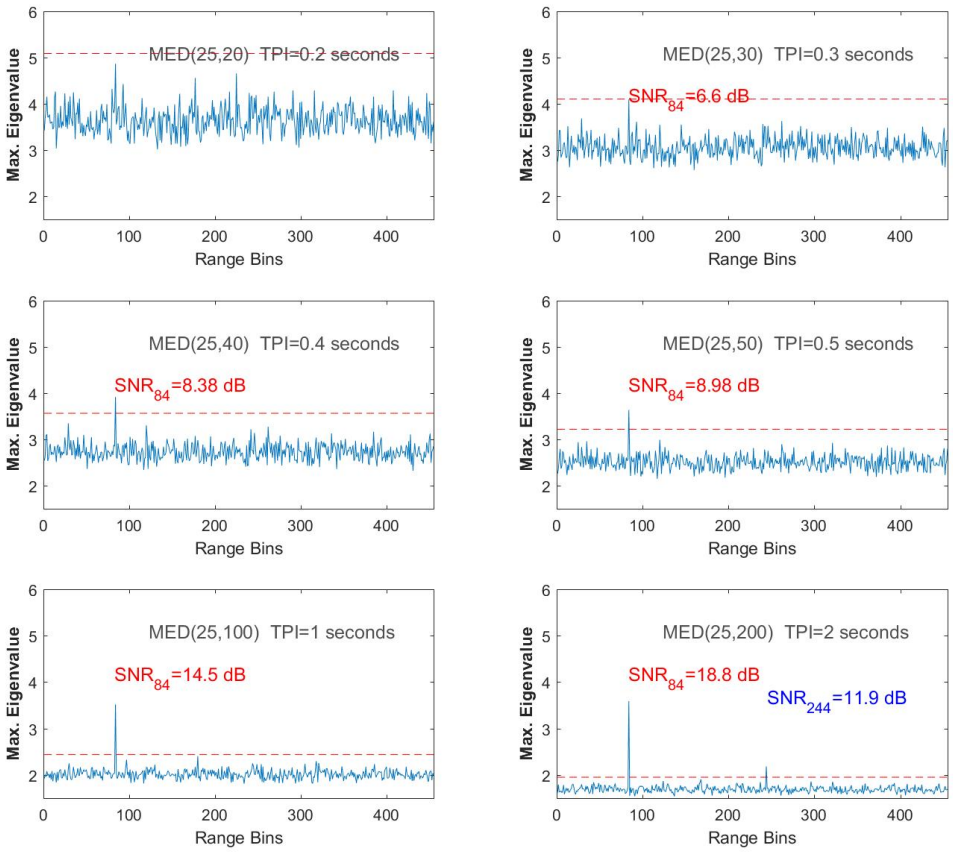


Figure 4.15. MED test results of the 4. data set with $M = 25$, $L = \{20, 30, 40, 50, 100, 200\}$ and $P_{FA} = 0.01$.

4.7. Empiricial Eigenvalue Distributions

Besides the binary detection problem analysis which is handled by the Tracy-Widom distributed largest eigenvalue, the bulk statistics of the eigenvalue distributions for different SCM structures were also analyzed to evaluate the MP Law with simulated ϕ-OTDR data. Some example results of this analysis are given for different data matrix sizes below. The eigenvalue distribution of a SCM estimated from noise-only range bins of the 1. data set is shown for two different segmentations in the upper row of Figure 4.16.

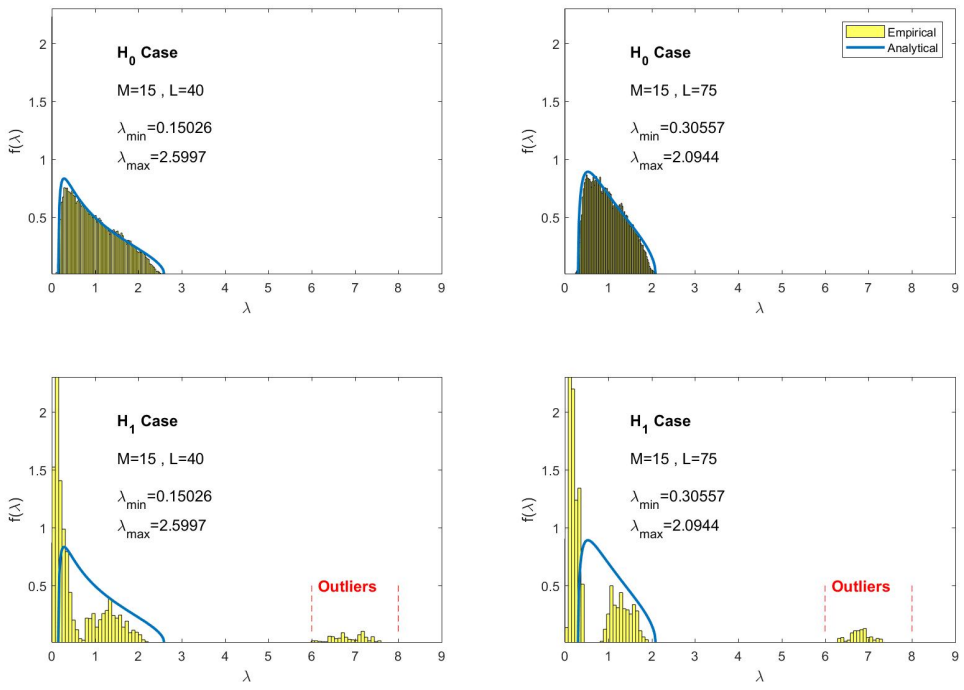


Figure 4.16. The MP bounds and the empirical eigenvalue distributions of the 1. data set computed from 50 consecutively TPI blocks at the 80. range bin (H_0 case) and the 84. range bin (H_1 case).

When the SCM is computed for the 84. range bin which includes a vibration, eigenvalues higher than the analytical maximum were observed as depicted in the lower row of Figure 4.16. The SCM computed at the 84. range bin is not a pure random matrix anymore, because it is contaminated by a vibration signal and this coherency in the SCM exhibits itself as an outlier in the eigenvalue distribution. The locations of the outliers are sufficiently far away from the upper bound due to the high SNR of the 1. data set. The outliers are expected to be observed closer the noise eigenvalues as the SNR is decreased which is verified simply by evaluating the MP Law for the 2. data set. As shown in the lower row of Figure 4.17, the outliers are located closer to the λ_{max} bound, which verifies the relative weakness of the vibration in the 2. data set.

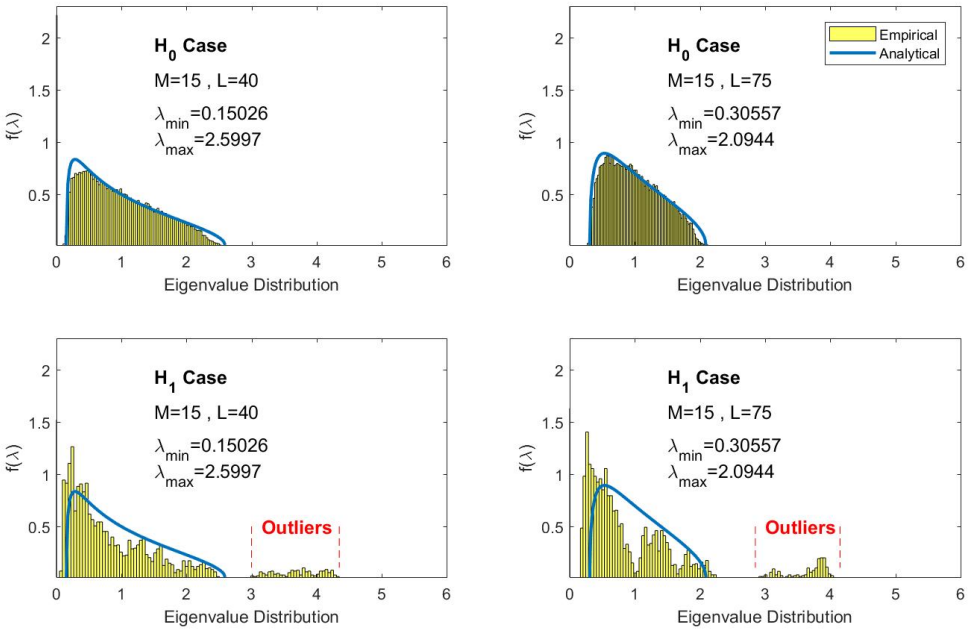


Figure 4.17. The MP bounds and the empirical eigenvalue distributions of the 2. dataset computed from 50 consecutively TPI blocks at the 80. range bin (H_0 case) and the 84. range bin (H_1 case).

The impact of the data segmentation for SCM estimation and the binary detection problem was presented in the previous section by evaluating the TW threshold for various (M,L) pairs. One example was given in Figure 4.14 that the MED(20,20) and MED(20,25) processing selections did not successfully detect the vibration of the 4. data set. When we evaluate the MP Law for MED(20,20) and MED(20,25) for this much weaker vibration, although very close the boundary, the outliers are predictable as shown in Figure 4.18. The separation of the noise and the vibration signal at the 244. range bin become more distinguishable when the TPI is increased similarly as done before. The empirical eigenvalue distribution compared with the analytical bounds for this case is shown in Figure 4.18.

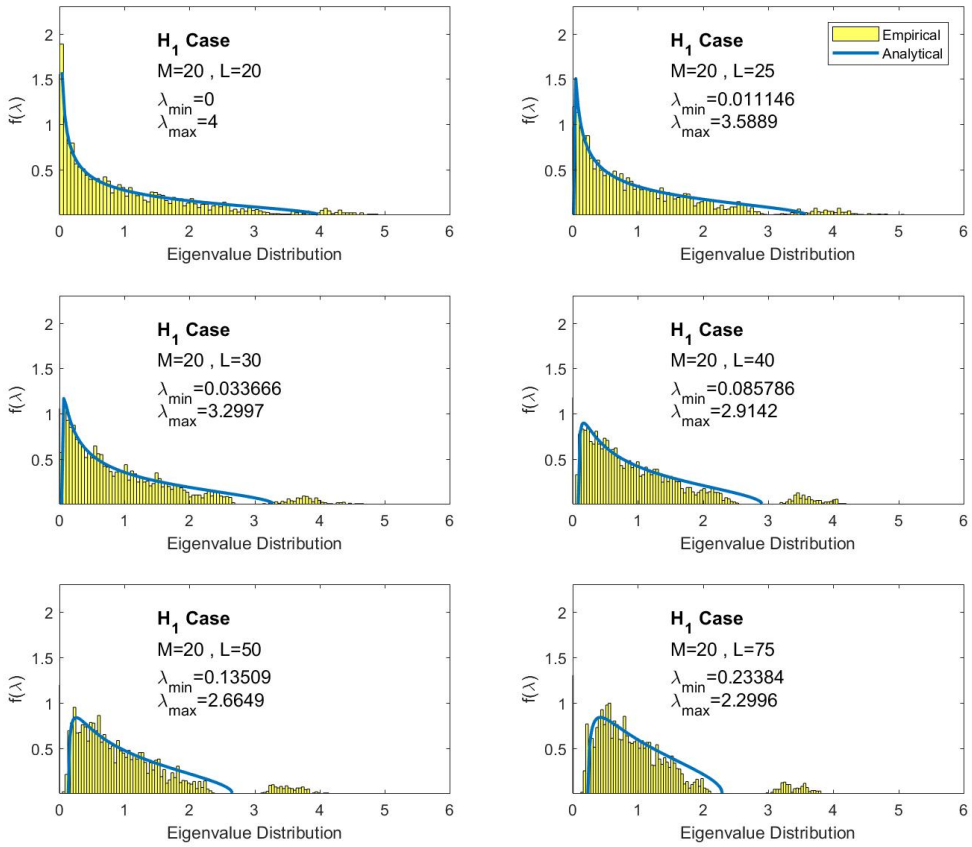


Figure 4.18. The outliers of the empirical eigenvalue distribution at the vibration range bin for the 4th dataset: 200 consecutive TPI's were processed with six different data matrix sizes with $M = 20$ and $L = \{20, 25, 30, 40, 50, 75\}$.

The fact that the MED(20,20) and MED(20,25) processing did not detect the vibration, but the bulk distribution presentations exhibited outliers with the same segmentation might be considered as suprising at first look. Actually, all the results presented for the binary detection problem in the previous section were the outputs of MED(M, L) selections for a single TPI duration. But, the MP Law was evaluated with several hundreds of TPIs selected consecutively; hence the distribution was evaluated by considering huge amount of data matrices. If the MED(M, L) processing with a predetermined threshold were processed consecutively or in a block-by-block basis, and the outcomes are coherently added for every TPI processed, then the detection

performance of the MED(M, L) would be improved as more TPI's are processed. Or vice versa, if the MP Law was evaluated for a single TPI and the distribution is plotted, for each TPI run, the outlier is expected to be visible in a random manner depending on the strength of the weak vibration.

4.8. Conclusions

It was experimentally verified by Monte Carlo simulated ϕ-OTDR data that the largest eigenvalue of the SCM can be used as a test statistic for detecting vibrations. Additionally, the recent results of RMT were also evaluated for DAS applications for the first time. The TW Law was visited to derive the necessary threshold for binary detection. The results of the MP Law were verified with experiments and it was shown that the bulk statistics of the optical SCM are well described by the MP Law; any outlier eigenvalue can be considered as a sign of coherency in the measured data.

When compared with the AMF approach, the main advantage of the MED technique is that it is capable of detecting the vibrations without any prior information about their frequency content. As it was the same with the AMF method, the performance is dependent on the observation (or the measurement) time, but it seems from the experiments presented in this chapter that there is no definite rule for a segmentation strategy.

5. LABORATORY EVALUATIONS WITH REAL ϕ-OTDR DATA

5.1. Introduction

Up to now, we proceeded with computer simulated ϕ-OTDR data to evaluate the performance of the proposed signal processing methods for DAS/DVS applications. In this chapter, we focus on the application of AMF and MED based techniques with real data gathered in the laboratory environment. For this purpose, a real ϕ-OTDR system developed is described in the first step. In the second step, the test results of AMF and MED techniques are given.

5.2. Real ϕ-OTDR System Development

Our experimental laboratory setup is shown in Figure 5.1 which is mainly composed of two sections: the DAS system and the fiber under test (FUT). The FUT is cascaded of two FO cable spools which are 7.3 km and 21.48 km long. Since we observed that the fiber optic cables were highly sensitive to acoustic noise in the environment, the FO spools were put in an enclosure with specially lined acoustic absorbers to relatively reduce to the impact of the environment. We extracted the cascading point of FO spools outside the box to induce an acoustic vibration on the 7300 m location. For this purpose, a PZT based linear phase-shifter is used to simulate the vibration and the frequency of the phase shifts is adjusted by a function generator. The output of the function generator is amplified which is necessary to drive the phase shifter due to the given specifications by the vendor. The order of the two spools can be interchanged to locate the perturbation closer to the far end of the FUT.

The experimental DAS system - which is called as FOTAS - consists of two main subsystems which are 19 inch rack mountable instruments. The first subsystem is called the Optical Interrogator Device (OID) which is based on CD-OTDR architecture

and the 10 percent of the laser power is used as OLO in the interrogator. It incorporates the all optical and electro-optical components of the DAS system as shown by the dashed lines in the upper left part of Figure 2.1 of Section 2.1.

The second subsystem is a Software Defined Radio (SDR) based receiver which is used to capture and analyze the ϕ-OTDR signals with DSP capabilities. The OID emits optical pulses and receives the backscattered optical power in a cooperative manner with the SDR receiver. The output of the OID is an IF stage of 100 MHz which is the modulation frequency of the AOM. The SDR device is also responsible for modulation waveform parameters that are managed by the user. It incorporates a radio frequency (RF) front end tunable up to 2.7 GHz. The IF & baseband processing section includes high-speed and high-resolution digitization and waveform synthesis blocks and a Field Programmable Gate Array (FPGA) with onboard memory.

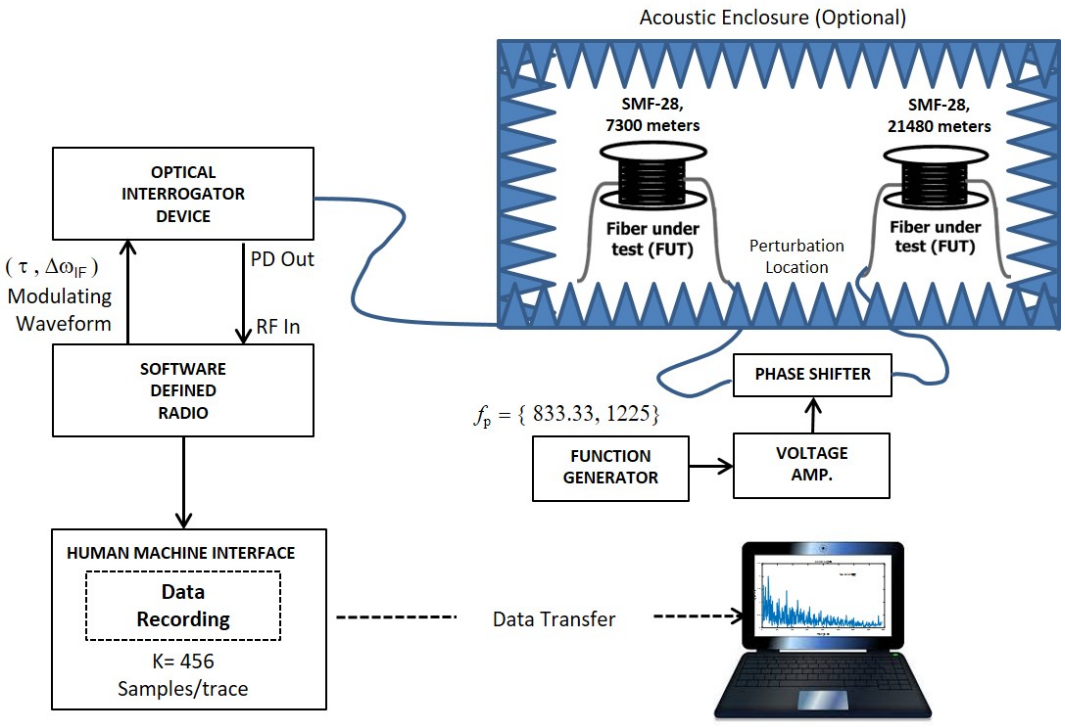


Figure 5.1. The general indoor setup for gathering real ϕ-OTDR data.

The SDR serves as the system control unit of the whole system which is operated via the Graphical User Interface (GUI) of the SDR receiver. After the electrical signal is routed to the super-heterodyning stage of the SDR, it is then digitized by a 2-channel 14-bit 250 MSPS A/D and processed by the FPGA block. The baseband I/Q signal is digitally down-converted and decimated for baseband conversion. The raw data is then transferred to another personal computer to be processed in MATLAB. In this research study, the SDR device is only used to capture real-time data and store it for subsequent processing. After decimation, the final number of I/Q data samples recorded by the system for the experiments can be adjusted by the system operator.

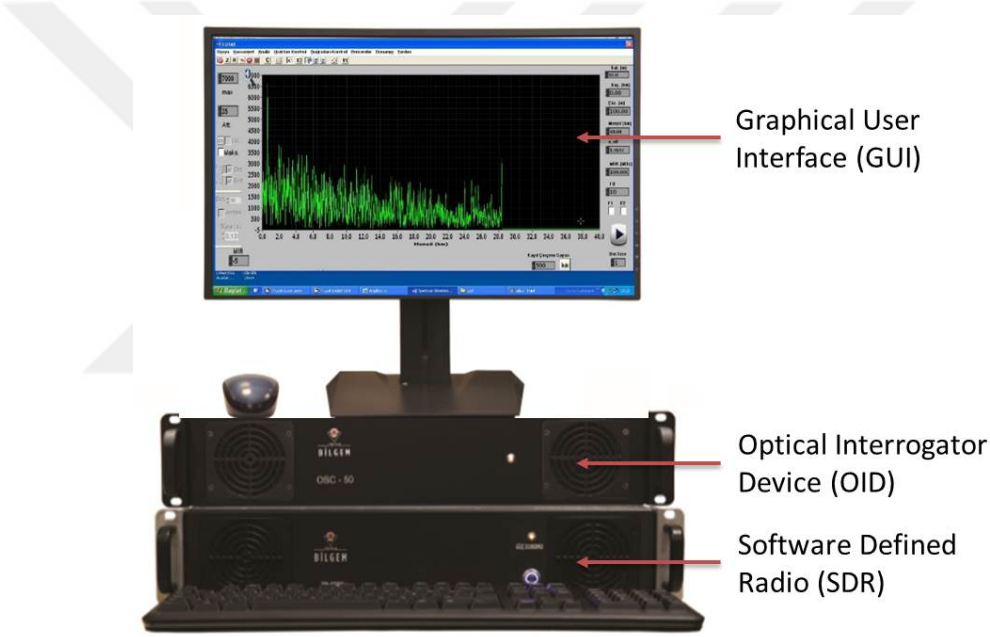


Figure 5.2. The FOTAS system: the 1st prototype developed for real data gathering.

5.3. Indoor Experiments

The FOTAS system parameters were set to fixed values before recording the data for different vibration frequencies. The technical specifications and the general system parameters used during experiments are summarized in Table 5.1.

Table 5.1. General test parameters used during indoor experiments.

System Component	**Smbol**	**Value**
Laser Wavelength	λ_t	1559 nm
Laser Source Power	P_t	22 mW
Laser Frequency Linewidth	Δ_f	$\leq$ 3 kHz
OLO Power	P_{LO}	2.2 mW
AOM Frequency	$\omega_{IF}/2\pi$	100 MHz
Type of FUT	-	Corning® SMF-28®
FO refractive Index	n_{eff}	1.467
Interrogation Frequency	PRF	2500 Hz
Pulse Width	τ	860 ns
Spatial Resolution	SR	Variable between 70 - 90 meters
Total Optical Insertion Loss	IL	6 dB
Length of the FUT	L_{FUT}	28780 m
Number of Range Bins/Trace	K	Variable between 228 and 4560
Vibration Location(s)	l_1, l_2	7300 meters (1st) , 21480 meters(2nd)
Range Bins of Vibration	p_1, p_2	84 or 104 (1st) and 307 (2nd)

Before the experiments, the connection between the detectable MR and the maximum SR definitions were experimentally verified by changing the system parameters such as the laser power, the pulse interrogation frequency, and the pulse width. Additionally, although a booster EDFA was not used at the transmitter stage, the peak optical power launched into the FUT was measured by an optical spectrum analyzer to verify that the system is free of the SBS impact which may have impairments on the received signal [31]. The display range of the GUI was taken as 40 km to observe the total Rayleigh scattering profile easily along the 28.78 km long FUT. A typical ϕ-OTDR trace gathered from the system is shown in Figure 5.3. The exponential decaying and speckle-like feature of the RBS up to the end of the FUT are obvious. As mentioned previously in Section 3.1, the total background noise is composed of the RBS related part plus the receiver noise term till the end of the FUT which is 28.78 km. The data record in the flat region of the plot corresponds only to the receiver noise terms.

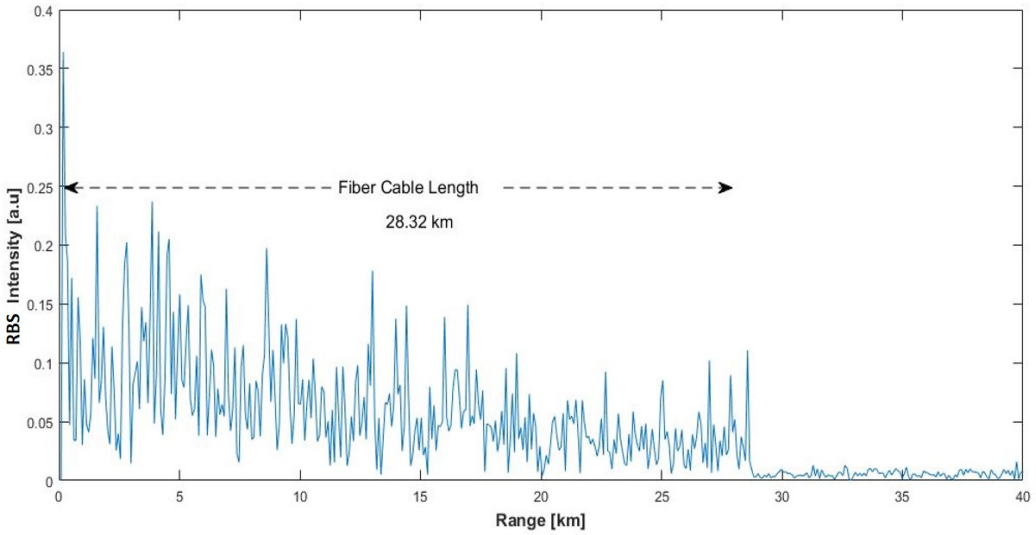


Figure 5.3. Typical ϕ-OTDR trace recorded indoor by the FOTAS system.

Table 5.2. Indoor data used during testing.

Data Index	**Frames Recorded**	**Duration [Seconds]**	**Vibration Information Frequency @Range Bin**	**Acoustic Isolation**	**SNR [dB]**
1	17,043	6,82	833.33Hz @84	Yes	30
2	13,753	5,44	1225Hz @84	Yes	30
3	25,588	10,23	833.33Hz @84	No	15
4	21,499	8,6	833.33Hz @84	No	0
5	13,500	5,4	1225Hz @104 /841Hz @307	Yes	25

The AMF and MED approaches introduced in the previous chapters were tested for several different vibration frequencies. The general parameters of the indoor datasets used in this thesis are summarized in Table 5.2. The first frequency selected is 833.33 Hz, and it is the same frequency used in our previous simulation study for comparison. The second vibration frequency is 1225 Hz, and it is selected close to the maximum detectable frequency, which is half of the PRF, i.e., 1250 Hz. These vibration frequencies are controlled by the function generator and is equal to the frequency of the sinusoidal waveform used to modulate the phase shifter. The first two data were recorded with the acoustic absorbers. The 3. and the 4. data sets were recorded with the acoustic absorbers removed to see the impact of the indoor environmental noise.

The last data set is the coherent sum of two different data records to analyze the multiple vibration case. For this data record, a second vibration was induced closer to the far end, at the 21.5 km of the FUT by interchanging the position of the FO spools and the ϕ-OTDR traces record were coherently added to the record of the 1. vibration at the 7.3 km. The total number of range bins of the 5. data set was modified as 570 and the SR was approximately 70 m. After raw ϕ-OTDR data are recorded by the SDR, they are subjected to off-line adaptive processing (i.e., MATLAB routines) with different processing times and the results are summarized below.

It is worth emphasizing that, before processing with either the AMF or the MED approach, the signal bearing time segments of all recordings were rigorously analyzed by observing the spectrograms of the vibration range bins. This was due to the SOP effects of the CD-OTDR based system and the vibration signal was sometimes observed to diminish. Because of these SOP effects, the TPI selections were carefully selected to assure vibration signal presence in the time interval of interest; and the same TPI's were used for both AMF and MED tests.

5.3.1. AMF Test results

The AMF algorithm was run several time with varying FIR filter sizes as done before in Section 3.1 for the Monte Carlo simulated data set. The first three range bins on the FUT axis were selected as the training region, i.e., $\#\{L_s\} = 3$. The filtered ϕ-OTDR traces were obtained with the AMF method for every range bin by setting the tuning frequency of the steering vector equal to the vibration frequency of
833.33 Hz. As mentioned earlier in this text, the performance criterion is to observe the SNR at the output of the filter along the FUT axis. Namely, remembering the equation (3.1), the fast-time SNR value for the 84. channel after AMF processing is calculated by the following equation:

$$SNR_{(k=84)} = 10log10\left(\frac{|z_{k=84}|^2}{(\#\{L_s\})^{-1}\sum_{\forall k\in L_s, k\neq 84}|z_k|^2}\right) \tag{5.1}$$

In the first four data recordings, L_{FUT} was computed as 327 according to the total length of the FUT and the spatial resolution, which were 28780 m and 88 m, respectively (i.e., $\lfloor 28,780/88 \rfloor = 327$, which is the number of effective channels worth monitoring). The DL level was fixed with $\rho = 10$ (10 dB). The computed results of (5.1) for various FIR sizes are displayed in Figure 5.4 ranging from the top to the bottom and left to the right for increasing N, which corresponds to increasing TPI.

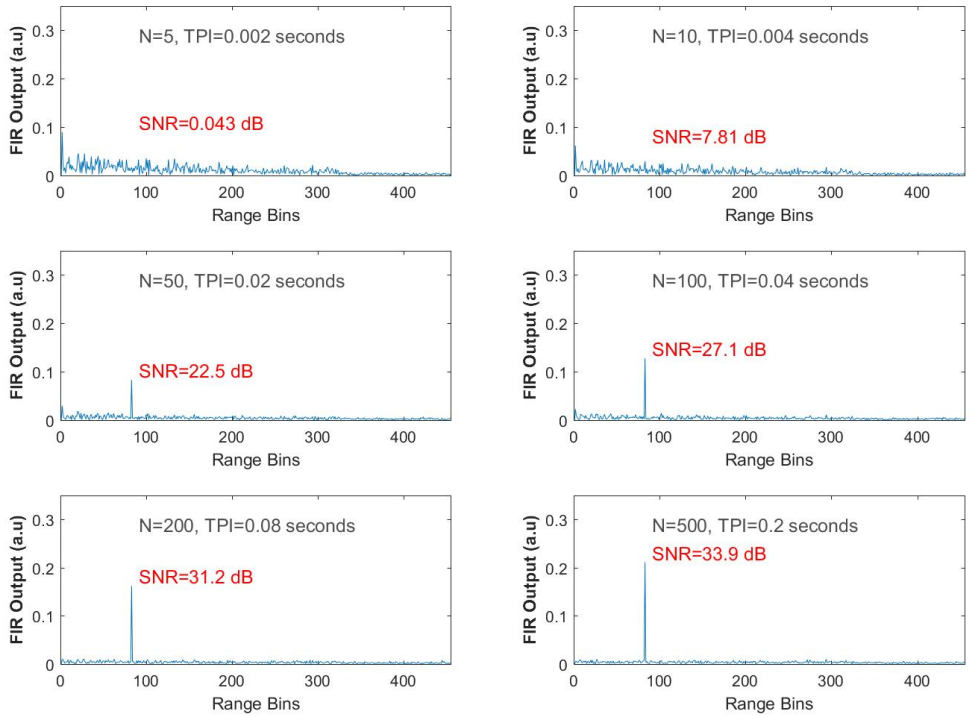


Figure 5.4. AMF test results for f_{p_1}= 833.33 Hz.

When the size of the input vector of the FIR filter is equal to five (i.e., $N = 5$ and TPI = 2 ms) there is significant speckle reduction in the ϕ-OTDR trace, but no peaks are observed along the range axis, while the calculated SNR is 0.043 dB for the vibration point. Although the SNR is calculated as 7.81 dB when is N increased two-fold, no peaks are observed. When $N = 50$ and the resulting TPI duration is increased to 0.02 seconds, the expected peak at the 84. range bin is revealed with a resulting SNR value of 22.3 dB.

For the last three trials, the size of the filter is selected as 100, 200 and 500, and the observed SNR values are 27.1 dB, 31.2 dB and 33.9 dB, respectively. As can be easily seen from the filter outputs given in Figure 5.4, the peak at the 84. range bin becomes obivous, while the background noise severely supressed with increasing N.

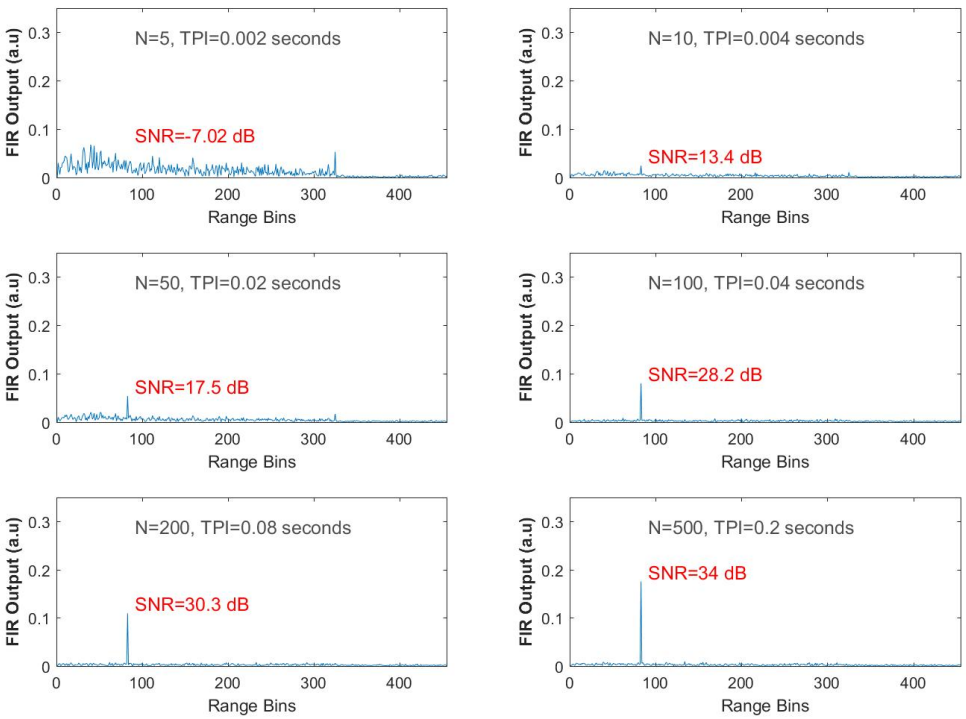


Figure 5.5. AMF test results for f_{p_2}= 1225 Hz.

The same steps were repeated for the second dataset recorded when $f_p = 1225$ Hz and with the test frequency equal to 1225 Hz. For $N = 5$, the SNR is -7.02 dB, and there is not enough reduction in the background noise to reveal the vibration signal. A high reflection is observed at the end of the FO cable. However, when the size of the FIR filter is increased to 10, the peak is observed with an SNR value of 17.4 dB. As N is increased to 100, 200 and 500, the results are similar to the ones obtained for the previous data with the fast-time SNR values of 28.2 dB, 30.3 dB and 34 dB, respectively,as shown Figure 5.5.

For the second dataset, the filter outputs were tested for some arbitrary frequencies that are not equal to the vibration frequency such as 171 Hz, 283 Hz, 497 Hz and 1180 Hz. It can be seen in Figure 5.6 that none of these tuned frequencies yielded spikes at the output of the FIR filter, as expected.

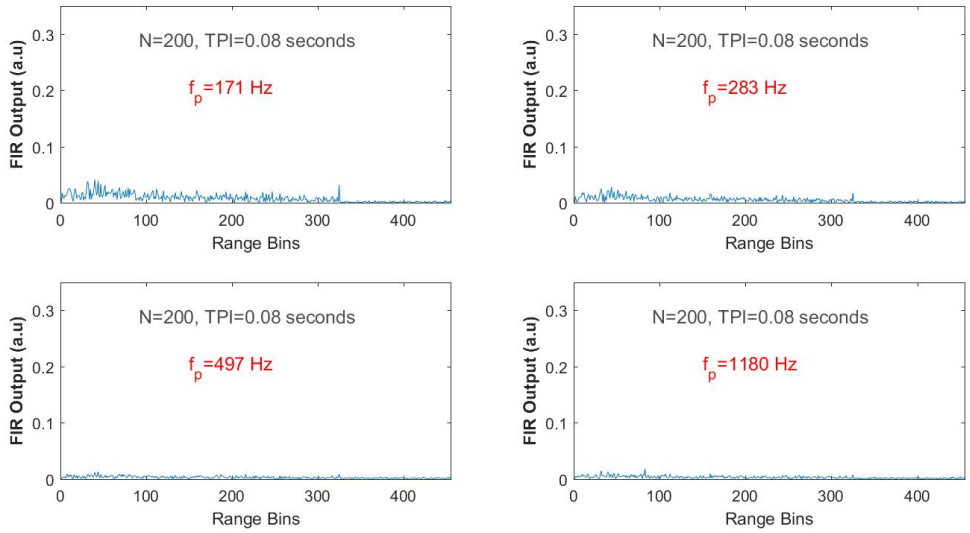


Figure 5.6. Indoor test results with the steering vector tuned to non-vibrating frequencies (i.e., $f \neq f_{p_i}, i = 1, 2$).

The measurement results were also compared with the simulated ϕ-OTDR data. For this task, firstly the simulated data having an almost identical slow-time SNR value was selected, and the power spectrum of the vibrations were compared for verification as can be seen in Figure 5.7 where both simulated and measured data yielded a peak at 0.833 kHz frequency and with the same signal strength. After observing the periodograms, the AMF algorithm was executed for both measured and simulated datasets for increasing N. The SNR versus N plots are given in Figure 5.8. It can be concluded in general that the simulation results are in good agreement with the experimental results. The main difference seems to be the relative asymptotic behavior in the measurement data compared to the simulation results. It is also quite obvious that for relatively such strong vibrations, there is no logical reason to increase the FIR size above 500 since the achieved noise suppression is sufficiently enough and the fast-time SNR reaches a limit.

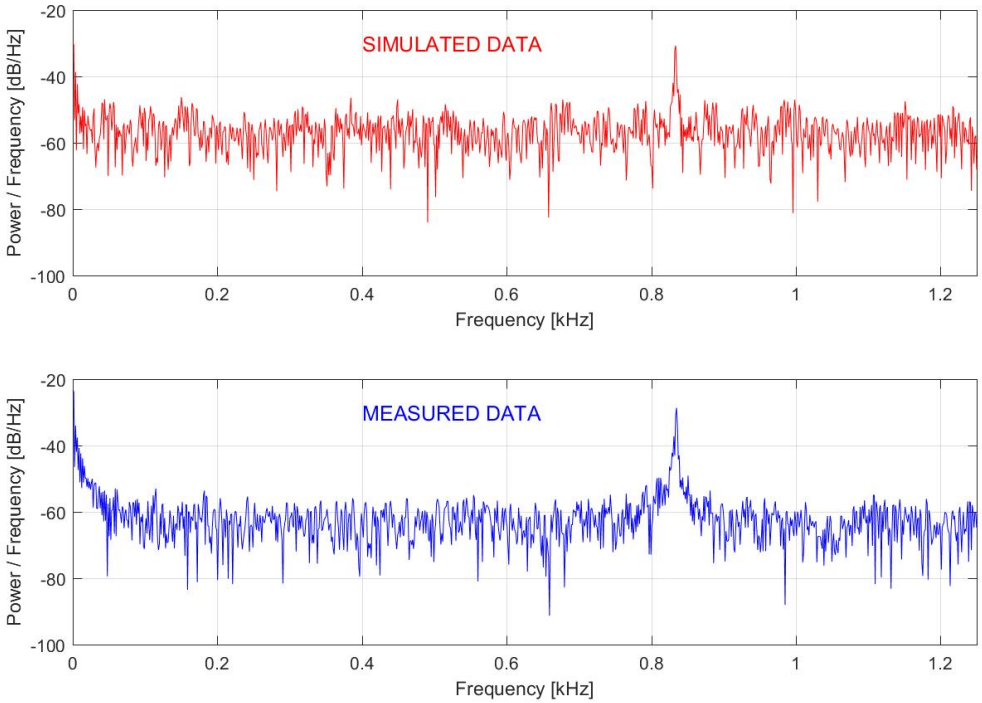


Figure 5.7. The slow-time power spectrum densities of the simulated and measured vibrations at the target range bins with 833.33 Hz vibration.

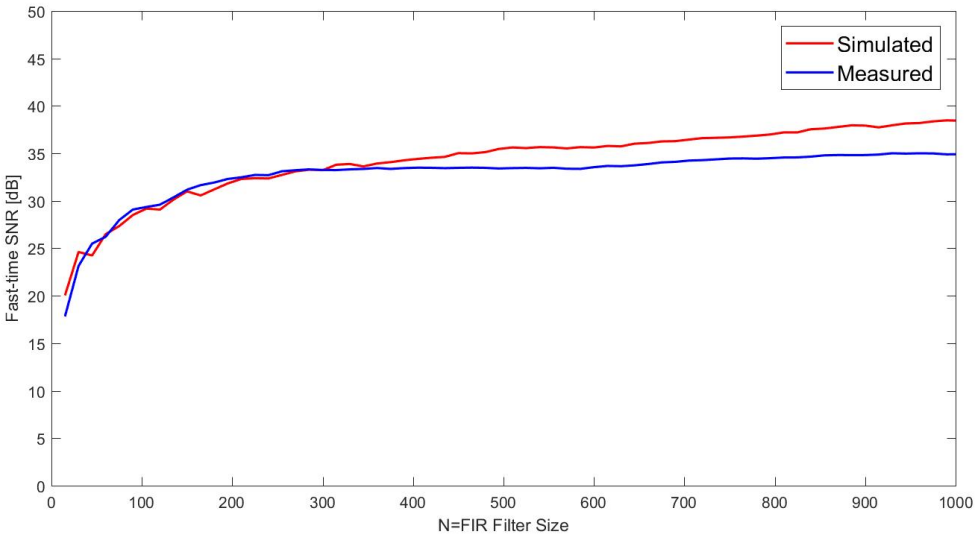


Figure 5.8. SNR vs N plots of the simulated and measured ϕ-OTDR data.

For the second phase of the indoor experiments, the AMF method was tested with the acoustic enclosure removed. Two different vibration datasets were recorded with the 833.3 Hz vibration turned on for approximately 3 s to emulate "strong" and "weak" non-continuous vibrations. For the strong vibration case, the amplitude of the signal generator was adjusted to a level that the spectrum of the signal exhibits a visible peak. For the weak vibration case, the amplitude was decreased to a level where the peak is obscured by the noise spectrum. The spectrogram plots computed in MATLAB are presented in the upper row of Figure 5.9.

The periodograms were computed for only the signal durations to verify the visibility of the signals before applying the AMF. As shown in the lower row of Figure 5.9, the SNR in the strong signal case is about 15 dB over the noise floor. From the graph presented in the lower right part of the figure, it is impossible to decide on the presence of a vibration signal with the power spectrum.

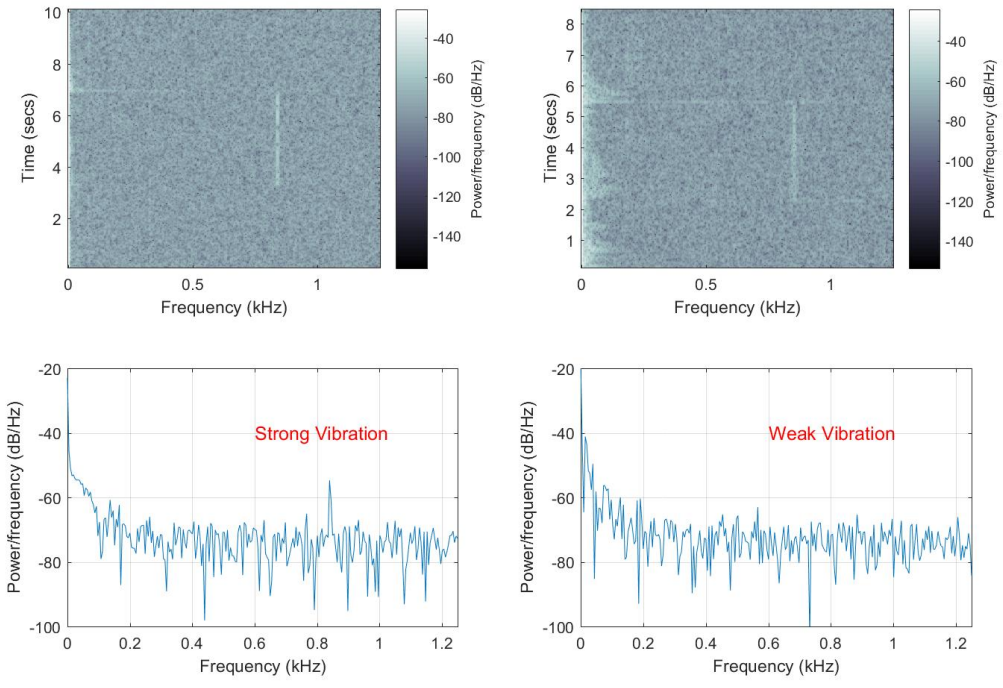


Figure 5.9. Power spectrum density plots for the last two data sets at the target range bin. The upper row is the time-frequency plot for the whole recording while the graphs in the lower designate the periodogram plots.

The fast-time SNR values versus FIR size were computed for both strong and weak signal conditions and shown in Figure 5.10. It can be seen from the figure that the trend of the SNR curve is similar to those obtained for the measured data with acoustic isolation. There is a negative offset in the SNR axis, which can be expected due to the reduced power of vibration signal; or vice versa, the increased impact of the background noise after removal of the enclosure. Thus, it can be thought that the observed results are consistent with the previous results obtained with stronger vibrations. An interesting point observed was the oscillating behavior of the SNR curve for the weak signal test.

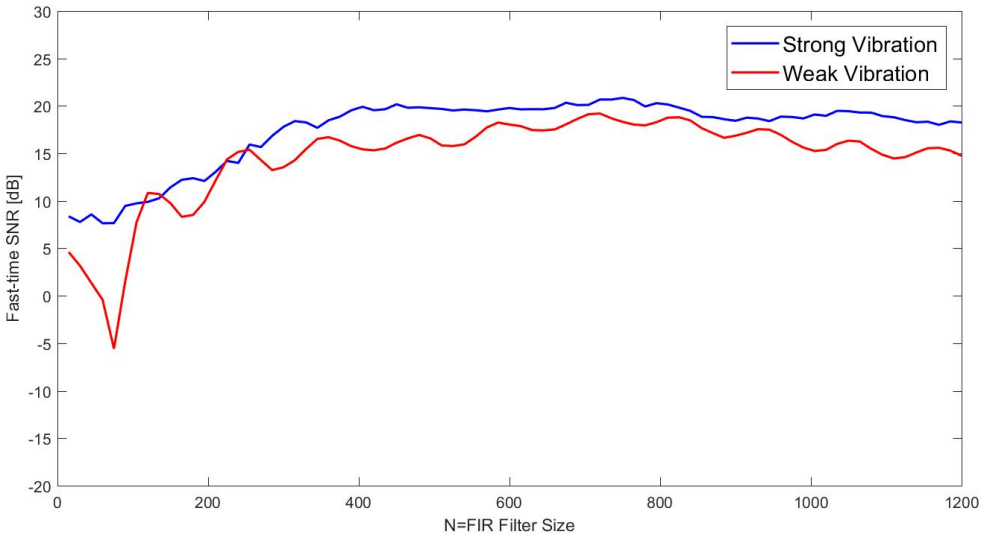


Figure 5.10. SNR vs N curve for the strong and weak vibration ϕ-OTDR data with $f_p = 833.33$ Hz.

Figure 5.11 presents AMF results obtained for the weakest vibration with the same set of N values as previously selected. The peak at the 84. range bin is not observed unless $N = 200$, but it is not very dominant since the next higher peak on the left seems to be another vibration candidate, which may cause a false alarm. Significant noise reduction is obtained, and the vibration is emphasized when $N = 500$.

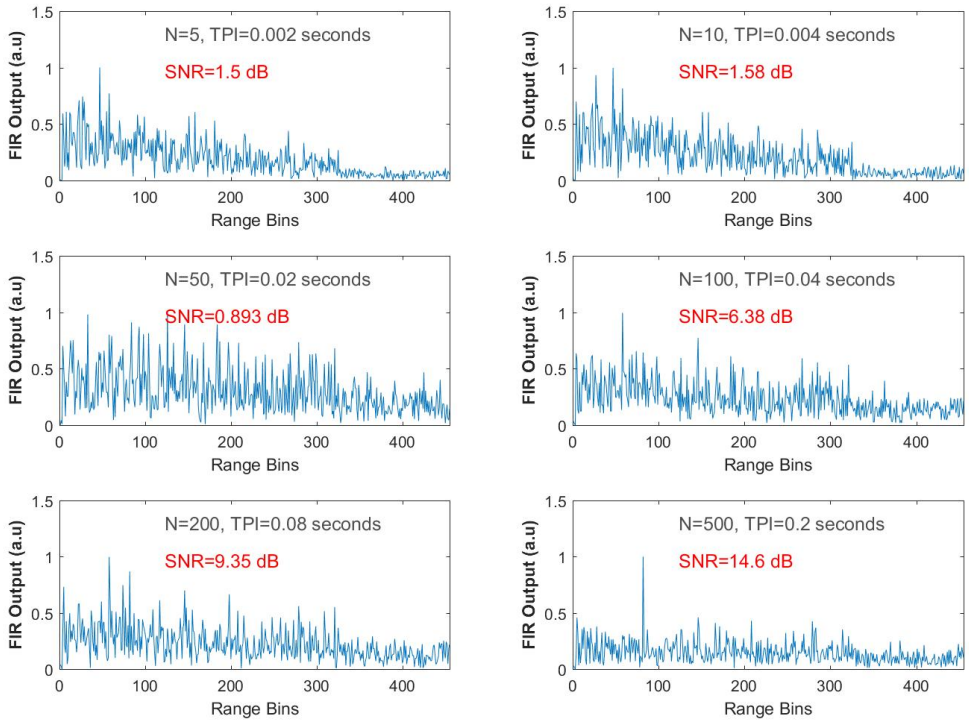


Figure 5.11. AMF test results for the weak vibration of 833.33 Hz without the acoustic isolation.

The last dataset included a 1225 Hz vibration at the 104. range bin and an 841 Hz vibration at the 307. range bin. The data was recorded with the acoustic enclosure. The results of AMF processing for the two vibrations are shown below with the FIR filter outputs normalized. The test results for the 1. vibration with the AMF tuned to 1225 Hz are shown in Figure 5.12. It can be seen that the fast-time SNR reaches a very high value even with N=100. According to the results obtained with the SNR vs N curves previously in Figure 5.8 and 5.10, no further test was required for $N > 500$. The vibration close to the far end of the FUT is also easily detected when the AMF is tuned to 841 Hz as can be seen in Figure 5.13. When compared to the vibration at the 104. cell, this is even detectable with less TPI durations. As it was depicted in Figure 2.3, the vibrations at short distances require more SNR for reliable detection. All the range bins along the FUT length were selected as the secondary data set (i.e.$\#\{L_s\} = 409$).

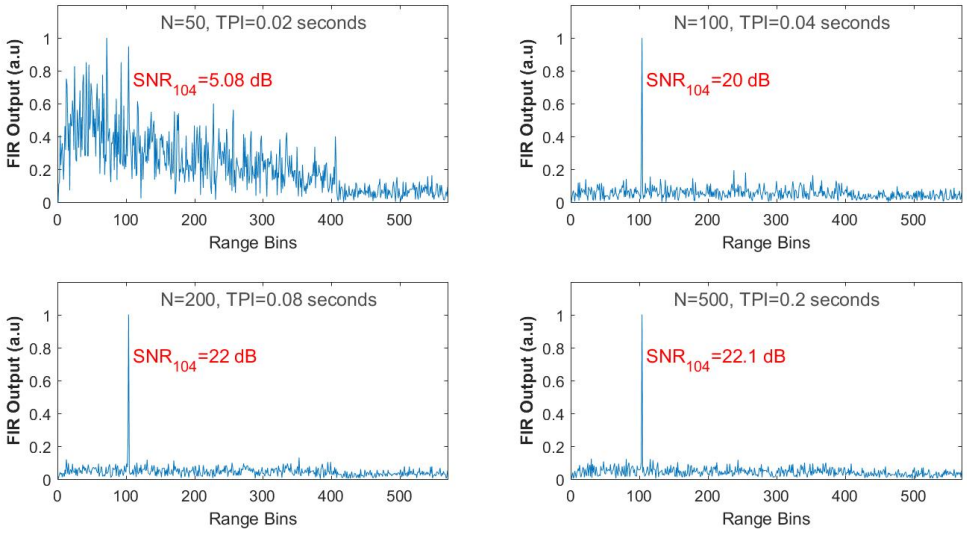


Figure 5.12. Normalized intensity at the AMF output for the vibration of 1225 Hz at the 104. range bin with the acoustic isolation.

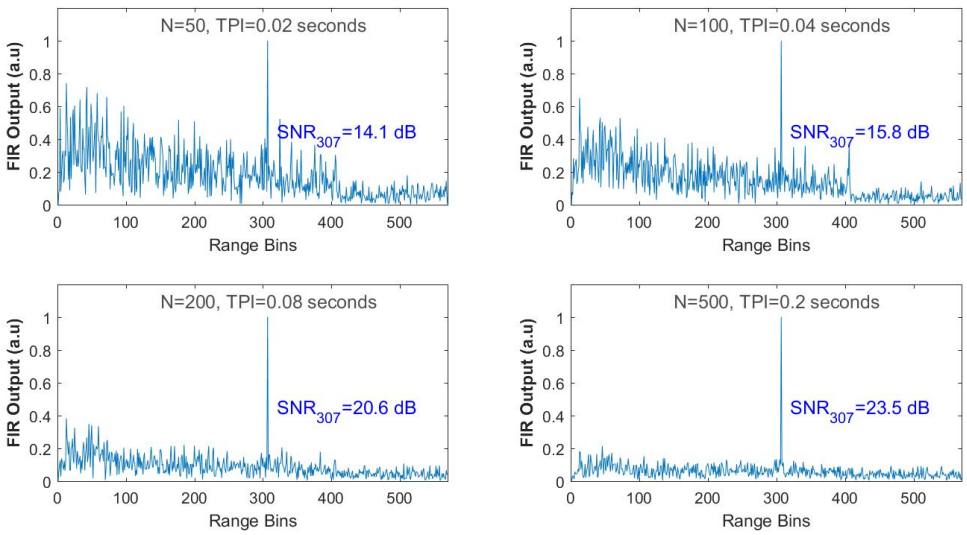


Figure 5.13. Normalized intensity at the AMF output for the vibration of 841 Hz at the 307. range bin with the acoustic isolation.

Shortly, the selectivity of the AMF method is once more emphasized in the results presented in Figure 5.12 and 5.13.

For the last case in AMF testing with the indoor data, the impact of the DL level was tested. Figure 5.14 and Figure 5.15 represent the AMF results for two different DL levels for $N = 200$ and $N = 500$ with the steering vectors tuned to 1225 Hz and 841 Hz, respectively. Although the 1. vibration is revealed for both of the DL selections, the impact of the DL level is obvious where the SNR is reduced for $\rho = 0.01$ as can be seen in the bottom right plot of the Figure 5.14. The fast-time SNR value was computed by considering all the range bins from 10 to 200 excluding the 104. cell as noise-only cells. For the second vibration at the 307. range bin, detection is possible only for $N = 500$ and $\rho = 100$. The vibration is missed when the DL level is reduced to $\rho = 0.01$. The fast-time SNR for this vibration was computed between the range bins 201 and 409 (i.e. the FUT ends at 410. range bin).

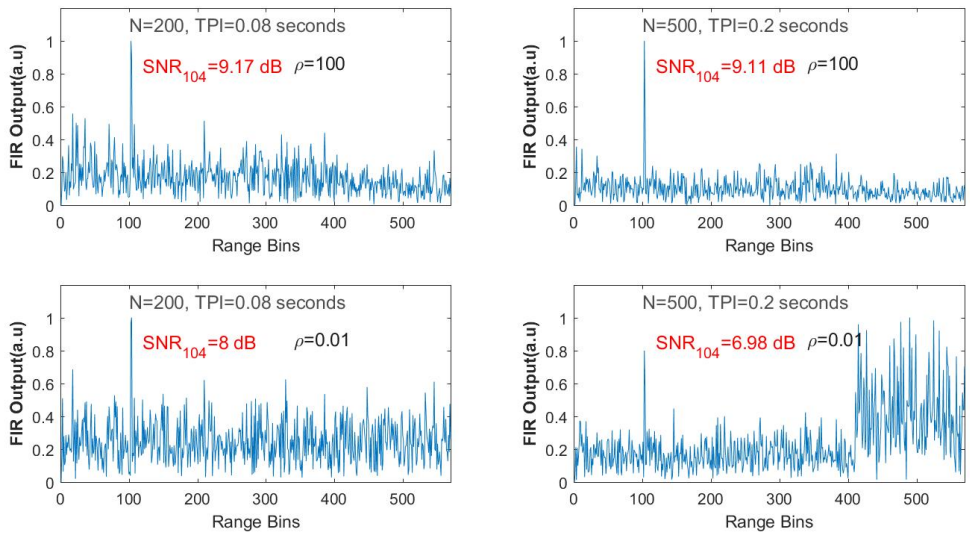


Figure 5.14. Normalized intensity outputs for the weak vibration of 1225 Hz with the acoustic isolation.

The impact of the DL level was additionally tested for two different L_s settings. AMF results for $\rho = 100$ and $\rho = 0.01$ are shown in Figure 5.16 and Figure 5.17 for both vibrations. For all trials, the FIR size was fixed as $N = 500$. For the 1. vibration at the 104. range bin with 1225 Hz, the first secondary data region was selected as $L_s = \{51, 52, ..., 59, 60\}$. As it is shown in the first column of Figure 5.16, the two extreme DL levels did not yield significant difference at the AMF outputs.

In the second test, the range bin with the vibration was included in the training set with $L_s = \{51, 52, ..., 159, 160\}$. In this case, the changes of the DL levels significantly affect the AMF output as can be seen in the right column of Figure 5.16.

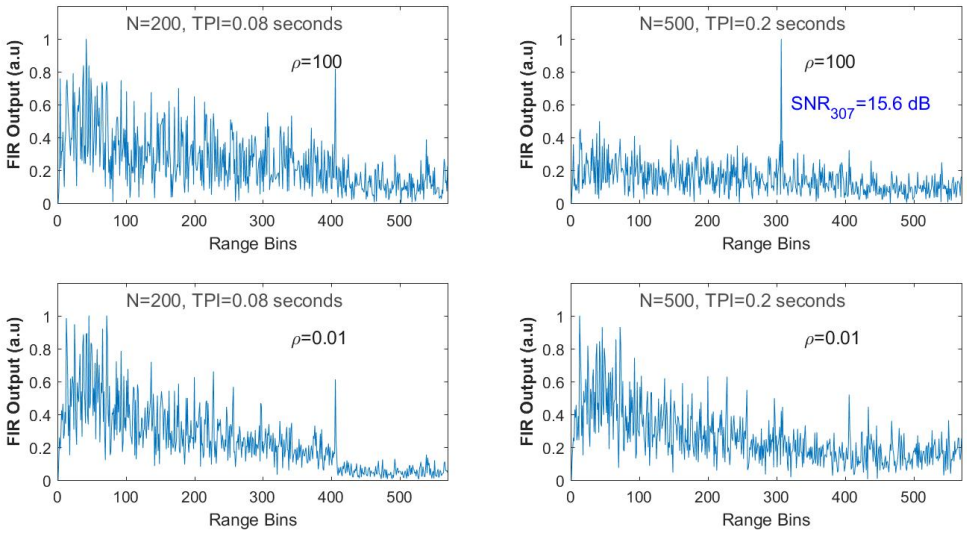


Figure 5.15. Normalized intensity at the AMF output for the vibration of 841 Hz with the acoustic isolation for two DL level selections.

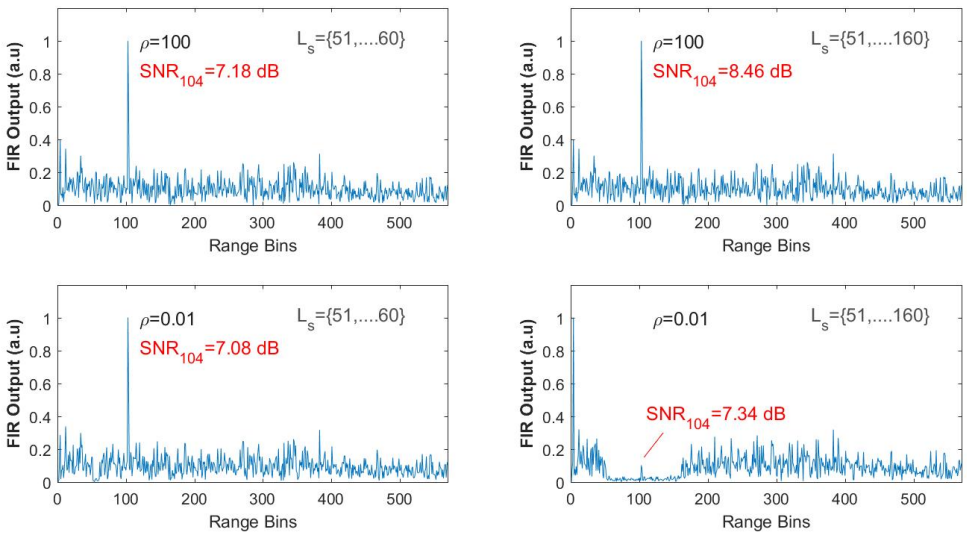


Figure 5.16. Normalized intensity at the AMF output for the vibration of 1225 Hz with the acoustic isolation for two DL levels and secondary data regions.

The same methodology was followed for the further vibration at the 307. range bin by tuning the AMF to 841 Hz. When the target vibration is not included in the secondary data, the AMF output is not affected by the training region size as shown in the left column of Figure 5.17 for the DL levels. When the vibration range bin is included in the noise-only assumed SCM estimation with $L_s = \{251, 252, ..., 359, 360\}$, then the AMF output significantly changes as can be seen in the right column of Figure 5.17. The vibration was not revealed clearly for this dataset when the DL is extremely reduced to -20 dB with respect to unity DL.

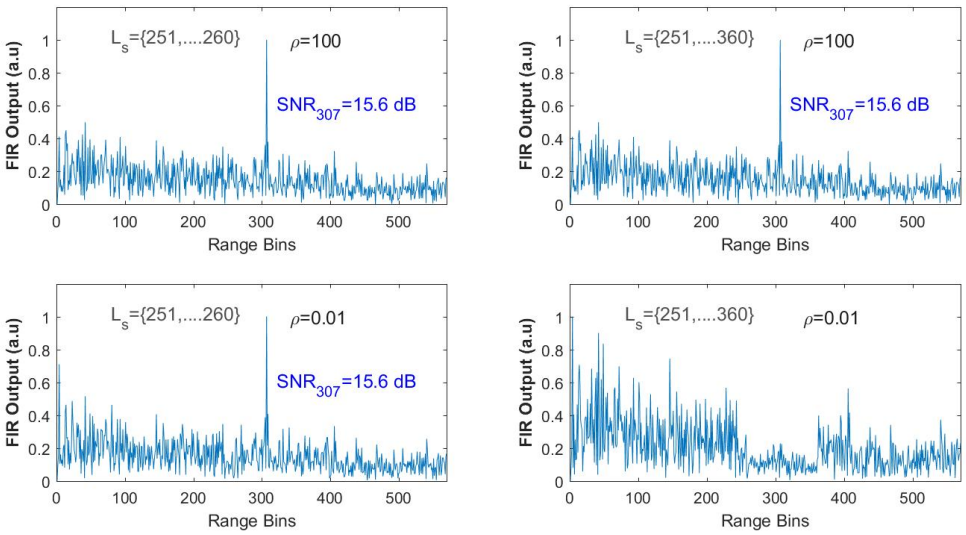


Figure 5.17. Normalized intensity at the AMF output for the vibration of 841 Hz with the acoustic isolation for two DL levels and secondary data regions.

The results show that the DL level selection is critical when there is signal contamination in the secondary data set. As mentioned earlier that there is not a definite rule for the selection of an optimum DL level and a general rule of thumb is followed usually by setting the DL level 5-10 dB above the noise floor in beamforming applications. On the other hand, although the optimum DL selection problem has not been studied for the temporal case, the steering vector errors under signal contamination in the secondary data have been studied for beamforming applications [91].

In [91], the direction of arrival of a weak signal is estimated under strong interference and it was analytically shown that the optimum value is always negative and it can be efficiently approximated by

$$\rho_{opt} \approx -(\sigma_w^2 + P_0 N) \tag{5.2}$$

where N is the size of the SCM (or the weight vector). The temporal frequency in our AMF approach corresponds to the sine of direction of arrival in [91] where a linear equi-spaced beamformer is of concern. Thus, with a heuristic approach and by interchanging the variables between the spatial domain and the temporal domain, the above approximation given by (5.2) can be easily derived by following the steps given in [61]. The details of this derivation can be found therein for interested readers. The equation states that the optimum value depends on the thermal noise level σ_w^2, the signal power P_0 and the selected FIR filter size N. We need to estimate the signal and the noise power to compute ρ_{opt} but a rigorous look at (5.2) reveals that this DL level is nothing but the negative of the largest eigenvalue of the data covariance matrix for a single complex sinusoid embedded in white noise with power σ_w^2, as we studied previously in Section 4.2. Thus, without any effort for estimating the noise and signal powers, an eigen-decomposition of the estimated SCM can be performed and take the negative of the largest eigenvalue as the optimum DL value for SNR maximization in the signal contamination case.

To validate the above statements, several tests with the indoor ϕ-OTDR data were carried out. In order to simulate the strong interference conditions of [91], sample ϕ-OTDR data traces with a relatively low SNR ($\leq$ 10 dB) were recorded additionally. The data set with the perturbation of 1225 Hz located at the 84. range bin was used for the experiments. For the diagonally loaded AMF test with this data, after tuning the steering vector to 1225 Hz, the following steps were followed:

(i) Select the training sample set including the vibration range bin and estimate the SCM ($\hat{\boldsymbol{R}}_k$) by (3.10)

(ii) Compute the weight vector by (3.6) and filter output with varying values of ρ at the target range bin (i.e. $z_{84} = \boldsymbol{w}_{dl}^{H}\boldsymbol{x}_{84}$).

(iii) Observe the peak and corresponding ρ_{opt} values on the results.

(iv) Compute the largest eigenvalue λ_1 of $\hat{\boldsymbol{R}}_k$ for comparison with the ρ_{opt} value.

(v) Repeat the above steps for varying FIR sizes.

According to the results obtained up to now, the speckle suppression performance of the AMF method is not significant unless the FIR filter size is at least 50 when working with fiber optic cables in the laboratory. Thus, the optimum DL investigation was carried out with a maximum value of $N = 50$ to emulate strong speckle-like noise and weak signal condition. Since the recorded data is quite large with thousands of frames, the above steps were executed for a large number of different temporal segments and for different training sizes. Only two sample results set are shown in Figure 5.18 and 5.19 below where the start of the segments were the 1000. and 2000. ϕ-OTDR frames. The vibration range bin was included in the sample data by selecting the training data set with $L_s = \{81, ..., 89\}$ and $\#\{L_s\} = 9$ in both sets.

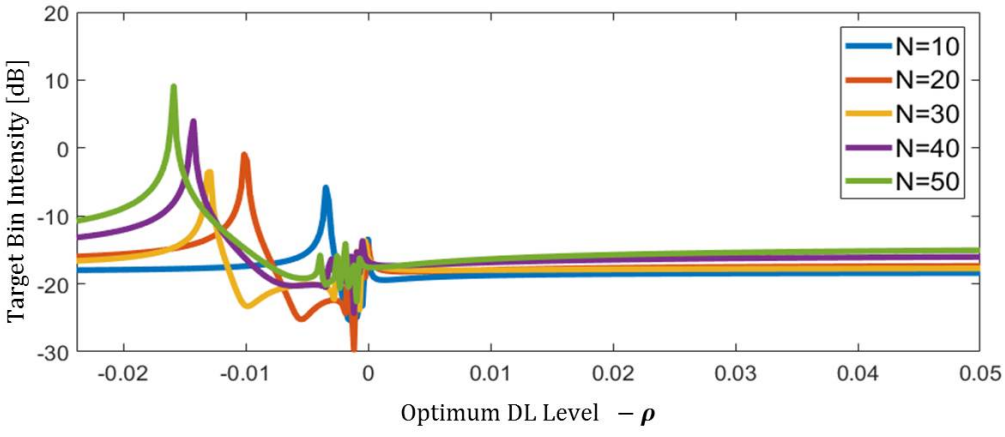


Figure 5.18. AMF filter output versus ρ values for the 1. test.

Table 5.3. ρ_{opt} and the largest eigenvalue comparison for the 1. test.

N	10	20	30	40	50
ρ_{opt}	-0.0035	-0.0101	-0.0129	-0.0146	-0.0159
λ_1	0.0034	0.0101	0.0130	0.0146	0.0166

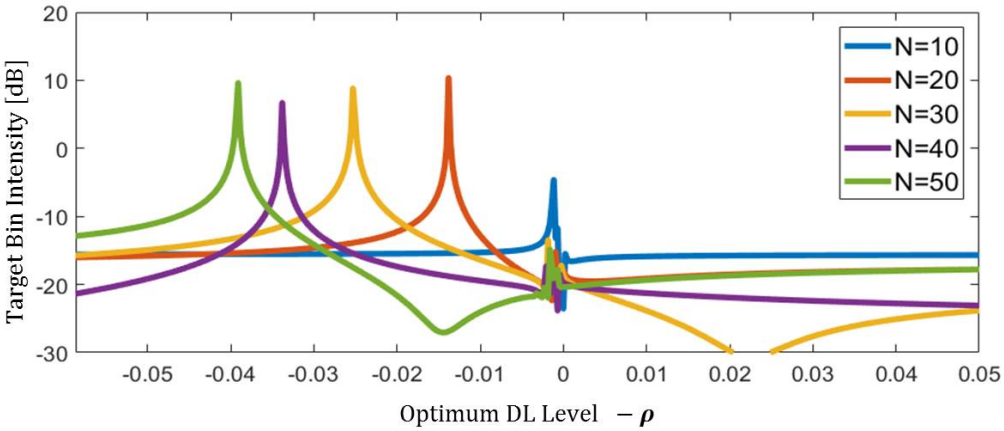


Figure 5.19. AMF filter output versus ρ values for the 2. test.

Table 5.4. ρ_{opt} and the largest eigenvalue comparison for the 2. test.

N	10	20	30	40	50
ρ_{opt}	-0.0012	-0.0138	-0.0253	-0.0338	-0.0391
λ_1	0.0018	0.0142	0.0257	0.0338	0.0394

It is clear from the figures that, for the signal contaminated SCM case, the ρ values which maximize the filter output at the vibration range are all on the negative axis and this verifies the negativeness of the ρ_{opt}. The measured ρ_{opt} values and the largest eigenvalues of the relevant SCMs computed for five different FIR filter sizes are summarized in Tables 5.3 and 5.4. The larger the filter size, the more negative the optimum DL value, which is again consistent with the analytical approximation given by (5.2). Additionally, the negative of λ_1 is almost the same with the measured ρ_{opt} for $N \leq 40$ and very close to it for $N \geq 40$.

5.3.2. MED Test results

The indoor data were subject to testing with the MED approach for various segmentation choices with a fixed $P_{FA} = 0.01$. Since the recorded ϕ-OTDR data used in this research were all intensity based, the relevant TW thresholds for the real-valued observations were computed for $\beta = 1$. The sample results presented here are given for a fixed selected MED(M, L) set for direct comparison of data recordings. The results for the 1. vibration are shown in Figure 5.20. Since the vibration is relatively strong with an average SNR of 30 dB over the recording time, all MED(M, L) choices resulted with the successful detection of the vibration as can be seen from the plots given in Figure 5.20.

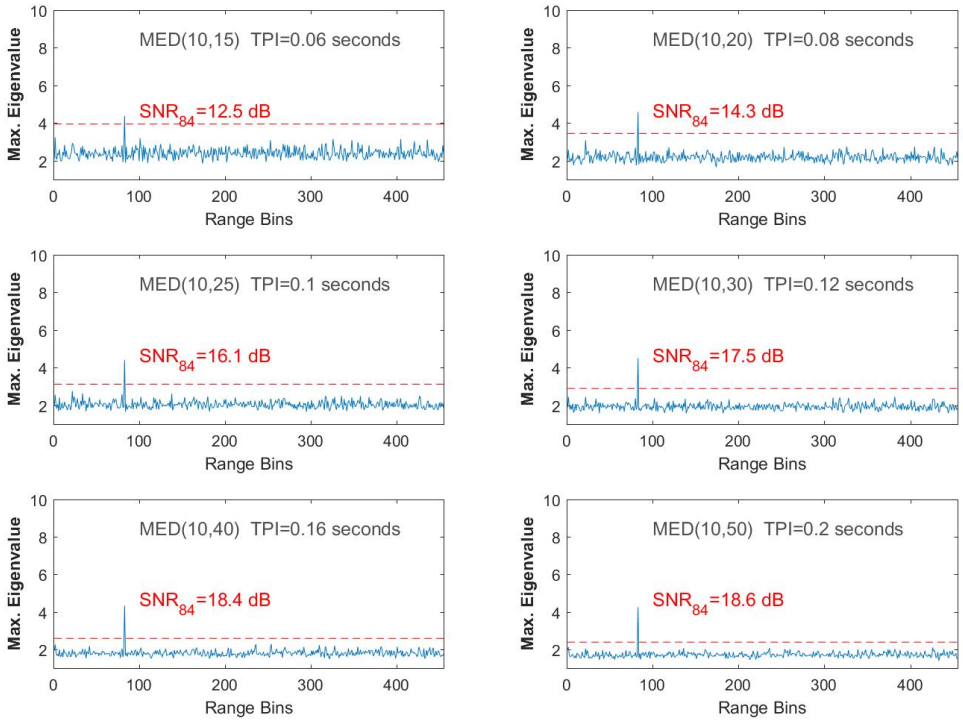


Figure 5.20. Sample test results of MED(M, L) for the strong 833 Hz vibration with $M = 10$ and $L = \{15, 20, 25, 30, 40, 50\}$.

Similar results were obtained when the MED approach was test with the 2. vibration of 1225 Hz, as shown in Figure 5.21. The fast-time SNR values were observed better than the 1. vibration. When the TPI durations for both vibrations were analyzed, the slow-time SNR of the 2. vibration was observed to be 3 dB higher than the slow-time SNR of the 1. vibration. This would result in 100% probability of detection even with more stringent FAR values. Hence, the MED processing of a vibration with relatively a higher SNR yielded a better detection performance which is an expected situation.

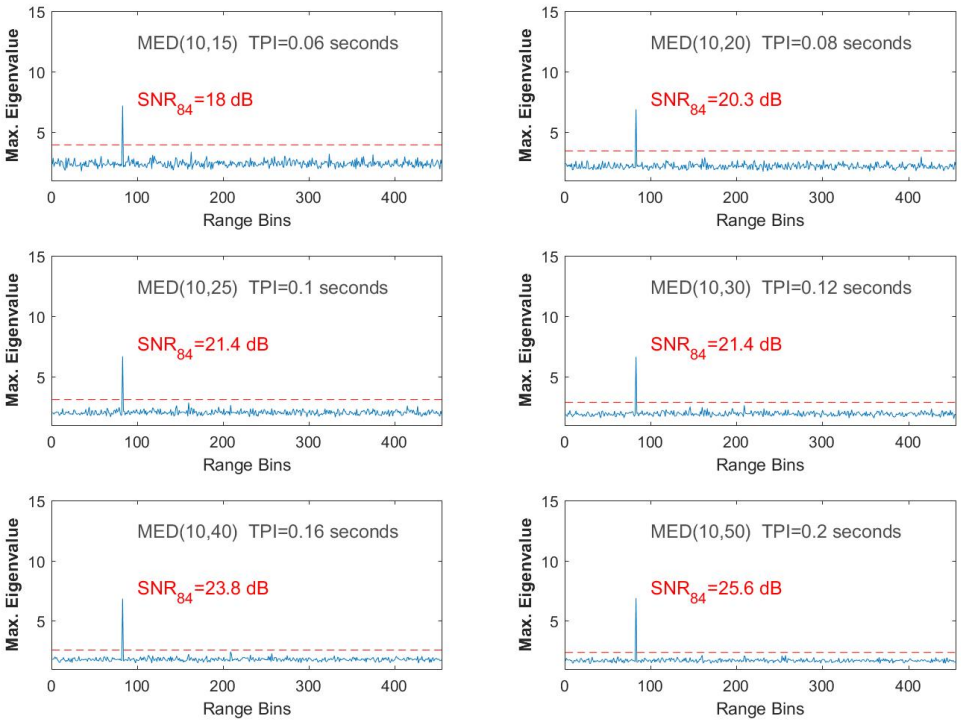


Figure 5.21. Sample test results of MED(M, L) for the strong 1225 Hz vibration with $M = 10$ and $L = \{15, 20, 25, 30, 40, 50\}$.

The impact of signal strength with real ϕ-OTDR data on the MED approach was also analyzed with the weaker 833.3 Hz vibration. None of the six TPI and segmentation choices yielded sufficiently high SNR for a reliable detection as presented in Figure 5.22. The MED$(10, L)$ selections failed to detect the vibration except for $L = 40$ and $L = 50$ which hardly detected the vibration as presented in Figure 5.22.

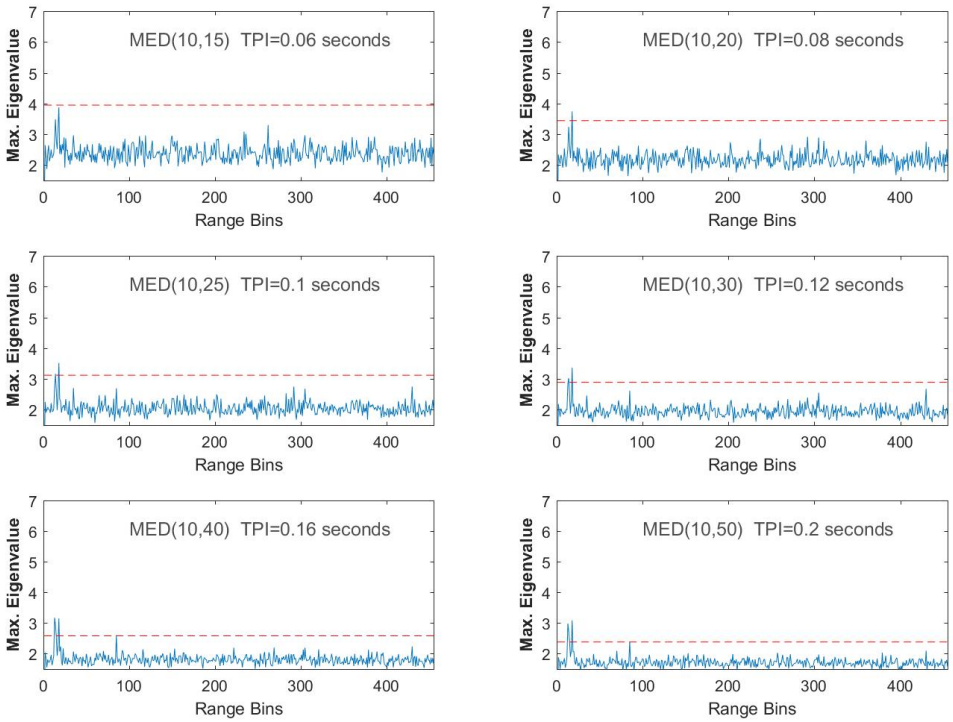


Figure 5.22. Sample test results of MED$(10, L)$ for the weaker 833 Hz vibration with $L = \{15, 20, 25, 30, 40, 50\}$.

The eigenvalues observed above the threshold at the beginning of the FO cable are due to the sensitivity of the ϕ-OTDR system to the environmental acoustic noise. Since the transmitted laser power through the FO cable is relatively much higher at close distances, the system response is high and the experimental system was very sensitive to indoor acoustic activities during tests, unfortunately these high peaks were unavoidable and such spurious responses were observed in some of the experiments. In commercial installations of DAS systems, it is known that additional 2-3 km long FO patches are usually used as a precursor to the real cable under inspection to alleviate this front end problem which is a learned lesson in this research study. These spurious peaks at the beginning of the FUT were also powered more when the TPI durations were increased for better sensitivity to detect the targeted vibrations as shown in Figure 5.23. When the segment size was increased to 15 samples to increase the TPI with the number of observations fixed, the results yielded higher peaks along the FUT

axis.

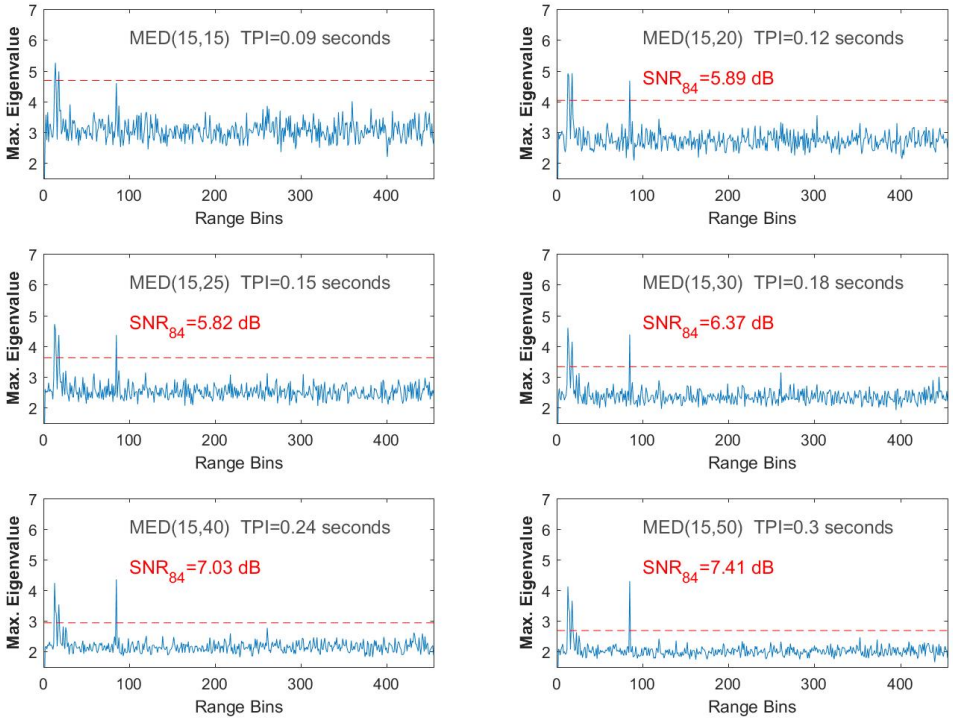


Figure 5.23. Sample test results of MED$(15, L)$ for the weaker 833 Hz vibration with $L = \{15, 20, 25, 30, 40, 50\}$.

The spurious responses were observed to occupy the frequencies close to the DC when the power spectra of the relevant range bins were computed in MATLAB. Hence, after high-pass filtering of the OTDR data with an appropriate cut-off frequency, (500 Hz in this example) the spurious peaks were easily eliminated as can be seen in Figure 5.24. On the other hand, the eigenvalues of the MED output are now relatively high, which will result in false alarms. This fact will be especially obvious for increased TPI durations which are depicted in the last row of the Figure 5.24. After high-pass filtering, the ϕ-OTDR samples are no more white but correlated; hence the SCM is not a white Wishart random matrix anymore. In this case, a whitening process can be employed before and then we can proceed with the MED approach.

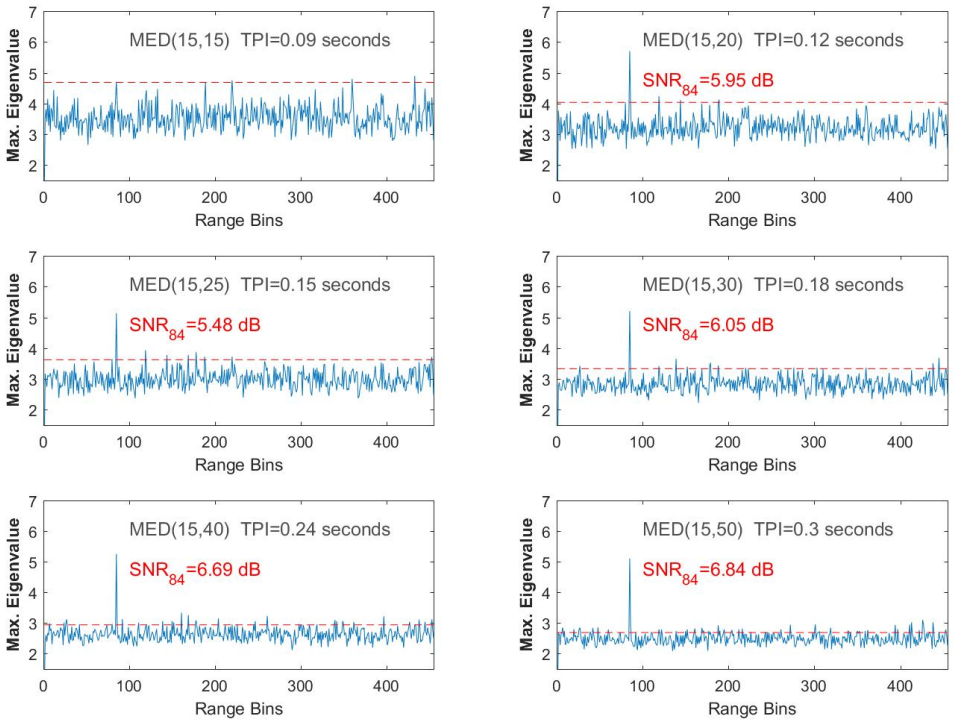


Figure 5.24. Sample test results of MED$(15, L)$ for the weaker 833 Hz vibration with $L = \{15, 20, 25, 30, 40, 50\}$ after high-pass filtering with a cut-off frequency of 500 Hz.

A commonly used whitening transformation for this objective is called Mahalanobis whitening which is analytically expressed as

$$\boldsymbol{Z} = \boldsymbol{R}_N^{-1/2}\boldsymbol{X} \tag{5.3}$$

where $\boldsymbol{R}_N$ is noise covariance matrix and $\boldsymbol{X}$ the data matrix given by (4.16) before. Here, since the true statistics cannot be known, the sample covariance of the noise denoted by data $\boldsymbol{R}_N$ is again estimated by noise-only assumed range bins. Hence, when we are dealing with the correlated noise case, we will be dealing with separate SCM estimations: one for the measured data to be whitened, the other for the transform matrix to whiten the observations. Thus, the matrix which can be defined by $\boldsymbol{R}_N^{-1/2}$ can be also called as the "whitening transform" for the correlated measurements and estimation of $\hat{\boldsymbol{R}}_N$ is no different from the SCM given by (3.9).

The result of the MED technique after the above whitening application is presented in Figure 5.25 for the 3. data set. For the whitening transform matrix estimation, all the range bins along the FUT were used to estimate the noise covariance (i.e.,$\hat{\boldsymbol{R}}_N = \frac{1}{327}\sum_{l=1}^{327} \boldsymbol{x}_l \boldsymbol{x}_l^H$). After computing the negative square root of $\hat{\boldsymbol{R}}_N$, the resultant matrix was used to decorrelate the pre-filtered ϕ-OTDR data by applying the transform given by (5.2). Finally, the MED approach was then applied for the SCM computed by $\hat{\boldsymbol{R}}_Z = (1/M)\boldsymbol{Z}\boldsymbol{Z}^H$. It is clearly seen in Figure 5.25 that the eigenvalues were more reliably computed to be comparable with the relevant TW threshold and the false alarms were canceled due to the filtering and decorrelation steps.

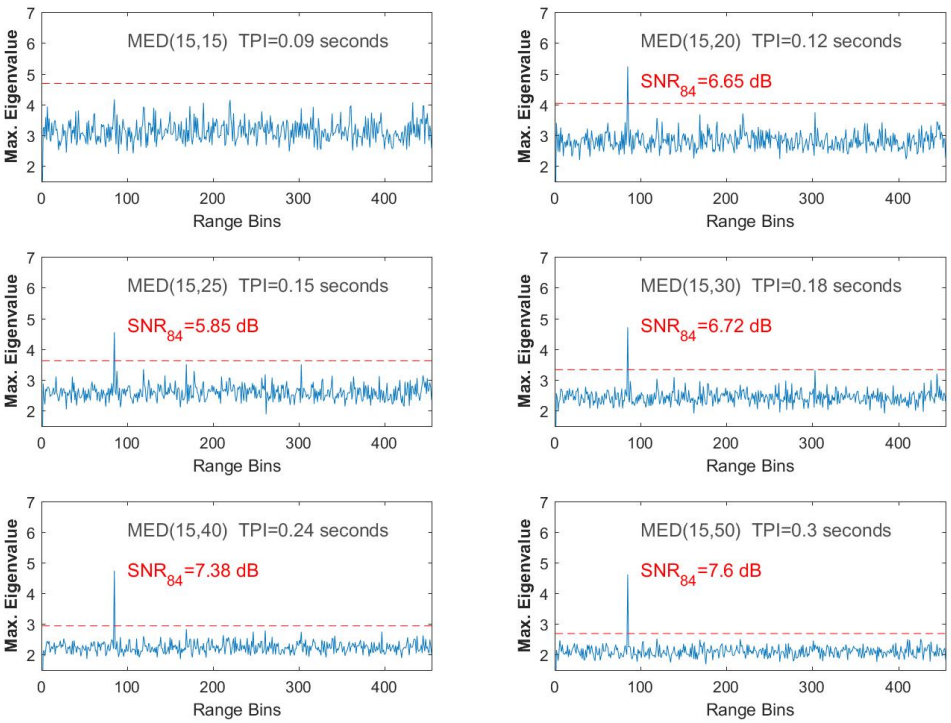


Figure 5.25. Sample test results of MED$(15, L)$ for the weaker 833 Hz vibration with $L = \{15, 20, 25, 30, 40, 50\}$ after high-pass filtering and whitening.

The indoor data with the multiple vibrations were tested with the MED approach similarly. The results for the last data set with MED$(10, L)$ are shown in Figure 5.26. There is a spurious peak at the very beginning of the FUT which is more evident as the TPI is increased over 0.2 seconds. When the MED approach was tested with square data structures, the MED technique resulted in more spurious responses as shown in Figure 5.27 and the number of the spurious peaks was increased significantly for $L > 20$. When compared with the previous segmentation strategy, it is clearly seen that the results are more dependent on the segmentation choice (or the data matrix structure) rather than the total TPI duration. The comparison of the MED$(10, 75)$ with MED$(25, 25)$ and the comparison of MED$(10, 100)$ with MED$(30, 30)$, etc., exhibit this impact very clearly. Although the TPI was 625 samples long for the MED(25,25) which is less than the TPI of MED(10,75), more spurious responses were observed with the MED(25,25). The situation is exactly the same for MED(10,100) since the TPI is longer than the one with MED(30,30) but only one spurious point was observed.

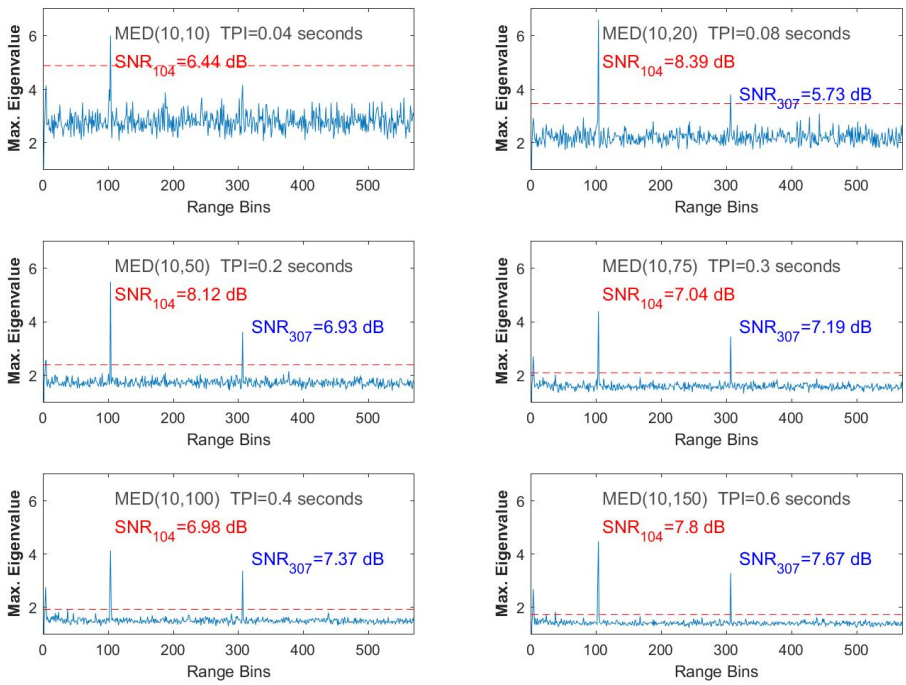


Figure 5.26. MED$(10, L)$ test results for the multiple vibrations with $L = \{10, 20, 50, 75, 100, 150\}$.

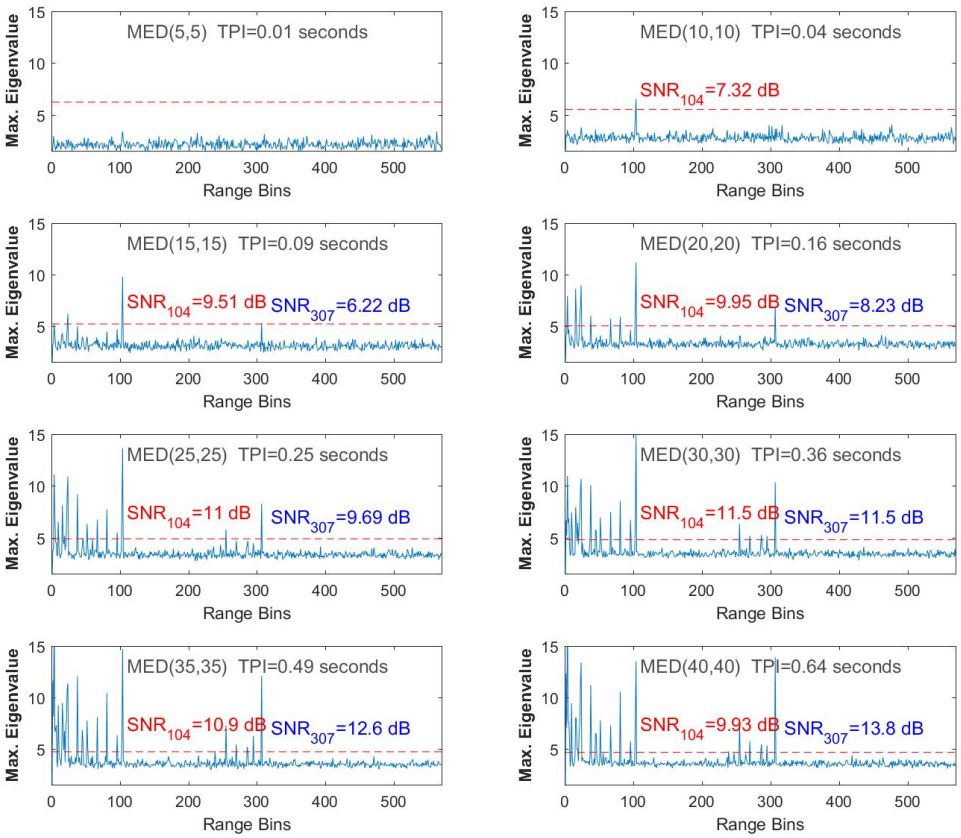


Figure 5.27. MED test results for the multiple vibration data with various square data matrix choices.

The previously tested pre-filtering and whitening steps were followed in the same manner to eliminate the spurious peaks and the results are shown in Figure 5.28. The high-pass filter used was a Chebyshev type of order 7 and with a cut-off frequency of 400 Hz and the whitening transformation was computed again with all the range bins considered for the noise-only snapshots. All of the eigenvalues due to the low-pass structure of the environmental noise were canceled as can be clearly seen in the Figure 5.28, except the spurious peak observed at the very beginning of the FUT for some MED choices. Such unconcealed peaks are probably due to the fact that their frequencies still fall into the pass-band region of the selected filter.

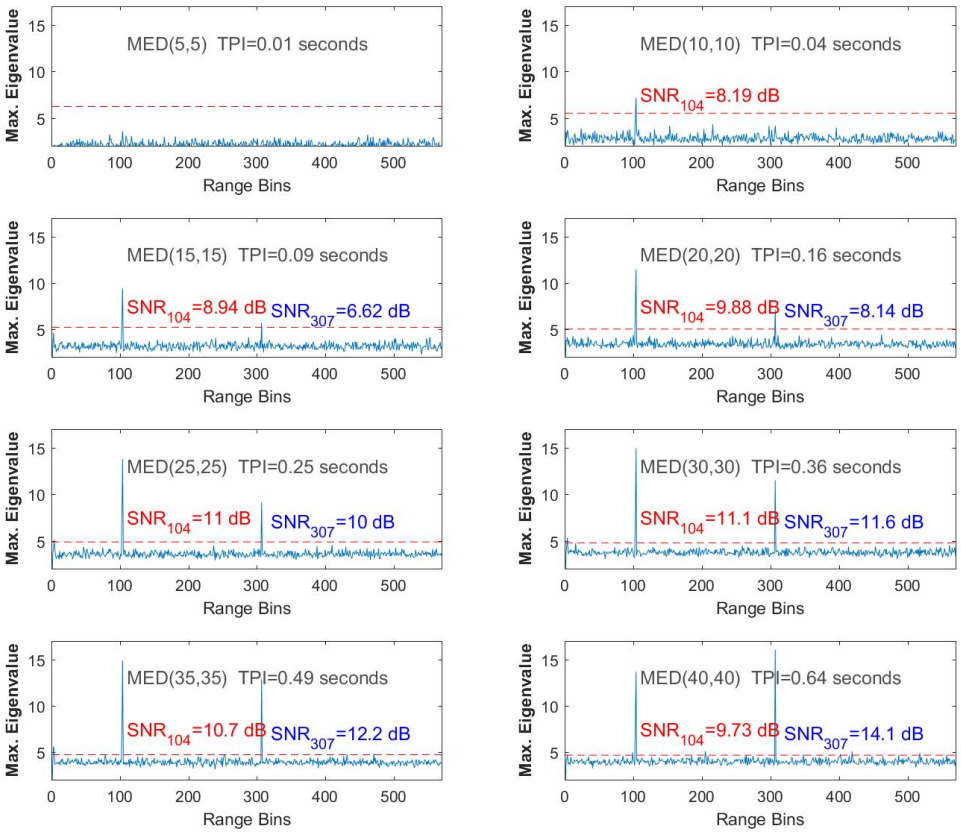


Figure 5.28. MED test results for the multiple vibration data with various square data matrix choices.

In order to observe the impact of TPI duration and segmentation choices and to draw a general picture, several batch processings were carried out with different slow-time SNR values. Since there are numerous segmentation possibilities, receiver operating characteristics (ROC) curves were computed for various segmentation choices and sample κ values to provide concise information about the proposed technique. Since it was observed during the experiments that the MED(M, L) outputs always yielded true positive for $M \geq 30$, the ROC curves were computed with segment sizes up to $M = 40$ starting from $M = 2$. The 1. and the 3. datasets were used for these experiments with the slow-time SNR values of 30 dB and 15 dB, respectively.

Figure 5.29 shows these ROC curves computed for the vibration range bins of the two data sets by sliding the TPI window incrementally for the whole data record duration. For each MED(M, L) processing, these temporal increments were appropriately selected to achieve at least 100 binary detection results for the evaluation of the probability of detection (P_D).

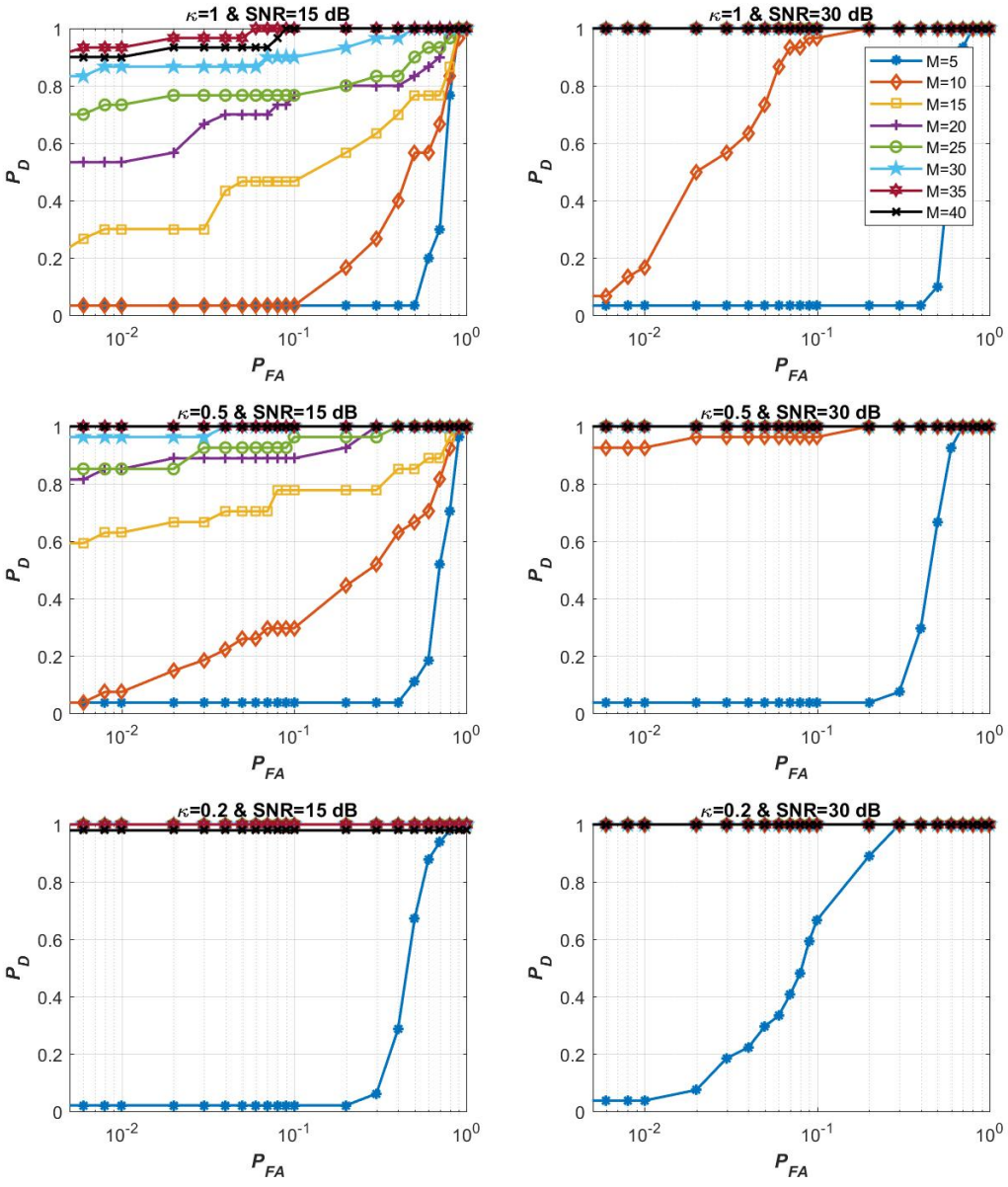


Figure 5.29. P_D versus P_{FA} for two different vibration strengths.

It can be deduced from the results that segment sizes as small as 5 are not sufficient for a reliable detection even the number of observations are increased. For the strong vibration, as can be seen at the right column plots of Figure 5.29, it seems to be a good choice to start with segment sizes at least 20 samples long not to miss the vibration. The P_D increases with processing durations for both SNR cases and the required TPI is shorter for the stronger vibration cases which are expected results. It is also clear from the upper row, especially from the left-most plot that square data matrix selections are not good choices since 100% P_D values are achievable only at high TPI durations and high P_{FA} rates.

5.3.3. Empiricial Eigenvalue Distributions of the Indoor Data

The MP Law was experimentally verified by Monte Carlo simulated ϕ-OTDR data in the previous chapter. To fulfill our experimental verification, one sample result of the MED(20,100) processing of the 833 Hz vibration data is included for the evaluation of the analytical bounds with the indoor data. The bulk eigenvalue statistics were computed with the MED(20,100) choice in a block-by-block basis for the 833.3 Hz vibration for 3 different vibration strengths. The TPIs were selected with overlapping time intervals of varying lengths. The bulk eigenvalue distributions computed at the vibration range bins for the 833.3 Hz vibration are shown in Figure 5.30. It is once more clear with the real ϕ-OTDR data that outliers exist in the spectrum for the H_1 case. These outliers are located further from the analytical confines if the vibration is strong enough (see the upper row of Figure 5.30 for the strongest vibration) and very close to the maximum eigenvalue bound if the vibration is so weak in strength (see the lowest row of Figure 5.30 for the weakest vibration).

Additional experiments were carried out for evaluating the TW Law with the empirical distributions of λ_1 of the noise-only SCMs. As the FOTAS system was able to record not only intensity but also complex (baseband I-Q samples) data, the TW_β distributions were tested for both $\beta = 1$ and $\beta = 2$ cases. The PDF for λ_1 computed with real intensity data was expected to be well described by TW_1 while the largest eigenvalue fluctuations of the complex SCM were expected to fit the TW_2 distribution.

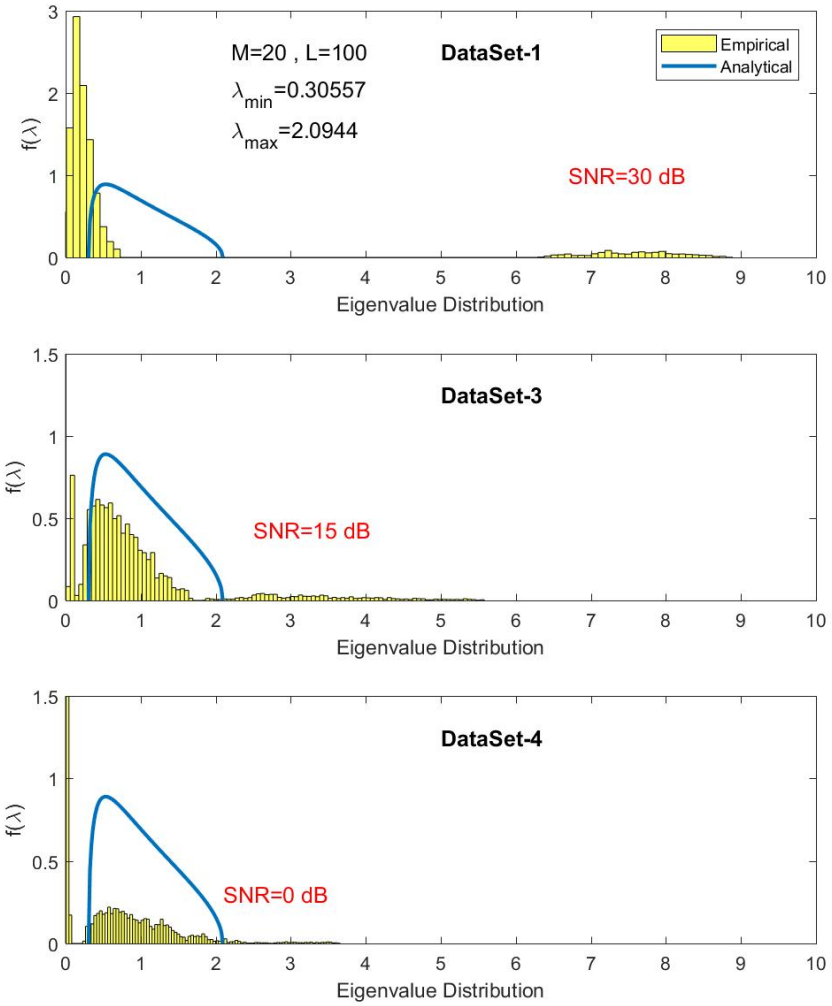


Figure 5.30. The bulk eigenvalue statistics at the vibration range bin for the 833 Hz vibration and their comparison with the relevant MP bounds.

Several noise-only SCMs were estimated by sequentially processing the non-overlapping time segments of the whole data. The largest eigenvalue distributions computed with the above approach are shown in Figure 5.31 and Figure 5.32 for the 1. and 3. data sets, respectively. Noise-only SCMs were computed by taking the range bins from the set $L_s = \{30, 31, ...80\} \cup \{90, 91, ..., 327\}$ leaving some guard bins on two sides of the vibration range. Since the raw data were all intensity related, the empirical distributions of λ_1 were expected to fit the TW_1 PDF.

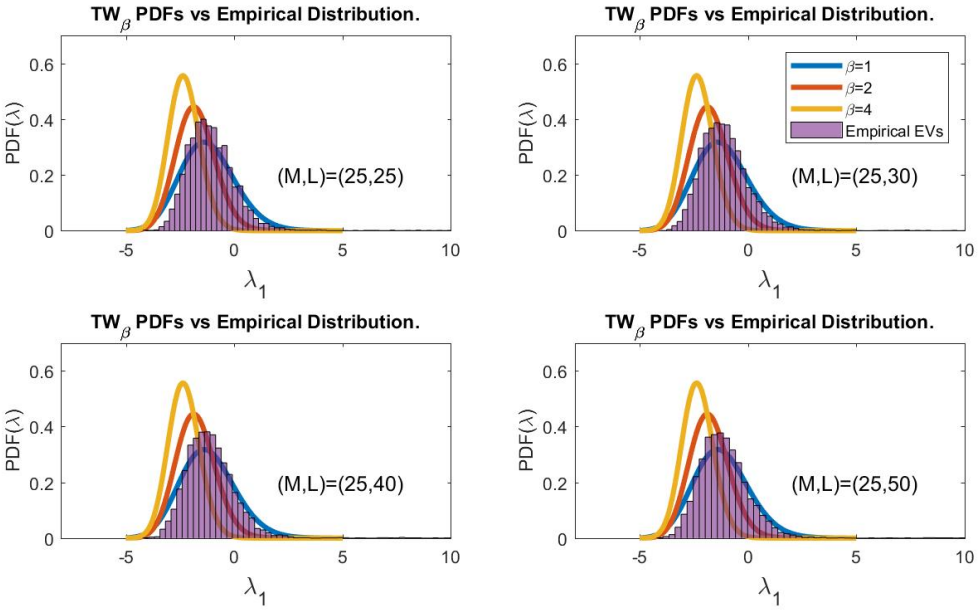


Figure 5.31. The empirical largest eigenvalue distribution of the noise-only SCMs computed for the 1. data set for $\beta = 1$ and various (M, L) choices.

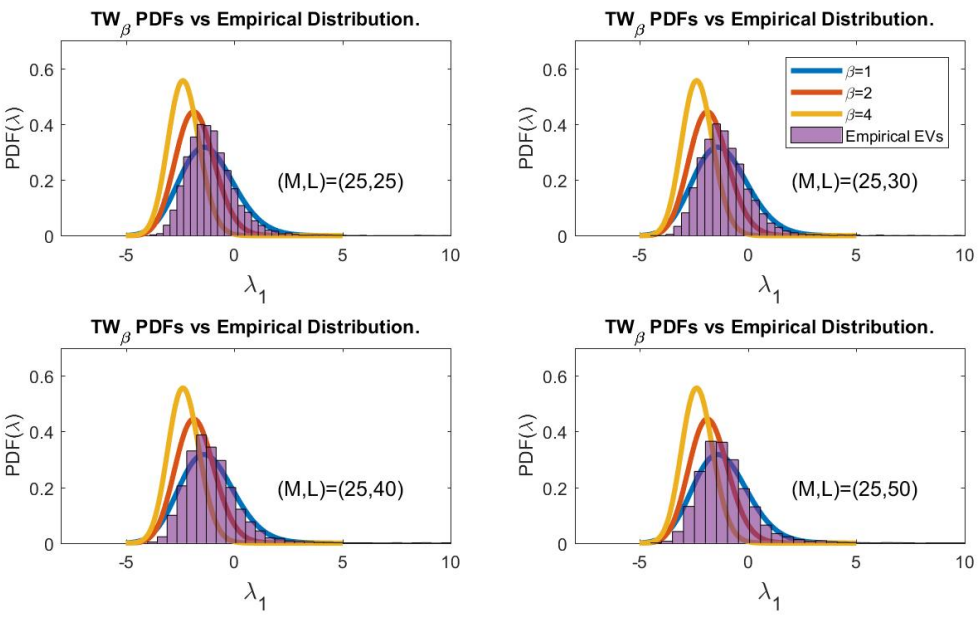


Figure 5.32. The empirical largest eigenvalue distribution of the noise-only SCMs computed for the 3. data set for $\beta = 1$ and various (M, L) choices.

It is clearly seen from Figure 5.31 and Figure 5.32 that the empirical distributions are in good agreement with the analytical results for both data sets.

The same steps were followed for an I-Q record as shown in Figure 5.33. The SCM's were computed from a different indoor data set where I-Q baseband samples were available in the recordings for the further vibration located at the 307. range bin. The noise-only SCM's were computed with the set $L_s = \{51, 52, ..., 300\} \cup \{310, 311, ..., 409\}$ by leaving again some guard bins around the vibration bin. The first 50 range bins were also omitted for the SCM estimation to prevent it from the weak spurious signals. It can be seen from the plots of Figure 5.33 that the empirical distributions of the largest eigenvalues are in good agreement with the TW_2 PDF. in the complex data case.

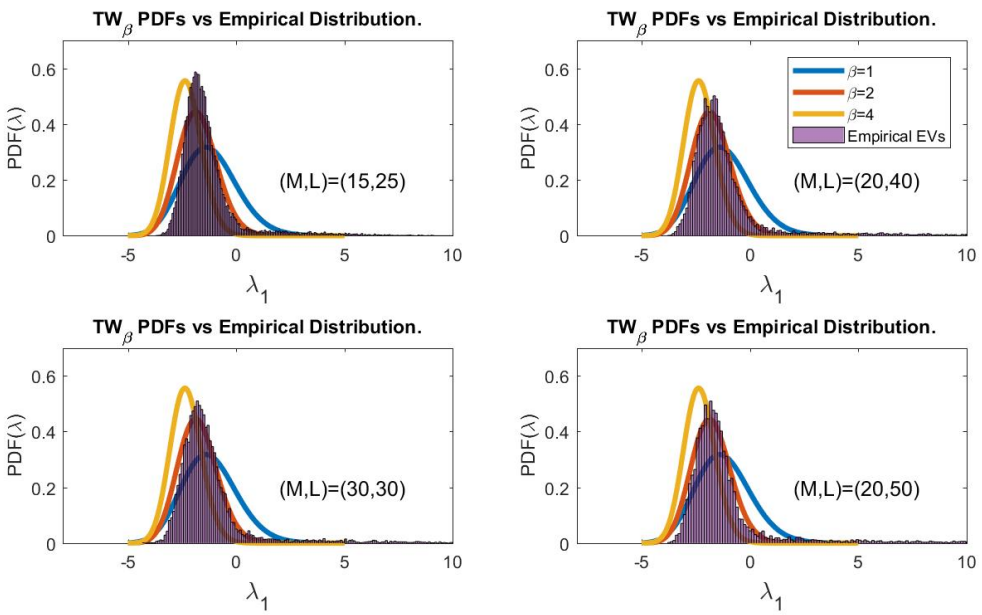


Figure 5.33. The empirical largest eigenvalue distribution of the noise-only SCMs computed with I-Q recorded data for $\beta = 2$ and various (M, L) choices.

5.3.4. Conclusions

In the previous two chapters, the proposed approaches of this research were tested with Monte Carlo simulated ϕ-OTDR data. The simulation studies did not take into account the noise from active devices such as the laser-induced noise and the noise of the receivers. In addition, the polarization effects were omitted from the theoretical model for simplicity. Under these conditions, the AMF and MED test results were observed to be promising methods.

In this chapter, both the AMF and the MED approaches were applied to real ϕ-OTDR data gathered in the laboratory environment to verify the results of the previous sections. The results of the AMF method were observed to be in good agreement with the simulation study outlined in Chapter 3. For the data gathered indoors with the acoustic isolation, it can be easily deduced from the results that the vibration is not solely observable unless the N, the FIR vector size (or in other words, the number of consecutive ϕ-OTDR traces processed at the same time), is over 50. When N is increased to values higher than 100, the amount of speckle reduction is significant, which means that the RBS term of the background noise is highly suppressed. When the acoustic isolation is removed and the vibration intensity is reduced to lower levels, more than a 10-dB speckle reduction can only be achieved with a much higher value of N. This is expected since both simulation data and the measurement data with isolation present the ideal situation. Without the isolation, the whole FUT is subject to more interference, and the vibration signal is highly obscured by the increased amplitude fluctuations of the ϕ-OTDR traces.

The results obtained by the MED method were also in good agreement with the results of Chapter 4. One important issue was the observation of the increase in the interference terms when the data matrices were in the square form. For $\kappa = 1$, the out of band vibrations due to the environment were more significant when compared to the segmentation choices of $\kappa < 1$. These results need more and deepened theoretical work on the detection tools borrowed from the RMT, to figure out why square matrix selections reveal weak signals better than the rectangular form of the data matrix.

On the other hand, the observation of other interference terms with the MED technique exhibits the main feature of this approach for detecting the vibrations without any prior knowledge about the environment. The AMF has the selectivity feature while the MED blindly detects every vibration knowing that both methods detection capability is dependent on the observation dimension: the more samples taken into account, the higher the SNR obtained.

One important result obtained with the AMF method was the severe impact of the DL level in the case of signal contaminated secondary data. If there is signal contamination in the noise-only assumed data, it will seriously affect the performance of the AMF method. It turns out that, the signal detection problem is encountered again from a different perspective. We can handle the signal detection problem once more for the selection of the "noise-only" range bins and to assure the absence of vibration content in the secondary data. In this regard, the MED technique may be used as a complementary solution to detect the presence of vibrations before the AMF method. In other words, a unified framework can be developed to deal with the covariance structure of the ϕ-OTDR data.

When the results of both methods were compared together for varying TPI durations, they were also observed to be consistent in dealing with the weak signal conditions as the TPI durations must be over 500 samples long for reliable detection. According to the results obtained up to now, a rule of thumb may be taken as at least 1000 samples long TPI selections can be a good selection to start with for real-time applications of the proposed methods for DAS systems. Weak signals embedded in a highly cluttered environment may need a further increase of the processing time in real-world conditions which is the topic of the next chapter.

6. FIELD TEST EVALUATIONS

6.1. Introduction

To validate the proposed approach in real-world conditions, field tests were carried out in the TÜBİTAK campus site with real fiber optic cables. Two different sensor installations were investigated. The first installation was dedicated to investigating the performance of proposed algorithms with buried fiber optic cables. At the second site, the fiber optic cables were installed on a fence which is more sensitive to the environment. The outdoor test sites and setup configurations are described below.

6.2. Outdoor Test Sites

6.2.1. Field Test Site-1: Buried Fiber Optic Cables

The acoustic enclosure and the fiber on reels were removed from the experimental setup and the ϕ-OTDR system was connected to the buried fiber through a patch panel in the office area. The general setup with the FUT for this test configuration is sketched up in Figure 6.1 and the route of the selected intranet line between two facilities is roughly depicted in Figure 6.2. The FO bundle consists of 12 different thick coated SMF cables that were installed with a protective sheath made of steel. One spare fiber cable was selected as the FUT to observe the sensitivity of this telecommunication line for a potential DAS application.

The exact length of the FUT was estimated as 987 m via our ϕ-OTDR system screen as shown in Figure 6.3. Before the field tests, several manholes were inspected to verify the route and to assign the correct range bin locations on the observed ϕ-OTDR traces. The starting and ending range of the effective buried length was estimated as 540 and 855 meters, respectively, by perturbing the manhole covers with footsteps.

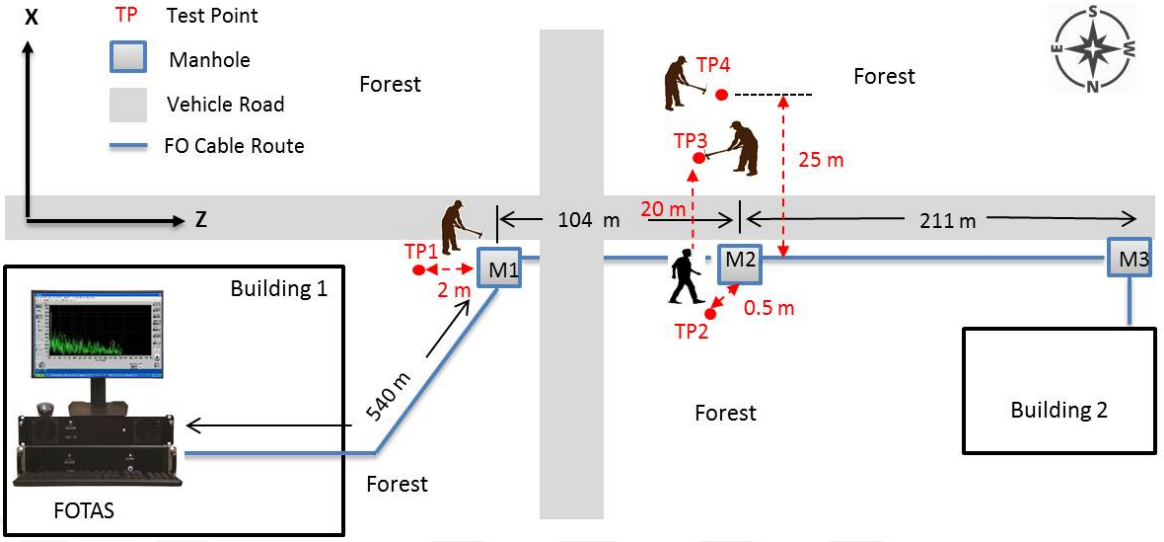


Figure 6.1. Generic schematic of the outdoor test points for the 1. test site.

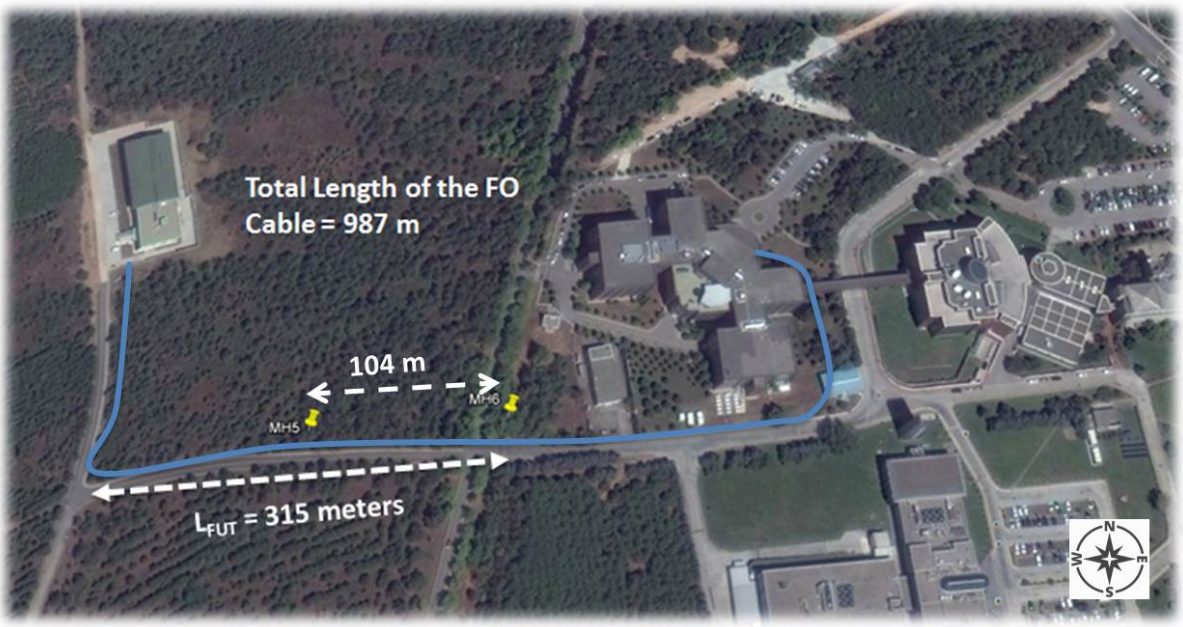


Figure 6.2. The route of the buried FO intranet line at 1. test site.

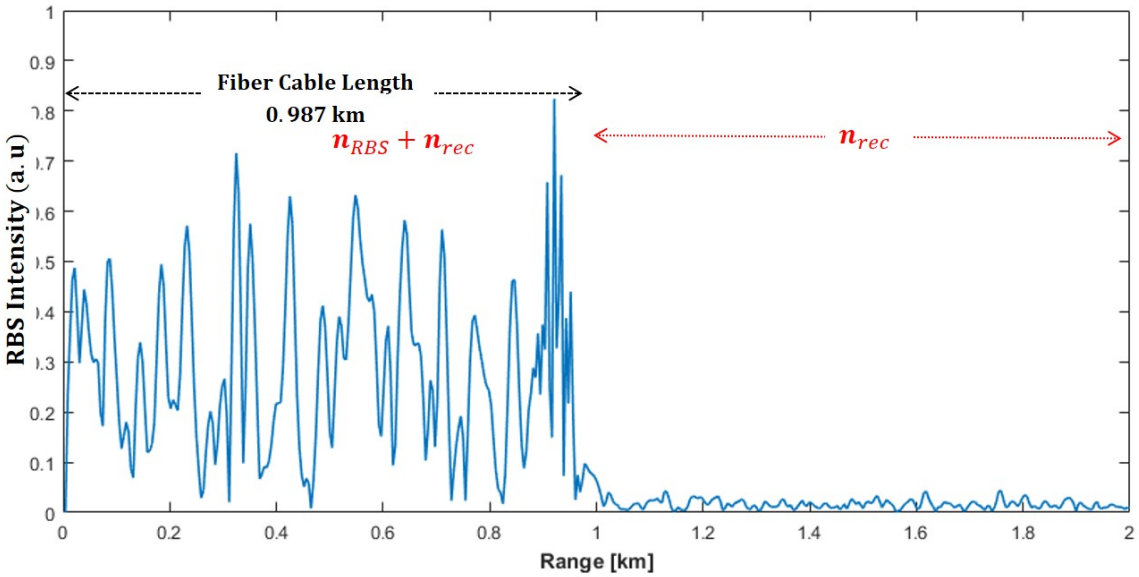


Figure 6.3. Typical raw ϕ-OTDR trace recorded by FOTAS at the 1. test site.

6.2.2. Field Test Site-2: Fiber Optic Cables on Fences

The second test field configuration has been set up in TÜBİTAK Gebze campus to study the environmental impact in fence installations. This new installation has been completed in the southern border of the campus where the high-speed train (YHT) route is approximately going in parallel with the security fences. Approximately a 2.4 km long FO cable – 2.4 km of a fiber reel which is totally 4.2 km long- was installed partially on the south border fence to study new environmental conditions. The fence installation has been commenced at the end of a buried 2.5 km long intranet line. The routing is completed up to point where the fence route diverges from the railway route and the retained section of the fiber reel which is 1.8 km long was left at the end. This new FUT is totally 6.7 km long which is sketched in Figure 6.4. This installation was done in purpose to study fiber optic cables without burying them and to investigate the severe impact of environmental effects. Another opportunity was the existence of a railway for exploring the capability of the proposed signal processing algorithms to detect and track the train during its pass. The fiber route in the close vicinity of the railway is depicted in Figure 6.5.

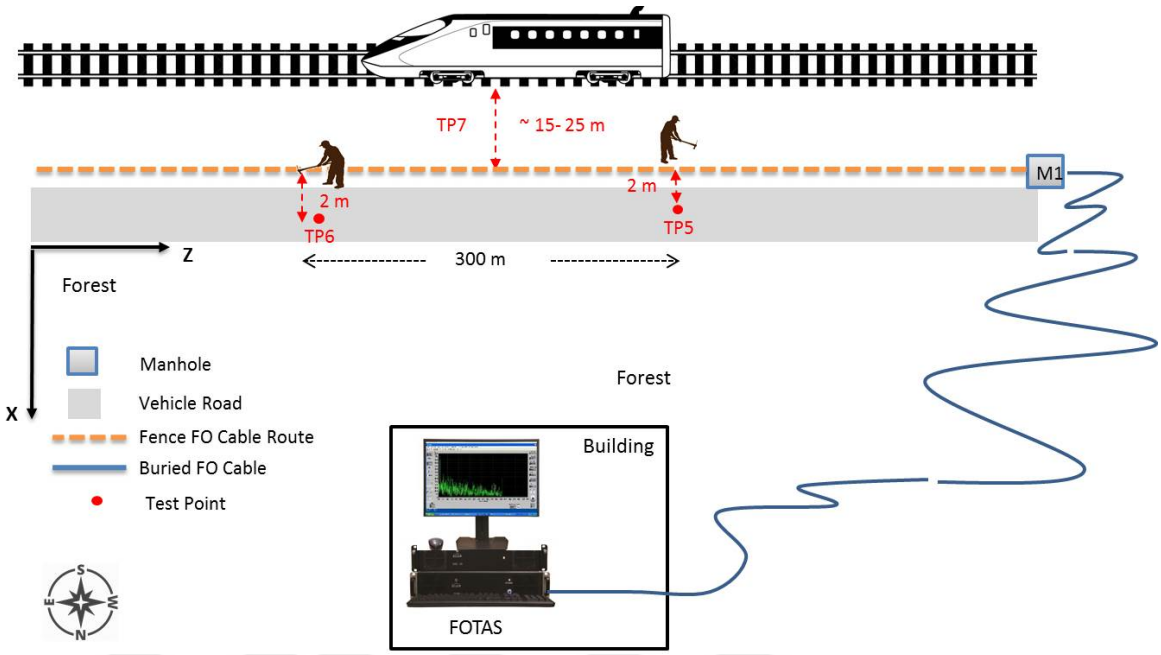


Figure 6.4. Generic schematic of the outdoor test points for the 2. test site.

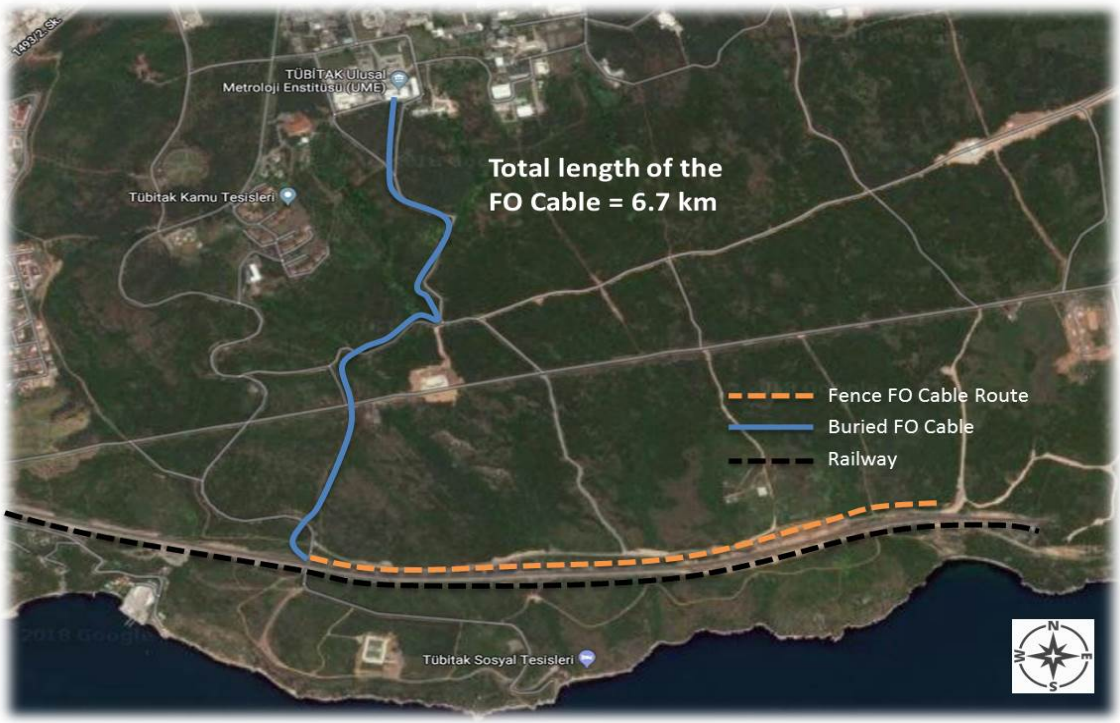


Figure 6.5. The route of the FO cables for the 2. test site.

This new configuration which possesses both underground and on ground conditions makes it available to observe two different kinds of RBS fluctuations. The attachment of the FO cable to the fence near the railway is shown in Figure 6.6. The distance between the new FUT route and the railway is at most 25 m.

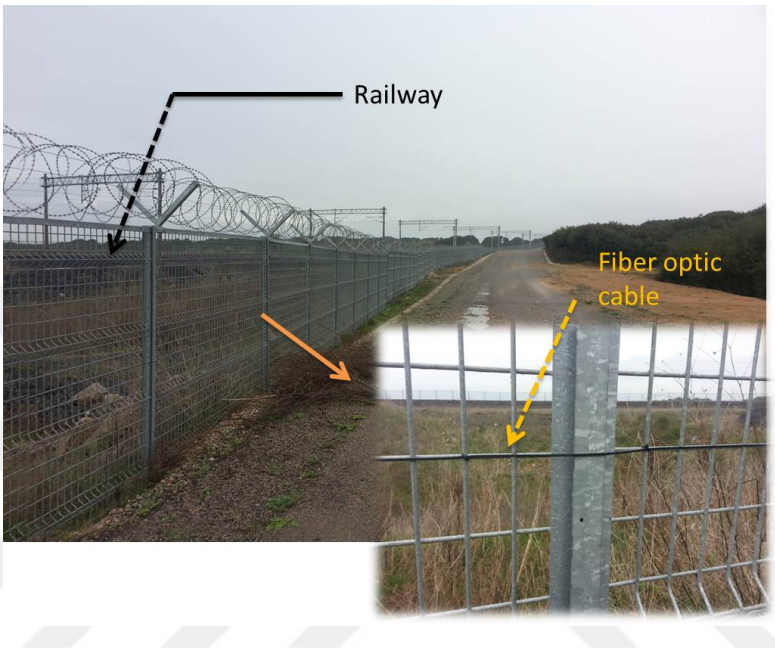


Figure 6.6. Aerial fiber installation near the railway.

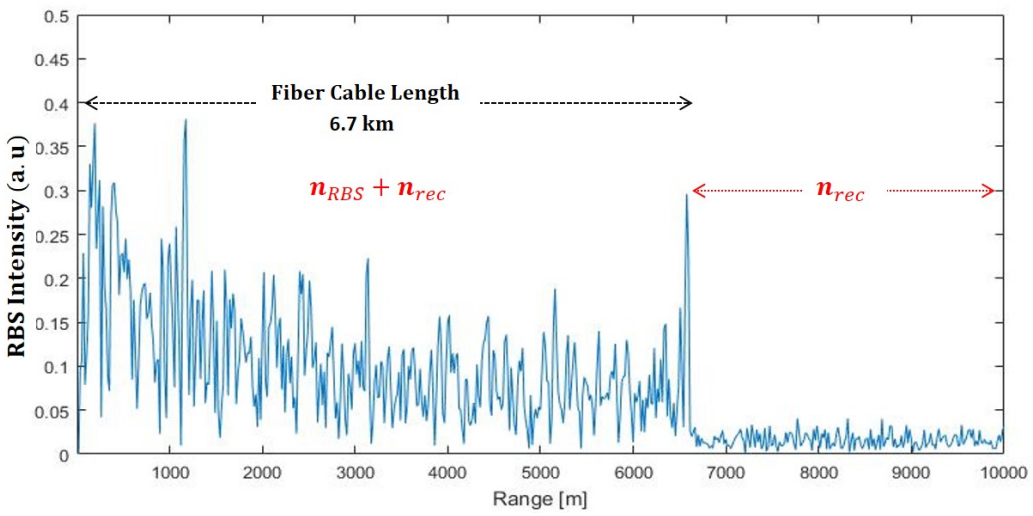


Figure 6.7. Typical raw ϕ-OTDR trace recorded by FOTAS at the 2. test site.

6.3. Applications at Test Site-1

More than 300 tests were conducted between the dates April 19, 2017, and March 14, 2018, with this buried FUT and real ϕ-OTDR data was recorded during two weeks at different locations and with varying distances from the FUT. Since the effective test length of the FUT was only about 315 m, by adjusting the OID and SDR parameters the spatial resolution was increased to 25 m by reducing the laser pulse width to 244 ns. Oversampling was performed for better visualization of the results on the range gates of interest. Due to the onboard storage limitations of the DSP hardware, the recorded range was fixed to 2 km with $K = 456$ samples per every OTDR trace independently from the adjusted PRF. The spatial sampling between the range bins is 2000 m /456 = 4.39 m is the result of the oversampling and it is higher than the actual SR of the system. Oversampling was preferred to synthetically increase the number of range bins and to evaluate more acoustic channels due to the short FUT at the 1. test site. The modified system parameters used during field tests are summarized in Table 6.1.

Table 6.1. Modified DAS system parameters for field tests.

System Component	**Symbol**	**Value**
Interrogation Frequency	PRF	10 kHz
Pulse Width	τ	244 ns
Spatial Resolution	Δz	25 m
Length of the FUT	L_{FUT}	987 m
Number of Bins per Trace	K	456
Vibration Center Locationfor TP1	l_1	541 m
Range Bins for TP1	p_i	$i = \{118, 119, ..., 127, 128\}$
Vibration Centers for TP2, TP3 and TP4	l_2	655 m
Range Bins for TP2, TP3 and TP4	p_j	$j = \{138, 139, ..., 155, 156\}$

Table 6.2. Test signals used at test site-1.

Data Index	**Frames Recorded**	**Duration [s]**	**Signal Type**	**# Hits / Rounds**	**Distance from FUT[m]**
DataSet-6	309,930	30.99	Hammer Hit	11	2 m
DataSet-7	251,005	25.10	Walking	3	0.5 m
DataSet-8	245,005	9.80	Hammer Hit	5	20 m
DataSet-9	222,030	8.88	Hammer Hit	5	25 m

The data sets recorded for testing are summarized in Table 6.7. The first two field data sets were recorded with strong perturbations to observe the spectrum of the vibrations in the close vicinity of the buried cable. The data shown in the left column of Figure 6.8 refers to intensities of the signal recorded at the first test point (TP1), while the gravel path 2 m away from the manhole was hit with a big hammer (2 kg).

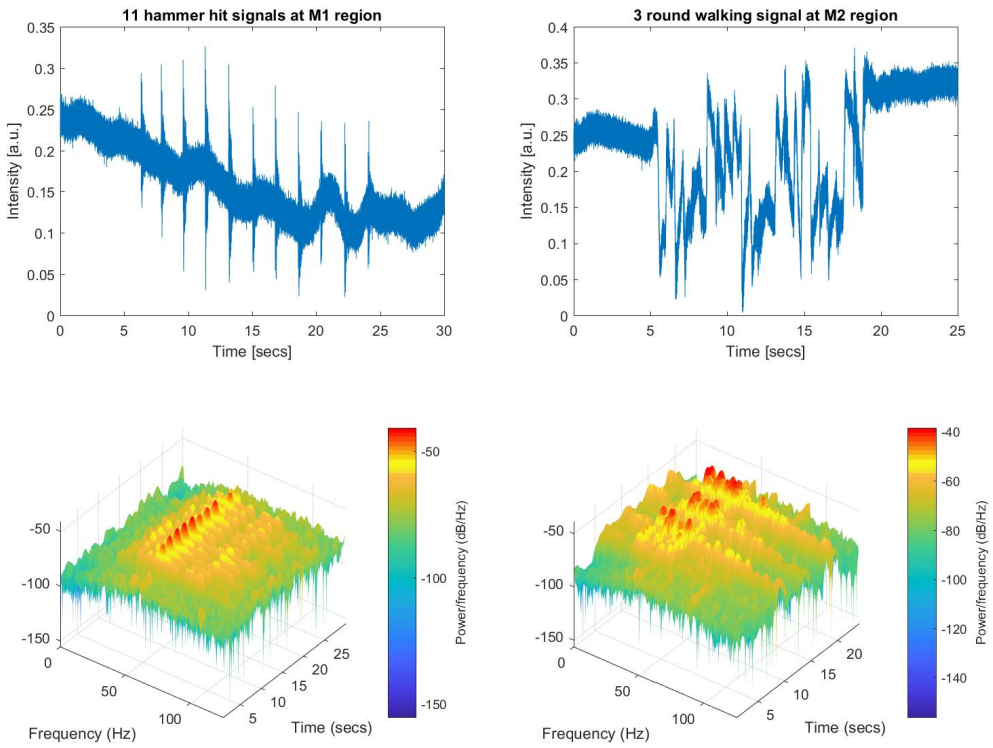


Figure 6.8. Two sample ϕ-OTDR intensity data recorded near TP1 and TP2.

As can be seen from the intensity profile in the graph, the total 11 sequential hits are considerably strong and easily "heard" by the FUT in the 123. range bin. The time-frequency analysis shows that most of the signal energy is concentrated in the 20 Hz – 100 Hz region giving a dominant peak around 50 Hz. The second field data set was recorded while a 90 kg weighted man circled the 2. manhole 3 times with uniform footsteps as illustrated in the right side of Figure 6.8. Similarly, the presence of a signal component is obvious both in the time domain and the frequency domain plots. Two dominant peaks were observed around 30 Hz and 45 Hz as can be seen from the spectrogram processing of the footsteps.

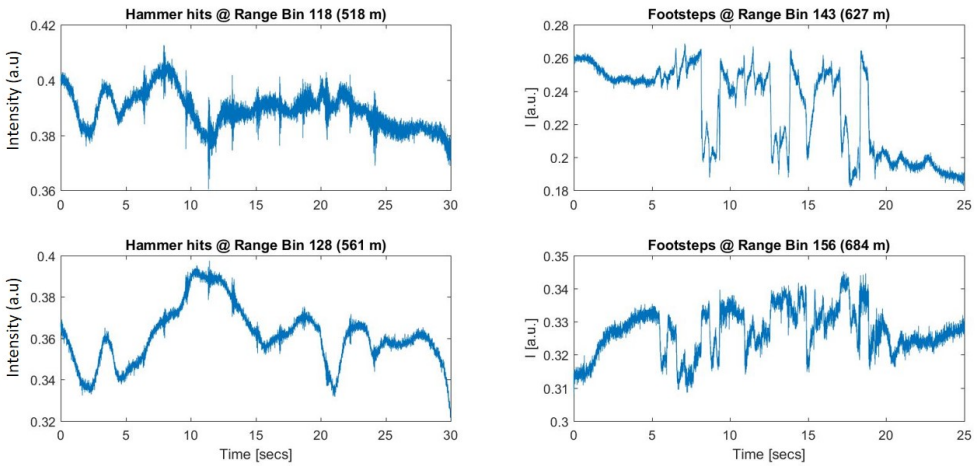


Figure 6.9. Recorded intensities at different range bins for TP1 and TP2.

The analysis of the raw ϕ-OTDR data for each range bin revealed that the induced acoustic energy on the FUT was spread on several range bins. The strongest signals recorded for the 1. and 2. manholes were observed on range bins 123 and 147, respectively. The distance between the two range bins is (147-123) $\times$4.39 m= 105.3 m which is consistent with the measured distance of 104 m as depicted in Figure 6.1. The range bins with the weakest contributions were noted along the FUT for every test during the experiments. All of the hammer hits are weakly heard by the 118. and 128. range bins. Thus, the vibrations were spread over a range of 10×4.39 m= 43.9 m . For the walking experiment at TP2, the footstep signals are visible along a range of 684 m - 627 m=57 m. as can be seen from Figure 6.9 with the corresponding range bins of 156 and 143.

6.3.1. AMF Test Results

The field data were subject to the AMF approach with similar steps followed before. The vibration signals recorded during field tests were no longer pure continuous sinusoidal signals. In order to analyze the noise only" and signal presence cases with the AMF approach, the start of and the end of the duration of each signal shot was noted, and the AMF method was tested with the observed peak frequencies.

For the sixth dataset, the first 400 ms of the record and the 400 ms duration of the sixth hammer hit were selected as inputs to the FIR filter in order to analyze the tested noise only (H_0) and signal presence (H_1) conditions. Since the dominant vibration was observed at 50 Hz for this dataset, the AMF method was tested with the FIR tuned to 50 Hz. The secondary data was set to $L_s = \{1, 2, ..., 109, 110\} \cup \{140, ..., 325\}$ and the diagonal loading was set as $\rho = 10$. As can be seen from the results shown in Figure 6.10, the speckle profile was still dominant, although the FIR size was set to $N = 500$. It has been observed after several trials that significant noise suppression was obtained when N reaches 1000. The vibrations observed at the beginning of the FUT and in the region of $200 \leq k \leq 230$ are due to the vibrations in the buildings which are unavoidable as the fiber is sensitive to all acoustic noise along its route. If there is vibration with the same frequency of the steering vector of the AMF, then it naturally exhibits itself in the path which is the case in the observed spectrum of vibrations in Figure 6.10.

Almost similar results were observed with the seventh dataset when the FIR was tuned 30 Hz, as shown in the right column of Figure 6.10. All of the range bins along the FUT were tested with the same FIR size and the test frequencies. The maxima of the FIR output were observed at the correct range bins as shown by the blue lines: the peak for the hammer hit signal (sixth data set) was observed around the 123. range bin, and the peak for the footsteps (seventh dataset) was obtained around the 147. range bin as expected. The FIR outputs responded with no peaks for the “noise only” cases, which are shown by red lines in the lower row of Figure 6.10.

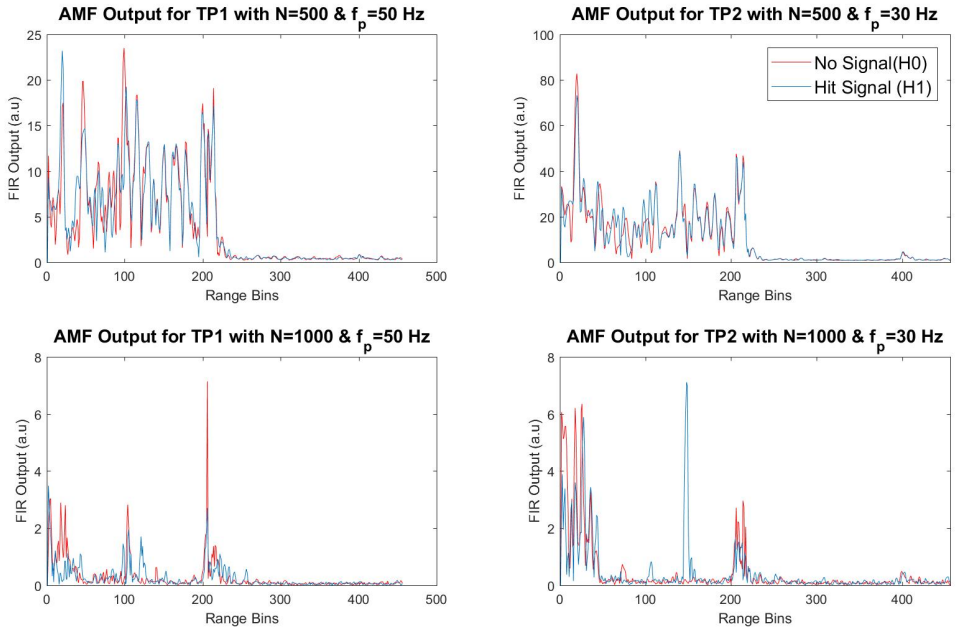


Figure 6.10. AMF outputs for TP1 and TP2 with $N = 500$ and $N = 1000$.

The impact of the DL level was analyzed for the field data set by with varying DL levels and observing the AMF outputs. The impacts of 3 different DL values are shown in Figure 6.11 with $N = 1000$ for the two field test data with the same TPI selections. The fast-time SNR values for the two target signals were computed for $100 \leq k \leq 200$ to assure that the range bins associated with the building locations are discarded. The vibration range bins during fast-time SNR computations were taken between the 118. and 128. for the hammer hit signal, and between the 143. and 56. for the footsteps. In Figure 6.11, the AMF processing results are given in the left column for the hammer hit, and in the right column for the footsteps, respectively. It is clearly seen that the decrease in the DL level improved the visibility of the hammer hit at the TP1 location significantly, but the same improvement was not observed for the footsteps at the TP2 location. The frequency selectivity of the AMF method and the impact of the DL selection in suppression the background noise were once more observed.

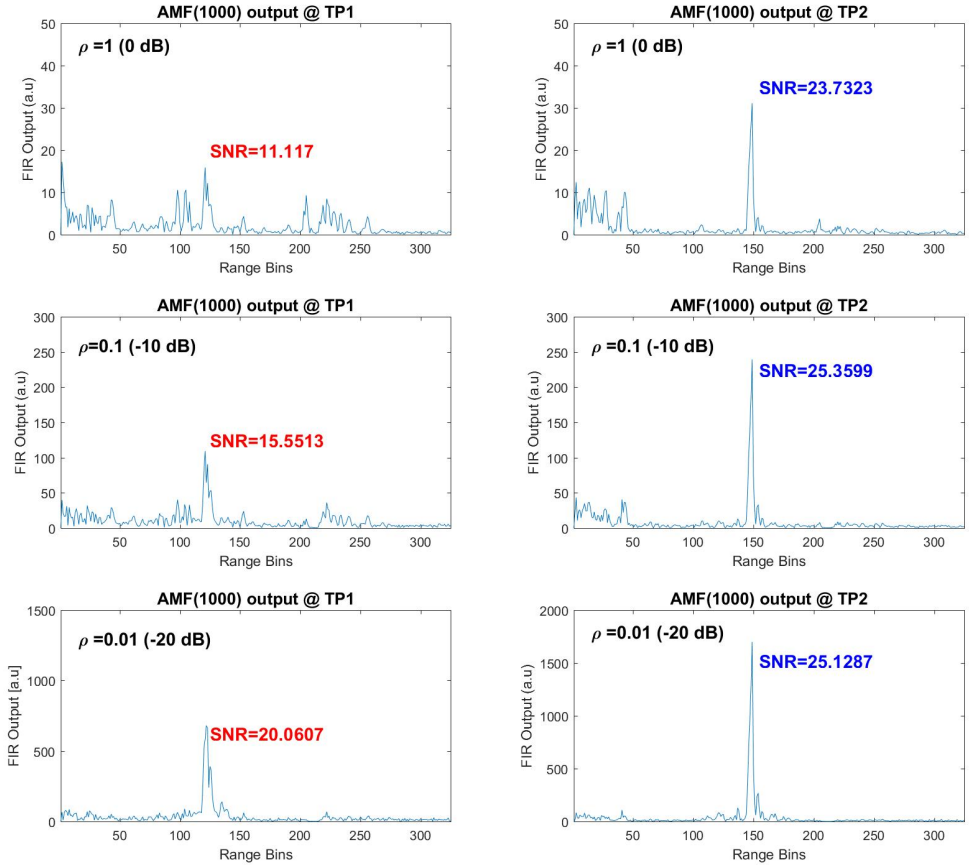


Figure 6.11. AMF(1000) outputs for TP1 and TP2 signals with for varying DL levels.

The last step with the AMF method was dedicated to the testing of the weak signal conditions. In order to meet the severe noise case, the test distances from the FUT were increased step by step where the acoustic vibrations on the relevant range bins became hardly seen by the operator on the system monitor. The system parameters were set to display the differences of averaged measured ϕ-OTDR traces with an averaging number ranging from 10 to 100, which was adjustable in real time by the operator. After the desired conditions were met, the tests were repeated to record the raw data without any averaging or any other preprocessing. The recorded raw data shown in Figure 6.12 below were gathered seven days after the above-presented recordings in almost similar environmental conditions. The first data shown in the left side of the figure are the recording of five sequential hammer hits at a lateral distance

of 20 m from the second manhole (eighth dataset), while the second data shown on the right side of the figure are the recording of another five sequential hammer hits gathered at a lateral distance of 25 m from the same manhole (ninth dataset).

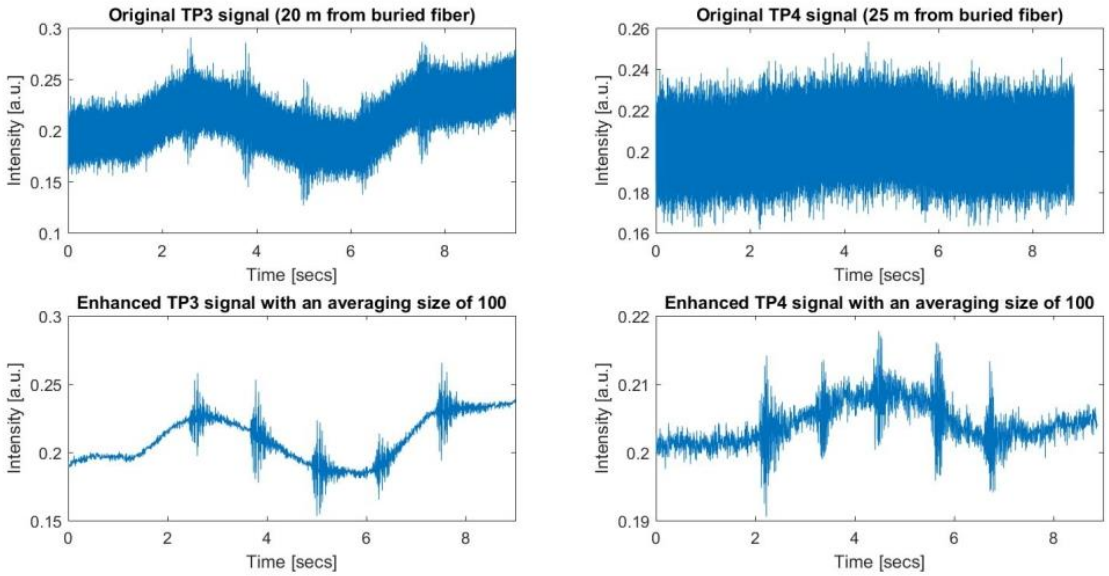


Figure 6.12. Raw and enhanced 147. channel data for TP3 and TP4.

The upper row presents the original record, while the lower row shows the processed versions of the same data with an average size of 100. As can be seen in the upper right plot, the hammer hits recorded at 25 m were highly obscured by noise that can be considered as a very weak signal condition since these signals became visible only when sufficient averaging was applied. Before applying the AMF to the whole raw data record, it was considered first to identify the time instants and the durations of the hit signals. After the start and the durations were identified for every hammer hit on the enhanced signal, these time indices and duration values were used to test the output of the AMF for all H_1 conditions. In order to compare the AMF with the conventional averaging techniques, the fast-time SNR was used as a figure of merit as mentioned before. The signal range bin set was collected with $k \in \{138, ..., 156\}$, while all other range bins along the "effective" FUT axis were taken for computation of the noise energy. Thus, the noise range bin set is the union of neighboring bins $\{110..., 137\}$ and $\{157, ..., 180\}$.

The noise only condition and the five different signal presence conditions were tested with three different FIR sizes when the steering vector was tuned to 50 Hz. The results obtained for the eighth dataset are given in Figure 6.13 and the comparison of SNR values with the SAM is listed in Table 6.3. The original raw data were processed with three different FIR sizes of 500, 1000 and 2000. The SAM method was applied with various pairs of (M, L) values, (i.e., SAM(M, L)) here M denotes the averaging block size and L is the lag value used in sequentially differentiating the ϕ-OTDR traces. After several trials with the eighth dataset, the relatively high SNR results obtained via SAM(100,50) and SAM(200,15) were selected for comparison, which are given in Table 6.3. For the proposed AMF method, the results show that significant SNR improvement is achieved with all selected FIR sizes, but a clear superiority to SAM(M, L) is achieved when N is increased over 1000.

The AMF outputs and the SNR values computed for the last data are shown in Figure 6.14 and listed in Table 6.4, respectively. For the SAM(M, L) method, after several trials with various average size and lag values, it was observed that the best SNR performance was achieved with SAM(200,15). The noise contamination for these data is more severe compared to the previous dataset, and significant noise reduction performance is achieved when N is selected over 2000 for the AMF method. More than 10 dB of SNR values can be easily obtained by increasing the FIR size over 3000 with penalties in processing times.

Besides the computed fast-time SNR values, the performance of the proposed AMF approach in comparison with the SAM(M, L) method for the weak vibrating signal condition can be better observed by having a closer look at their outputs on the fast-time. For example, for the detection of all five hammer hits in the 8. dataset, the acoustic vibration profiles obtained from both SAM(100,50) and AMF(5000) processings are presented in Figure 6.15 and Figure 6.16, respectively. It can be easily seen from Figure 6.15 that the differentiated intensity profile obtained from SAM(100,50) does not provide clear visibility for the hitting activities.

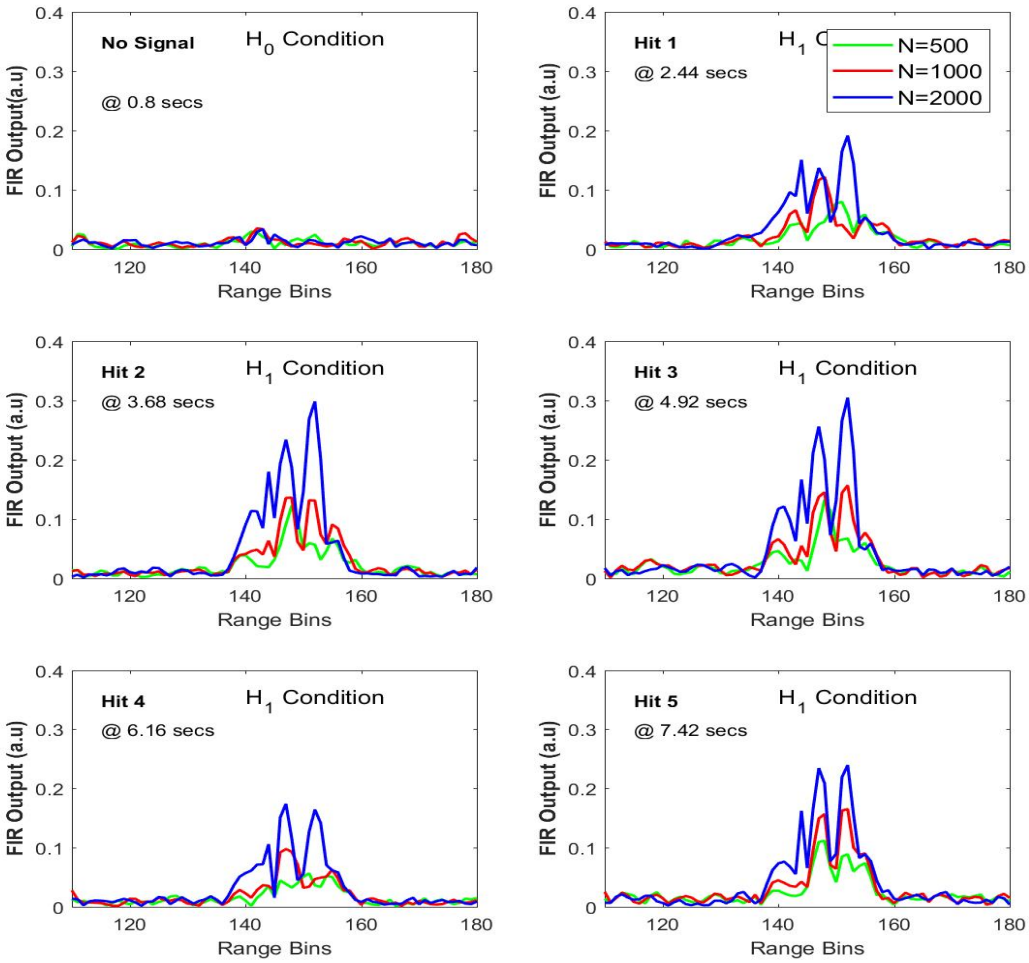


Figure 6.13. Acoustic vibration profiles along the FUT axis after AMF processing with three different N values for the 8. dataset.

Table 6.3. SNR results with $f_p = 50$ Hz at TP3.

Method	**Hit-1**	**Hit-2**	**Hit-3**	**Hit-4**	**Hit-5**
SAM(100,50)	9.52	10.39	11.69	8.82	17.71
SAM(200,15)	12.79	7.22	8.71	10.19	12.40
AMF(500)	11.52	13.17	12.26	10.77	12.27
AMF(1000)	12.12	15.53	14.65	13.59	15.21
AMF(2000)	17.75	23.22	20.77	18.47	18.99

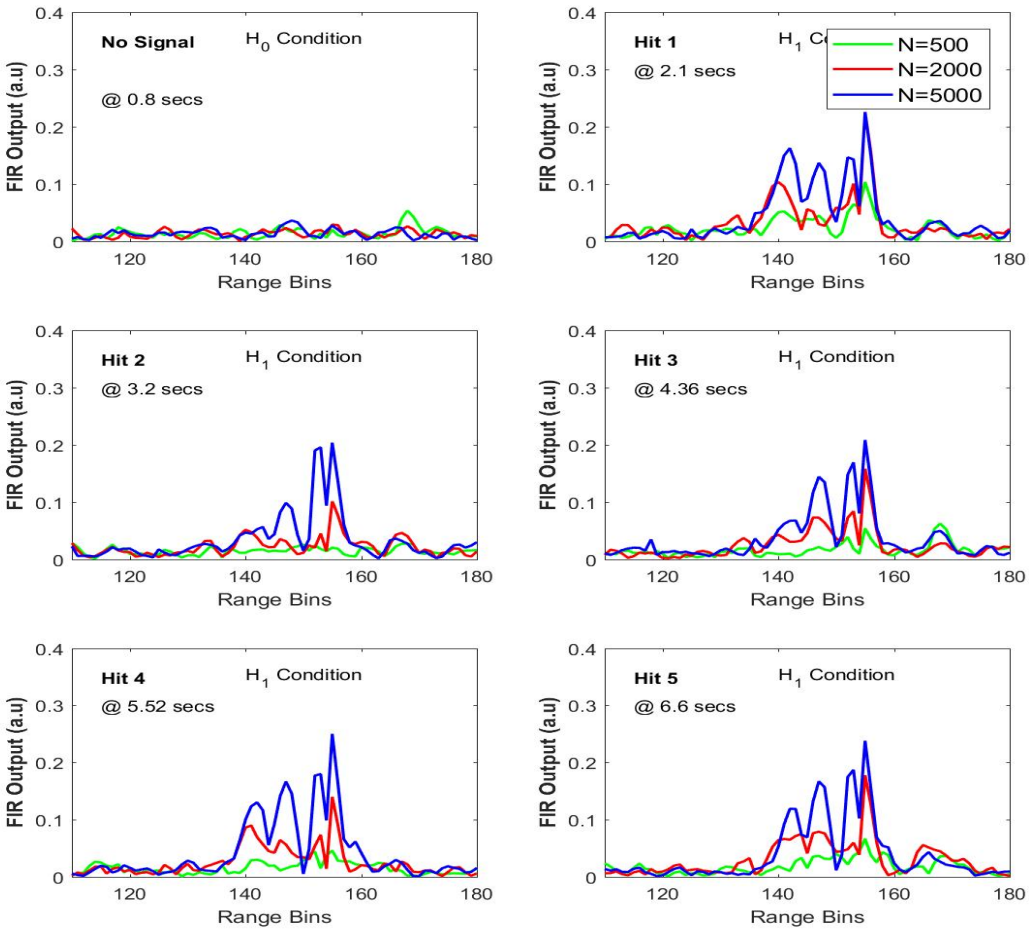


Figure 6.14. Acoustic vibration profiles along the FUT axis after AMF processing with three different N values for the 9. dataset.

Table 6.4. SNR results with $f_p = 50$ Hz at TP4.

Method	**Hit-1**	**Hit-2**	**Hit-3**	**Hit-4**	**Hit-5**
SAM(100,50)	8.13	2.79	4.47	4.55	0.35
SAM(200,15)	9.72	4.47	10.70	10.50	7.35
AMF(500)	8.17	0.40	1.17	5.61	5.91
AMF(1000)	11.81	5.10	5.98	8.37	6.18
AMF(2000)	11.81	6.36	11.55	12.42	10.69
AMF(5000)	14.77	14.35	13.94	15.45	15.93

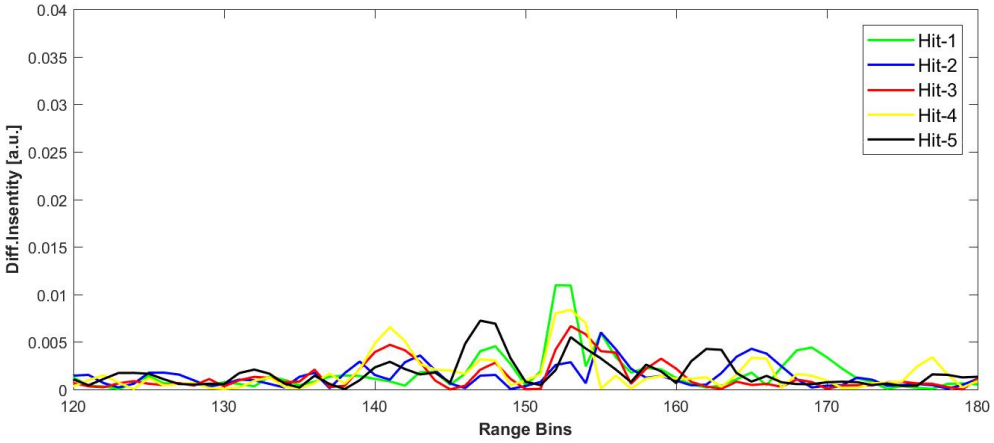


Figure 6.15. SAM(100,50) processing results for the five hammer hit durations of the 9. dataset.

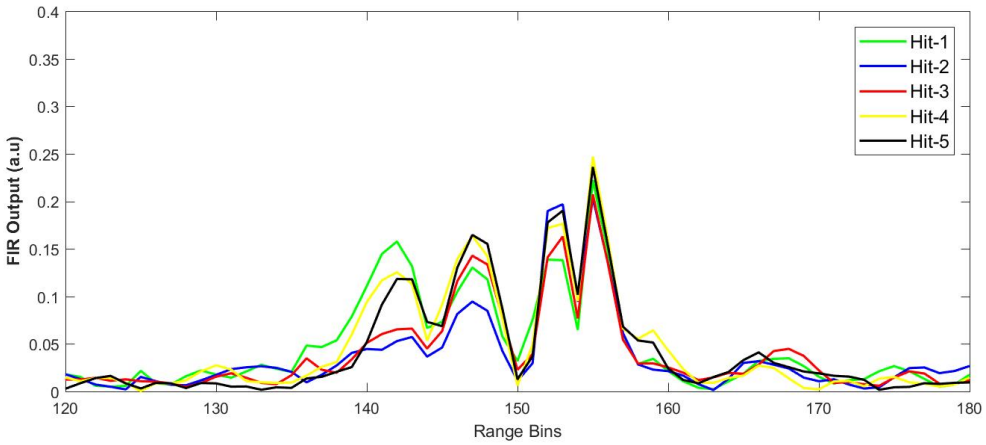


Figure 6.16. AMF(5000) processing results for the five hammer hit durations of the 9. dataset when the FIR is tuned to 50 Hz.

Figure 6.16 is the processed version of the same dataset with the AMF approach with an FIR size of 5000 and $f_p = 50$ Hz. As the noise background is highly suppressed, the acoustic vibrations distributed along the FUT axis are better visualized, and all of the five hammer hits designated by different colors in Figure 6.16 are clearly detectable. The training region size selected during processing was 20 (i.e., $L_s = \{231, 232, ..., 249, 250\}$).

6.3.2. MED Test Results

To avoid redundancy, the tests with the MED approach were mostly focused on the detection of weak vibrations and comparison with the AMF technique. For this purpose, the 8. data set was selected and subject to the MED approach with varying segmentation choices. The largest eigenvalue plots were computed for all the time segments covering the instants of all of the five hammer hits. The 1. hammer hit signal was analyzed with six different segmentation choices which are presented in Figure 6.17. The fast-time SNR values were computed for the range bins between 120 and 180 where the acoustic vibrations induced on the range bins between 138 and 156.

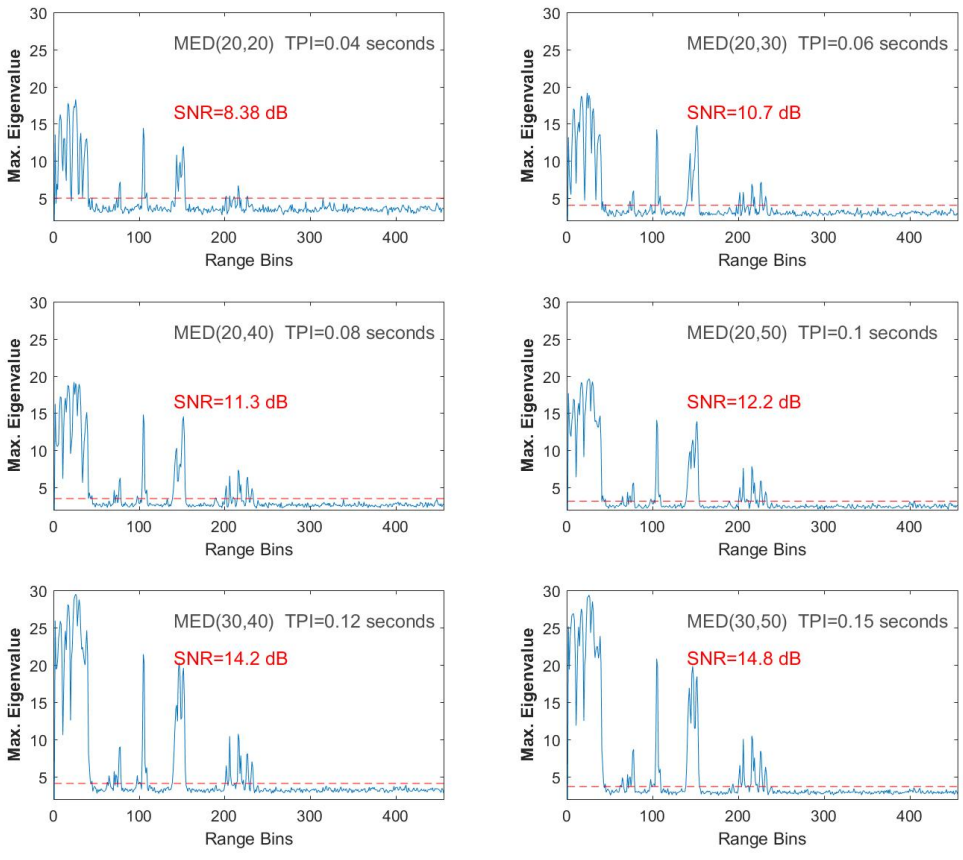


Figure 6.17. The MED test results at a time segment including the 1. hammer hit at 20 m distance from the FUT.

As we did with the AMF method in this chapter, we are interested in the range bins roughly between 100 and 200 since the other range bins are due to the vibrations in the buildings, sensitive manhole locations along the FUT route, etc. It is clear that the hammer hit activity which is 20 m away from the FUT route is clearly detectable for all TPI and segmentation choices. A huge amount of processing were executed for various (M, L) pairs. Instead of plotting the other MED outputs and associated range profiles for the other hammer hit signals, the fast-time SNR values were computed and summarized for various some selected MED(M, L) choices in Table 6.5. It is very clear that the locations of the peaks observed with the MED method are in agreement with AMF outputs for all range bins. Hence, two AMF test results were included in the last rows of Table 6.5 for comparison. Additionally, the average processing time required for a single TPI was computed for both the AMF and the MED methods. The average processing times were computed in a laptop computer running MATLAB R2017a with Intel® Core i7 CPU and 16 GB RAM.

Table 6.5. SNR [dB] comparison for MED and AMF methods with varying TPIs.

Method	**CPU Load [s]**	**Hit-1**	**Hit-2**	**Hit-3**	**Hit-4**	**Hit-5**
MED(20,20)	0.42	8.37	8.39	9.03	7.19	10.18
MED(30,30)	0.44	12.97	13.45	13.29	10.42	10.19
MED(40,50)	0.45	15.57	16.83	16.82	14.25	16.24
MED(50,60)	0.48	17.41	17.89	17.94	15.32	17.53
MED(60,60)	0.54	18.11	19.15	18.99	16.06	17.96
MED(60,70)	0.57	18.21	19.17	19.09	16.22	18.32
MED(60,80)	0.57	18.31	19.32	19.46	16.24	18.63
MED(80,100)	0.62	18.71	17.81	19.81	16.70	18.96
AMF(1000)	1.7	12.12	15.53	14.65	13.59	15.21
AMF(2000)	5.70	17.75	23.22	20.77	18.47	18.99

From the results presented in Table 6.5, it can be deduced that the MED outperforms the AMF when the computational costs are of concern. SNR values above 20 dB with the AMF can only be achieved for $N > 2000$ at the expense of increased computation time. On the other hand, high SNR rates close to 20 dB can be achieved with the MED technique with almost 10 times less computational power. As an example, although the TPI duration of MED(80,100) is 4 times longer than the one in AMF(2000), the required time to process all the FUT profile is about 0.6 seconds while it takes more than 5 seconds to process the whole range profile with the AMF for $N = 2000$. These results only give a general idea about the performance of the two methods because the total processing time required in DAS applications depends on several other parameters or processing tasks which are out of the scope of this research.

As examples for the acoustic vibration profiles, the MED processing results for detecting the 2. hammer hit by four different MED(M, L) choices are presented in Figure 6.18 below. The increase in the visibility of the acoustic vibrations is clearer within this zoomed region of the FUT. The increase in the fast-time SNRs correlated with the increased TPIs are already summarized in the first four rows of the 4. column of Table 6.5..

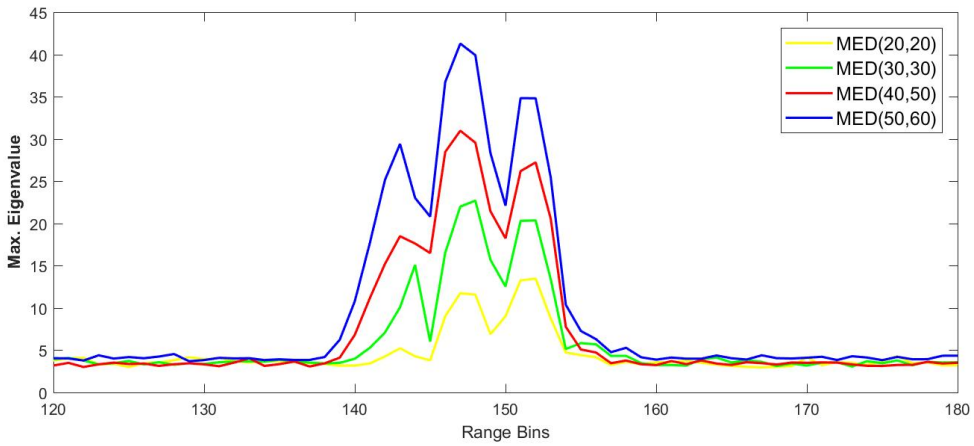


Figure 6.18. The acoustic vibration profiles detected by the MED processing at TP3 for the time segment including the 2. hammer hit.

6.4. Applications at Test Site-2

The AMF and MED methods were tested with buried fibers underground in the previous section. In this section, we consider an aerial configuration where the FUT is installed on a fence. Since the fence location is parallel to both a vehicle road and a railway along most of its route, it provides the opportunities to detect physical activities such as vehicle and train motion. The FOTAS system which is located in the laboratory was connected to the new FUT and several experiments were executed. Test data were mostly recorded for 2 main test scenarios: hammer hits on the ground surface close to the fence and during a vehicle passing such as a train or a car. The results of one dataset recorded during hammer hit activity and two datasets recorded during the vehicle passing (train and a car) are reported here. The system parameters used for these 3 datasets are summarized in Table 6.6.

The FOTAS GUI was set up for proper data collection before the tests: the maximum range for data recording was fixed to 10 km, but the PRF and the fast-time sampling rates were varied during the experiments. The total number of samples along the 10 km range was 570 for the first two data sets used for detecting the vehicle motion. Thus the synthetically enhanced resolution due to oversampling was 10 km/570= 17.54 m. For these two data sets, the data collections were started as the train or the car passed the southern border. The end of the FUT was observed to be at the 377. range bin and the fence configuration starts at the 147. range bin (i.e., at 2510 m) as can be seen from the averaged ϕ-OTDR frames with an average size of 100. It is also obvious that the background noise of the aerial section is significantly different from the buried section.

Table 6.6. Summary of the test signals used at test site-2.

Data	**Frames**	**Duration[s]**	**Signal**	**Dist. from FUT [m]**
Dataset-10	213394	42.7	Train Pass	15 - 25
Dataset-11	250531	125.3	Train& Car Pass	10 - 25
Dataset-12	106710	21.34	Hammer Hits	2

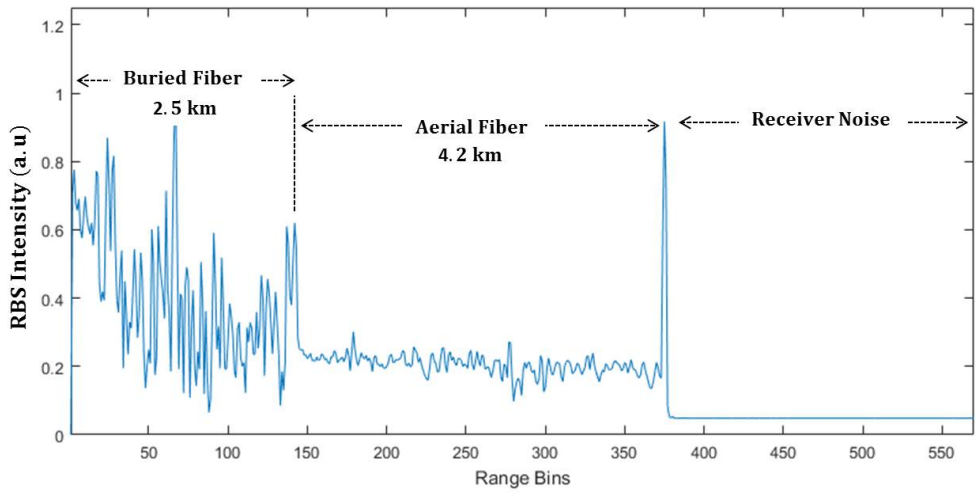


Figure 6.19. The normalized FOTAS range profile of the second test site with 100 OTDR frames averaged. The aerial background is significantly different from the buried section.

It was observed during the measurements that the amplitude fluctuations in the middle section of the FUT were higher than the fluctuations associated with the buried section of the FUT. This observation inherently verified that the clutter or the environmental interference impact was much more severe than the noise considered in the buried case.

In the second data set, in addition to the train pass, a car has also started a short cruise with a speed of 40 km/h in reverse direction of the train on the road. The objective of this experiment was to verify if two different physical activities at the same time could be detected and tracked with the AMF and the MED techniques. In the last data set, digging activities which were executed at two different locations along the FUT were recorded. During the experiments, two different groups were started to dig at the same time at the locations which are approximately 300 m apart. The first hammer hit activity was started at the beginning of the aerial configuration – which is at 2600 m - while the second test group worked at around 2900 m location near the road and 2 meters close to the fence. The FOTAS system parameters are summarized in Table 6.7.

Table 6.7. Test signals and associated FOTAS system parameters used.

Sys. Component	**Symbol**	**Dataset-10**	**Dataset-11**	**Dataset-12**
Pulse Frequency	PRF	5	2	5
Pulsewidth [ns]	τ	492		246
Spatial Resolution	Δz	50		25
Enhanced SR	N/A	17,54		8,77
Tot. Range Bins	K	570		1140
Vib. Locations	l_i	2600 m – 4000 m		2650-3000
Vib. Range Bins	p_i	$i \in \{148, ..., 230\}$		$i \in \{300, ..., 342\}$.
Length of FUT	L_{FUT}	10000 m		

The range profiles obtained at the output of the AMF and the MED techniques are presented together in this section. Figure 6.20 shows two sample results of the AMF and MED processing for the train pass of the 10. data set. At the beginning of the 10. seconds, 100 samples of raw ϕ-OTDR was subject to AMF with a fixed diagonal loading of 10 dB with a steering vector tuned to 150 Hz. At the AMF output the background is suppressed significantly and the train is located at 2.65 km as can be seen from the upper plot of Figure 6.20, presented by the blue lines. 30 seconds later the AMF processing locates the train approximately at 3.65 km with a clear peak observance as seen in the same plot with the red lines. Since the difference is approximated by 1 km in 30 seconds time interval, the speed of the train is estimated as 120 km/h roughly.

One sample MED processing of the same record with $M = 20$ and $L = 30$ is shown in the lower plot of Figure 6.20 in a similar manner. The MED(20,30) test result at the 10. seconds which is presented by the blue line is consistent with the AMF processing for the same time interval since the peak is observed at the same range bin location. The same consistency is observed for the MED(20,30) processing at the 40. seconds where the acoustic vibrations are spread among several range bins around 3.6 km location. The varying distance of the railway to the fence is a possible source for this change in the acoustic vibration energy detected at different locations.

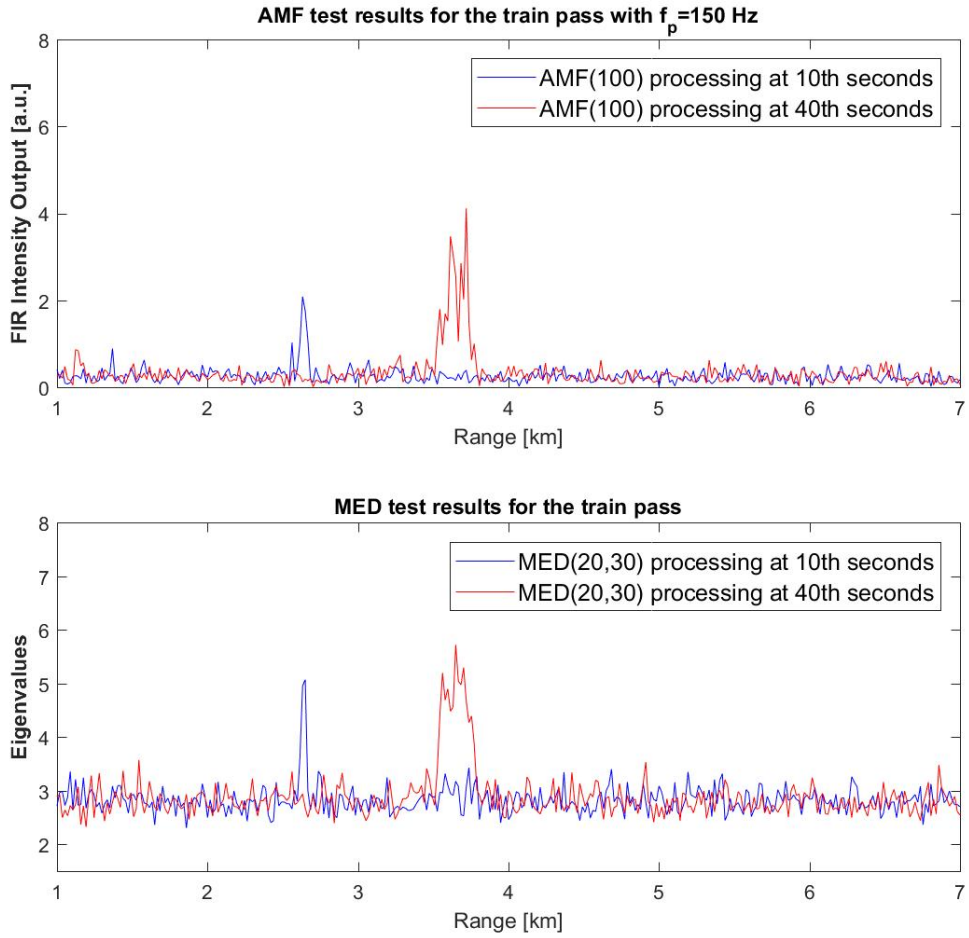


Figure 6.20. AMF and MED processing results for the train pass at 10. and 40. seconds of the record.

The AMF(100) with $f_p = 150$ Hz and $\rho = 10$ dB settings, and MED(20,30) processing were sequentially repeated from the beginning to the end of the record with non-overlapping time segments. The results are presented in waterfall format in Figure 6.21. The first two rows present the AMF and MED results without any preprocessing of the data. It is clear that the two-dimensional processed MED results are quite noisy when compared with the AMF(100) output shown in the upper row. The plot in the lowest row is the output of the MED(20,30) processing when the raw-data is high-pass filtered with a cut-off frequency of 20 Hz before applying the MED approach. If the tracking of train motion is of concern then it can be deduced that appropriate pre-filtering enhances the MED output.

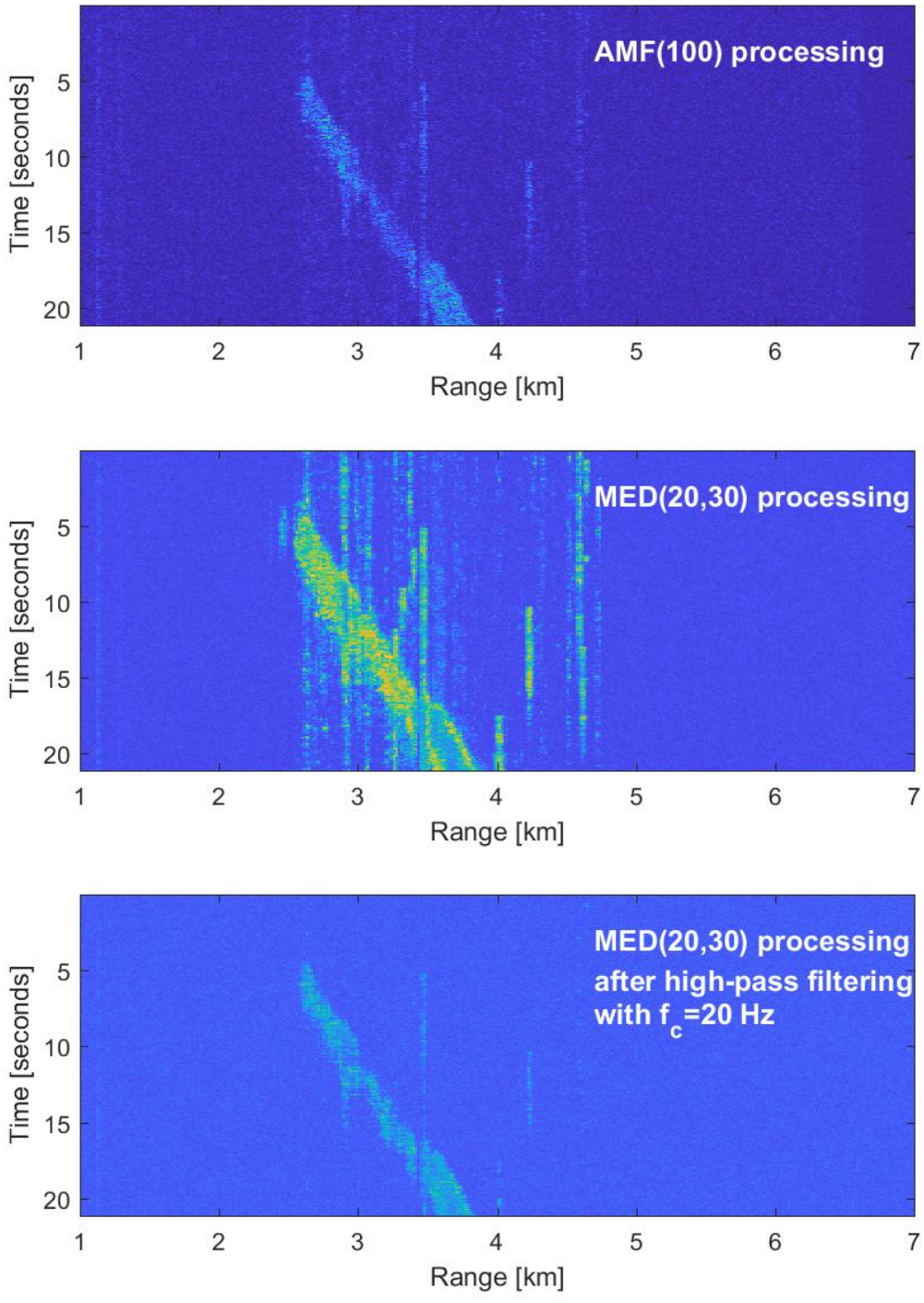


Figure 6.21. Waterfall presentation of AMF and MED processing of the whole data record for the train pass of the 10. data set.

The 11. data set was tested in a similar manner where both car and train motions were to be detected at the same time. During AMF(100) test results with the steering vector tuned to 150 Hz and DL level set to 10 dB are shown in the upper plot of the Figure 6.22. The results of the AMF processing did not yield clearly visible peaks as observed in the previous data set. Another interesting phenomenon was observed with the MED(20,30) test where the aerial section of the FUT exhibited high amplitude fluctuations as can be seen in the lower plot of Figure 6.22, which obscures the vehicle related vibrations in the range profile. To observe the impact of the background noise better, the waterfall processing was employed and observed as shown in Figure 6.23.

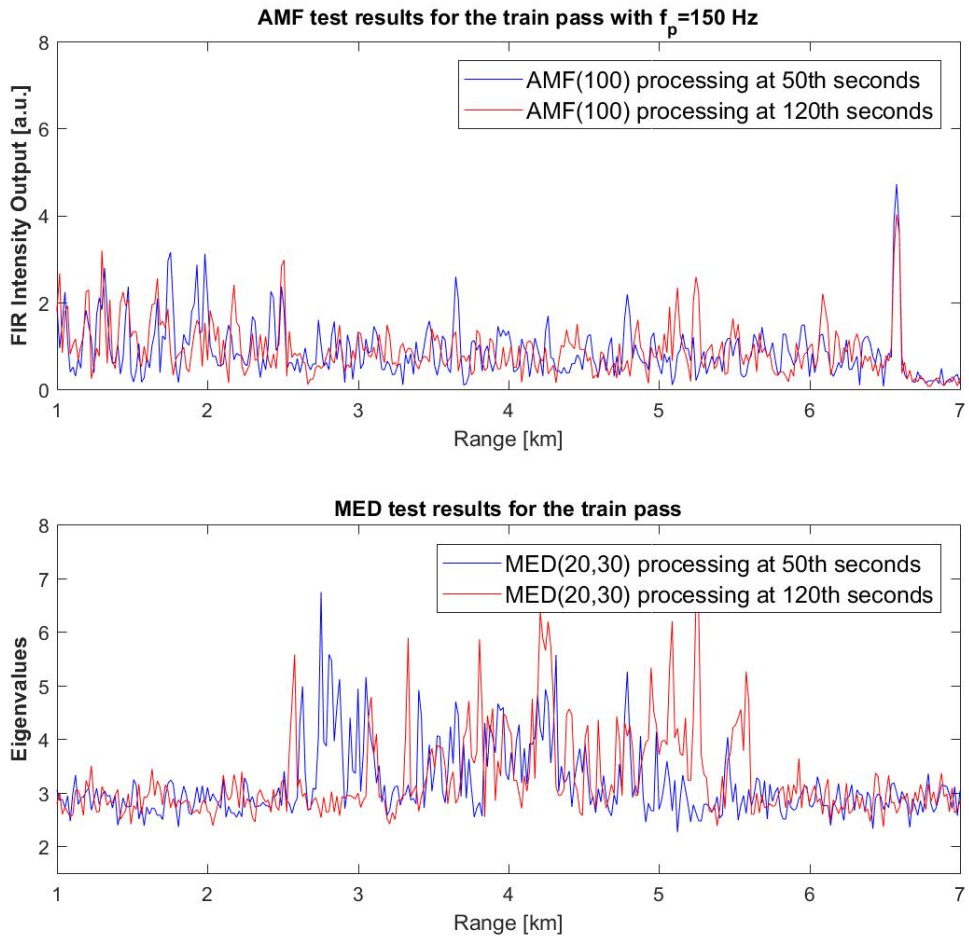


Figure 6.22. AMF and MED processing results for the vehicle passes at 50. and 120. seconds of the record.

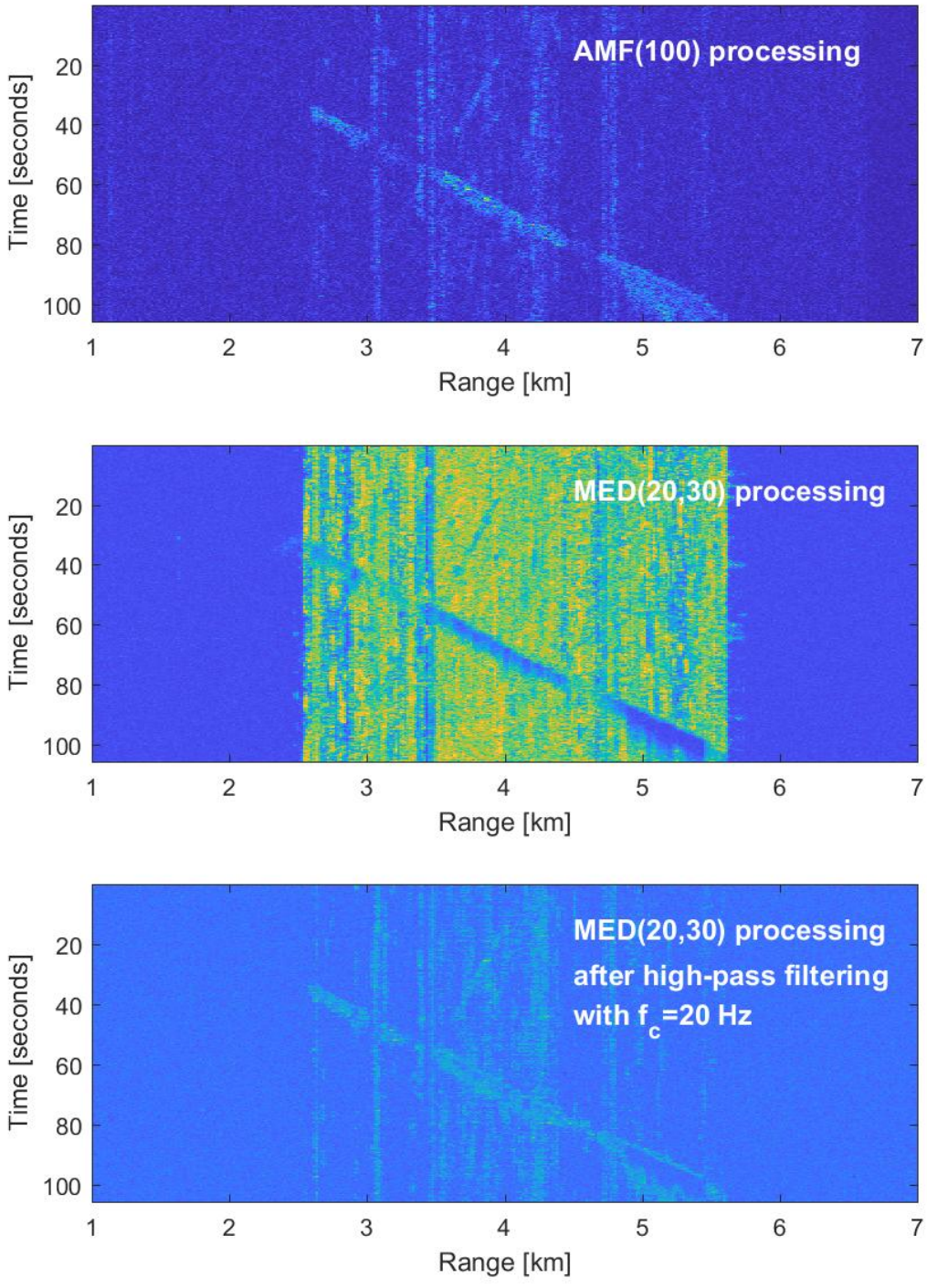


Figure 6.23. Waterfall presentation of AMF and MED processing of the whole data record for the vehicle passes of the 11. data set.

It is obvious that the waterfall format is better for visualizing the results and the acoustic signatures in a dynamically changing environment. Additionally and explicitly, the AMF method seems to better cope with the background noise due its inherent band-pass structure as can be seen from the upper plot of Figure 6.23. Apart from the signature of the high-speed train which is very clearly observed, the signature of the car moving in the reverse direction was also detected between 20. and 40. seconds in the temporal axis, and between the 3.5 and 4.0 km locations in the range axis.

The highly fluctuating acoustic vibrations of the aerial section are still very dominant after the MED(20,30) processing of the raw data, as shown in the middle plot of Figure 6.23. Without any filtering, the vehicle motions are embedded in the strong interference but they were exhibited very differently by the waterfall display from the previous results obtained with the 10. data set. The main suspect for the strong vibrations observed in this dataset was the windy weather as it probably induced additional vibrations on the fence which are easily detected by the MED technique.

The last plot of the Figure 6.23 is dedicated to the filtered version MED(20,30) processing. The raw data was subject to the same high-pass filter with a cut-off frequency of 20 Hz. When compared with the filtered version of the previous data set the train signature is more distorted with background noise, but the pre-filtered MED technique again performs better than its unfiltered version. From the slope of the line associated with the train pass, the speed of the train was estimated as 130 km/h. From the slope of the car signature, the speed of the car was roughly estimated as 37 km/h which is close to the actual cruise speed of the car during the experiment.

The last experiment was conducted with two test staff located 300 m apart from each other. One team was located at the beginning of the aerial fence configuration which approximately corresponds to the 2.65 km (TP5) in range, and the other staff was located at around 2.95 km (TP6). Both staff was requested to start digging at the same time but randomly, stop for a while a few seconds later, and then start the digging the soil for a few seconds more. This experiment can also be considered as the

counterpart of the multiple vibration case studied in the previous chapters. The AMF was applied for the 100 samples of ϕ-OTDR data at the beginning of the record and at 4.8 seconds (with $N = 100$, $f_p = 150$ Hz and $\rho = 10$), respectively. The results are shown in the upper plot of Figure 6.24. The peaks expected at around 2.65 and 2.95 km locations for the H_1 case are visible as can be seen in the plot depicted by the solid red line. The H_0 (null case) was also verified as no vibration peaks were observed as shown by the solid blue line.

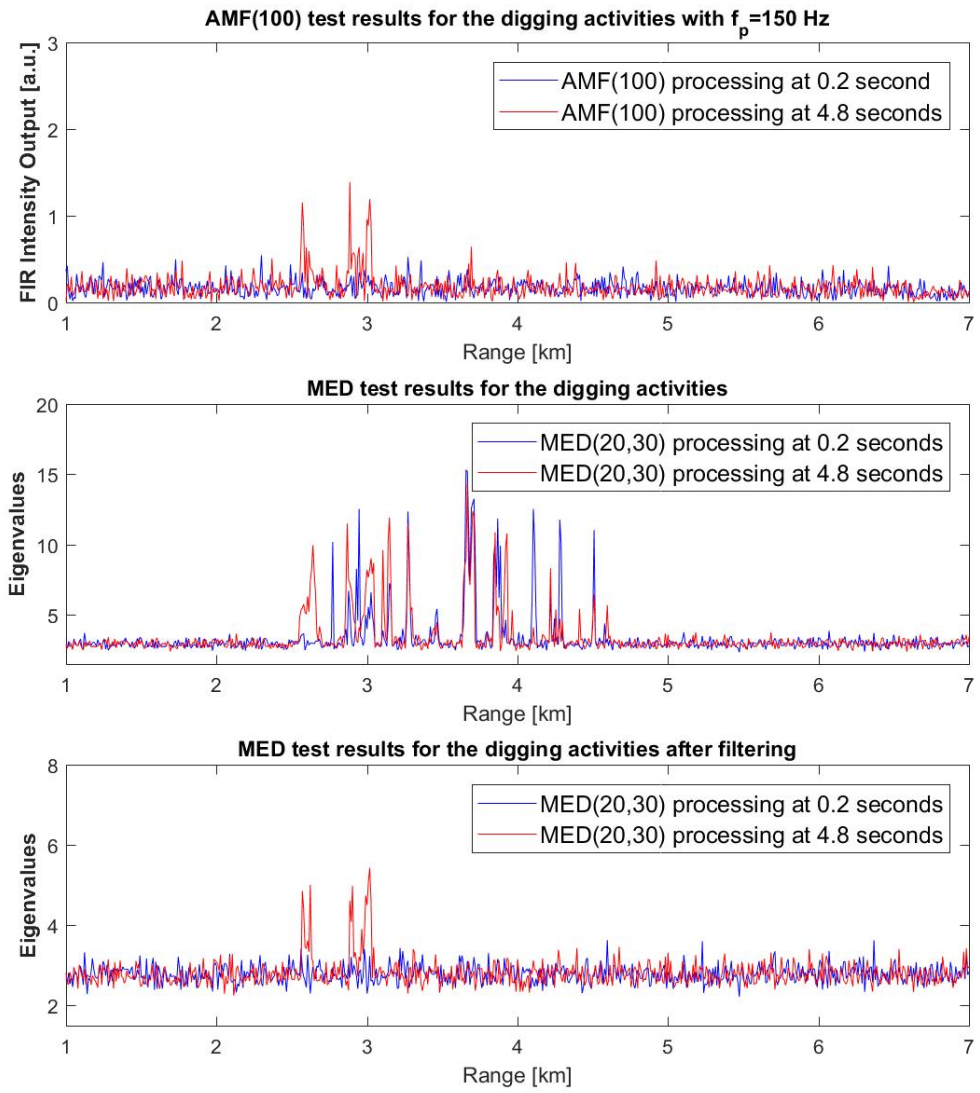


Figure 6.24. AMF and MED processing results for the vehicle passes at 0.2 and 4.8 seconds of the record.

The MED(20,30) test results are shown in the middle and lower plots of Figure 6.24 for the non-filtered and filtered cases, respectively. It is explicitly clear that the MED processing of the non-filtered ϕ-OTDR traces yields spurious responses over the entire aerial section of the FUT. The high sensitivity problem due this section was compensated again by high-pass filtering the data first and then applies the whitening transform to the filtered data before proceeding with the MED technique. The resulting range profile of this pre-filtering and whitening preceded MED(20,30) is shown in the last plot of Figure 6.24. The result of pre-filtered MED(20,30) is consistent with the result of AMF(100) tuned to $f_p = 150$ Hz.

The above mentioned AMF(100) and MED(20,30) were sequentially applied to the whole record for 2-D processing and visualization of the results as shown in Figure 6.25. The AMF(100) results are in good agreement with the start, break and re-start points of the conducted digging experiment at the two test points. On the other hand, it is quite difficult to distinguish these activities from the interfering vibrations points observed at the other range bins after the MED(20,30) processing, as depicted in the middle plot of the Figure 6.25. Although the start of the activity is clear, the break and re-start points were obscured and not clear. Bu after pre-filtering and whitening stages, the MED (20,30) results are significantly improved as shown in the last plot of Figure 6.25. The two activities started at two test points are again very clearly discriminated and the silence of other locations is clarified.

6.5. Conclusions

The applicability of the AMF and MED techniques for DAS applications were validated by field tests in this chapter. A huge amount of data has been collected with numerous tests executed in TÜBİTAK Gebze campus since March 2017. Due to the limited space of this document, sample test results were selected for the presentation of the efficiency of the proposed methods. Both buried and aerial fiber configurations were tested during the field experiments.

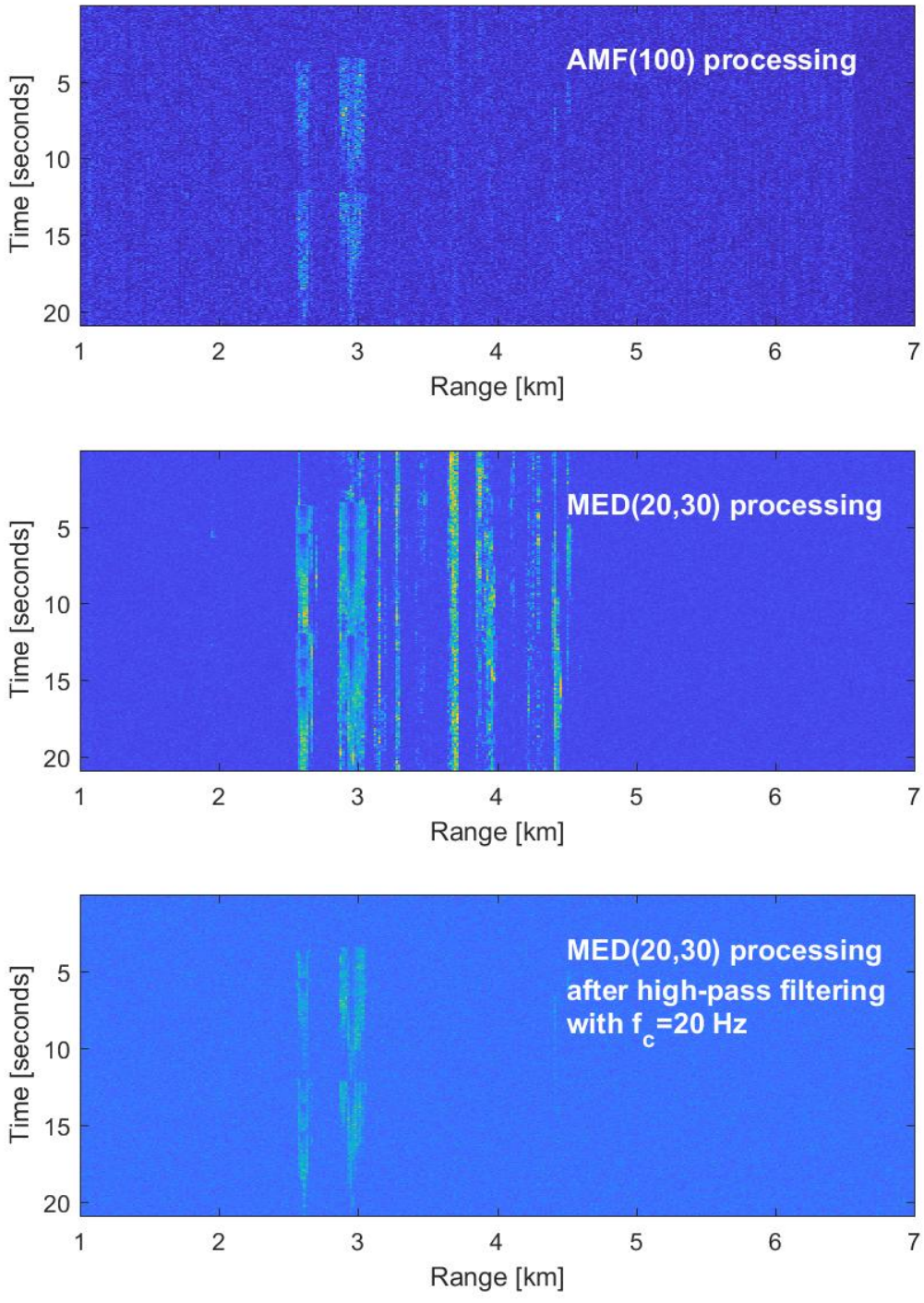


Figure 6.25. Waterfall presentation of AMF and MED processing of the whole data record for the digging activities of the 12. data set.

All field tests yielded the same results, and the vibration energy is better visualized when the TPI duration is increased for both methods. For the AMF technique, we can say that the longer the FIR size interval, the better the speckle reduction is. It is very clear from the test results that an FIR size of at least 1000 seems to be mandatory to reach a minimum SNR level of 10 dB for weak signal detection. On the other hand, when we are only concerned with the real measurement data, the processing time and the computational load will increase with increasing N; it will not be more beneficial to use very large values of N for real-time applications. The same situation is also valid for the MED technique since the detection performance is improved with the increased TPI durations. Some interesting results were obtained for the $\kappa = 1$ case where the spurious vibrations were significantly enhanced. It was observed that these vibrations were mostly below 20 Hz. The AMF inherently cancels all the vibrations that are out of its tuned band, but the MED technique is sensitive to all vibrations in the environment. Hence, it was experimentally verified that the performance of the MED technique can be improved with appropriate pre-filtering and pre-whitening stages.

It is worth emphasizing that no precautions were taken to avoid SOP changes in the setup. Thus, when much longer TPI observations were tested the fast-time SNR values were observed to be inconsistent. The inherent SOP changes due to the "coherent detection" employed in our DAS system were unavoidable. But this situation does not have a negative impact on the applicability of the AMF and MED techniques.

As a conclusion, the proposed AMF and MED approaches are superior to all conventional averaging techniques due to their capability of detecting the vibration location and frequency at the same time while preserving the frequency response of the system. Hence, both AMD and MED seem to be promising techniques which can be easily implemented for real-time DAS applications for remote detection of events along a fiber optic cable route.

7. CONCLUSION AND FUTURE WORK

Within the scope of this research, we have outlined two distinct signal processing techniques to improve the detection performance of ϕ-OTDR based DAS systems, and developed a real ϕ-OTDR based DAS system from scratch to validate these techniques in real-world conditions. Our general findings and the main contributions to the field of distributed fiber sensing are summarized as follows.

7.1. Sample Covariance of the Optical Backscatter

In this dissertation, two novel signal processing approaches for DAS systems were presented: the adaptive matched filtering (AMF) and the maximum eigenvalue detection (MED). Both methods utilize the covariance structure of the measured optical backscatter for the detection of events along the fiber route. The second order statistics are estimated from a limited number of measurements to compute the weights of a linear FIR filter in the AMF approach, while the sample covariance is used for computing the spectrum of eigenvalues in the MED technique. The former approach was borrowed from the classical optimum noise filtering problem, and successfully tailored for conditioning the outputs of the virtual microphones of the DAS system. Each acoustic channel is considered separately to filter out the speckle-like profile of the ϕ-OTDR traces. The FIR weight computation has an inherent adaptive structure which can adapt the DAS system to environmental changes if the sample covariance is updated on a regular basis.

In the latter approach, the received samples are segmented to form a data matrix and the largest eigenvalue of the measured covariance is computed to distinguish the vibration from the background noise. Eigenvalue based detection is a blind detection technique since it does not require the prior knowledge which the AMF is dependent on for optimum performance. It was demonstrated by extensive experimental work that both AMF and MED can be efficiently used for detecting the acoustic perturbations on the fiber cable; hence the physical events along a fiber route. The outputs of either

method can be used to detect and track vibrations with a predetermined threshold but the output range profile must be carefully processed to handle the false alarm rate (FAR) of the DAS system. In this regard, it is important to remark that both methods possess the constant false alarm rate (CFAR) property which is of importance to maintain the FAR for real applications. However, the CFAR property of the proposed approaches is only for "event detection" but not for "event classification". The classification of events detected for DAS application was deemed to be another research topic which should be focused separately and the proposed AMF and MED methods have pertained to the vibration detection problem herein.

It is worth emphasizing that the SCM based methods presented in this dissertation do not limit themselves to any specific DAS interrogation or detection scheme employed for analyzing the optical backscatter. Both AMF and MED techniques presented are still applicable methods that are independent of the optical and electro-optical components used in the DAS system. Although the baseline of the experimental system considered here was a lean system excluding many additional optical stages, the outlined covariance based techniques are directly applicable to any kind of optical backscatter measurements given that the mathematical model considered is realistic. Additionally, since the data handled in the experimental work was all intensity related, it is obvious that both AMF and MED will be directly applicable to direct detection based ϕ-OTDR systems as well.

7.2. Random Matrix Theory

One important topic visited in this research is the random matrix theory (RMT) which studies the eigenvalue distribution of random matrices. The SCM which is estimated from noise-only measurements is a Wishart random matrix which has been well-studied in the context of statistical signal processing. It has been analytically shown that the largest eigenvalue fluctuations can be described by the Tracy-Widom distribution. The recent results of the RMT which have been successfully exploited in radar and cognitive radio applications were also tested and experimentally verified in this context as promising tools for acoustic vibration detection. It was successfully

demonstrated that the maximum eigenvalue of the optical covariance matrix can be used to detect the acoustic vibrations along the fiber cable. In order to establish an appropriate method for adjusting the false alarm rate, the fluctuations of the largest eigenvalues of the covariance have been studied via RMT assistance. To the best of our knowledge, the experimental results presented in this dissertation are the first study that the optical Rayleigh backscatter measurements, when properly scaled, can be evaluated by the tools of the RMT. We have heuristically applied the outcomes of the two main theorems of the RMT: the Marcenko-Pastur and the Tracy-Widom. These two theorems seem to provide very useful tools in understanding the fundamental limits of signal detection embedded in noise. Surprisingly enough, there is no need to understand the deep theoretical basis for these theorems to apply them for a specific detection problem. Generating a data matrix and normalizing the data to have zero mean and unit variance will suffice before computing the eigenvalues of the measured data. If the computed eigenvalues of the measurements are similar to those computed for pure random matrices of the same size, then we can conclude that there is no coherency or correlations in the data. If the computed eigenvalues are well beyond the boundaries predicted by the RMT, then we can say that there is correlation in the data. This simple methodology in our research provided a compelling understanding of the RMT without going into too much theoretical work and handling our vibration detection problem in a practical way. Additionally, it also increased our motivation to improve our understanding of the theory for other signal processing applications.

7.3. Real DAS System Development

Significant effort was dedicated to working with real DAS data in the field. For this purpose, a system engineering approach was followed to develop a real system from scratch. Coherent detection based ϕ-OTDR system was developed in the electro-optics and laser systems laboratory of TÜBİTAK in the first term of the research. Desktop instrumentation was used to configure the necessary setup and data acquisition. After all proof-of-concept demonstrations were completed in the laboratory, the system was decided to be sectioned as two main subsystems: the optical interroga-

tor and the electrical signal analysis subsystem. The optical interrogator section was assembled in a 19-inch rack-mountable instrument which can also be operated as a standalone unit or tested with alternative desktop instrumentation. The RF instrumentation and the data acquisition hardware used during laboratory studies were later replaced by a single desktop instrument as the electrical signal subsystem. This was achieved by modifying a software defined radio (SDR) device which is used to monitor the RF spectrum.

Besides the SDR's capability of analyzing the signals in the RF spectrum like a conventional spectrum analyzer, it includes a high-speed ADC and a Xilinx Virtex-6 FPGA. Thanks to its reconfigurable hardware, the embedded software of the device was modified to handle the electrical output of the interrogator. An additional RF modulator component was developed and included in the RF section of the SDR to drive the AOM of the interrogator which operates with the pulse triggered from the FPGA. A graphical user interface was written in C language for the operation of the DAS system. The above-mentioned development process was supported by the FOTAS project under the contract SAP00253 of TÜBİTAK BİLGEM, which was commenced in February 2016 by the project proposal of the author.

The first field deployment and tests of the FOTAS systems were done in two different cities in the countryside in 2016. During these field tests, the maximum range and the lateral sensitivity of the system were tested at two different military sites with buried intranet fiber cables. The maximum detection range and the lateral sensitivity of the system were observed to be approximately 50 km and 2 m, respectively. By the end of 2016, the lateral sensitivity of the system for digging activities was improved to 25 m after slight modifications in the electrical subsystem.

Since 2017, the field tests of FOTAS and the new algorithms were limited to the confines of the TÜBİTAK Gebze campus site. First validation tests with the AMF method was completed in the middle of 2017, and the results were published in the Sensors journal. Real-data gathering and field tests are continuing to improve the performance of the FOTAS system.

By the author's initiative and the approval of the executive office of TÜBİTAK BİLGEM, a new contract was signed in 2018 between our industry partners SAMM Teknoloji A.Ş., and ESEN Sistem Entegrasyon to proceed with the industrialization of the research efforts and commercialization of FOTAS. In addition to these cooperative actions, the author also prepared another project proposal and submitted to TÜBİTAK Defence and Security Technologies Research Grant Group (SAVTAG) for the development of a military version of FOTAS for homeland security with an expanded consortium of four industry partners.

7.4. Future Prospects

There is still a long way to go to further extend the AMF and eigen-analysis based methods to improve the performance and the sensitivity of DAS systems for practical applications. The author has noted the following issues as new research fields, which all possess a great potential for integration of the adaptive and eigen-analysis based methods for further performance improvement in DAS systems.

7.4.1. Spatial Beamforming

One important topic is the beamforming concept, which can be efficiently utilized by considering the acoustic channels as an array of microphones. Beamforming is a widely used concept in radar and sonar array processing field and a great deal of interest is still devoted to robust detection and estimation of multiple targets under strong interference and jamming conditions. A similar situation naturally arises in the DAS application when we are focused on detecting and localizing the events for various purposes under strong interfering events in the field. The accurate estimation of an approaching or a static acoustic source might be of importance for PIDS applications. 2-D and 3-D processing of the acoustic fields will provide useful information especially for imaging of the seismic events. The AMF and the MED techniques were only applied in the temporal domain in this context. Spatial array processing will open the door for the application of several emitter location algorithms in a straightforward manner.

In this regard, the adaptation of the AMF and MED or the derivatives for the spatial case, and coupled with the popular sub-space based estimation methods will be one of the key topics to extend our research study.

7.4.2. Frequency Diversity Techniques

Adhering to the resemblance of the receiver to many radar counterparts, including the frequency diversity techniques which have not yet considered herein, both AMF and MED techniques reserve the integration capability in conjunction with much more different optical receiver and waveform topologies. The main philosophy actually relies on the simplest mathematical model that fits the application in hand. One such potential application is the pulse compression technique or the stepped-frequency approach which is popularly employed for radar and lidar systems. The processing of the backscatter by sweeping a large amount of frequencies enables several high-resolution spectral estimation methods which have not been studied for DAS in the open literature. The conventional linear frequency modulated (FM) waveforms or stepped frequency techniques are popular in radar field adhering to their high-range resolution capability. Since the SCM based techniques have not been considered in the frequency domain for the OTDR scheme, the FM interrogation methods were noted by the author as potential research topics for the integration of the SCM based approaches.

7.4.3. Machine Learning

The DAS system in this research was limited to the analysis of Rayleigh based ϕ-OTDR architectures. One of the main challenges of Rayleigh based DAS systems is their sensitivity to both strain and the temperature. The total influence of the two is unpredictable in advance which adds complexity to the system design and efficiency of any signal processing method. In case of a dispersive medium where the fibers are mounted or installed, the frequency selectivity of the AMF method will become a more complex issue to overcome, as it relies on the prior knowledge of the environment for optimum detection performance which is not met in most of the industrial applications. Spectral signatures of various vibrations of should be gathered or should they already

be available in advance for a reliable application in the field. This issue may not be so critical for SHM related use but it is the capability of a DAS system which makes it an indispensable choice for security related applications when it comes to deciding among several physical events. Although the alternative blind MED technique overcomes such uncertainties, as mentioned above, the covariance structure of the optical backscatter was exploited for the detection but not for distinguishing between several activities. This necessitates a training period of the system for optimizing itself for the specific application. Nevertheless, it is a known fact that the commercial DAS systems are deployed in the field assisted with long training and calibration periods to learn the environment they intended to work with. Hence, machine learning algorithms seem to be perhaps the most critical part of the DAS systems that the designers have to deploy for the final application. Given that the DAS is trained with extensive field tests under various environmental conditions and distinctive spectral features are gathered for targeted vibrations, the AMF and MED approaches will have more potential use, but they still need more detailed and deepened research. In this regard, the application of the AMF and MED techniques will be considered in conjunction with machine learning algorithms for real-world applications.

As a conclusion on the overall, distributed acoustic vibration sensing by means of fiber optic cables is gaining popularity in several industrial applications. The proposed AMF and MED techniques presented in this research contribute to the detection of various acoustic vibration sources for different DAS applications and expected to be efficient tools for the design and implementation of fiber optic sensor technologies.

REFERENCES

1. Hartog, A. H., *An Introduction to Distributed Optical Fibre Sensors*, CRC Press Inc, Boca Raton, USA, 2017.

2. Fang, Z., K. K. Chin, R. Qu and H. Cai, *Fundamentals of Optical Fibre Sensors*, John Wiley & Sons Inc, New Jersey, USA, 2012.

3. Bao, X. and L. Chen, "Recent Progress in Distributed Fiber Optic Sensors", *Sensors*, Vol. 12, No. 7, pp. 8601–8639, 2012.

4. Taylor, H. F. and C. E. Lee, "Apparatus and Method for Fiber Optic Intrusion Sensing", *US Patent 5194847*, 1993.

5. Shatalin, S. V., V. N. Treschikov and A. J. Rogers, "Interferometric Optical Time-Domain Reflectometry for Distributed Optical Fiber Sensing", *Applied Optics*, Vol. 37, No. 24, pp. 5600–5604, 1998.

6. Park, J. H., W. K. Lee and H. F. Taylor, "Fiber Optic Intrusion Sensor with the Configuration of an Optical Time-domain Reflectometer Using Coherent Interference of Rayleigh Backscattering", *Proc. SPIE 3555, Optical and Fiber Optic Sensor Systems*, 1998.

7. Juarez, J. C., E. W. Maier, K. N. Choi and H. F. Taylor, "Distributed Fiber-optic Intrusion Sensor System", *Journal of Lightwave Technology*, Vol. 23, No. 6, pp. 2081–2087, 2005.

8. Kimbell, J. F., *History and Analysis of Distributed Acoustic Sensing for Oilfield Applications*, M Sc. Thesis, Texas A&M University, 2013.

9. Tejedor, J., H. F. Martins, D. Piote, J. Macias-Guarasa, J. Pastor-Graells and S.-M. Lopez, "Toward Prevention of Pipeline Integrity Threats Using a Smart Fiber-

Optic Surveillance System", *Journal of Lightwave Technology*, Vol. 34, No. 19, pp. 4445–4453, 2016.

10. Peng, F., H. Wu, X.-H. Jia, Y.-J. Rao, Z.-N. Wang and Z.-P. Peng, "Ultra-long High-Sensitivity ϕ-OTDR for High Spatial Resolution Intrusion Detection of Pipelines", *Optics Express*, Vol. 22, No. 11, pp. 13804–13810, 2014.

11. Lindsey, N. J., E. R. Martin, D. S. Dreger, B. Freifeld, S. Cole, S. R. James, B. L. Biondi and J. B. Ajo-Franklin, "Fiber Optic Network Observations of Earthquake Wavefields", *Geophysical Research Letters*, Vol. 44, No. 11, pp. 11792–11799, 2017.

12. Li, Z. and Z. Zhan, "Pushing the Limit of Earthquake Detection with Distributed Acoustic Sensing and Template Matching: A Case Study at the Brady Geothermal Field", *Geophysical Journal International*, Vol. 215, No. 3, pp. 1583–1593, 2018.

13. Jreij, S., *The Feasibility of Using Distributed Acoustic Sensors in Surface Seismic Application*, M Sc. Thesis, Colorado School of Mines, 2018.

14. Duckworth, G. L. and E. M. Ku, "OptaSense Distributed Acoustic and Seismic Sensing Using COTS Fiber Optic Cables for Infrastructure Protection and Counter Terrorism", *Proc. SPIE 8711, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense*, 2013.

15. Daley, T. M., B. M. Freifeld, J. Ajo-Franklin, S. Dou, V. S. Roman Pevzner and, S. Kashikar, D. E. Miller, J. Goetz, J. Henninges and S. Lueth, "Field Testing of Fiber Optic Distributed Acoustic Sensing (DAS) for Subsurface Seismic Monitoring", *The Leading Edge*, Vol. 32, No. 6, pp. 699–706, 2013.

16. Riedel, M., C. Cosma, N. Enescu, E. Koivisto, K. Komminaho, K. Vaittinen and M. Malinowski, "Underground Vertical Seismic Profiling with Conventional and Fiber-Optic Systems for Exploration in the Kylylahti Polymetallic Mine, Eastern Finland", *Minerals*, Vol. 8, No. 11, pp. 538–558, 2013.

17. Hicke, K., M.-T. Hussels, R. Eisermann, S. Chruscicki and K. Krebber, "Condition Monitoring of Industrial Infrastructures Using Distributed Fibre Optic Acoustic Sensors", *25th Optical Fiber Sensors Conference (OFS)*, pp. 1–4, 2017.

18. Villaba, S. and J. R. Casas, "Application of Optical Fiber Distributed Sensing to Health Monitoring of Concrete Structures", *Mechanical Systems and Signal Processing*, Vol. 39, No. 1-2, pp. 441–451, 2013.

19. Leung, C. K. Y., "Fiber Optic Sensors in Concrete: The Future?", *NDT & E International*, Vol. 34, No. 2, pp. 85–94, 2001.

20. Taylor, H. F. and C. E. Lee, "Traffic Sensing and Monitoring Apparatus", *US Patent 7652245*, 2010.

21. Wang, Z., B. Lu, H. Zheng, Q. Ye, Z. Pan, H. Cai, R. Qu and Z. Fan, "Novel Railway-Subgrade Vibration Monitoring Technology Using Phase-sensitive OTDR", *25th Optical Fiber Sensors Conference (OFS)*, pp. 1–4, 2017.

22. Peng, F., N. Duan, Y.-J. Rao and J. Li, "Real-Time Position and Speed Monitoring of Trains Using Phase-Sensitive OTDR", *IEEE Photonics Technology Letters*, Vol. 26, No. 20, pp. 2055–2057, 2014.

23. Melvin, W. L., "A STAP Overview", *IEEE Aerospace and Electronic Systems Magazine*, Vol. 19, No. 1, pp. 19–35, 2004.

24. Edelman, A. and Y. Wang, *Random Matrix Theory and Its Innovative Applications*, 2014, `https://math.mit.edu/~edelman/homepage/papers/rmtia.pdf`, accessed in June 2019.

25. Barnoski, M. K. and S. M. Jensen, "Fiber Waveguides:A Novel Technique for Investigating Attenuation Characteristics", *Applied Optics*, Vol. 15, No. 9, pp. 2112–2115, 1976.

26. Koyamada, Y. and H. Nakamoto, "High Performance Single Mode OTDR Using Coherent Detection and Fibre Amplifiers", *IET Electronic Letters*, Vol. 26, No. 9, pp. 573–575, 1990.

27. Izumita, H., *Highly Developed Coherent Detection OTDR Technology and Its Applications to Optical Fiber Networks Monitoring*, PhD Thesis, Waseda University, 2008.

28. Adany, P., C. Allen and R. Hui, "Chirped Lidar Using Simplified Homodyne Detection", *Journal of Lightwave Technology*, Vol. 27, No. 16, pp. 3351–3357, 1990.

29. Yang, G., X. Fan, Q. Liu and Z. He, "Distributed Vibration Sensor Based on Space-Division Multiplexed Reflectometer and Interferometer in Multicore Fiber", *Journal of Lightwave Technology*, Vol. 36, No. 4, pp. 5764–5772, 2018.

30. Iida, D., K. Toge and T. Manabe, "High-frequency Distributed Acoustic Sensing Faster than Repetition Limit with Frequency-Multiplexed Phase-OTDR", *2016 Optical Fiber Communications Conference and Exhibition (OFC)*, 2016.

31. Kobyakov, A., M. Sauer and D. Chowdhury, "Stimulated Brillouin Scattering in Optical Fibers", *Advances in Optics and Photonics*, Vol. 2, No. 1, pp. 1–59, 2010.

32. Chen, W. and Z. Meng, "Effects of Phase Modulation Used for SBS Suppression on Phase Noise in an Optical Fibre", *Journal of Physics B: Atomic, Molecular and Optical Physics*, Vol. 44, No. 16, pp. 1–6, 2011.

33. Healey, P., "Fading in heterodyne OTDR", *IET Electronics Letters*, Vol. 20, No. 1, pp. 30–32, 1984.

34. Qin, Z., T. Zhu, L. Chen and X. Bao, "High Sensitivity Distributed Vibration Sensor Based on Polarization-Maintaining Configurations of Phase-OTDR", *IEEE Photonics Technology Letters*, Vol. 23, No. 15, pp. 1091–1093, 2011.

35. Juarez, J. C. and H. F. Taylor, "Polarization Discrimination in a Phase-sensitive Optical Time-domain Reflectometer Intrusion Sensor System", *Optical Letters*, Vol. 30, No. 24, pp. 3284–3286, 2005.

36. Juarez, J. C. and H. F. Taylor, "Field Fest of a Distributed Fiber-optic Intrusion Sensor System for Long Perimeters", *Applied Optics*, Vol. 46, No. 11, pp. 1968–1971, 2007.

37. Martins, H. F., S. Martin-Lopez, P. Corredera, M. L. Filograno, O. Frazao and M. Gonzalez-Herraez, "Phase-sensitive Optical Time Domain Reflectometer Assisted by First-order Raman Amplification for Distributed Vibration Sensing over 100 km", *Journal of Lightwave Technology*, Vol. 32, No. 8, pp. 1510–1514, 2014.

38. Martins, H. F., *Distributed and Remote Fiber Sensing Assisted by Raman Effect*, PhD Thesis, University of Porto, 2014.

39. Cedilnik, G., G. Lees, P. E. Schmidt, S. Herstrøm and T. Geisler, "Pushing the Reach of Fiber Distributed Acoustic Sensing to 125 km Without the Use of Amplification", *IEEE Sensors Letters*, Vol. 3, No. 3, pp. 1–4, 2019.

40. Liu, X., B. Jin, Q. Bai, Y. Wang, D. Wang and Y. Wang, "Distributed Fiber-Optic Sensors for Vibration Detection", *Sensors*, Vol. 16, No. 8, pp. 1164–1194, 2016.

41. Lu, Y., T. Zhu, L. Chen and X. Bao, "Distributed Vibration Sensor Based on Coherent Detection of Phase-OTDR", *Journal of Lightwave Technology*, Vol. 28, No. 22, pp. 3243–3249, 2010.

42. Zhu, T., X. Xiao, Q. He and D. Diao, "Enhancement of SNR and Spatial Resolution in ϕ-OTDR System by Using Two-Dimensional Edge Detection Method", *Journal of Lightwave Technology*, Vol. 31, No. 17, pp. 2851–2855, 2013.

43. He, H., L. Shao, H. Li, W. Pan, B. Luo, X. Zou and L. Yan, "SNR Enhancement in Phase-sensitive OTDR with Adaptive 2D Bilateral Filtering Algorithm", *IEEE*

Photonics Journal, Vol. 9, No. 3, 2017.

44. Wang, Y., B. Jin, Y. Wang, D. Wang, X. Liu and Q. Bai, "Real-Time Distributed Vibration Monitoring System Using ϕ-OTDR", *IEEE Photonics Technology Letters,*, Vol. 17, No. 5, pp. 1333–1341, 2017.

45. Zhang, Z. and X. Bao, "Distributed Optical Fiber Vibration Sensor Based on Spectrum Analysis of Polarization-OTDR System", *Optics Express*, Vol. 16, No. 14, pp. 10240–10247, 2008.

46. Yue, H., B. Zhang, Y. Wu, B. Zhao, J. Li, Z. Ou and Y. Liu, "Simultaneous and Signal-to-Moise Ratio Enhancement Extraction of Vibration Location and Frequency Information in Phase-sensitive Optical Time-Domain Reflectometry Distributed Sensing System", *Optical Engineering*, Vol. 54, No. 4, 2015.

47. Zhang, G., S. Li, Y. Qin and Z. Zhang, "Distributed Fiber Sensor Using a Wavelet Transform", *Proc. SPIE 3541, Fiber Optic and Laser Sensors and Applications*, 1999.

48. Rao, Y. J., J. Z. Li, Z. L. Ran and K. L. Xie, "Distributed Intrusion Detection Based on Combination of ϕ-OTDR and POTDR", *Proc. SPIE 7004, 19th International Conference on Optical Fibre Sensors*, 2008.

49. Qin, Z., L. Chen and X. Bao, "Wavelet Denoising Method for Improving Detection Performance of Distributed Vibration Sensor", *IEEE Photonics Technology Letters,*, Vol. 24, No. 7, pp. 542–544, 2012.

50. Qin, Z., L. Chen and X. Bao, "Continuous Wavelet Transform for Nonstationary Vibration Detection with Phase-OTDR", *Optics Express*, Vol. 20, No. 18, pp. 20459–20465, 2012.

51. Wang, Z. N., J. J. Zeng, J. Li, M. Q. Fan, H. Wu, F. Peng, L. Zhang, Y. Zhou and Y. J. Rao, "Ultra-long Phase-sensitive OTDR with Hybrid Distributed Am-

plification", *Optics Letters*, Vol. 39, No. 20, pp. 5866–5869, 2014.

52. Shi, Y., H. Feng and Z. Zeng, "A Long Distance Phase-Sensitive Optical Time Domain Reflectometer with Simple Structure and High Locating Accuracy", *Sensors*, Vol. 15, No. 9, pp. 21957–21970, 2015.

53. Wu, H., S. Xiao, X. Li, Z. Wang, J. Xu and Y. Rao, "Separation and Determination of the Disturbing Signals in Phase-Sensitive Optical Time Domain Reflectometry (ϕ-OTDR)", *Journal of Lightwave Technology*, Vol. 33, No. 15, pp. 3156–3162, 2015.

54. Huang, N. E., "Introduction to the Hilbert-Huang Transform and Its Related Mathematical Problems", *Interdisciplinary Mathematical Sciences: Volume 5, Hilbert-Huang Transform and Its Applications*, pp. 1–26, 2005.

55. Bueno-Lopez, M., M. Molinas and G. Kulia, *Understanding Instantaneous Frequency Detection: A Discussion of Hilbert-Huang Transform versus Wavelet Transform*, 2017, `http://hdl.handle.net/11250/2490658`, accessed in June 2019.

56. Hui, X., S. Zheng, J. Zhou, H. Chi, X. Jin and X. Zhang, "Hilbert–Huang Transform Time-Frequency Analysis in ϕ-OTDR Distributed Sensor", *Journal of Lightwave Technology*, Vol. 26, No. 23, pp. 2403–2406, 2014.

57. Jiang, F., H. Li, Z. Zhang, Y. Zhang and X. Zhang, "Localization and Discrimination of the Perturbation Signals in Fiber Distributed Acoustic Sensing Systems Using Spatial Average Kurtosis", *Sensors*, Vol. 18, No. 9, pp. 1–16, 2018.

58. Qu, S., J. Chang, Z. Cong, H. Chen and Z. Qin, "Data Compression and SNR Enhancement with Compressive Sensing Method in Phase-sensitive OTDR", *Optics Communications*, Vol. 433, pp. 97–103, 2019.

59. Ölçer, İ. and A. Öncü, "Adaptive Temporal Matched Filtering for Noise Suppres-

sion in Fiber Optic Distributed Acoustic Sensing", *Sensors*, Vol. 17, No. 6, pp. 1288–1314, 2017.

60. Wojcik, A. K., *Signal Statistics of Phase Dependent Optical Time Domain Reflectometry*, PhD Dissertation, Texas A&M University, 2006.

61. Healey, P., "Statistics of Rayleigh Backscatter From a Single-Mode Fiber", *IEEE Transactions on Communications*, Vol. 35, No. 2, pp. 210–214, 1987.

62. Howard, R. M., "Statistics of Coherently Detected Backscatter and Range Performance of Coherent OTDRs", *Optical and Quantum Electronics*, Vol. 19, No. 3, pp. 145–168, 1987.

63. Taylor, H. F., *Investigations of Fiber Optic Delay Line Signal Processos and a Fiber Optic Intrusion Sensor System*, Final Technical Report, Texas A&M University, 1997.

64. Haykin, S., *Adaptive Filter Theory*, Prentice Hall, New Jersey, 1996.

65. Poularikas, A. D. and Z. M. Ramadan, *Adaptive Filtering Primer with MATLAB*, CRC Press, Boca Raton, Florida, 2006.

66. Brennan, L. E. and I. S. Reed, "Theory of Adaptive Radar", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 9, No. 2, pp. 237–252, 1973.

67. Benesty, J. and J. Chen, *Optimal Time-Domain Noise Reduction Filters:A Theoretical Study*, Springer, New York, 2011.

68. Rugini, L. and P. Banelli, "On the Equivalence of Maximum SNR and MMSE Estimation: Applications to Additive Non-Gaussian Channels and Quantized Observations", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 64, No. 23, pp. 6190–6199, 2016.

69. Rabiner, L. R., J. F. Kaiser, O. Herrmann and M. T. Dolan, "Some Compar-

isons Between FIR and IIR Digital Filters", *American Telephone and Telegraph Company, The Bell System Technical Journal*, Vol. 53, No. 2, pp. 305–331, 1974.

70. Trees, H. L. V., *Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing*, John Wiley & Sons, New York, 2002.

71. Stoica, P. and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, New Jersey, 2005.

72. Schmid, H., *How to Use the FFT and MATLAB's pwelch Function for Signal and Noise Simulations and Measurements*, 2012, `http://www.schmid-werren.ch/hanspeter/publications/2012fftnoise.pdf`, accessed in June 2019.

73. Robey, F., D. Fuhrmann, E. Kelly and R. Nitzberg, "A CFAR Adaptive Matched Filter Detector", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 28, No. 1, pp. 208–216, 1992.

74. Farina, A. and F. Struder, *A Review of CFAR Detection Techniques in Radar Systems (in: Optimized Radar Processors)*, Peter Peregrinus Ltd, Stevenage, 1987.

75. Gandhi, P. and S. Kassam, "Analysis of CFAR Processors in Nonhomogeneous Background", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 24, No. 4, pp. 427–445, 1988.

76. Zeng, Y. and Y.-C. Liang, "Eigenvalue-based Spectrum Sensing Algorithms for Cognitive Radio", *IEEE Transactions on Communication*, Vol. 57, No. 6, pp. 1784–1793, 2009.

77. Zeng, Y. and Y.-C. Liang, "Eigenvalue Based Double Threshold Spectrum Sensing under Noise Uncertainty for Cognitive Radio", *Optik*, Vol. 127, No. 15, pp. 5968–5975, 2016.

78. Tulino, A. M. and S. Verdu, *Random Matrix Theory and Wireless Communications*, now Publishers Inc., Hanover, 2004.

79. Mehta, M. L., *Random Matrices*, Elsevier, Oxford, 2004.

80. Bun, M. J., *Application of Random Matrix Theory to High Dimensional Statistics*, PhD Thesis, Universite Paris-Sud, 2016.

81. Harding, M. C., *Essays in Econometrics and Random Matrix Theory*, PhD Thesis, Massachusetts Institute of Technology, 2007.

82. Rojkova, V., *Features Extraction Using Random Matrix Theory*, PhD Thesis, University of Louisville, 2010.

83. Marcenko, V. A. and L. A. Pastur, "Distribution of Eigenvalues for Some Set of Random Matrices", *Mathematics of the USSR-Sbornik*, Vol. 1, No. 4, pp. 457–483, 1967.

84. Tracy, C. and H. Widom, *The Distribution of the Largest Eigenvalue in the Gaussian Ensembles: β= 1, 2, 4: in van Diejen J.F., Vinet L. (eds) Calogero—Moser—Sutherland Models. CRM Series in Mathematical Physics*, Springer, New York, 2000.

85. Johnstone, I. M., "On the Distribution of the Largest Eigenvalue in Principal Components Analysis", *The Annals of Statistics*, Vol. 29, No. 2, pp. 295–327, 2001.

86. Bejan, A., "Tracy Widom and Painleve II: Computational Aspects and Realization in S-plus with Applications", *Buletinul Academiei de Stiinte A Republiicci Moldova Matematica*, Vol. 1, No. 44, pp. 4160–4164, 2004.

87. Nadler, B., "On the Distribution of the Ratio of the Largest Eigenvalue to the Trace of a Wishart Matrix", *Journal of Multivariate Analysis*, Vol. 102, No. 2, pp.

363–371, 2011.

88. Neyman, J. and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses", *Phil. Trans. R. Soc. Lond.*, Vol. 231, pp. 289–337, 1933.

89. Melsa, J. L. and D. L. Cohn, *Decision and Estimation Theory*, McGraw-Hill, Tokyo, 1978.

90. Zeng, Y., C. L. Koh and Y.-C. Liang, "Maximum Eigenvalue Detection: Theory and Application", *IEEE International Conference on Communications*, pp. 4160–4164, Beijing, 2008.

91. Vincent, F. and O. Besson, "Steering Vector Errors and Diagonal Loading", *IEE Proceedings - Radar, Sonar and Navigation*, Vol. 151, No. 6, pp. 337–343, 2004.

APPENDIX A: DERIVATION OF THE MAXIMUM SNR EQUATION

Consider the SNR definition given by Equation (3.5a). For simplicity, we will assume a unit amplitude perturbation signal (i.e., $\sigma_p = 1$), and with some matrix manipulations and exploring the Hermitian symmetry of the covariance matrix, for an observation dimension of N, we can write:

$$SNR_{(slow-time)} = \frac{|(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)^H(\boldsymbol{R}_k^{-1/2}.\boldsymbol{s})|^2}{(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)^H(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)} \tag{A.1}$$

where $\boldsymbol{R}_k$ is the covariance matrix of order $N \times N$, $\boldsymbol{w}_k$ is the column vector of complex filter coefficients of size $N \times 1$, and $\boldsymbol{s}$ is the vibration signal vector of size $N \times 1$. $(\cdot)^H$ denotes the complex conjugate transpose operation. To find an upper bound for the above expression, we can use the well-known Cauchy–Schwarz inequality and write:

$$SNR_{(st)} = \frac{|(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)^H(\boldsymbol{R}_k^{-1/2}.\boldsymbol{s})|^2}{(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)^H(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)} \leq \frac{\left[(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)^H(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)\right]\left[(\boldsymbol{R}_k^{-1/2}.\boldsymbol{s})^H(\boldsymbol{R}_k^{-1/2}.\boldsymbol{s})\right]}{(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)^H(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)} \tag{A.2}$$

Thus, the above equation can be simplified as:

$$SNR_{(slow-time)} = \frac{|(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)^H(\boldsymbol{R}_k^{-1/2}.\boldsymbol{s})|^2}{(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)^H(\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k)} \leq \boldsymbol{s}^H.\boldsymbol{R}_k^{-1}.\boldsymbol{s} \tag{A.3}$$

We can achieve the upper bound if we choose:

$$\boldsymbol{R}_k^{1/2}.\boldsymbol{w}_k = \eta.\boldsymbol{R}_k^{-1/2}.\boldsymbol{s} \tag{A.4}$$

where, η is an arbitrary real number.

We can verify this upper bound condition by plugging it in the Equation (A.4) and obtain:

$$SNR_{(slow-time),MAX} = \frac{|\boldsymbol{s}^H.\boldsymbol{R}_k^{-1}.\boldsymbol{s}|^2}{\boldsymbol{s}^H.\boldsymbol{R}_k^{-1}.\boldsymbol{s}} = \boldsymbol{s}^H.\boldsymbol{R}_k^{-1}.\boldsymbol{s} \tag{A.5}$$

Thus, the optimal matched filter is:

$$\boldsymbol{w}_{opt,k} = \eta.\boldsymbol{R}_k^{-1}.\boldsymbol{s} \tag{A.6}$$