WAVELET-BASED ADAPTIVE ARRAY SIGNAL PROCESSING FOR GUNSHOT DETECTION AND DOA ESTIMATION ON UNMANNED AIR VEHICLES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

MURAT YILMAZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

SEPTEMBER 2023

WAVELET-BASED ADAPTIVE ARRAY SIGNAL PROCESSING FOR GUNSHOT DETECTION AND DOA ESTIMATION ON UNMANNED AIR VEHICLES

submitted by **MURAT YILMAZ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Information Systems Department, Middle East Technical University** by,

Date: 04.09.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Murat Yılmaz

Signature :

ABSTRACT

WAVELET-BASED ADAPTIVE ARRAY SIGNAL PROCESSING FOR GUNSHOT DETECTION AND DOA ESTIMATION ON UNMANNED AIR VEHICLES

Yılmaz, Murat

Ph.D., Department of Information Systems

Supervisor: Prof. Dr. Banu Günel Kılıç

September 2023, 91 pages

This study aims for detection and Direction of Arrival estimation of muzzle blast and shock wave components of gunshot sound onboard a drone, despite the excessive ego-noise of the vehicle. The method depends on using the whole array data, Array Correlation Map, for improved detection and adaptive usage of Continuous Wavelet Transform scales for tuning to transient events of varying frequency. Although studied specifically for the processing of gunshot sounds on drones, the three novelties this study offers may be generalized to other array processing applications. The first is that low signal-to-noise ratio can be remedied in a detection problem with the help of the directionality of the sound source as compared to the ego noise. This is achieved by introducing the "Array Correlation Map" of the microphone array and using it to emphasize the unanimity among the array sensors. Using a simple mean value of the correlation map revealed successful results for very low signal-to-noise-ratio muzzle and shock wave scenarios, although other geometries and array processing problems may use the correlation map differently. Secondly, the help of CWT analysis is maximized by a self-adaptive selection of CWT scales. Thirdly, the tune-like feature of scales-selection is presented, which is demonstrated by automatically focusing on either muzzle or shock wave scales/frequencies. Results reveal signal-to-noise-ratio enhancement, successful muzzle and shock wave signal detection, and DOA estimation performance improvement.

Keywords: Acoustic event detection, Gunshot DOA estimation, Drone acoustic characteristics, Signal denoising, Wavelet analysis

ÖZ

İHA ÜZERİNDE SİLAH ATIŞI TESPİTİ VE YÖN TAHMİNİ İÇİN DALGACIK TEMELLİ UYARLANABİLİR DİZİ İŞARET İŞLEME

Yılmaz, Murat

Doktora, Bilişim Sistemleri Bölümü

Tez Yöneticisi: Prof. Dr. Banu Günel Kılıç

Eylül 2023, 91 sayfa

Bu çalışma, silah sesine ait namlu patlaması ve şok dalgası bileşenlerinin, insansız hava araçları üzerinde, yüksek gürültüye rağmen fark edilmesi ve yönünün belirlenmesini hedeflemektedir. Yöntem mikrofon dizisi verisinin bütüncül kullanımıyla (Dizi Bağlılaşım Haritası) tespitte başarı sağlamaya; CWT ölçeklerinin anlık uyarlanabilir kullanımıyla da farklı sinyallerin frekanslarına odaklanmaya dayanmaktadır. İnsansız hava araçları ve silah atışına yönelik çalışılmış olsa da, önerilen yöntemin üç yenilikçi yanının başka dizi işleme problemlerine de genellenebilir olması beklenmektedir. Bunlardan ilki, ilgilenilen sinyalin, sensör dizisinin etrafındaki gürültüye göre yönlü ve uzak olmasının avantajını ön plana çıkararak, düşük sinyal-gürültü-oranına rağmen başarı elde edilmesidir. Dizi Bağlılaşım Haritası konseptinin ortaya atılması ve sensörler arası hemfikirliği ön plana çıkaracak şekilde bir ortalama değer alınarak kullanılması namlu patlaması ve şok dalgası tespitinde başarım sağlamaktadır. Farklı geometriler ve problemler için Dizi Bağlılaşım Haritasını farklı şekillerde kullanmak mümkündür. İkinci olarak, CWT ölçeklerinin seçilerek kullanımı, dalgacık dönüşümünün bilinen faydalarını gürültü temizleme katkısıyla artırmaktadır. Üçüncü olarak, CWT ölçek seçiminin ilgilenilen sinyallere otomatik ayarlandığı/odaklandığı görülmektedir. Sonuç kısımlarında, sinyal-gürültü-oranı iyileşmesi, namlu atışı ve şok dalgasının başarıyla tespiti ve yön belirleme başarım iyileşmesi sunulmaktadır.

Anahtar Kelimeler: Akustik olay tespiti, Silah atış yönü belirleme, İnsansız hava aracı akustiği, Sinyal gürültüsü temizleme, Dalgacık analizi

To peace...

“War must be, while we defend our lives against a destroyer who would devour all; but I do not love the bright sword for its sharpness, nor the arrow for its swiftness, nor the warrior for his glory. I love only that which they defend: the city of the Men of Númenor; and I would have her loved for her memory, her ancientry, her beauty, and her present wisdom. Not feared, save as men may fear the dignity of a man, old and wise."

— J.R.R. Tolkien, The Lord of the Rings

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Banu Günel Kılıç for her guidance, advice, and understanding throughout the development of this thesis.

I appreciate committee members Hasan Şakir Bilge, Altan Koçyiğit, Mustafa Sert, and Özden Özcan Top for sharing valuable experiences during the development of this thesis.

I owe my wife, Burcu Kılıç Yılmaz, a debt of gratitude for her support throughout this study. Without her understanding, I could not have done any of this with our two kids at home. I hope to do the same for her during her upcoming Ph.D. thesis preparation.

I thank my daughters, İpek and Piraye. The elder one, a four-year-old at the time of writing, says, "Dad, why do you always have to study after I go to sleep?"... This thesis was written with love.

I thank my friends Murat Vural, Aydın Pehlivanlı, Halit Gölcük, and Kerem Furkan Çiçek for their help in collecting the initial sound samples, gunshot recordings during a drone flight.

I also owe to my colleagues at the workplace, who supported my efforts and motivated me from the beginning up to the last day.

Last but not least, I would like to express my regards for my greater family, especially my father, Selami Yılmaz, one of the primary motivators throughout the long efforts of this study.

TABLE OF CONTENTS

ABSTRACT iv

ÖZ v

DEDICATION vi

ACKNOWLEDGMENTS vii

TABLE OF CONTENTS viii

LIST OF TABLES xii

LIST OF FIGURES xiii

LIST OF ABBREVIATIONS xvi

CHAPTERS

1 INTRODUCTION 1

1.1 Real-life Scenarios and Challenges 1

1.2 Aims and Objectives 4

1.3 Contributions of the Study 4

1.4 Outline of the Thesis 5

2 LITERATURE REVIEW 7

2.1 Background on Gunshot Sound Characteristics 7

2.2 The Denoising Problem and Background on Drone Sound Characteristics 10

2.3 Gunshot Detection and Localization Problem 11
2.4 Wavelet-based Methods 12
2.5 Related Work 14
3 METHODOLOGY 17
3.1 Examination of the Problem Geometry 17
3.2 Preparation of Datasets in Accordance with Problem Geometry 17
3.3 CWT Analysis of Shock-wave, Muzzle Blast, and Drone Sound 19
3.4 Introducing a Novel Method: Array Correlation Map with Adaptive CWT for Gunshot Detection on Drones 22
3.4.1 Array Correlation Map 24
3.4.2 Adaptive-CWT based Detection 24
3.4.3 Auto-tuning Feature 28
4 SNR IMPROVEMENT RESULTS 33
4.1 Visualizing SNR Improvement by the Proposed Method 33
4.2 SNR Calculations Before / After the Proposed Method 35
4.3 Conclusions on SNR Improvement 36
5 DETECTION RESULTS 39
5.1 Performance Assessment Metrics for the Detection Problem 39
5.2 Detection Performance wrt Muzzle Range 40
5.3 Detection Performance for Different Muzzle Direction of Arrivals 41
5.4 Improvement of Detection Performance with Automatic Tuning of Scales 42
5.5 Receiver Operator Characteristics (ROC) Curves for the Detection Performance 44
6 DOA ESTIMATION RESULTS 47

6.1 Reference DOA Estimation Methods 47
6.1.1 The MUSIC Algorithm 47
6.1.2 The MVDR Algorithm 48
6.2 Performance Metric for the DoA Estimation Problem 48
6.3 DoA Estimation Performance for the Muzzle Blast Component 49
6.4 Muzzle Range 50
6.5 Direction of Arrival 53
6.6 DOA Method 55
6.7 Gun Type 59
6.8 Drone (Noise) Type 59
6.9 DoA Estimation of the Shock Wave Component 60
7 DISCUSSION 67
7.1 Discussions on Muzzle and Shock Wave Detection 67
7.2 Notes for a Real-Life Drone Implementation 69
7.3 On Possible Challenges Not Covered by the Dataset Scenarios 70
7.4 On Generalizability of the Proposed Method to Similar Problems 71
8 CONCLUSIONS 73
8.1 Key Features of the Proposed Method 73
8.2 Future Work 74
REFERENCES 77
APPENDICES
A ROC THRESHOLD PERFORMANCES 83

B CV 87

LIST OF TABLES

Table 1 Factors that affect acoustic source localization outdoors 3

Table 2 Domain-specific factors that may affect gunshot localization on drones 3

Table 3 Characteristics of the gunshots used in the dataset 18

Table 4 Reference drone properties for popular small-scale drones 19

Table 5 Details of Dataset Scenarios 20

Table 6 Example SNR values for original and denoised array data: MB308r shot01 muzzle range: 400 m, azimuth of 60 degree, elevation of 60 degree, shock miss distance of 24 m, shock azimuth of 30 degree and elevation of 30 degree 36

Table 7 Mean SNR improvement calculated over several trials of different scenarios 37

Table 8 Muzzle Blast Detection Details for Different Ranges 42

Table 9 Muzzle Blast Detection Details for Different DOA Angles 42

Table 10 Detection Performance for all data / Muzzle blast and Shock Wave components 44

Table 11 Area Under the Curve for Window Length of 512 Samples 45

Table 12 Area Under the Curve for Windows Length of 256 Samples 46

Table 13 Dataset Scenario Summary for the Parameters of the Problem 51

Table 14 Multivariate Analysis Table for Muzzle Blast DOA Estimation 51

Table 15 Multivariate Analysis Table for Shock Wave DOA Estimation 61

Table 16 Comparison of Results and Performance with Related Studies 69

Table 17 Muzzle Detection Strength ROC Threshold Performances for Window Length of 256 Samples 83

LIST OF FIGURES

Figure 1 Spectrogram of an example gunshot sound signal windowed with a 1024 sample Hamming window and FFT size of 1024. Gunshot occurs at 0.1 s. 9

Figure 2 Simple sound field geometry with plane wave assumption. 12

Figure 3 Representative problem geometry and the two main signal components. 18

Figure 4 Gunshot components travel time through geometry. 19

Figure 5 Initial CWT Analysis of a Gunshot + Drone sound signal with Haar wavelet using MATLAB Wavelet Toolbox™ 21

Figure 6 Absolute value of the CWT output for an example mixture of muzzle blast, shock wave and drone sound. It is known that there is one gunshot instance at 10^4 and three instances between 2.5×10^4 and 3×10^4. 21

Figure 7 Comparison of drone, muzzle blast, and shock wave audio samples' frequency content - Each signal is scaled to 1 to be visually comparable. The method and visualization are incorrect here, suggesting that time window size should be selected in accordance with typical muzzle and shock wave component duration. 22

Figure 8 Comparison of drone and muzzle blast over the time duration of muzzle blast, for 9 ms. 23

Figure 9 (1) Absolute value of the CWT output for an example mixture of muzzle blast, shock wave, and drone sound. The shock component is dominant since the d_{shock} is 12 m. (2) Shock waves will be there only in certain geometric scenarios. Scale range 38-111 visualized; muzzle blast is at 2.507 s, with a peak at scale of 53, but not dominant. 23

Figure 10 (a) Output of Array Correlation Map, calculated for consecutive 5 ms windows (b) How the pairs agree, calculated through the mean of the Array Correlation Map output for each window. 25

Figure 11 (a) Correlation Map coherence over all scales. Muzzle blast, shock wave, and drone sound are all distinguishable with the CWT + ACM method (b) Detection instant and detection strength pairs for all scales so that the optimal scales can be tuned. 26

Figure 12 Flow diagram of the proposed method: Array Correlation Map with adaptive CWT for drone gunshot detection. 27

Figure 13 Steps of the CWT+ACM based detection, for an example scenario with muzzle blast and shock wave. 28

Figure 14 Illustration of tuning capability by using scales of interest: Ability to selectively reject muzzle and shock wave components in the same sample. 29

Figure 15 ACM-CWT Output & detection strengths example: For shock wave (left) and muzzle blast (right) components of the same shot. 30

Figure 16 Detection strength over all scales. Zoomed in time to muzzle detection instant (left) and a no-signal instant (right). 31

Figure 17 Original microphone data and improved muzzle and shock wave signal visibility with precise detection and automatic CWT scales selection. 34

Figure 18 Original microphone data and improvement on muzzle signal visibility for the range of 800 m with precise detection and automatic CWT scales selection 35

Figure 19 Mean SNR Improvement vs Muzzle Ranges 38

Figure 20 Mean Shock Wave SNR Improvement vs Bullet Miss Distances 38

Figure 21 Muzzle Blast Detection Performance wrt Muzzle Range in m. 41

Figure 22 Muzzle Blast Detection Performance wrt Muzzle DoA in Degrees 43

Figure 23 ROC Curve for Muzzle Detection with Window Length of 512 Samples 45

Figure 24 ROC Curve for Muzzle Detection with Window Length: 256 Samples 46

Figure 25 Arrival path and geometric parameters of the muzzle blast component. 50

Figure 26 DOA Estimation Performance Improvement by the Proposed Processing wrt Muzzle Range 53

Figure 27 Mean Muzzle DOA Estimation Errors wrt Range per Gun Type with Original Array Data 54

Figure 28 Mean Muzzle DOA Estimation Errors wrt Range per Gun Type with Proposed Processing 54

Figure 29 DOA Estimation Performance Improvement by the Proposed Processing wrt DOA 55

Figure 30 DOA Estimation Performance Improvement for Different DOA Algorithms 56

Figure 31 Mean DOA Estimation Errors for Different DOA Methods Against Range - with Original Array Data 57

Figure 32 Mean DOA Estimation Errors for Different DOA Methods Against Range - with Proposed Processing 57

Figure 33 Mean DOA Estimation Errors for Different DOA Methods Against Angle of Arrivals - with Original Array Data 58

Figure 34 Mean DOA Estimation Errors for Different DOA Methods Against Angle of Arrivals - with Proposed Processing 58

Figure 35 DOA Performance Improvement for Different Gun Types 60

Figure 36 DOA Performance Improvement for Different Noise Samples 61

Figure 37 Arrival path and geometric parameters of the shock wave component. 62

Figure 38 DOA estimation error performances for different DOA methods 63

Figure 39 DOA estimation error performances for different miss distances and gun types . . . 63

Figure 40 DOA Estimation Performance Comparison with the Proposed Method with respect to the Miss Distance 64

Figure 41 DOA Estimation Performance Comparison with the Proposed Method with respect to the DOA Method 65

Figure 42 DOA Estimation Performance Comparison with the Proposed Method with respect to the Gun Type 65

LIST OF ABBREVIATIONS

ACM	Array Correlation Map
ANOVA	Analysis of Variation
ASL	Acoustic Source Localization
AUC	Area Under Curve
CWT	Continuous Wavelet Transform
dB	Decibel
DOA	Direction of Arrival
DWT	Discrete Wavelet Transform
ER	Error Rate
GCC	Generalized Cross Correlation
MUSIC	Multiple Signal Classification
MVDR	Minimum Variance Distortionless Response
RMSE	Root Mean Square Error
ROC	Receiver Operator Characteristics
RPM	Revolution per Minute
SAF	Small Arms Fire
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
SRP	Steered Response Power
TDOA	Time Delay of Arrival
ULA	Uniformly Spaced Linear Array
wrt	with respect to
3-D	Three Dimentional

CHAPTER 1

INTRODUCTION

Despite the popularity of drone-mounted camera solutions, auditory information is not used as much, primarily due to the excessive ego-noise of drones. This ego noise, along with the effects of airflow, vibration, and the fact that the microphones on a drone must be positioned close to the motors, makes this a very low-SNR problem. Gunshot localization in different environmental conditions is a real-life problem for surveillance and protection of battlegrounds and public places. There is academic work, as well as some products that are deployed at public places such as city halls or critical military locations for ground-based gunshot localization of Small Arms Fire (SAF) and snipers [1] and [2]. However, such solutions are highly stationary in the way they are set up in the form of a stationary microphone array (or several microphone arrays), and there are only a few studies of onboard applications for drones or other aircraft.

Common to all these studies, microphone arrays are used as sensory equipment to collect data from the sound field. Then, the contributions of the muzzle blast and the shock wave (or a combination of them) are used to detect the gunshot and localize its position. The outputs are further enhanced by the use of multiple arrays and by applying geometric solutions, such as triangulation. Although giving precise coverage of a specific area, such solutions are not portable and must be integrated into a pre-determined area of interest. Also, they are not helpful in any moving mission scenario.

Use cases for microphone arrays would be extended if they were made portable on drones. The use of drones so far has been through the effective use of onboard cameras. Studies with drone-mounted microphone arrays are now trying to pave the way for making microphones usable on drones, despite these vehicles' ego noise and wind noise.

This study analyzes gunshot recordings collected on a drone and proposes an algorithm, initially focusing on the detection problem. Considering the non-stationary noise, an adaptive low Signal-to-Noise Ratio (SNR) solution is sought so that the time instant and the frequencies to focus are autonomously extracted. The solution's performance is examined in three aspects: SNR enhancement, detection performance, and Direction of Arrival (DOA) estimation error enhancement performance.

1.1 Real-life Scenarios and Challenges

To begin with, some real-life mission scenarios are given below to better understand the problem domain.

- Drones and helicopters operate at lower velocities (typically below 200 km/h) than fixed-wing aircraft, performing vertical take-off and landing, making them more open to ground attacks with SAF.
- Excessive noise levels of motor, exhaust, etc. make it hard, if not impossible, for regular auditory setup to perform any audio surveillance or any hearing aids
- Ultra-sonic projectile speeds cause extraordinary sound wave characteristics that depend on weapon characteristics and vary even among specific bullets, requiring specific consideration.
- Although there are real-life examples of single-shot cases such as sniper or riffle attacks, low-altitude mission conditions require an onboard application to consider conventional battle situations, where multiple gunshots can be incident from different directions, simultaneously to human perception.
- Since a gunshot is a highly impulsive sound, multiple reflections represent a severe problem. Even on the ground, military personnel. Besides seeing during the night, infrared vision devices scan possible directions to determine the actual sound source. Still, it is possible to mislead infrared vision devices in specific environmental conditions, and, in some cases, localization with infrared has not been possible.
- A realistic target range should include SAFs weapons such as rifles, handguns, and snipers.
- A realistic application should cover as many weather conditions where the aircraft is operational.
- A realistic system should be reliable such that it has to either output results within an acceptable error range or notify if it cannot. This requires assessing its own error range and/or deciding whether current conditions are operational.
- There are stationary solutions that can localize and classify gunshot instances. By their nature, these can cover only a pre-determined area.

Although there is plenty of research and implementation in the literature concerning Acoustic Source Localization (ASL) or gunshot source localization, our problem's extreme carrier vehicle noise conditions of our problem add up to the usual detection and DOA estimation problem, which can be summarized as follows.

- the extreme noise level of drones generated by the motors
- rapid changes in frequency domain characteristics of the noise due to changing RPM
- onboard vibration (especially on multi-rotor type drones)
- airflow generated by the propellers and the wind

Although Sound Pressure Level (SPL) levels of weapons are considerably high, even up to 160 dB in some cases [3], a significant amount of signal energy is lost due to gunshot distance, as explained in detail in the following chapters of this study. Furthermore, typical drones can carry a payload in grams to a few kilograms. The electrical power is also very limited since batteries supply it, and any

Table 1: Factors that affect acoustic source localization outdoors

Factors that Affect Gunshot Localization		Signal Processing Parameters	
Geometrical Spreading Factors	Atmospheric Absorption Factors	ASL Algorithms Parameters	Microphone Array Parameters
Distance Incidence Angle Reflections Wind Meteorological Events	Temperature Humidity Pressure	Window Length Sampling Frequency Simultaneous Sources	Number of Microphones Microphone spacing Array geometry

Table 2: Domain-specific factors that may affect gunshot localization on drones

Drone Sound Characteristics	Gunshot Sound Characteristics
Sound pressure level Blade passage frequencies Blade tip speed Airspeed Maneuvers	Sound pressure level Gun type Barrel dimensions Explosive type Projectile speed Miss distance Bullet caliber Direction of trajectory

increase in power demand directly reduces the mission duration of the drones. Indeed, the payload and the maximum mission duration are the leading performance specifications of all drone products.

If these challenges are successfully handled, and the signals of interest are successfully discriminated, there are still factors that could misguide DOA estimation algorithms. Characteristics of disturbances such as reflections, shockwaves of supersonic projectiles, and simultaneous gunshots are also considered the most critical of these factors, as depicted in previous research [4].

To summarize, the acoustic DOA estimation of gunshots on drones is a challenging problem with many factors and parameters in place. In a broader sense, Table 1 and 2 summarize the factors that may affect acoustic detection and localization on drones. These two tables present a comprehensive list of the factors that affect gunshot acoustics on drones to better illustrate the dimensionality of parameters for such a system. It should be noted that the simulator of this study is capable of simulating distance (range), incidence angle, temperature, humidity, pressure, window length, sampling frequency, number of microphones, microphone spacing, and array geometry. Also, the dataset samples cover different gun and drone SPLs, drone maneuvers, miss distances, gun types, and projectile speeds.

1.2 Aims and Objectives

This thesis aims to detect gunshot events on low-SNR drones and obtain enhanced array signals for applying signal processing tasks with improved performance, primarily gunshot DOA estimation.

With this aim, the following objectives were achieved.

- To record gunshot sound samples on a drone during flight to investigate signal and noise characteristics, revealing the difficulty of detecting gunshot signals in low SNR.
- To generate a dataset using a reliable sound field simulator for different values of well-defined parameters and realistic array recordings of drone noise.
- To propose an algorithm for adaptive, autonomous detection of gunshot components despite low SNR and nonstationary noise characteristics.
- To quantify the SNR enhancement by the proposed detection and reconstruction method.
- To quantify the detection performance with suitable event detection metrics and to show that the detection phase has decent binary classifier characteristics.
- To check if the enhancement suggested by the proposed pre-processing improves DOA estimation performance as compared to conventional methods.
- To quantify the DOA estimation error improvement with the proposed pre-processing

1.3 Contributions of the Study

Considering the complex sound field of the problem, precise examination of dependable and realistic sound recordings was first targeted. For this purpose, gunshot samples were collected on a drone during flight. Initial examination revealed comparisons of drone noise and gunshot signal components, challenging aspects of the detection problem, but still, suitable nature of wavelet-based analysis for this kind of problem.

With an approach that focuses on the challenges of the problem, especially the varying nature of the drone sound and frequency content of signal and noise, an algorithm was proposed that extracts agreement of array channels and is self-adaptive in scale (frequency) selection. A realistic and reliable dataset was needed to assess the proposed method's generalizability and performance. A previous study by the author of this thesis examined gunshot localization on helicopters by performing sound field simulations that consider the effects of outdoor environmental propagation [5] and [6]. Using the sound field simulator, gunshot signals were mixed onto complete array recordings of drone noise to obtain realistic samples. Also, the simulator covered the effects of range, DOA, temperature, air pressure, humidity, and frequency-dependent power attenuation.

Using this dataset, the performance was assessed in three different aspects. First, the enhancement on the signals was quantitatively reported. Secondly, the detection performance was given in terms of F-score, Error Rate, and ROC curves. Thirdly, two conventional DOA estimation algorithms were applied with and without the pre-proposed reconstruction. This way DOA estimation performance improvement with the proposed signal enhancement was revealed.

1.4 Outline of the Thesis

This thesis was organized in a manner that first defines the breakdown of the problem of array-based gunshot detection and DOA estimation on drones. In this regard, Chapter 1 explains the real-life aspect of the problem and the many parameters in effect. Then, Chapter 2 gives the background information on drone and gunshot characteristics and related studies and techniques in the literature. Challenges and the problem being yet unresolved in the literature are emphasized.

Having listed sound field characteristics and challenges, Chapter 3 starts turning these challenges and advantages of the field into a solution strategy. To use the whole array data and to take advantage of the fact that the signal is distant-directional-impulsive, while the drone noise is close-directionless-continuous, a new concept, Array Correlation Map (ACM), is introduced. Then, an adaptive way of using Continuous Wavelet Transform (CWT) scales is proposed for the varying nature of the drone noise and for having a tune-like feature to a signal of interest. The outcome of the method's steps is exemplified with samples and figures.

Following the concept and examples of Chapter 3, three results chapters followed, reporting on the performance of the proposed solution with representative datasets. Specifically, Chapter 4 begins with claiming and measuring the SNR improvement for the automatically focused signal-of-interest regions. Then, Chapter 5 reports on detection performance, including ROC analysis, proving that the offered method yields decent classifier performances for both muzzle blast and shock wave detection. Then, Chapter 6 utilizes multivariate ANOVA analysis from SPSS and graphical DOA estimation error results for different muzzle blast and shock wave DOA estimation parameters. The results are given in a way that compares the performance of reference conventional DOA estimation algorithms with and without the proposed pre-processing.

Then, Chapter 7 discusses the SNR enhancement, detection, and DOA estimation results, together with how the method's performance would generalize for similar problems. Moreover, notes on the real-life implementation on drones are shared. Finally, Chapter 8 concludes the thesis by commenting on scope and limitations, possible enhancements, and future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Background on Gunshot Sound Characteristics

A successfully recorded gunshot sound can contain information corresponding to three major components of a typical gunshot sound, which are the muzzle blast, the shock wave component (in case of a projectile of supersonic speed), and the firearm's mechanical action [4]. Usually, the mechanical action of the gun has the lowest amplitude. It is not expected to be observable for the purposes of this study due to the noise level of the drone and the firearm-to-drone distance. On the other hand, the muzzle blast and the shock wave components are highly related and will be examined concerning both SPL and frequency characteristics. [7] has examined the characteristics of gunshot recordings, giving these three major sound components. Another approach is [2], where muzzle blast and shock wave components are examined in an integrated manner to take advantage of sensor fusion techniques to use detectable characteristics of both components together.

A firearm (such as a rifle, a sniper etc.) uses an explosive charge inside a barrel to throw out the projectile with a very high velocity [7]. This explosive action and the accompanying sound is called the muzzle blast. This can be considered a regular sound signal of high amplitude that spreads out of the barrel in all directions. Another study [3] measures a 20 dB difference in the sound pressure levels caused by this explosion depending on the measurement direction. As may be expected, the SPL is highest in the direction of the firearm. Another important indication of the study is that a typical muzzle blast lasts 3 to 6 ms and propagates through the air at the speed of sound. On the other hand, the shock wave typically lasts in the order of 200 ms [7].

Naturally, this regular sound signal reduces amplitude as it propagates in the air. [7] stated that, for the measurements close to the firearm, muzzle blast could be a significant sound of interest and is very useful for purposes of detection and localization. As the gunshot distances increase to several hundred meters, the muzzle blast sound is reduced, as any other sound would. The significant factors for this attenuation are energy reduction by spherical spreading and the effects of air absorption (Harris, 1966). It is also more challenging to use this signal since it is to be distorted by reflection, reverberation, and obstacles in the sound field [8].

A shock wave is generated for projectiles with higher velocity than sound speed. This is a cone-shaped shock wave that expands behind the bullet. The speed of expansion of the cone outwards is the same as the speed of sound. This signal reaches the microphones before the muzzle blast [1]. The inner angle

of the cone is called the Mach Angle and is given by:

$$\theta_M = \sin^{-1}(1/M) \tag{1}$$

where M is defined as Mach number:

$$M = V/c \tag{2}$$

where V is the bullet speed, c is the speed of sound.

Thus, the shock wave's arrival direction can be calculated if the projectile speed is known. Also, the Mach angle widens as the bullet slows down during its travel in the air [3]. Both [1] and [3] reported that the shock wave of a gunshot appears as an N-shaped signal of typically 200 ms. That way, it can be discriminated from the succeeding muzzle blast and used for detecting and positioning firearms.

For retrieving more detailed information, such as the complete bullet trajectory, firearm characteristics, and miss distance, prior knowledge of bullet speed and caliber is required, which are hard to know beforehand [4]. Still, some studies search for similarities between the detected gunshot signal and a set of pre-recorded known gunshot signals. For instance, [9] employed two parallel GMM classifiers trained with 47 audio features, obtaining a false rejection rate of 8%.

Regarding amplitude and frequency characteristics of typical firearms, [10] has reported on characteristics of a list of gunshots containing five revolvers, six pistols, and three rifles. The range covered barrel lengths of 6.4 cm to 56.4 cm and projectile speeds of 250.2 m/s to 889.1 m/s for five different bullet types. SPL levels varied between 151.4 dB and 160.8 dB; all three rifles had larger SPL (around 160 dB) than revolvers and pistols. All three rifles had supersonic bullets. The SPL measurements were taken at 2 m from the barrels. The authors also reported that both muzzle blast and shock wave components were highly affected by shooting direction, and high-frequency losses of the muzzle blast component were significant at long ranges.

Regarding timing characteristics, [4] obtained the time difference between shock wave and muzzle blast components from gunshot signals of supersonic bullets for different target distances. For example, for a .308 caliber Winchester, the particular bullet speed was 831.5 m/s, yielding a Mach Number, M, of 2.54 and a Mach Angle, θ_M, of 23.2°. The shock-wave-to-muzzle-blast time difference was around 100 ms for a downrange gunshot recording at 91 m. The two components were in the same time window of 10 ms for an 8 m gunshot distance. A critical observation from the measurements is that, although the muzzle blast component dramatically loses its amplitude above a range of 300 m, the shock wave component is not significantly reduced and still easily observable, so long as the miss distance is kept constant.

Regarding the frequency characteristics of gunshot signals, [11], [12], and [13] observed wide-band characteristics in gunshot recordings. The details of spectra, however, varied considerably among different gunshot models. Calculating 1/3 octave band spectra of a 4-inch barrel DWM pistol with a 7.65 Luger P LNR cartridge, [12] have reported a 10 dB to 25 dB increase in SPL levels for frequencies above 1 kHz. [13] observed similar frequency spectra with a .22 Carbine; however, the dominant frequencies were below 500 Hz for a .12 shotgun and below 1500 Hz for a 9mm Pistol, all recorded at 30 m from the firearm.

Figure 1 gives a spectrogram of an example gunshot signal for visualizing frequency characteristics. The gunshot occurs at 0.1 s. Both impulsive nature and wide-band characteristics are observable, and the spectrogram is in conformance with previous studies in the literature.

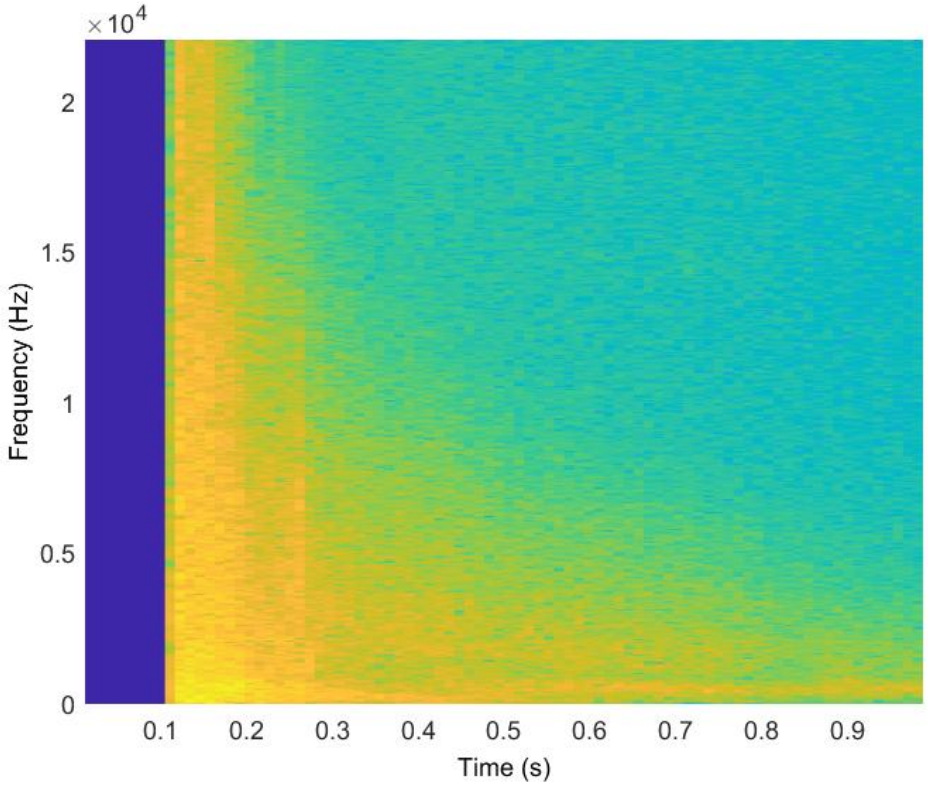


Figure 1: Spectrogram of an example gunshot sound signal windowed with a 1024 sample Hamming window and FFT size of 1024. Gunshot occurs at 0.1 s.

To conclude, previous work in the literature explains the geometry and characteristics of gunshot sound and gives expected SPL levels of different gunshots at different distances and directivity. In the frequency domain, despite having some commonalities, spectral characteristics of weapons and bullets vary for different firearms and bullets.

2.2 The Denoising Problem and Background on Drone Sound Characteristics

Like many array processing implementations, acoustic gunshot detection and localization systems must utilize algorithms to improve the signal of interest while eliminating the noise. Various signal processing techniques are utilized, such as signal processing solutions (spectral analysis, band-pass filtering, filterbanks), learning-based solutions such as CNNs [14] and k-means, and methods of pattern recognition to enhance SNR or improve performance despite low SNR at hand.

Several studies try different denoising methods for drones' onboard excessive noise problem. An example is [14], which performs sound source separation (SSS) for SNR enhancement. Specifically, the authors report GHDSS-AS and RPCA as the most advanced multi-channel and single-channel methods. A basic and popular approach is deployed by [15] for the DOA estimation side of the problem; however, for coping with the noise and the dynamic nature, an alternative approach is taken, namely particle filtering and utilization of Doppler shifting.

Work of [16] derives a noise model for assessing the feasibility of microphone-array techniques for noise reduction. The study suggests that drone noise consists of narrow-band harmonic and broadband noise, whose spectra vary dynamically with the motor rotation speed. An important assertion is that for a P-motor drone, P+2 microphones are required to obtain signal denoising. Another highlight is that due to fixed relative locations between the motors and the microphones, the acoustic mixing network of the ego-noise tends to remain stationary during the drone's movement. This is a staging point for proposing Array Correlation Maps (ACM) usage by this thesis: The noise spectral positioning should remain fixed. At the same time, the external signal of interest is comparatively outstanding and sudden.

Another way of handling drone ego noise is the usage of spatially informed filtering [17]. Like the previous example of [16], the fixed physical location of noise sources compared to the array is critical, which can be computationally eliminated. As expected, the limitation arises when the source of interest coincides with such directions.

Methods that depend on collecting a noise database should also be reported here. One such implementation is the work of [18], where to discriminate between its own sound and an external sound, a robot (not necessarily a drone, but a robot in the generic sense) collects templates of its own sound in real-time, instead of training offline. The method is reported to suppress the ego noise contaminating multiple sound sources for ground bases robot applications.

One of the significant challenges with drone noise is that it is highly un-stationary due to how multi-rotor drones are controlled. A helicopter has a motor and possibly multiple rotors revolving with pre-defined rates, i.e. RPM, and the control and maneuvering are achieved by changing the angles of control surfaces. Despite having significant mechanical and electrical differences, fixed-wing aircraft also have motors of fixed RPM and control surfaces for maneuvering. On the other hand, multi-rotor drones have surfaces with fixed angles and obtain maneuvering by setting the RPMs of the motors differently from each other. The implication of this fact on sound characteristics is twofold: First, there

are different motor sound components (which are considered noise in this case). Secondly, the frequency characteristics of the noise change rapidly since RPMs of all motors are changed continuously and differently from each other. These two factors incur rapidly varying noise characteristics onboard a drone.

An additional characteristic reported by [19] is that, since the microphones are mounted on the drone itself, they must be close to the noise sources, leading to extremely low SNR: 15 dB. The authors compare three microphone array algorithms to enhance the target sound captured by a micro aerial vehicle: time-frequency spatial filtering, beamforming, and blind source separation. It is reported that by exploiting the time-frequency sparsity of the acoustic signal, this spatial filter works robustly for sound enhancement in the presence of strong ego-noise. This implies that a solution that suppresses the sounds around the array and emphasizes the distant and directional sound may be adequate.

Another limitation arises from the moving characteristics of drones. Drones (multi-rotor drones in particular) characteristically have swift movements that rapidly change the UAV's bearing relative to potential sound sources. As stated by [20], ASL must be performed using relatively short time windows leading to near real-time requirements for real-life ASL scenarios. This nature of the problem means it is essential to adopt a solution that is adaptive to momentary changes.

Finally, other than the wind occurring naturally in outdoor scenarios, the airflow produced by the rotating propellers is another contribution of the drone to the noise components. This makes mechanical positioning of the microphones critical and measures for preventing signal clipping and may require specific attention at denoising methods.

2.3 Gunshot Detection and Localization Problem

Acoustic gunshot detection uses audio recordings to decide whether a gunshot event exists in a given sound wave (or a specified portion of a continuous sound wave). A simple Acoustic Source Localization (ASL) application locates a sound source by applying acoustic localization algorithms on signals collected from the sound field. This position information may be only the direction of the source, in which case it is called direction-of-arrival (DOA). If the range of the sound source is also determined, the location of the sound source is determined in 3-D, i.e., acoustic source localization.

With the acoustic far-field assumption, if the sound source and sensors are positioned far enough from each other, the wavefront appears as plane waves. Therefore, the geometry of the problem with acoustical far-field assumption appears to be as given in Figure 2.

Using this more straightforward form of the sound field geometry, direction-of-arrival (DOA) algorithms compute the angle of arrival, θ, by using the time difference between the two signals, τ. The methods for obtaining τ and extracting the angle of arrival, θ, vary among different DOA algorithms. Moreover, specific implementations may combine knowledge of microphone array geometry and amplitude differences at separate receiver signals with this DOA information to determine optimal source location [21].

Sound signals can provide information about the detection [2], DOA, and distance of the gunshot as well as the speed and trajectory of the projectile [22]. Gunshot forensics has methods [2] that geometrically derive conclusions by using arrival times, arrival angles, and typical frequency characteristics

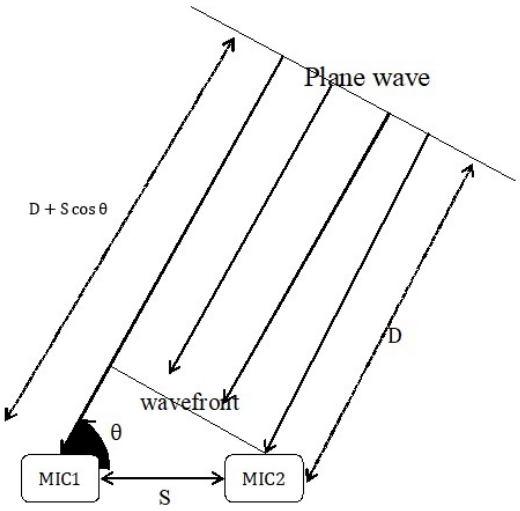


Figure 2: Simple sound field geometry with plane wave assumption.

of the muzzle and shock wave with respect to each other. With classification purposes, the weapon type may also be determined [23].

Apart from academic research or prototype studies, there are some commercially available systems for localization and identification of sniper shots [22]. Although all aspects are not all in the scope of this study, gunshot signal characteristics are to be examined extensively for a better understanding of the signal domain. Traditionally, since gunshot events are high amplitude, impulsive signals, it is pretty straightforward to apply denoising and detect/extract the high-pitch portion of audio information. However, the problem domain of interest (battlefield or onboard air-craft) dictates noisy and complex audio environments. Traditionally, most research focuses on the mathematical analysis of predefined feature vectors. Some examples are wavelet sub-band features or zero-crossing features [24]. Recent academic studies employ ML algorithms, representing the authors' choice of ML methods. Examples from the literature include the usage of SVM, K-means, or Hidden Markov Models [25].

It should be noted that the scope of this study is focused on acoustic detection and localization, while there are optical-based gunshot detection systems, too. Such systems use cameras or infrared sensors to detect optical transients such as muzzle flashes at shooting instant. By their nature, they may be susceptible to visual challenges, such as low lighting conditions. More critically, fog, clouds, or even a minor physical obstruction in the line of sight means failure to detect despite an impulsive sound.

2.4 Wavelet-based Methods

Since we are interested in analyzing temporal signals at multiple resolution units of interest, the two main methods classes are wavelet analysis and time-frequency analysis [26]. Traditionally, the Fourier transform provides frequency information over the entire signal, with the well-known trade-off between precision in time and frequency. On the other hand, the wavelet transform allows for a variable time-frequency resolution for signals with nonstationary frequency characteristics. This precision is

achieved thanks to the windowing function, the wavelet, which is scaled and translated over the input signal.

Wavelet-based methods are used for many signal processing problems, including audio and speech processing, image and video processing, medical signal analysis, seismic data analysis, etc. It uses wavelets and mathematical functions for transforming signals from the time domain to the time-frequency domain. This process represents signals with a sharper time-frequency response [27]. This is achieved mainly by the wavelet transform. The input signal is transformed into constituent frequency components at different time scales (or resolutions). These time scales correspond to frequencies of interest. In this study, an analysis of initial array samples with Continuous Wavelet Transform (CWT) was experimented with as a starting point.

Continuous Wavelet Transform (CWT) yields representations of the input signal by varying the translation and scale parameters of the wavelets [27]. The continuous wavelet transform of a function x(t) at a scale, (a > 0), and transitional value, b, is expressed by the following integral:

$$Xw(a,b) = \frac{1}{|a|^(0.5)} \int_{-\infty}^{\infty} x(t)\rho * \left(\frac{t-b}{a}\right) dt \tag{3}$$

where ρ is a continuous function in both the time domain and the frequency domain called the mother wavelet, and the * represents a complex conjugate operation. The mother wavelet is selected according to the signal and application characteristics. It provides a source function to generate the daughter wavelets. The daughter wavelets are simply the translated and scaled versions of the mother wavelet [28].

CWT is more beneficial for feature extraction and time series analysis due to its fine resolution, whereas DWT is a fast transform that represents a time series in a more redundant form and is often used for data compression and noise attenuation of two-dimensional signals or images [26]. The continuous wavelet analysis can study how spectral features evolve over time and identify common time-varying patterns in two signals. It is also possible to perform time-localized filtering. The continuous wavelet transform (CWT) is ideal for analyzing non-stationary signals, that is, the signals with varying frequency-domain representation over time.

Discrete Wavelet Transform (DWT) applications on audio signals include compression, denoising, feature extraction, watermarking, equalization, time or pitch shifting, and transient detection. Denoising and transient detection features are promising for the problem of this study. DWT-based analysis can detect change points, discontinuities, and other events not readily visible in raw data.

It is reported by [29] that to cope with nonlinear and non-stationary signals, such as EEG, the classical frequency methods have strict restrictions, suggesting that time–frequency techniques eliminate them. Types of such techniques for signal decomposition are listed as Empirical Mode Decomposition (EMD), Discrete Wavelet Transform (DWT), and Wavelet Packet Decomposition (WPD) by the same study. An example using DWT for automatic detection is [30] for detecting patterns on electroencephalogram (EEG) signals. The study reports classification accuracy of 90% and sensitivity and specificity of 88% and 92%, respectively.

An interesting usage style of the wavelets is reported by [31], for considering multi-scale relationships among several variables. The study offers new statistical tools that help to determine the overall correlation for the whole multivariate set on a scale-by-scale basis. Using wavelets and cross-correlation,

and offering to use the whole set of signals in a combined manner, are the two aspects of [31] that are similar to the methodology of this thesis, as detailed in Section 3.4.1.

2.5 Related Work

Despite the popularity of camera-mounted applications on all aircraft scales, only a few studies of onboard audio applications exist. One such study is [32] that investigates pre-processing and different direction-of-arrival (DoA) estimation on drones, deploying and comparing two methods, namely The Data Selection Least Squares with Median Filtering and Multichannel Blind Source Separation (MBSS) toolbox. The authors report promising DoA results, stating limitations of max 120 m distance and that the accuracy of the estimates is sensitive to the drone position relative to the shooter. Also, detection is done manually, and the study scope is localization only.

Another study with drone-mounted microphones is [25], which offers a Generalized Eigenvalue Decomposition Multiple Signal Classification (GEVD-MUSIC) algorithm with adaptive estimation of noise correlation matrix. The paper reports that the adaptive nature helps to cope with the nonstationary ego noise of the drone, and the method offers an improvement in localization performance for high SNR scenarios. A learning-based example is [14], which aims to classify sound events into eight categories with a microphone array on a drone. The paper reports combining Geometric High-order Decorrelation-based Source Separation (GHDSSAS) and CNN successfully classified 80% of the sound events. The sound source to microphone distance was small, though, 3 m away from a hovering drone. [19] and [17] suggest and compare methods for coping with the ego-noise of drones, too. The methods suggest using the geometry of drones and onboard sound sources and sparsity of the sound sources and offer ways of improving noise reduction performance when the direction of interest is known. Despite claiming reasonable performance, the suggested method does not apply to the problem addressed with this study since the direction of interest is not already known.

As an example of sound localization onboard a drone, [15] aimed at localizing emergency signals of predetermined narrow-band characteristics. Although different from this study in terms of target sound characteristics (since gunshot sound is not narrow-band), work of [15] and [33] are examples of coping with ego-noise and wind effects on a drone. [34] implemented a prototype system that can localize multiple human voice sources in an outdoor environment, with the limitation that the sound sources are only a few meters away from the drone. [35], [36], and [37] are other examples where experimental systems are proposed that are characteristically small in size so that they can be deployed on small UAV systems.

Very similar to the subject of this study, [34] studies sound source localization using an aerial vehicle with a microphone array in an outdoor environment for rescue operations. A use case is reported as finding distressed people in a disaster situation, and the high levels of ego-noise are emphasized. The authors deploy incremental Generalized EigenValue Decomposition (iGEVD-MUSIC), reporting that the selection is convenient for the low-SNR problem. Also, similar to this study varying nature of ego noise is pointed out in this paper, offering an incremental estimation of the noise correlation matrix to adapt to dynamic changes in noise. Although the key challenging points and resolution approaches are similar to this thesis, the paper reports that the solution is convenient for SNRs slightly below 0 dB. This level may be enough for rescue operations where the geometry of the use case suggests the

microphone array moves close to the sound source. However, the SNR level expectancy is much lower for gunshot scenarios with typically large ranges, as given in Section 2.1.

The DREGON dataset reported by [20] should be noted here, which is a rich resource for research on drone sound characteristics and onboard DOA estimation. DREGON is a publicly available dataset concerning DOA estimation using a microphone array embedded in an unmanned aerial vehicle (UAV). The dataset contains clean and noisy in-flight audio recordings annotated with the 3D position of the target sound source.

Extending over the DREGON dataset [20] and the IEEE SPCUP challenge [38], [39] aims at possible search and rescue implementations with drones and reports on improving the baseline methods' performance. The authors propose enhancements over the baseline methods of the challenge by reducing the UAV rotor noise effects with a weighting envelope based on the rotor noise PSD and a post-processing algorithm to improve localization accuracy. The algorithm is reported to bring SNR improvement and localization accuracy. Despite being more precise than the baseline methods, the scenarios are comparable to several meters of range. It is also reported that the proposed method struggled to deliver noticeable performance improvement for the flying drone and broad-band signal scenario.

The SPCUP challenge [38] calculates DOA angle's root mean square error RMSE and the DOA estimation success rate (SR). Apart from the apparent RMSE of the DOA estimation, the success rate gives a basic measure of whether the system could predict the DOA within the correct one-sixth horizontal plane. This metric is important from a functional point of view. That is, it may be used as a basic measure by a potential future application, where the DOA estimation will be used to steer a camera to the direction of interest, similar to the case of [40].

A recent study by [41] reports on direction of arrival estimation on a drone, not for gunshot signals but a similar signal: explosion sound from fireworks. The authors used spectral subtraction to denoise drone ego noise and beamforming with the 32-channel microphone array for DOA estimation. DOA estimation error that starts from 1.4 degrees in the vicinity of the drone going up to 10 degrees for 150 m of range is reported, which is a reasonable performance for this range. However, the errors increase with range, so enhancements are required in the application for a range larger than such a scale. Moreover, the study does not report on automatic detection performances, which is crucial for the gunshot detection problem, given that instant changes in ego noise and wind gusts frequently result in false positives. In contrast, extremely low SNR frequently results in false negatives.

A recent publication by [42] uses similar algorithms to this study, with a different approach, obtaining promising results. With an emphasis on extra noise and false positives due to instantaneous drone movements, the authors first aim at denoising focused on drone sound characteristics. Noise suppression is aimed at the multi-channel recordings through the measured current-specific drone noise spectrum. This is applied with a pre-recording and pre-processing operation, which is a limitation of the study. The proposed method of TDOA-based localization with noise angular spectrum subtraction performs with a fair angular error rate, but does not have a specific focus on the crucial detection aspect other than intrinsically improving false positives performance through denoising.

Similar implementations are not limited to academic studies. Various applications deploy microphone arrays on drones. An initiative is between US company ShotSpotter, and an Israel-based manufacturer of autonomous drones, Airobotics. The former also has stationary microphone array-based solutions for gunshot detection and localization, and the drones are intended to help by moving toward the

detected sound source to collect more information. Therefore, the microphones and the operation are not onboard.

An onboard product is METRAVIB Defence's PILAR system for helicopters. The microphone array-based system performs real-time gunshot threat detection and graphically supplies gunshot location information to the flight crew. The subject of interest is similar to the previous work of the author of this thesis study as given by [6], [43] and [5]. The noise characteristics and scenarios differ between a helicopter and a drone, although the successful operation and the reported importance of the product's real-life use cases are encouraging. A similar product is ATARGET's AHELI. Details on performance are unavailable, although it is reported to be a production-ready helicopter solution.

Another company, Microflown, with its commercial products and technology, should be mentioned here. Microflown has products for ground vehicles as well as helicopters. The technology depends not on microphone arrays as most other implementations do, but rather on the company's own sensor called acoustic vector sensor [44]. This sensor utilizes acoustic particle velocity information instead of regular microphones' sound pressure. Hostile Fire Locator for helicopters was reported to be functional in 2013, and [45] reports on possible uses of the Acoustic Vector Sensor (AVS) technology on drones, including impulsive sound source localization. It is reported that mortar firing with approximately 500 m of range could be detected and localized on a UAV. It should be noted that the mentioned drone is a fixed-wing UAV, which is less noisy, does not offer fixed position (hovering) flights, and is not representative of the popularly known multi-rotor drone concept.

Another ongoing initiative is SmartSound™ which reports its roadmap as controlling the drones with just voice commands in the future, not needing a controller anymore. The company's approach to the problem claims that lowering the propeller noise and improving MEMS microphone performances will pave the way. However, there is no current product offered.

An implementation from the onboard sound processing field is the acoustic Miss Distance Indicators (MDI), designed to determine how close a missile (or similar projectile) was compared to the reference target. A typical MDI is designed to indicate miss distance and the angular position in N-sectors of a passing supersonic projectile. The miss distance is determined by the amplitude of the shock wave, while the angular position is selected from the hit order between the different pressure sensors in the MDI nose. Such devices are already commercial products, and examples include products by the Sweden-based AIR TARGET, QinetiQ Target Systems, Meggitt, and many more. The technology depends solely on the geometric limitation that the bullet is aimed at the microphone array carrier device. Moreover, most of the time, the shock wave for the supersonic projectile is determined since, again, the shock wave cone is known to cover the microphone array due to problem geometry. Therefore, the technology already commercialized in the MDI devices is related to, but does not answer the purposes of this study.

CHAPTER 3

METHODOLOGY

3.1 Examination of the Problem Geometry

The scenario for detecting gunshot signals on drones is not limited to cases where a shooter aims to shoot the drone itself. A drone-mounted system would be helpful if it could locate a ground-to-ground shooter, too. Urban security applications for crowded places would be an example of public security and advantage through superior environmental awareness, and shooter localization is an example from the battlefield. As such, use cases with different geometries are possible. The sound field components and characteristics also depend on the field geometry. Figure 1 visualizes the gunshot fired near a flying drone. Two of the three gunshot sound components are accounted for, giving their paths and speeds. As stated in 2.1, the third component (the mechanical action of the firearm) is not accounted for since it is the least loud and is not considered observable due to drone noise.

As evident from this geometry and as suggested by 2.1, there are two separate signal components that (i) arrive at different time instants, (ii) from a different angle of arrivals, and (iii) have different frequency characteristics. First, examining the muzzle blast component, Figure 3 gives the muzzle blast's propagation path, distance, and DOAs. The signal energy is inversely proportional to the square of the distance d_{muzzle}, so this distance should be used while preparing the data sets to compare algorithm performances.

The total time for arrival of the shock wave component is given by 3.1. As long as the SNR of the shock wave component is concerned (i.e., how strongly the shock wave is audible against the drone noise), the SPL level depends on d_{shock}. In case one is interested in the arrival times of the muzzle and shock wave signals in the time domain, the rest of the components in $t_{shock-arrival}$ are used.

$$t_{shock-arrival} = d_{bullet}/v_{bullet} + d_{shock}/v_{sound} \tag{4}$$

3.2 Preparation of Datasets in Accordance with Problem Geometry

A previous study from the author of this thesis examined gunshot localization on helicopters by comparing performances of traditional source localization algorithms and performing sound field simulations that take into consideration effects of outdoor environmental propagation [5] and [6]. In the

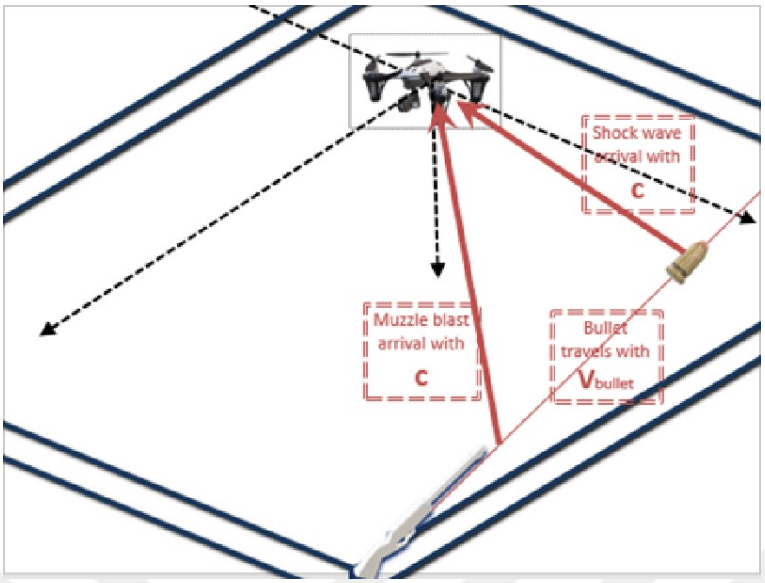


Figure 3: Representative problem geometry and the two main signal components.

literature, such algorithm comparisons are made against different SNR levels. However, comparisons should better be made in terms of geometries and scenarios, considering atmospheric effects, instead of only comparing against different-SNR mixes since the environmental effects result in significant differences in resulting array signals, especially for long ranges. This study uses the environmental sound simulator of [5] so that, instead of just comparing SNR performances, distances and geometries are taken into account together with effects of outdoor atmospheric sound attenuation under outdoor wave attenuation characteristics of ISO9613 [46] and the sound absorption characteristics per [47]. While building the array signals, the geometry given in Figure 4 was considered.

Typical SPL values of gunshots and drones are used, taking into account the SPLs reported by [4], as given in Table 3 and Table 4. Note that the referenced authors measured the SPL over a 9 ms window, beginning approximately 1 ms samples before the detected muzzle blast, to calculate the RMS sound pressure for each shot, with the 94dB recorded calibration tone as the reference. Then, the muzzle blast component is shifted right to account for the time muzzle blast sound travels through the distance d_{muzzle}. On the other hand, the shock wave component is shifted, considering both $t_{bullet} + t_{shock}$.

Table 3: Characteristics of the gunshots used in the dataset

Gunshot type	Projectile speed	Effective range	SPL at 3 m
Colt 45 handgun	<300 m/s)	<30-80 m	137 dB
Rimfire 22 rifle	<330 m/s)	<140 m	119 dB
AR15 M4 Carbine	<900 m/s	<500 m	144 dB
308 Winchester rifle	<860 m/s	<800 m	145 dB

To study algorithm performances, gunshot + drone sound mixes were prepared not by simply making SNR-based mixes, but using the simulator of [5]. The simulation parameters are selected to represent the realistic problem geometry of Figure 3. Different values of d_{shock}, d_{muzzle} and direction of arrival,

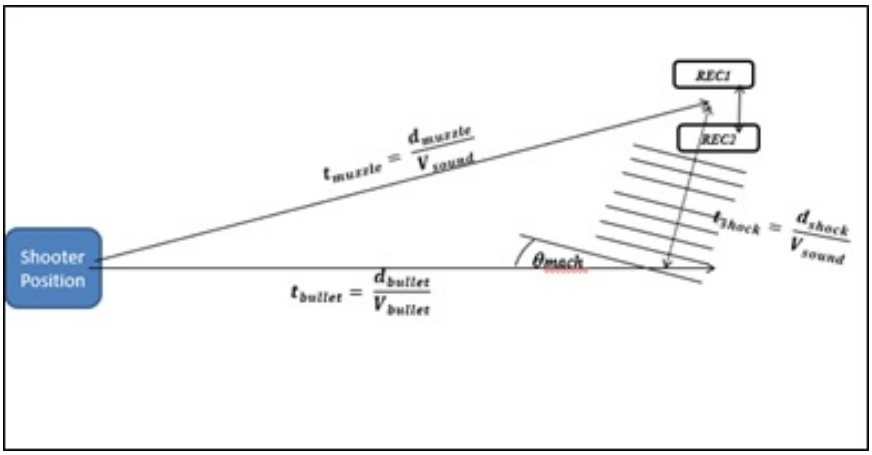


Figure 4: Gunshot components travel time through geometry.

Table 4: Reference drone properties for popular small-scale drones

Drone model	Noise level (dB*)	Drone Diameter (cm)
DJI Mavic Pro	98	21.3
DJI Mavic Pro Platinum	93	33.5
DJI Mavic Air	99	33.5

(*) Measured at 15 cm below drones, similar to an onboard array use case

and different gunshot and drone signals are mixed for a fixed value of environmental parameters, temperature, humidity, and air pressure. Mixes contain:

- drone sound samples from IEEE DREGON challenge [20]. Three different portions of flight data were used: Samples from free flight, hovering and spinning flight
- muzzle blast samples from the gunshot dataset [48]
- shock wave samples from the same gunshot dataset [48]

Table 5 lists details of the dataset concerning geometry and parameter scenarios.

3.3 CWT Analysis of Shock-wave, Muzzle Blast, and Drone Sound

As a starting point, an initial analysis was performed using the MATLAB™ Wavelet Toolbox on the four signal samples collected on a DJI drone with actual rifle gunshots during flight. The data is one channel, the instant of gunshots were manually extracted at post-processing. That was possible since they were faintly audible to human ears, the gunshot distance to drones being 5 to 20 m. Moreover, the video recordings from the onboard camera give visual clues regarding gunshot instant, too. Figure 5 gives the CWT analysis screen on MATLAB™.

Table 5: Details of Dataset Scenarios

Muzzle range (m)	Shock wave miss distance (assuming mach angle = 90 degrees)	Drone samples from the DREGON dataset ([,] represents sample intervals in s)
200 400 600 800 (Maher's dataset is collected at 3 m) (max effective range of guns from Table 3 is 800m)	12 24 (bullet near-misses the mic#1 of Maher's dataset with $\sim$1.5 m)	Freeflight#1: free-flight [13,16] Hoverinig#1: hovering [20,23] Spinning#1: spinning [15,18]

Let's examine the magnitude of the CWT analysis of a gunshot (with shock wave) + drone sound mix, given in Figure 6. The gunshot includes both muzzle blast and shock wave and is known to be at the following sample numbers: one around 1×10^4, three at 2.5×10^4 to $3 \times *10^4$. The instances are visible and distinguishable at CWT output at scales smaller than 10. Despite being a good starting point, this has only limited usage capacity in terms of the following:

• The signal components (both from muzzle blast and shock wave) coincide with drone sound at larger scale values (i.e., lower frequencies)

• Even where it is visible, the magnitudes of CWT outputs for the gunshot signal and the drone noise are comparable. Indeed, such outputs for the signal are smaller than that of the noise for some of the simulations

• What differentiates gunshot acoustics from any other impulsive sound (N-shape signal, positioning of Shock and muzzle, frequency characteristics, etc.) are not distinguishable

• Drone noise has rapidly varying frequency characteristics. The visibility of gunshot components can not be counted on for different drones, gunshot types, SNR levels, as well as different portions of the flight.

Taking the CWT-based analysis of Figure 6 based on CWT, now examine the signal and noise components of the problem domain with a similar methodology. Figure 7 shows frequency characteristics of drone noise from the DREGON challenge's dataset, obtained by summing all data on the time axis and revealing all scales. Doing this for several gunshots and drones, the CWT scale (and so the frequency) content of the signals are compared. The same method is used for a healthy comparison of all sound samples (signal or noise).

The gunshot samples of Figure 7 are gunshots fired towards a microphone and collected by [49]. In [49], it is reported that both shock wave and muzzle blast components are apparent in the audio. However, Figure 7 does not reveal the expected wide-band characteristics of the signals. The problem with this kind of analysis arises because the muzzle blast component is typically 3 to 6 ms, and the shock wave is 0.5 ms [3]. Summing signal and noise for 1 s and comparing them gives faulty results.

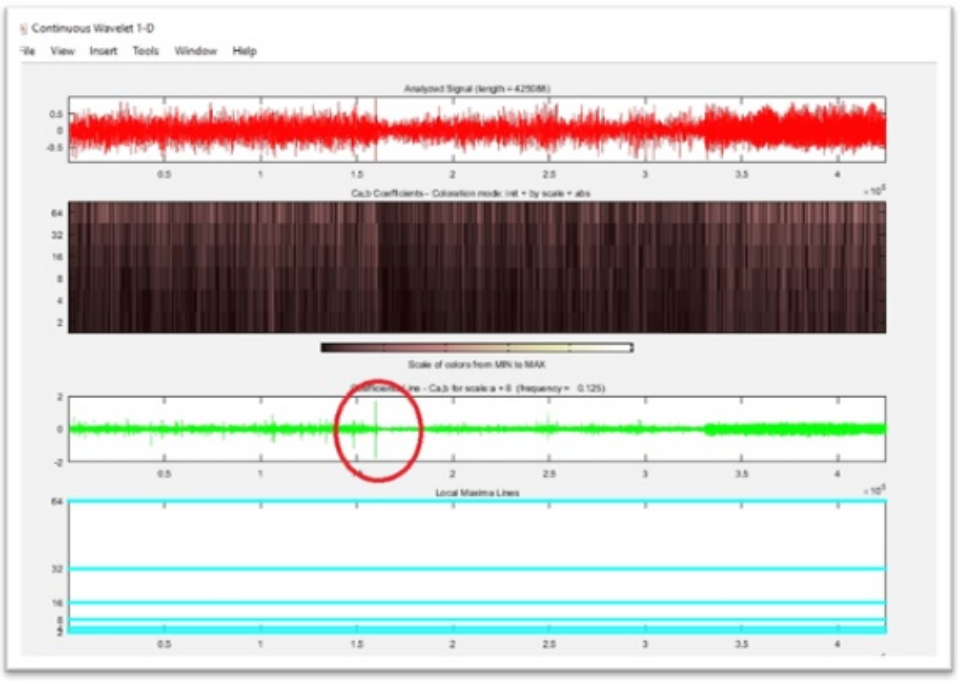


Figure 5: Initial CWT Analysis of a Gunshot + Drone sound signal with Haar wavelet using MATLAB Wavelet Toolbox™

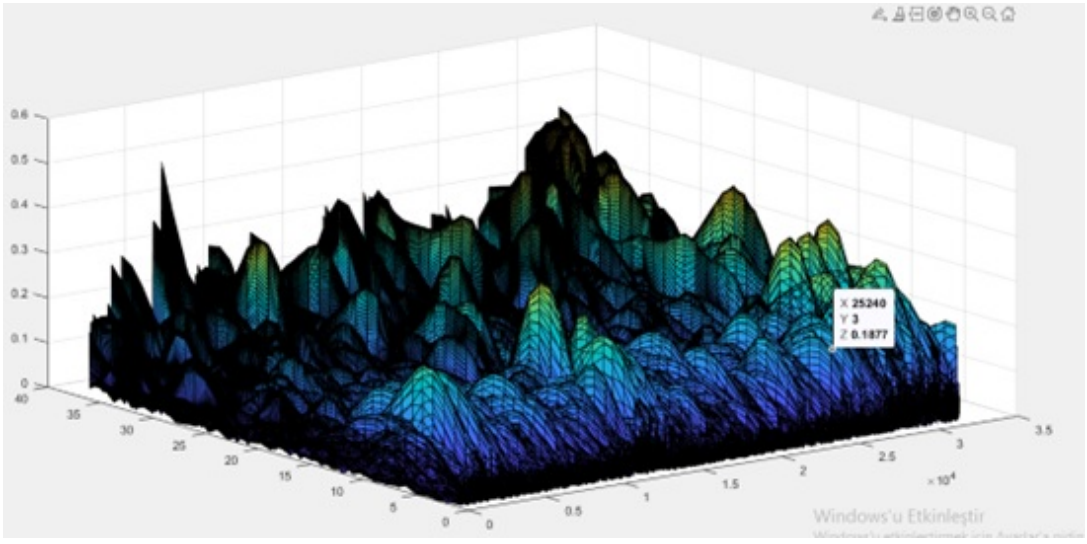


Figure 6: Absolute value of the CWT output for an example mixture of muzzle blast, shock wave and drone sound. It is known that there is one gunshot instance at 10^4 and three instances between 2.5×10^4 and 3×10^4.

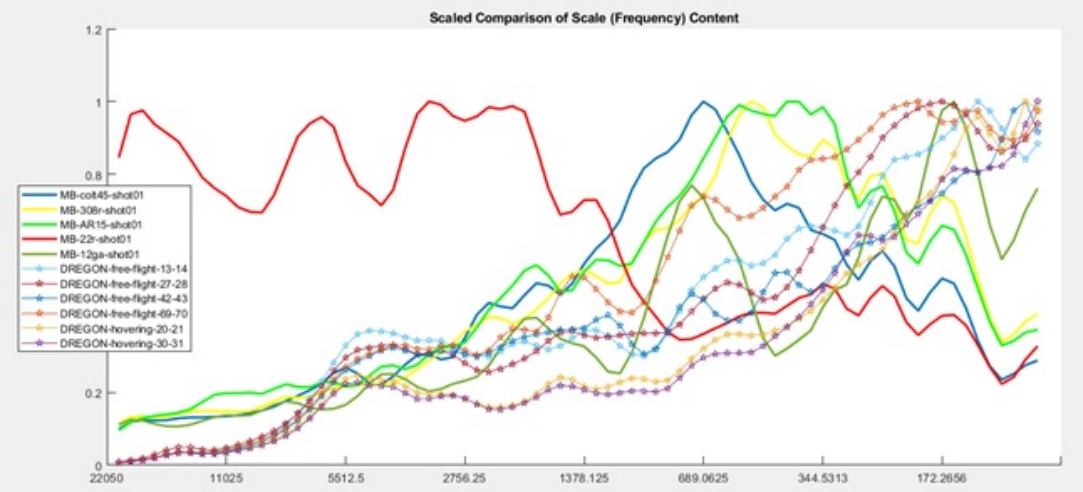


Figure 7: Comparison of drone, muzzle blast, and shock wave audio samples' frequency content - Each signal is scaled to 1 to be visually comparable. The method and visualization are incorrect here, suggesting that time window size should be selected in accordance with typical muzzle and shock wave component duration.

This suggests that the window length should be kept smaller for better performance. This also aligns with the advantageous nature of CWT methods for better time resolution. Repeating the comparisons for both muzzle blast signal vs. drone noise and shock wave signal vs. drone noise for window lengths of 1 ms and 20 ms, we get Figure 7 and Figure 8. In these figures, we observe the advantage of selecting the window size in line with the signal of interest. The CWT outputs corresponding to gunshot signals are far more dominant than drone noise for some frequencies, promising an advantage with CWT scale selection. In case CWT scales are used selectively, the frequencies of interest will be focused without using extra frequency selection filters. Finally, the peaks at the gunshot signal curves of the figure are at expected locations since the muzzle blast component is typically at a frequency interval of 100-1000 Hz.

3.4 Introducing a Novel Method: Array Correlation Map with Adaptive CWT for Gunshot Detection on Drones

To understand the phases of the solution by the proposed method, first, the CWT output of the gunshot + drone sound sample with large d_{muzzle} (and so low SNR) is given in Figure 9. Since different scales of the CWT output correspond to different frequencies, the shock wave 1.167 s is dominant around scales 6 to 40 in Figure 9-1. To better visualize the muzzle blast, Figure 9-2 gives scales that leave the shock wave out. Muzzle blast is at 2.507 s; however, due to the small SNR with d_{muzzle}: 762 m, it can be seen that the detection decision is not confident. Manual selection of scales is possible, though the frequencies (so scales) vary depending on gunshot type, drone type, and between different instants of flight. Moreover, the amplitude and non-stationarity of the drone sound scale/frequency content can be observed in Figure 9, too.

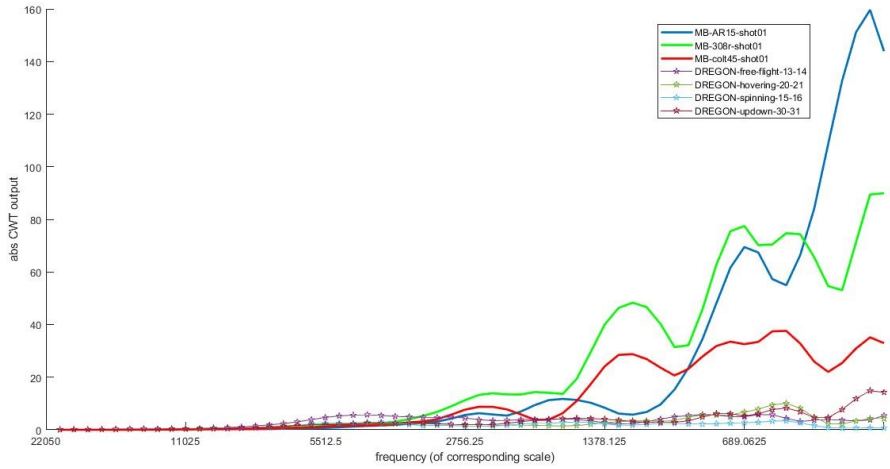


Figure 8: Comparison of drone and muzzle blast over the time duration of muzzle blast, for 9 ms.

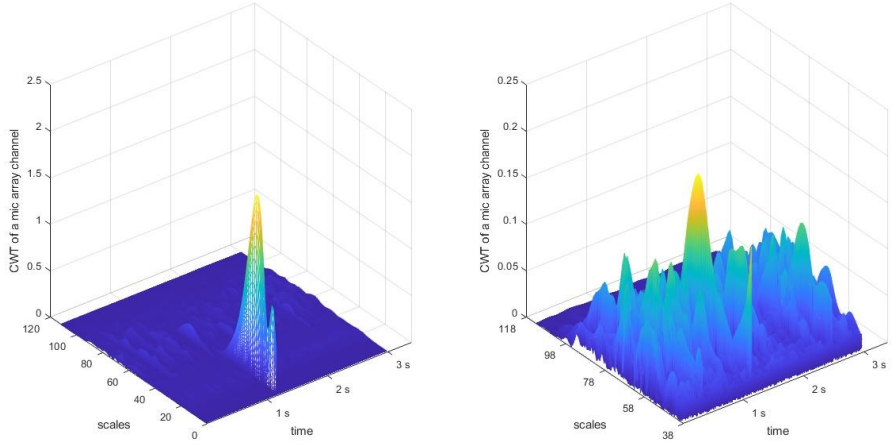


Figure 9: (1) Absolute value of the CWT output for an example mixture of muzzle blast, shock wave, and drone sound. The shock component is dominant since the d_{shock} is 12 m. (2) Shock waves will be there only in certain geometric scenarios. Scale range 38-111 visualized; muzzle blast is at 2.507 s, with a peak at scale of 53, but not dominant.

3.4.1 Array Correlation Map

Cross correlation-based techniques such as GCC estimate the time delay between a pair of spatially separated microphone signals by taking the cross-correlation of the two signals [50]. Since the signals are shifted versions of each other, the cross-correlation function will have peaks at points corresponding to the time delay in between [51]. This method is used for array-based localization, yielding reasonable performance for impulsive signals. If the array has more than two microphones, the microphone data is used in pairs, and a conclusion is made using the multiple-pair outputs. The idea behind correlation maps is to form a matrix corresponding to all pairs. That is, for the N-microphone array, the N-by-N correlation map matrix carries the information for all the microphone pairs of the array. Depending on the application, one can choose between using the maximum of cross-correlation values or the shifts corresponding to the cross-correlation maxima and what to do with the correlation map output.

As stated in Section 2.2, for the microphone array on the noisy drone and directional gunshot sound with far-field assumption, a solution that suppresses the sounds around the array and emphasizes the distant/directional sound may be effective. So, the N-by-N correlation map is formed and used following this requirement, though one may select differently for other array processing applications.

Usually, the shift at cross-correlation outputs is used since GCC uses the correlation for DoA estimation. Here the array correlation map is formed by taking the maximum value of cross-correlation for the whole N x N pairs, and the mean of this maximum is calculated for the whole pairs. This method takes advantage of the directionality and momentary nature of the signal of interest. Since the drone sound is around the microphone array and the remote signal is directional for any pair, there is a consensus regarding the signal component, but not for the noise component. For the sample given in Figure 10-a, when this calculation is applied consecutively for 5 ms windows with 50% overlap, the correlation map of the array reveals that most pairs have consensus only at the muzzle blast instant. The mean value of the pairs is taken as a simple way for making this decision in Figure 10-b. The typical duration of muzzle blast is 1 to 3 ms, and the window size of 5 ms is selected as comparable to typical muzzle blast duration on purpose, making the peaks more obvious.

3.4.2 Adaptive-CWT based Detection

As stated in Section 2.1 and Section 2.2, it is important to have a solution that is adaptive to

- the rapidly changing nature of drone noise
- differences in sound characteristics of different drones
- differences in sound characteristics of different firearms

For the CWT+ACM methodology of this study, this means using the frequency (scales) content adaptively instead of hardcoding the scales specific to a portion of drone flight and specific gunshot sound. First, remember that the correlation map of Figure 10 uses a manually selected scale range of 51 to 55. This range corresponds to frequencies 689 to 522 Hz, which is reasonable from Figure 8: strong with gunshot sound and avoiding the drone-noise-dominant scales.

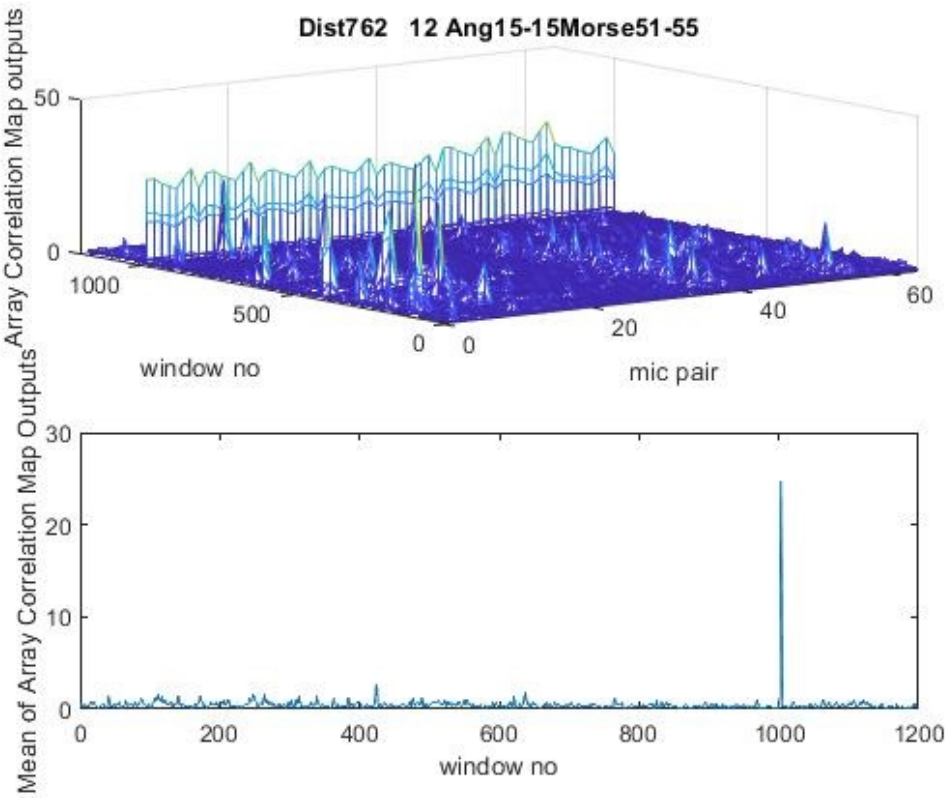


Figure 10: (a) Output of Array Correlation Map, calculated for consecutive 5 ms windows (b) How the pairs agree, calculated through the mean of the Array Correlation Map output for each window.

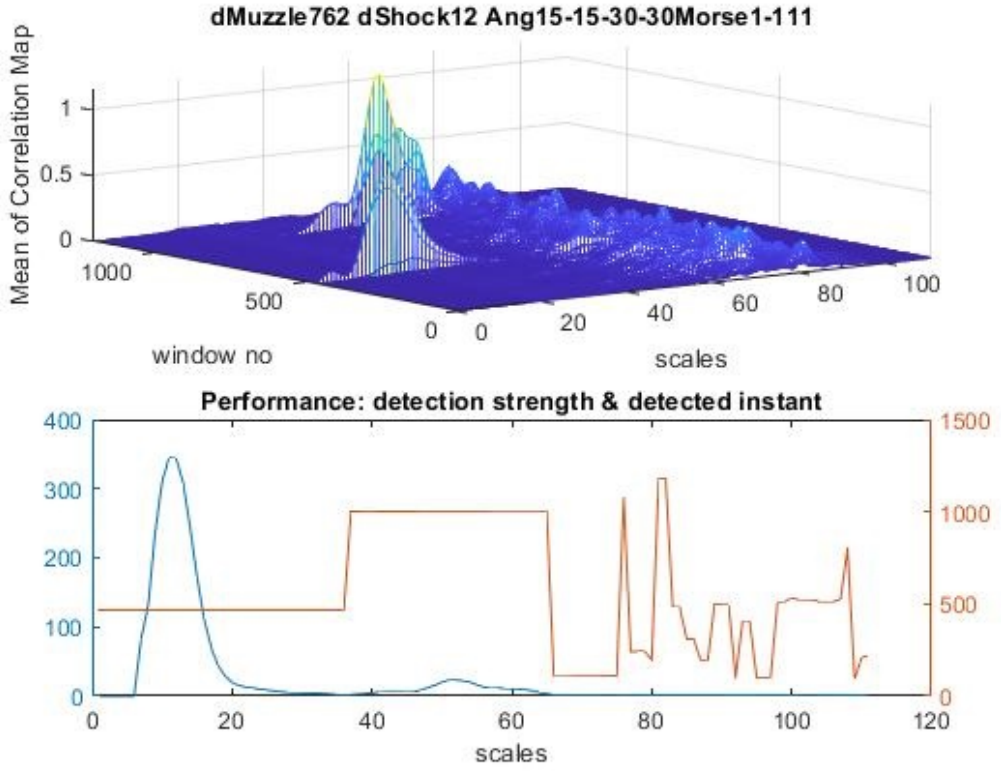


Figure 11: (a) Correlation Map coherence over all scales. Muzzle blast, shock wave, and drone sound are all distinguishable with the CWT + ACM method (b) Detection instant and detection strength pairs for all scales so that the optimal scales can be tuned.

Now, running the same algorithm in a loop for the whole scales individually, we get the output given in Figure 11-a. One can quickly identify the scales emphasizing muzzle blast, shock wave, and drone sound on this figure. To automatize the selection of scales, we need an index that quantifies how dominant is the gunshot-containing window as compared to the neighboring windows. Calculating this index for each scale gives Figure 11-b, which can be used for two purposes:

- which scale/scales should be used for this neighborhood of windows, i.e., tuning to specific scales
- how confident are we with the detection decision, i.e., a metric for detection strength

Figure 12 summarizes the whole method. From its black-box boundaries, the algorithm takes microphone array data of a major frame length (M: 8 microphones, major frame: 3 s for this case), and returns a detection instant and detection strength pair. The tuned scale can be returned, too, which is the optimal scale or scale range the algorithm has tuned to. This may be used not to repeat the tuning process and use the scale determined for once. The whole method can be considered a specific transform applied on array signals and returns these three pieces of information. The "correlation maps over time" and "correlation map coherence over scales" are what was visualized by Figure 10 and Figure 11, respectively. Note that the significant frame is taken as 3 s here, determined by a sim-

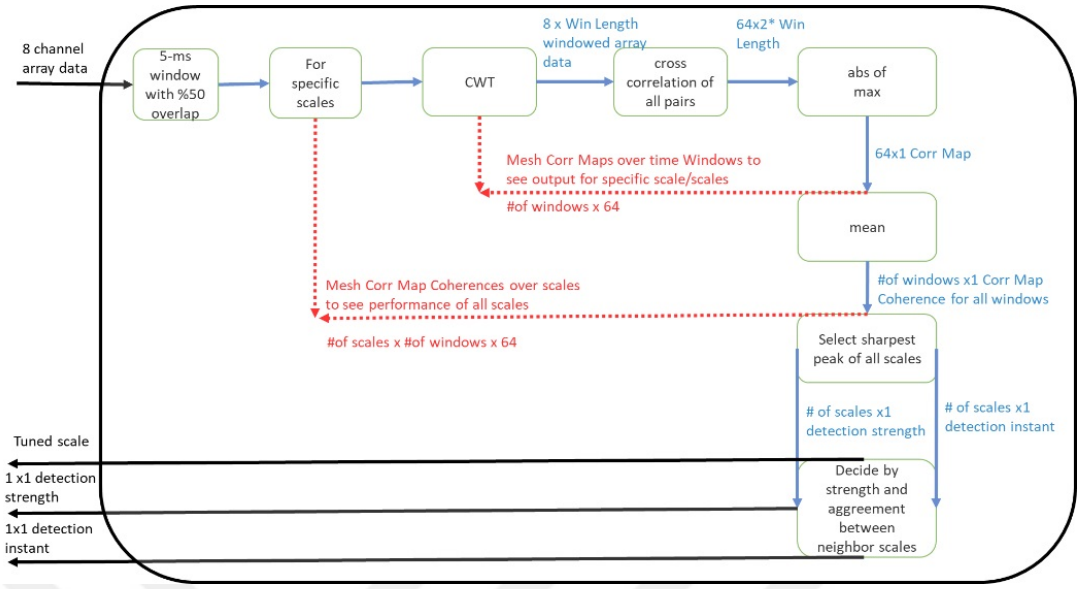


Figure 12: Flow diagram of the proposed method: Array Correlation Map with adaptive CWT for drone gunshot detection.

ple geometric calculation that the arrival time between muzzle blast and shock wave can be 1.32 s at maximum for a sample geometry maximum d_{muzzle} of 800 m and a minimum d_{shock} of 3 m. For the decision iteration, 5 ms windows with 50% overlap are used, comparable to a typical muzzle blast duration of 1 to 3 ms. A window of 50% overlap is applied first, and the remaining parts are done for 5 ms windows. The CWT operations with this smaller window size are suitable for real-time operation, while the sharpness of the detection decision can still be compared among all the scales within each processing window.

A supersonic projectile scenario with a large muzzle range illustrates the advantages introduced by the steps of the proposed method. d_{muzzle}: 762 m , d_{shock}: 12 m, gunshot type: AR15 M4 Carbine, muzzle DOA: elevation-15°, azimuth-15°, shock DOA: elevation-30°, azimuth-30°. In Figure 13-1, the original array signal is seen first. The shock wave is visible at 1.167 s (sample: 41486) since the d_{shock} is small. The muzzle blast is at 2.507 s (sample: 110580). Due to the large d_{muzzle}, there is a large delay in time until the muzzle sound arrives, and the signal is not distinguishable. Using only the shock wave is insufficient since it would be observed only in cases where a bullet passes near the drone, and the cone outward from the bullet trajectory covers the drone.

In Figure 13-2, taking CWT of the signal, the shock wave decision is obvious, but the muzzle is not. Then we apply a scales-selection in Figure 13-3, instead of using a predetermined frequency range of interest, from Figure 11, we learn that the detection strength for muzzle signal has a peak at scale: 53. Using the five scales with best detection strength, we obtain an improved performance at Figure 13-4, where the muzzle blast is first observable. Still, other candidates are at this output due to low SNR and partly overlapping frequency (scale) characteristics of drone and muzzle sound. Adding an ACM-based method to take advantage of the whole array and directional characteristics of the signal of interest, we obtain a clean decision in Figure 13-5, despite a large range (d_{muzzle}) and low SNR.

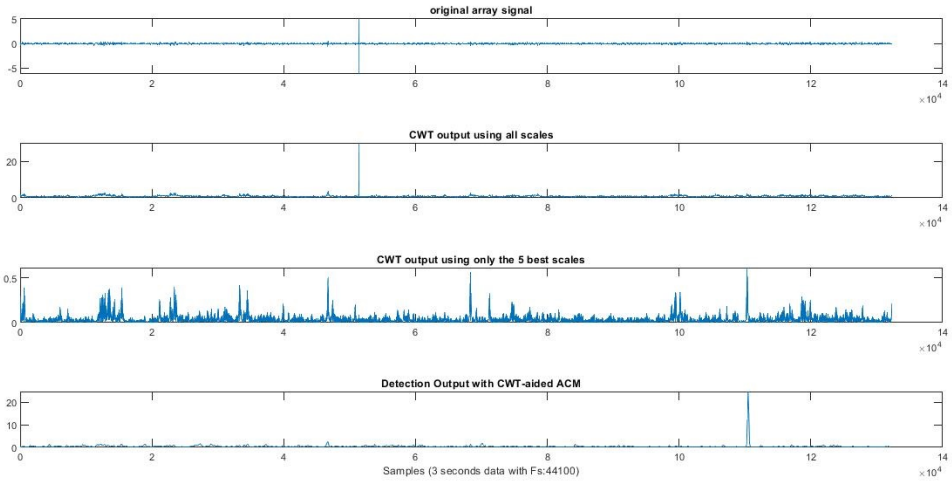


Figure 13: Steps of the CWT+ACM based detection, for an example scenario with muzzle blast and shock wave.

Apart from the ability to distinguish muzzle blast from drone noise, the tune-like feature of the method should be emphasized. In Figure 13-5, the overall output focused on the faint muzzle signal and rejected the strong shock wave signal. This characteristic of the method is demonstrated in more detail in Figure 14. Traditionally, the range of frequencies would be determined in accordance with gunshot and drone frequency content, selecting a range of scales. Figure 14-2 reveals the result of such a selection. The method first tunes to muzzle blast in Figure 14-3 and then to shock wave in Figure 14-4. As reported by [2], gunshot forensics has methods that geometrically derive conclusions by using arrival times, arrival angles, and typical frequency characteristics of the muzzle and shock wave with respect to each other. In other words, one can comment on whether two impulsive sounds are the muzzle and shock wave of the same shot.

For the specific gunshot problem of this thesis, this tune-like feature may be used to decrease the false positives, wind gusts, rapid drone movement, and non-gunshot impulsive sounds. In a generic sense, such a feature may be helpful in tuning to different components of interest, given that they are even slightly distinct in frequency. Simply put, ACM used in this way distinguishes directional and transient event candidates if there is a coherence in the Array Correlation Map and selection on scales of the CWT selects/tunes the candidate that matches with frequency of interest.

3.4.3 Auto-tuning Feature

Of the two main aspects of the proposed method (namely, ACM and CWT), the mentioned adaptivity comes from how CWT scales are used. It is standard implementation to determine the scales of interest in a CWT-based method. The scales correspond to the frequency of interest, and one can hard-code the scale selection in an application. Ideally, a good choice of scales means frequencies where drone noise is weaker and gunshot sounds are strong. However, only estimations of this selection can be made for varying flight portions, drone types, or gunshot types. In case there is a way of selecting the

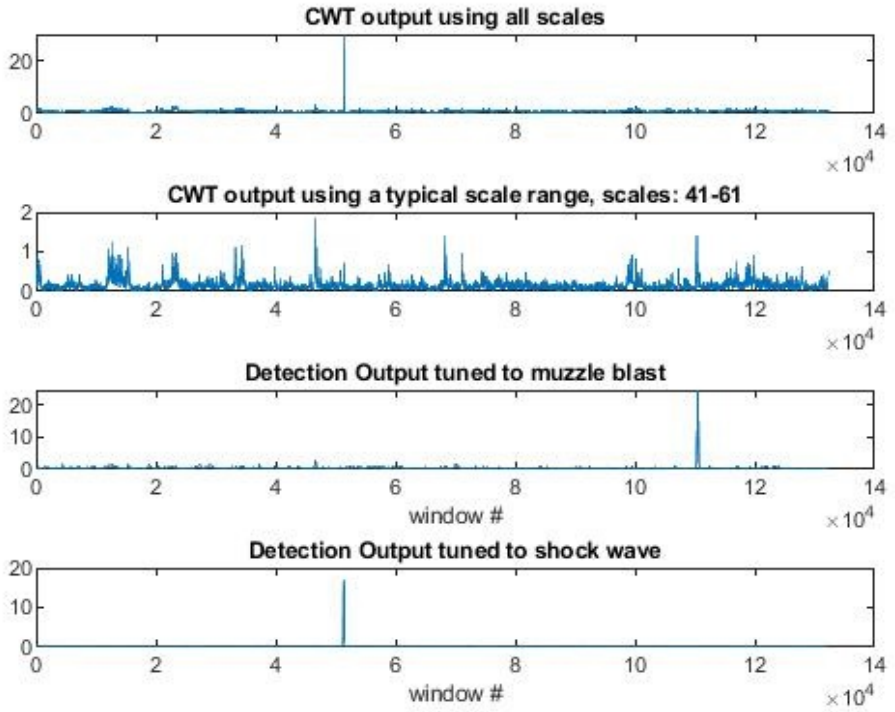


Figure 14: Illustration of tuning capability by using scales of interest: Ability to selectively reject muzzle and shock wave components in the same sample.

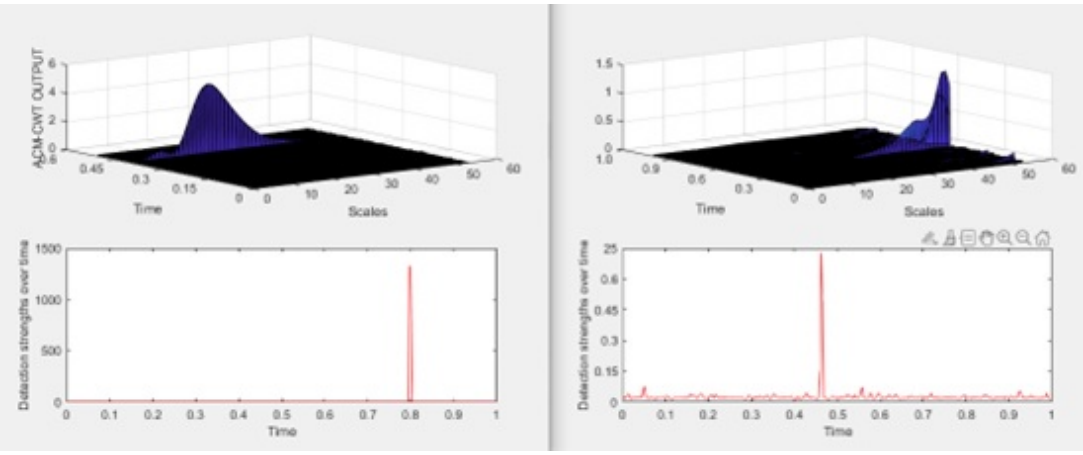


Figure 15: ACM-CWT Output & detection strengths example: For shock wave (left) and muzzle blast (right) components of the same shot.

optimum scales instantly, then this means the method is tuning to scales for the best SNR. As shown in previous sections, the proposed method has a metric on "how certain the detection decision is". Since the ACM and CWT are applied in 5 ms windows, yielding ACM-CWT results for all scales per 5 ms calculation window, this information can be used within a sliding window. At any calculation frame, the ACM-CWT outputs of the last 9 frames are used to calculate the detection strengths for all scales. If the mean strength of all scales is above a threshold, that instant is the detection instant, and the corresponding strength is the detection strength. The threshold is apparent thanks to the sharpening effect of ACM calculation and cleansing of the signals with CWT. Moreover, since all scales' detection strength value is known, the most potent scales are determined, too.

The following figures exemplify the process. Figure 15 reveals algorithm outputs and resulting detection strengths calculated for whole 1 s samples. As expected, the scale differences between shock wave and muzzle blast components and a stronger detection calculation for the shock wave can be seen. Now, zooming into the above-mentioned 9 frames corresponding to the muzzle instant of Figure 15, Figure 16 reveals what each scale had calculated as detection strength. The overall detection strength per time frame is the mean of all these scales. The two sub-figures of Figure 16 are parts of the muzzle sample of Figure 15, taken at different times. These two instants in Figure 16 reveal the difference in ACM-CWT outputs and detection strengths when a signal of interest is present.

Naturally, scale tuning is achieved where the maximum mean value of detection strengths for all scales gives detection instant, and the scale yielding the maximum strength at that instant is the scale center to tune. This scale represents an automatic frequency selection that changes in time to yield maximum performance.

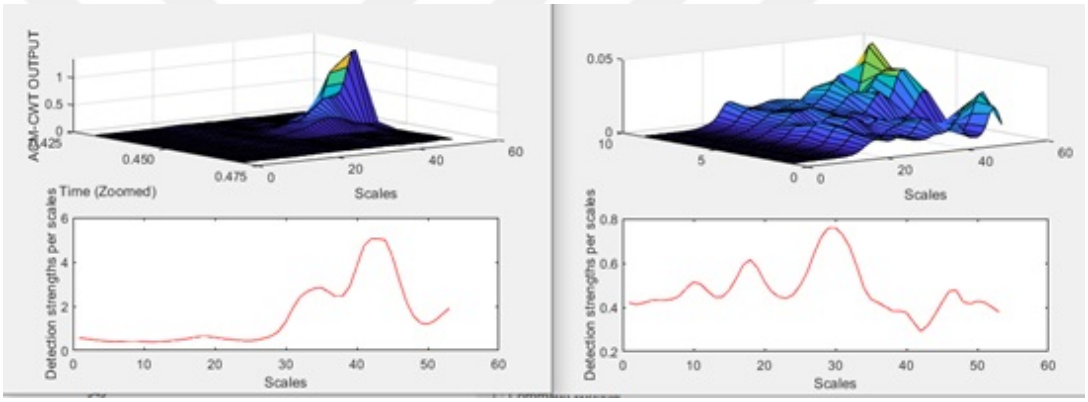


Figure 16: Detection strength over all scales. Zoomed in time to muzzle detection instant (left) and a no-signal instant (right).

CHAPTER 4

SNR IMPROVEMENT RESULTS

This study aims to implement gunshot sound detection and DOA estimation from the sensors integrated into a drone, despite the extreme noise conditions. A novel solution is offered, claiming that the proposed method makes the originally unavailable detection decision certain and precise in time. This precision improves detection and DOA estimation error performance. Despite being focused on the detection problem, the outcome of the proposed solution is an adaptively enhanced signal. Therefore, the gain is expected to be general to the detection and DOA estimation problems. To concretize this claim, the improvement in the signal is represented as pre/post-SNR calculations.

Despite being focused on the detection and DOA estimation problems, the gain from the proposed solution is actually through adaptively enhancing the input signal. In other words, the benefit is expected not to be specific to the detection and DOA estimation, but is a way to obtain an SNR-improved signal focused on impulsive directional signal components. In this regard, this chapter reports on SNR improvement that was suggested by the proposed method. The improvement is visually revealed on time domain signals and through SNR calculations.

4.1 Visualizing SNR Improvement by the Proposed Method

The SNR improvement is visually observable on noisy signal samples. As an example, Figure 17 illustrates how the proposed method's original array signal is improved. The data corresponds to the scenario of MB308r shot01 with d_{muzzle} of 400 m, azimuth of 30°, elevation of 30° and d_{shock} of 24 m. The whole 3 s array data is given first. The shock wave is already visible for this scenario at 0.55 s, and the muzzle blast is known to be at 0.88 s, but is not visible. Thanks to the proposed method, the precise impulsive signal location and scales are automatically determined. Now, one can focus on the 10 ms window (comparable to the gunshot muzzle blast signal duration) around the known detection instant. We take the inverse CWT of the array signal for the scales corresponding to muzzle detection to obtain the SNR improvement by selecting the CWT scales. Remember that the proposed method outputs the scales corresponding to muzzle detection. This is a reconstruction of the array signal focused by the proposed processing on which the muzzle blast is visible. Compare this to the same instant in the original array data, where the muzzle was not visible.

Similarly, the same Figure 17 reveals enhancement for the shock wave signal. The selected scales are different for this part since the algorithm tunes to strong scales for shock wave signals. The shock wave is visible. It can be asserted that it was also clear at the original array signal. However, it should be remembered that the miss distances are small in the simulation scenarios. Due to gunshot sound

geometries, one can not always depend on a shock wave component. More importantly, apart from the visibility of the shock wave peak, eliminating the noise around is important for rejecting false positives. Functionally speaking, these all are evident in the resulting F-score and ER metrics in the detection results.

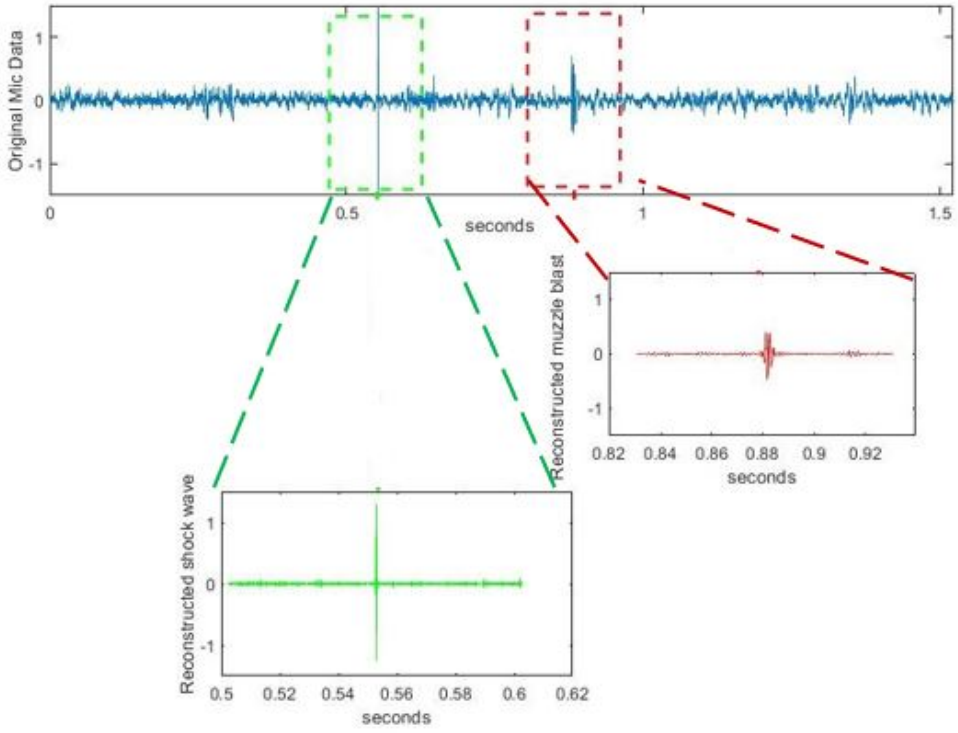


Figure 17: Original microphone data and improved muzzle and shock wave signal visibility with precise detection and automatic CWT scales selection.

Figure 18 reveals another example with d_{muzzle}: 800 m. This time, the muzzle is known to be at 2.61 s. Since the muzzle range is very large at 800 m, this is a more challenging scenario, and the muzzle is neither visually seen nor audible in the original array signal. The shock wave is visible in the figure at 1.23 s, since the simulation scenarios only cover small bullet miss distances, which is d_{shock}: 24 m. Notice that the known truth table entries for muzzle and shock wave times are larger in this scenario due to the bullet travel time and sound wave propagation time. Also, compare the magnitudes of the signal components with that of Figure 17. The smaller magnitudes are expected due to the propagation of the signal energy with a larger muzzle range. The "cleaned" muzzle wave at the method's output has a smaller amplitude, too; but, more importantly, both the muzzle and the shock wave signals are visible in Figure 18, despite the extensive range of d_{muzzle}: 800 m.

Another interesting aspect arises when examining the two figures together. Since in both scenarios, the bullet misses the array with a small distance; the shock waves are visible in both Figure 17 and Figure 18. The visibility of muzzle blast and the amount of noise is different in that, as expected, the d_{muzzle}: 800 m case has much higher noise. However, the resulting "cleaned" muzzle blast signal is apparent in both figures. This means promising performance, eliminating or reducing the effects of low SNR caused by larger muzzle ranges.

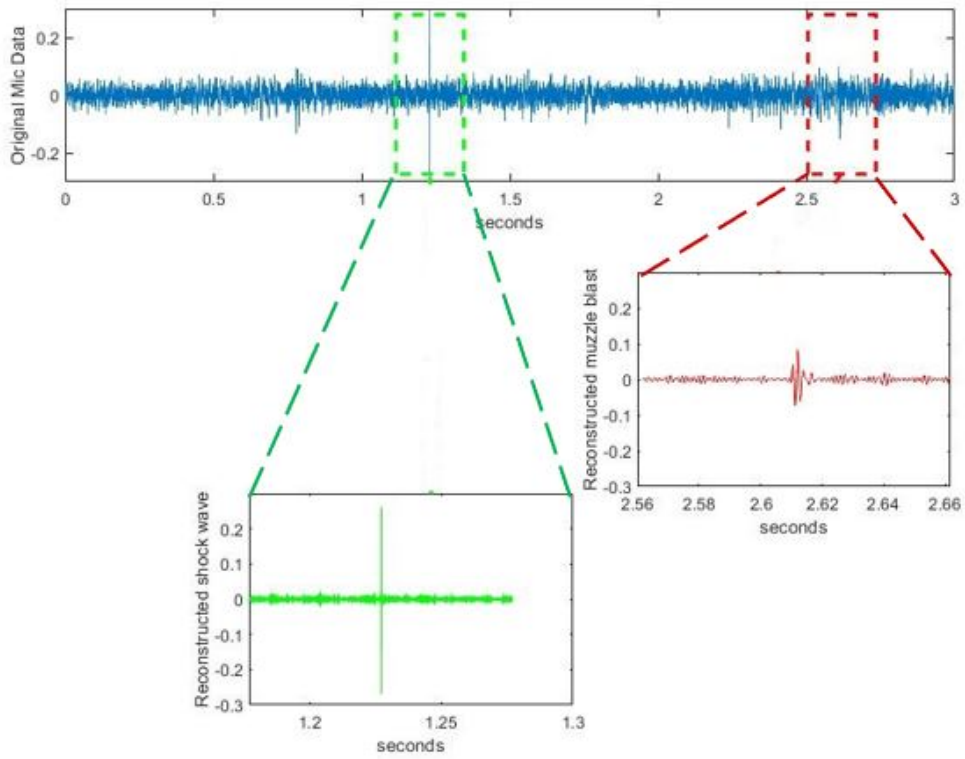


Figure 18: Original microphone data and improvement on muzzle signal visibility for the range of 800 m with precise detection and automatic CWT scales selection

4.2 SNR Calculations Before / After the Proposed Method

Figure 17 and Figure 18 are singular examples of the SNR and range performance for visualization purposes. The previous section reports promising aspects of range performance in a more generic sense. Apart from visually examining samples of the mixture signals, SNR values before and after applying the proposed selective CWT analysis were calculated as proof of denoising success.

Table 6 lists SNR values for different scales before/after the application of the proposed method for the same scenario of Figure 17 above. Since the typical muzzle signal length of interest is 3 to 10 ms, and again in line with the SNR calculation methodology of Maher's data set, the SNR is calculated for 9 m. Given the sampling frequency of 44100 Hz, starting with 90 samples before the known signal (muzzle/shock wave) time and ending at 310 samples afterward. Therefore, even the SNR at the original data (i.e., the first row) is high once the detection instant is known. Note that the proposed method offers detection time with high resolution in time.

Examined channel-wise, all SNRs are increased by applying the method with tuned scales (which is also an output of the method). The table also reveals that when improper scales (eliminated by the method) are used on purpose, there is a significant drop in the SNR values. While considering the table's proper and improper scale ranges, note that the detection algorithm tuned to the scale of 52 for muzzle blast and the scale of 20 for the shock wave for this scenario.

An interesting point in Table 6 is the difference between the SNR values of different channels. Taking the first row as an example, the SNR values of the eight channels vary between 0.66 to 10.11 dB. As another example, at the row corresponding to the scale selection appropriate for muzzle blast, the SNR values vary from 11.7 dB to 35.7 dB. A similar case holds for other scenarios, too. This is due to different noise levels of the microphones since the noise (i.e., drone sound) signals are actual array recordings on drones. Therefore, uncorrelatedness and SPL level differences amongst different microphones are expected for a real array recording. Indeed, the situation is more realistic, which is a strong attribute of how the sound simulator of this study is arranged. The whole array data is not synthesized by generating N-signals for N microphones. Instead, the gunshot sounds are mixed into the channels of actual drone noise microphone array recordings. This brings real noise data, more realistic noise+signal microphone array signals, realistic physical noise differences at microphones, and actual recordings of uncorrelated noise between channels instead of just assumptions. In this aspect, the important point to remember is to use the data in Table 6 to compare different calculations of the same channel while deriving conclusions. That is, comparing values on the same column to compare SNR values at the same array channel under different scenarios.

Table 6: Example SNR values for original and denoised array data: MB308r shot01 muzzle range: 400 m, azimuth of 60 degree, elevation of 60 degree, shock miss distance of 24 m, shock azimuth of 30 degree and elevation of 30 degree

Array channels	**ch-1**	**ch-2**	**ch-3**	**ch-4**	**ch-5**	**ch-6**	**ch-7**	**ch-8**
Muzzle blast SNR at the original array data	0.6607	4.0525	5.4328	7.4373	5.7964	10.1193	7.8049	9.5281
SNR for the scales tuned to muzzle blast (42-62)	11.7047	35.7193	31.8628	27.1084	25.5669	20.5034	17.7906	32.3320
SNR for improper scale range selected on purpose (67-87)	-8.6603	-11.445	-10.677	-10.329	-12.426	7.853	1.860	-3.383
Shock wave SNR at the original array data	1.2158	-0.9824	4.3190	4.9655	6.0606	6.0522	2.2919	4.3270
SNR for the scales tuned to shock wave (15-25)	17.3078	17.7613	39.4572	26.1174	22.2668	23.7322	26.7810	14.4575

Table 6 gives the SNR improvement for only a single simulation scenario. However, quantizing the SNR improvement with a single metric is important so that it can be calculated for many simulation scenarios. This way, a more generic SNR improvement conclusion can be achieved. The SNR improvement for muzzle wave can be defined as the mean (overall array channels) of the difference between the SNR of the array data reconstructed with muzzle-tuned scales and the SNR of the original array. The SNR improvement for the shock wave is defined in the same manner. In Table 6, the mean SNR improvement for muzzle blast is 18.7 dB, and the mean SNR improvement for shock wave is 19.9 dB.

4.3 Conclusions on SNR Improvement

Now that we have a single number that quantifies SNR improvement for a simulation sample, the SNR improvement was calculated over 54 different simulation scenarios. This way, SNR improvement characteristics over different factors and scenarios are obtained. Table 7 gives the mean SNR improvement results.

Examining Table 7, the SNR improvement results are in line with the detection performance results given for the muzzle range and the muzzle DOA, Figure 21 and Figure 22 respectively. First, the SNR improvement performance does not depend on the DOA of the impulsive signal for the proposed method and this array geometry. Regarding the range, for the long-range scenarios with low SNR, there is significant mean SNR improvement (6 to 11 dB for the muzzle blast, around 19 dB for the

shock wave component). This enhancement is what lies behind the promising long-range performance reported in the detection performance result in Section 5.2.

Table 7: Mean SNR improvement calculated over several trials of different scenarios

All scenarios with...	Mean SNR improvement (dB)
Muzzle range: 200m	6.9344
Muzzle range: 400m	11.8901
Muzzle range: 800m	10.4543
Shock miss distance: 12m	19.5622
Shock miss distance: 24m	19.3174
Muzzle DoA: 45	10.7078
Muzzle DoA: 60	10.0033
Muzzle DoA: 75	10.5619
Muzzle DoA: 90	10.7457

Using the same mean SNR improvement metric of Table 6, Figure 19 and Figure 20 reveal SNR improvements for many simulation scenarios. Each data point in these figures represents an eight-channel mean of SNR improvement in a simulation scenario. The cases marked with SNR improvement = 0 dB in the figure are those without detection, so SNR improvement is undefined. They are included in the figures for completeness so that all the scenarios' results are accounted for in these figures. Examining Figure 19, we can assert the scenarios with a 200 m range resulted in less SNR improvement. This may be because, for 200 m, the SNR is already not too low. Still, it seems that for any range, the SNR improvement varies between 5 dB to 20 dB and is related to the simulated scenario (the specific gunshot, noise, etc.) rather than the muzzle range. Similarly, Figure 20 visualizes SNR improvement for all samples if the algorithm could tune to CWT scales (i.e., there was successful detection). Entries with SNR improvement : 0 dB in these figures represent cases where the method did not tune to CWT scales (did not detect the signal). The amount of improvement is scattered and is not correlated directly with whether the bullet miss distance is 12 m or 24 m.

Although both figures reveal scattered scenarios regarding SNR improvement, comparing the two figures may be interesting. We can conclude that all scenario runs resulted in SNR improvement for both the muzzle blast and shock wave components. We can also conclude that the improvement at the shock wave component is more significant at 15 dB to 35 dB compared to that of the muzzle blast, 5dB to 20 dB. This can be expected since the shock wave is more distinct in frequency that, once you find the scales (or frequencies) to focus, the signal of interest is much more "focused" in the shock wave case.

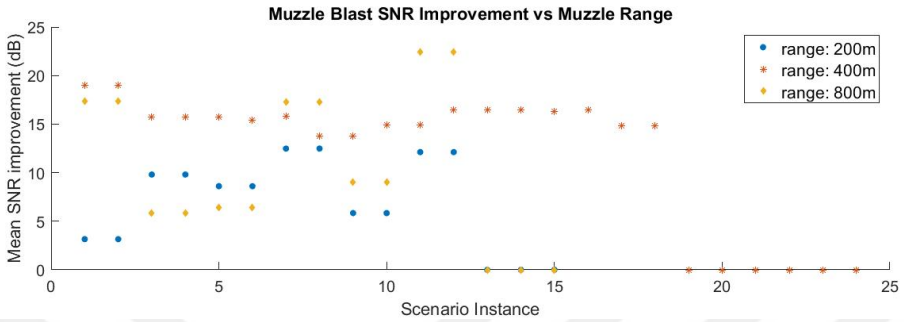


Figure 19: Mean SNR Improvement vs Muzzle Ranges

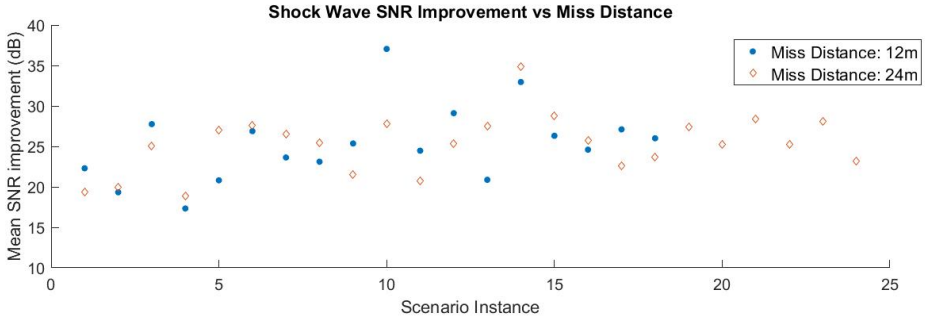


Figure 20: Mean Shock Wave SNR Improvement vs Bullet Miss Distances

CHAPTER 5

DETECTION RESULTS

5.1 Performance Assessment Metrics for the Detection Problem

In such a noisy environment as drones, gunshot forensics is primarily a detection problem. With the proposed method, the detection part should work successfully for the DOA estimation to be successful. More specifically, the performance of DOA estimation (or any other future application) depends on the fact that the proposed method results in precise resolution in detection time and good detection decisions despite extreme noise conditions. Therefore, due to the nature of the problem, the first critical performance assessment should be done on the detection.

It should be noted here that this thesis uses two different metrics for the two main problems it addresses, namely the detection and DOA estimation parts. Although this is the usual approach due to the different nature of the two problems, it is also possible to utilize a single metric for the two aspects in a combined manner. A novel metric is offered by [52] to jointly evaluate the localization and detection performance of systems aimed at sound event localization and detection. Concerning metrics selection, this study follows the usual approach and uses two different metrics for detection and DOA estimation problems.

For the detection problem, metrics of the IEEE Detection and Classification of Acoustic Scenes and Events (DCASE) challenge, Task 2: Detection of Rare Sound Events, are considered a convenient example [53], [54].

Event-based metrics detail how to assess the detection performance in an advanced and comprehensive manner. An event is considered correctly detected if its position in time overlaps the position of an event with the same label in the ground truth. A tolerance will be allowed for the onset and offset, which will be determined later. Amongst the detection and classification metrics described in [55], the following intermediate metrics are suitable for the gunshot detection problem of this study for both muzzle blast and shock wave detection:

• True positives TP: correctly detected gunshot event. The event was detected for a 1 ms length sample, including a gunshot, and the timing is right within 50 ms.

• False positives FP: falsely decided that a gunshot event existed. The 1 ms length sample did not include a gunshot, but the algorithm decided there was one.

• False negatives FN: The 1 ms length sample included a gunshot, but the algorithm decided there was none.

• True negatives TN: The 1 ms length sample did not include a gunshot, and the algorithm decided that there was none.

• reference events N: number of events in the ground truth

Using these metrics, the assessment outputs error-rate (ER) given in Eq. 5 and F-score given in Eq. 6 can be calculated:

$$ER = \frac{FP + FN}{N} \tag{5}$$

$$F - SCORE = \frac{2PR}{P + R} \tag{6}$$

where

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN} \tag{7}$$

5.2 Detection Performance wrt Muzzle Range

As discussed earlier, one of the main challenges of the problem is low SNR. Despite the high amplitude gunshot sounds, the factor that brings low SNR is the range, the distance between the muzzle blast origin and the microphone array on the drone. The detection performance is expected to naturally depend on SNR, which is directly related to muzzle range or the shock wave miss distance. In this aspect, it is equivalent to examining the detection performance with respect to either SNR or these ranges. Handling the problem from a functional point of view, the method's performance is reviewed in terms of distances: the muzzle range and the miss distance.

To reveal performance change with the muzzle range, samples with different ranges are compared, and each range includes the same pool of the remaining parameters, shock distance, DoA, flight portion, and gunshot type. Figure 21 gives muzzle blast detection performance with F-score and Error Rates for different muzzle ranges. It should be noted while comparing muzzle range performance, the dataset includes a wide range of 960 samples. The dataset details used are given in 13, and the same dataset is used in the detection and DOA estimation results chapters.

The first impression from the figure is the expected performance decrease trend with increasing range. The ER values increase, and the F-score values decrease with increasing range. Still, the detection has a promising performance since the F-score values are fairly high for all ranges, considering the extra-low SNR. Considering the spherical signal attenuation and the SNR drops that naturally follow, this is a promising result. It should be remembered that typical gunshot scenarios cover ranges of several hundred meters, and there is plenty of work in the literature with successful detection, DOA estimation, or localization methods for small-range applications. It is important that the technique increases the range and gives a stable performance, irrespective of the range, for several hundred meters.

Examining the unsuccessful scenarios, we know from a detailed examination of the False Negatives that the unsuccessful samples belong to the Rimfire 22 rifle. This results from the weapon's extra low SNR, since the typical SPL of Rimfire 22 rifle is significantly lower than the other weapons, as given by Table 3. Since this SPL difference directly reflects on SNR, the small ranges for this rifle are more comparable to the larger ranges for the other weapons. As a justification for this, once the detection

performance of the Rimfire 22 rifle is examined, we see that its detection metrics are also improved with smaller muzzle ranges.

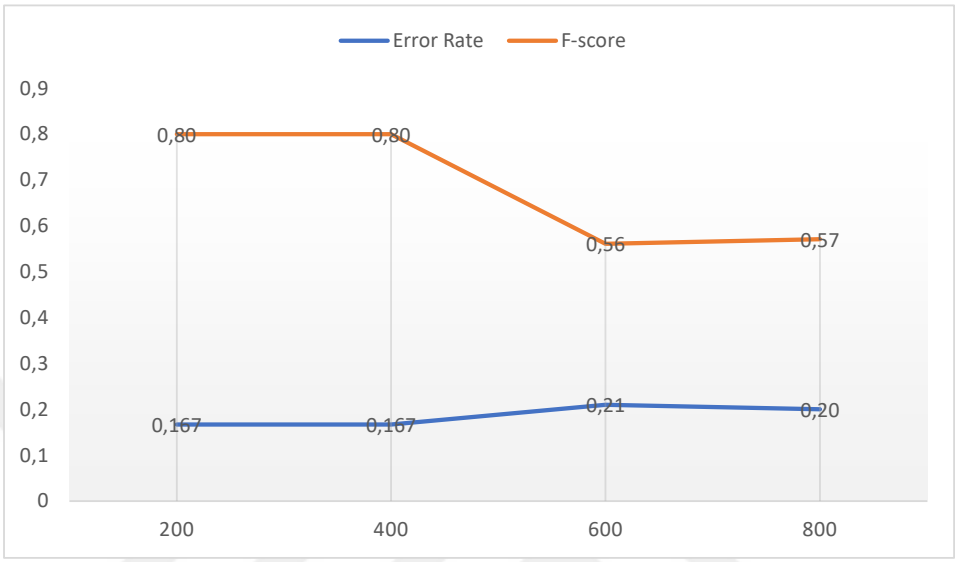


Figure 21: Muzzle Blast Detection Performance wrt Muzzle Range in m.

The details of TP, TN, FP, and FN values are given in Table 8. Note that the rejection performance for the negative samples is better than the values of the table suggest, considering that the method runs in 5 ms. windows with 50% overlap and for a true rejection in the 1 s. long sample, all 400 windows produce negative results. That is:

- A muzzle or shock wave sound is considered detected (TP) if the detection instant is within 50 ms of the known truth table
- FP is decided if there was not a signal **or the detection instant is not correct**

5.3 Detection Performance for Different Muzzle Direction of Arrivals

While examining the effect of DoA on detection performance, first, it should be noted that the offered detection method does not depend on the usage of DoA information. Moreover, the array geometry does not have a specific orientation for any DoA. Still, the detection method uses GCC and takes advantage of signal phase differences. To assess the correlation of the muzzle detection performance with DoA, the muzzle detection performance with respect to DoA is visualized in Figure 22.

Table 8: Muzzle Blast Detection Details for Different Ranges

Range (m)	200	400	600	800	TOTAL
TP	80	80	32	32	224
TN	120	120	158	160	558
FP	40	40	42	40	162
FN	0	0	8	8	16
ER	0.167	0.167	0.21	0.20	0.19
F-SCORE	0.80	0.80	0.56	0.57	0.72

Table 9: Muzzle Blast Detection Details for Different DOA Angles

DOA (Degrees)	45	60	75	90	TOTAL
TP	56	56	56	56	224
TN	138	140	140	140	558
FP	42	40	40	40	162
FN	4	4	4	4	16
ER	0.19	0.18	0.18	0.18	0.19
F-SCORE	0.71	0.72	0.72	0.72	0.72

It should be noted while comparing DOA performance, the dataset includes a range of 960 samples. Each DOA covers the same range pool, shock miss distance, flight portion, and gunshot-type scenarios. The dataset details used are given in Table 13, and the same dataset is used in the detection and DOA estimation results chapters.

Figure 22 reveals nearly the same ER and F-score values for different DOA scenarios, and Table 9 gives the TP, TN, FP, and FN details that result in these results. It can be concluded that the muzzle detection performance does not depend on the direction of arrival of the muzzle signal for this array geometry. Considering the characteristics of the proposed method, this is as expected. The ACM usage of the technique emphasizes the highly directional signal of interest instead of the noise of the drone propellers, which are positioned in every direction. However, while doing so, the specific directions are unimportant; the method does not depend on or use any direction information. Besides that, the selective usage of CWT scales corresponds to a frequency selection and is unrelated to the signal's DOA.

5.4 Improvement of Detection Performance with Automatic Tuning of Scales

Other than reporting the performance of the proposed method for different ranges and DOA angles, it is important to report the effect of automatically selecting scales. With this purpose, as an additional analysis, the detection phase was applied to a dataset with the following details, with and without the scale selection of this study. The comparative detection performance was reported regarding F-SCORE and Error Rate as defined by Section 5.1.

- There are 90 samples, 1 s. long each, 30 of which contain muzzle blast

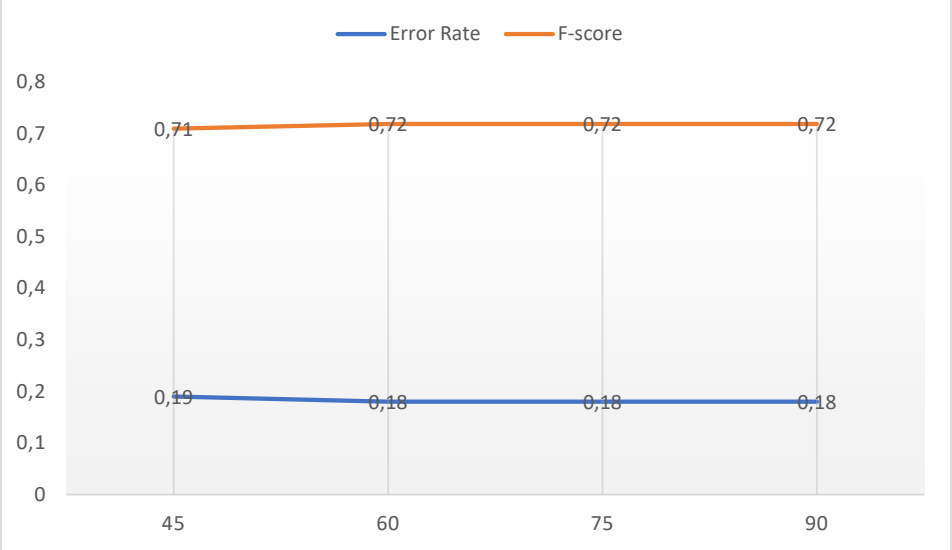


Figure 22: Muzzle Blast Detection Performance wrt Muzzle DoA in Degrees

- 54 of the 1 s. long samples contain shock waves, too
- DoA of muzzle blast is [15, 45, 60] azimuth & elevation
- DoA of the shock wave is [30] azimuth & elevation

Repeating the method with scales suitable for the specific gunshot type, the muzzle blast and shock wave detection performance is calculated. Table 10 gives the detection metrics for 90 samples with 30 known muzzle blasts and 54 samples with 18 known shock wave signals. The columns "Selective scales" reveal the improvement in the scale selection of the proposed method. This is more evident if the hard-coded scales align with drone and gunshot characteristics. That is, the range of 30-50 corresponds to the frequency range of 2954 Hz to 738 Hz, which is suitable for typical muzzle frequency characteristics and is as far as possible from the drone spectrum. Similarly, 5-25 corresponds to 16710-4177 Hz and is suitable for covering different gunshots' shock waves. Still, their performance is limited, and the proposed method's scale selection brings significant improvement.

The comparisons of Table 10 are also crucial in showing the significance of the CWT side of the CWT+ACM-based solution. Although the ACM concept would suppress the sounds around the array and emphasize the distant/directional sounds, the frequency-characteristics-based selection is crucial in gunshot detection. This leads to another critical aspect of the CWT-based solution. 54 of the 90 samples include muzzle blast and shock wave content at different instants. This is a challenging factor, especially for False Positives. Still, the method's False Positives performance is fairly high, thanks to the scales selection, verifying the claims of Figure 21 and Figure 22.

Table 10: Detection Performance for all data / Muzzle blast and Shock Wave components

	90 muzzle blast samples		54 shock wave samples	
	Manual scales 30-50	Selective scales	Manual scales 5-25	Selective scales
TP	15	22	0	18
FP	1	1	12	0
TN	59	59	27	36
FN	15	8	15	0
ER	0.178	0.1	0.5	0
F-SCORE	0.65	0.83	0	1

5.5 Receiver Operator Characteristics (ROC) Curves for the Detection Performance

Since previous sections reveal promising performance for detecting gunshot components, it is important to prove that the method is helpful in all aspects expected from a detection classifier. As introduced by Section 5.1, for a detection problem, a method's success at detecting true positives is not sufficient. A detector may be 100% successful at detecting true positives, but also return positives for so many negative cases that it is not useful. A well-known method for assessing the actual performance regarding concerns such as this is using receiver operator characteristics (ROC) curves.

ROC curves are frequently used to graphically show the trade-off between sensitivity and specificity for every possible cut-off for a detection or classification algorithm. As stated by [56], a ROC graph is used for visualizing, organizing, and selecting classifiers based on their performance. An early adoption of ROC curves for machine learning applications was [57], which utilizes the analysis for comparing algorithm performances. For the purposes of this study's detection part, the benefits may be summarized as follows:

- Visualize the trade-offs between sensitivity and specificity
- In case of a threshold-based decision in the algorithm, the ROC curve reveals performance for all thresholds. It makes it possible to select the threshold value.
- An additional by-product is the area under the ROC curve (AUC). It gives an idea about how good a classifier the proposed method is. Also, the AUC may be used to compare different versions of the classifier method.

Figure 23 gives the ROC curve for a specific sample size selection: 512 samples. That is, as given in Section 3.4.1, the proposed algorithm runs on sliding windows. Different window sizes can be utilized, while a size comparable to that of a typical muzzle blast component is expected. A window size selection is 512, corresponding to 512 / 44100 = 11.6 ms. Theoretically speaking, this window size is expected to yield reasonable performance since it is comparable to typical muzzle blast duration, and it should yield a sharp result at GCC output. Moreover, since the method looks for sharp peaks within N-consecutive calculations, the 11.6 ms window is expected to yield a maximum once it coincides with the muzzle blast. Checking Figure 23, one can assert that the method yields reasonable performance for a window size of 512 samples. Depending on the application, one could select the threshold that yields a sensitivity around 0.6 and 1-specificity around 0.1. Alternatively, 1- specificity of 0.25 could

Table 11: Area Under the Curve for Window Length of 512 Samples

Test Result Variable(s):	Muzzle Detection Strength Output	
Area	Asymptotic 95% Confidence Interval	
	Lower Bound	Upper Bound
0.841	0.812	0.870

be traded for increased sensitivity at 0.9. However, before going into more detail, let us obtain the ROC for another classifier configuration for the window size of 256 samples.

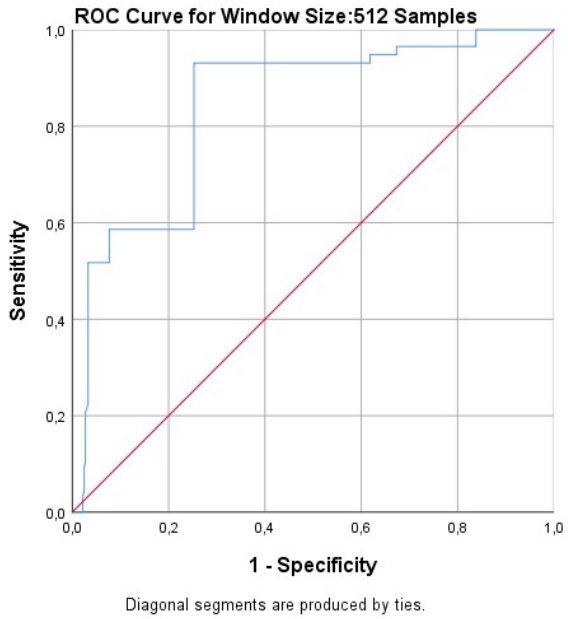


Figure 23: ROC Curve for Muzzle Detection with Window Length of 512 Samples

Similarly, a window size of 256 samples corresponds to 5.8 ms. However, we do not theoretically know if a 5 ms window or a 10 ms window is a better selection, but we only know that the selection should be comparable to a typical muzzle blast, in this case, both of which are. Figure 24 gives the ROC curve for the proposed method with the same parameters, except for the window size, which is 256 for this case. The detector with a window size of 256 samples has better sensitivity and specificity characteristics.

Although the performance seems pretty apparent in Figure 23 and Figure 24, such a comparison should still be made with the area under the curve (AUC) metric. Table 11 reveals the AUC details for the window size: AUC for both cases, calculated using SPSS. The AUC values comparison is in line with the sample case.

Similarly, Table 12 gives the AUC information for window size: 256 samples case. The AUC metric is in line with the ROC figures presented.

Having a better detection performance, we can focus our ROC analysis on the window length of 256 samples. Appendix A gives all thresholds' sensitivity and 1-specificity details. Such data makes it

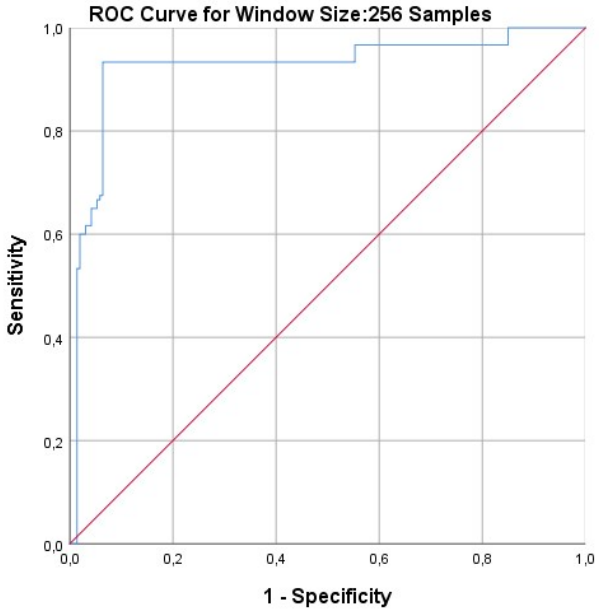


Figure 24: ROC Curve for Muzzle Detection with Window Length: 256 Samples

Table 12: Area Under the Curve for Windows Length of 256 Samples

Test Result Variable(s):	Muzzle Detection Strength Output	
Area	Asymptotic 95% Confidence Interval	
	Lower Bound	Upper Bound
0.925	0.901	0.949

very obvious to select the threshold value. The algorithm's "detection strength" calculation used as the threshold at 3.43 for muzzle blast detection decision with a window length 256 seems an optimal point. Thresholds above this value yield better specificity, but have less sensitivity since more true positives are lost. Thresholds below this value yield better specificity at the cost of lower sensitivity. Although specific implementations may select different thresholds on this table, depending on the application-specific requirements, the above-mentioned corner point at ROC is considered convenient in this study, with detection threshold: 3.43, sensitivity: 0.933, and 1-specificity: 0.064.

CHAPTER 6

DOA ESTIMATION RESULTS

The novel method proposed by this thesis is claimed to improve SNR in challenging sound field scenarios in an adaptive manner. The useful output of such a system should enhance the performance of signal-processing tasks. Chapter 4 revealed such improvements for SNR level as a product of adaptively adjusting to frequencies of interest to achieve better detection. Similarly, Chapter 5 reveals detection performance with F-score, error rate, and precise resolution in detection time. Another widely used feature of microphone array-based systems is the problem of the direction of arrival (DOA) estimation. Improved SNR and time resolution on detection output are expected to improve DOA estimation. The following sections seek to visualize and prove this improvement by examining DOA estimation errors against different parameters for both original array signals and signals enhanced by the novel method of this study.

Since the DOA estimation method affects the performance, its selection is a parameter. There is a variety of DOA estimation methods, and comparing them is not the subject of this study. Still, to avoid deriving conclusions depending only on a specific DOA method, two different DOA estimation methods were utilized: The multiple Signal Classification (MUSIC) algorithm and the Minimum Variance Distortionless Response (MVDR) algorithms. These algorithms were implemented using the MATLAB Phased Array System Toolbox's built-in DOA functionality. The toolbox supports custom array geometries used together with these algorithms. The DOA part of the problem is achieved by running the detection and pre-processing parts and then utilizing the toolbox with this study's custom array geometry.

6.1 Reference DOA Estimation Methods

6.1.1 The MUSIC Algorithm

The Multiple Signal Classification (MUSIC) is selected as one of the reference DOA algorithms. The MUSIC method, proposed by [58], estimates the DOA of sound sources in various applications, including music production, acoustic surveillance, and telecommunications. The algorithm is based on the eigenvalue decomposition of the covariance matrix of an array. Its major distinctive abilities are being precise regarding angle and power to localize multiple instantaneous signals. It also offers insights into the spatial characteristics of signals utilizing an array of microphones or sensors to estimate the DOA of sound sources with high resolution, even in complex acoustic environments.

The method is based on computing a spatio-spectral correlation matrix and decomposing it to signal and noise components. That way, vectors corresponding to (possibly multiple) sound sources and those corresponding to noise components are determined. Then, the vector, which can be expressed as a function of the direction of arrival, is steered for different incidence angles to find the direction that results in minimum Euclidian distance with the noise components [21]. Depending on the assumption that the signals of interest are uncorrelated with the noise components, minimum Euclidian distance corresponds to orthogonality between signal space and noise space.

Variants of the MUSIC algorithm deploy the same basic principle with different characteristics. The root-MUSIC algorithm proposed by [59] can reduce this calculation complexity for practical purposes. Another one, Estimation of Signal Parameters using Rotational Invariance Technique (ESPRIT), was proposed by [60], that overcomes the uniformly spaced linear array (ULA) restriction that applies for both TDOA and Steered Response Power (SRP) techniques. These advantages arise from the MUSIC algorithm making no assumptions about the array geometry. Moreover, [58] states that the receivers can have different amplitude/phase directionality characteristics. Finally, by its definition, the MUSIC algorithm can be employed for multiple simultaneous signals without performance loss.

6.1.2 The MVDR Algorithm

The Minimum Variance Distortionless Response (MVDR) method is used for sound source localization and beamforming. The method is based on adaptively adjusting the spatial filters based on the estimated covariance matrix for the array. It aims to estimate sound sources' Direction of Arrival (DOA) by utilizing spatial information captured by a microphone array, computing optimal weightings to enhance the desired source signals while attenuating the others. It is suited for complex sound field environments and reverberations. Sound separation study by [61] applies MVDR and concludes that a higher source localization accuracy is achieved by identifying and utilizing single-source targets compared to reference beamforming methods. As for the low-SNR performance, [62] utilizes MVDR for the DOA side of their solution and reports improved localization accuracy and resolution for weak sources in terms of SNR. Therefore, like the MUSIC algorithm, MVDR is also suited for this study's complex, low-SNR problem.

Moreover, having both algorithms from the same resource, the MATLAB Phased Array System Toolbox's DOA package is appropriate. Since the emphasis of this study is the detection and denoising performance and since the DOA methods are just a parameter to comment on the performance of the other phases, it is essential to have validated packages for the DOA side of the problem. The comparative results of this study regarding the DOA method's effect on DOA estimation error are given in the following sections, together with other parameters.

6.2 Performance Metric for the DoA Estimation Problem

The DREGON challenge [20] DOA estimation task has 300 8-channel audio recordings, roughly 2 s each. All recordings were obtained by adding together a clean recording of a static loudspeaker emitting random utterances from the TIMIT database [63] from an unknown (azimuth, elevation) direction in the UAV microphone arrays frame and a recording of UAV noise mixed with -20 to 5 dB. The goal is to find the azimuth and elevation angles of the sound source for the 300 recordings. A source was

considered correctly localized when the great-circle distance between the estimated and ground truth directions was less than 10°. 1 point was given for each correctly localized static file, for 300 points.

The SPCUP challenge [20] uses two performance assessment metrics for the second part of the problem, DOA estimation. These metrics are the DOA angle's root mean square error RMSE and the DOA estimation "success rate". The success rate gives a basic measure of whether the system could predict the DOA within the correct one-sixth of the horizontal plane. The "success rate" metric is a convenient use case for a final user application. That is, it may be used as a basic measure by a potential future application, where the DOA estimation will be used to steer a camera to the direction of interest, similar to the case of [40]. The Root Mean Square Error (RMSE) yields the basic measure of estimation performance.

The DOA estimation output is represented by two angles, the elevation and the azimuth, representing the DOA in 3-D space. Considering this on top of the RMSE metric for DOA estimation accuracy, the angle between the estimation and the known truth table must be calculated. In other words, the estimation and the truth table value have an azimuth-elevation pair in 3-D. The angle between representative unit vectors of these two directions gives how accurate the estimation was in 3-D, in degrees. This thesis reports this angle difference for different scenarios, algorithm parameters, and algorithms.

6.3 DoA Estimation Performance for the Muzzle Blast Component

To look over the DOA estimation performance with the proposed method, one can start with a representative problem geometry. Figure 25 visualizes the problem geometry regarding muzzle blast sound components. The muzzle sound originates from the barrel and propagates in the air with the speed of sound. The smallest distance between the barrel and the drone (or the microphone array) is defined as the range. Also, azimuth and elevation angles are defined for the direction of arrival of the muzzle signal onto the microphone array with the far field assumption, as visualized in the figure.

DOA estimation on the typical complex sound field is affected by many components. Examining their effects brings insight into which parameters affect DOA estimation error and makes it possible to see if the proposed pre-processing eliminates or reduces any effects on DOA estimation. For example, the DOA error estimation performance is naturally expected to fall with increasing range for muzzle detection. If the proposed processing method reduces the negative effect of range on DOA estimation performance, that would be promising for long-range, low SNR applications.

The parameters that are expected to affect DOA estimation are as follows:

- range (between muzzle and microphone array)
- direction of arrival
- the DOA estimation method used
- the gun type
- the noise signal (changing momentarily)

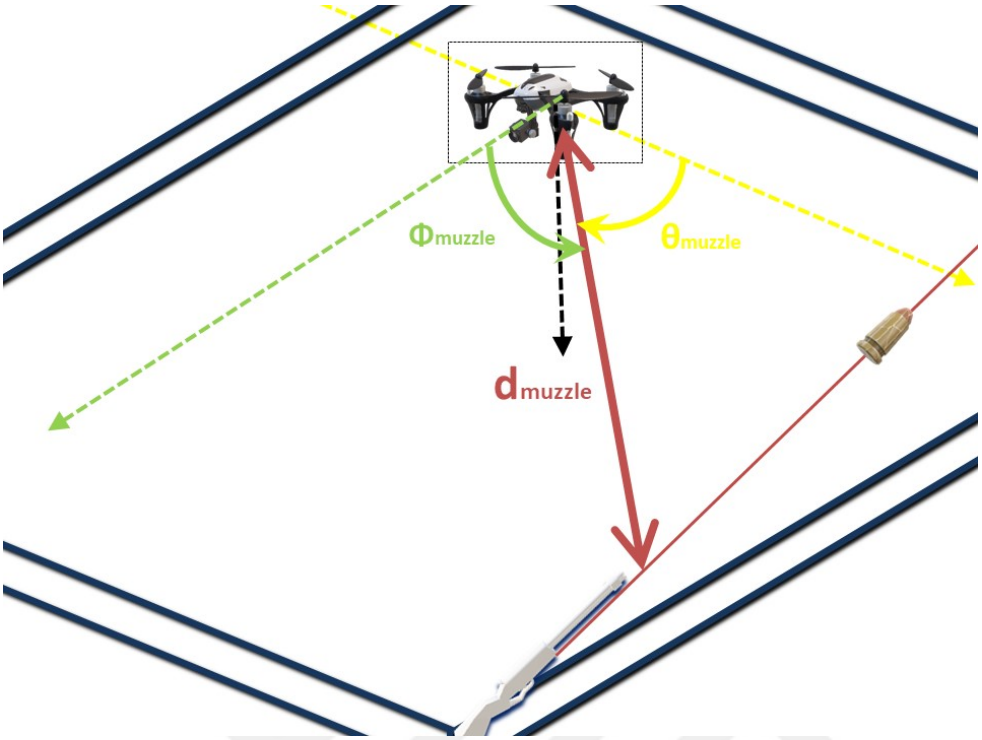


Figure 25: Arrival path and geometric parameters of the muzzle blast component.

The DOA estimation performance with the proposed method was examined by running the two DOA estimation methods on a dataset built as follows: 3 different gun types x 2 different noise signals x 4 different DOA x 4 different ranges x 2 different DOA algorithms = 192 scenarios. Table 13 summarizes the scenarios corresponding to the factors.

Multivariate (generalized linear model) analysis estimates the relationship between two or more independent variables and more than one dependent variable. As an introduction to the effects of different parameters on DOA estimation of the muzzle component, Table 14 F and significance values can be examined. The table lists the effects of parameters both individually and in a co-dependent manner.

Apart from giving the effects of parameters, another aspect of Table 14 should be emphasized. For each scenario, the table has two rows corresponding to the original array data and the signal "cleaned" by the auto-tuned CWT scales of this study. In other words, before performing DOA estimation, the tuned-scales knowledge obtained automatically at the detection phase is used as a pre-processing. If this pre-processing reduces the effects of parameters on muzzle DOA estimation, then this means an extra benefit at the DOA estimation phase. The following sections give detailed examinations of parameters and their effects.

6.4 Muzzle Range

As mentioned in earlier chapters, the main challenge of the problem of this study is the extremely low SNR. As given in the Related Work section 2.5, example implementations where the microphone

Table 13: Dataset Scenario Summary for the Parameters of the Problem

SCENARIOS & PARAMETERS

		N
Gun Type	MB_308r_shot01	384
	MB_AR15_shot01	384
	MB_colt45_shot01	192
Noise Type	DREGON_free-flight_nosource_room2_13_16	480
	DREGON_hovering_nosource_room2_20_23	480
AOA (Az. & Elv.)	45	240
	60	240
	75	240
	90	240
Muzzle Range	200	240
	400	240
	600	240
	800	240
DOA Method	MUSIC	480
	MVDR	480
Shock Miss Distance	12	384
	24	384
	N/A	192

Table 14: Multivariate Analysis Table for Muzzle Blast DOA Estimation

Tests of Between-Subjects Effects

Source		T. III Sum of Sq.	df	Mean Sqr	F	Sig.
Gun Type	DOA Err Orig.	3011.518	2	1505.759	5.570	0.004
	DOA Err Cleaned	1214.989	2	607.494	3.204	0.042
Noise Type	DOA Err Orig.	1842.832	1	1842.832	6.817	0.009
	DOA Err Cleaned	307.132	1	307.132	1.620	0.204
Muzzle DOA	DOA Err Orig.	14461.290	3	4820.430	17.831	0.000
	DOA Err Cleaned	2788.281	3	929.427	4.903	0.002
Range	DOA Err Orig.	2144.766	3	714.922	2.645	0.049
	DOA Err Cleaned	956.890	3	318.963	1.682	0.170
DOA Method	DOA Err Orig.	6577.425	1	6577.425	24.330	0.000
	DOA Err Cleaned	0.422	1	0.422	0.002	0.962

to sound source distance is in the order of a few meters are already successful. The range of hundreds of meters makes this problem an extremely low SNR one. More precisely, examining algorithm performance against the range is equivalent to examining it against the SNR.

First, examining Table 14, the significance value of 0.049 for the range parameter in the case of the original (not cleaned) array signal indicates that it affects DOA performance. This is as expected since increasing range means decreasing the SNR, and since there is no adaptive scale selection, the performance is likely to drop with decreased SNR.

Checking the second row entry for the range parameter in Table 14, the error for the signal cleaned by the proposed method, the significance value is 0.17 (>0.05), depicting that there is not a significant effect of Range on DOA estimation error with the proposed method. This is a critical claim since eliminating the impact of range means improving the range of a potential implementation, at least up to a reasonable range. Recalling the scenario summary table of Table 13, the ranges simulated are 200, 400, 600, and 800 m. If it can be concluded that the proposed method yields DOA estimation errors not dependent on the range, although they were dependent on the original array data, this means a promising performance aspect.

The overall mean error and performance comparisons can be applied in Figure 26. In Figure 26, the mean DOA estimation error for the ranges 200, 400, 600, and 800 m is revealed. The DOA error is again defined as the differences between the 3-D vectors of the actual direction of arrival for the muzzle blast component (truth table) in azimuth and elevation and the azimuth-elevation output of the DOA algorithm. On the linear scale of the figure, the increase in DOA error with increasing range is visible and is as expected. The error values are between 11 and 18 degrees. Note that the DOA estimation is strong even for the 800 m case, with a mean error of 18 degrees. This is because the DOA algorithm is run only if there is detection, and with the proposed detection method, the detection is precise within 5 ms. Therefore, the DOA estimation algorithm is run on a 5 ms window, meaning a strong SNR at the short-time window. Still, the expected decrease in DOA performance with increased range is observed for the range of 800 m.

Now, compare this with the DOA estimation results after the proposed pre-processing. The mean error values are between 5 to 8 degrees, which is not only an improvement over the original array (no pre-processing) case, but also is not proportional to the range. This is a promising result for the proposed methodology's range vs. SNR performance. It can not be asserted that the DOA methods deploying the adaptive selected CWT scales as pre-processing result in DOA performance independent of SNR. Still, one can claim that the proposed methods result in performance nearly independent of range in terms of muzzle DOA estimation error. This sequence is the proposed adaptive detection method + preprocessing with the adaptively-tuned scales + the regular DoA algorithm.

Further examination is given in Figure 27 and Figure 28. Visualizing the effects of range and gun type in the same image is possible. Both figures reveal the impact of gun type on DOA error. Moreover, the range effect for original and pre-processed data is observed similarly when clustered by gun type. One point to emphasize on these figures is that the adequate performance of the proposed method, in terms of range, holds for different gun types. In other words, the method's promising performance concerning nearly range-free operation is observed for three different gun types.

Apart from the expected performance characteristics with respect to range, let us examine the difference between the curves for different gun types. Recalling the SPL levels of the three weapons given

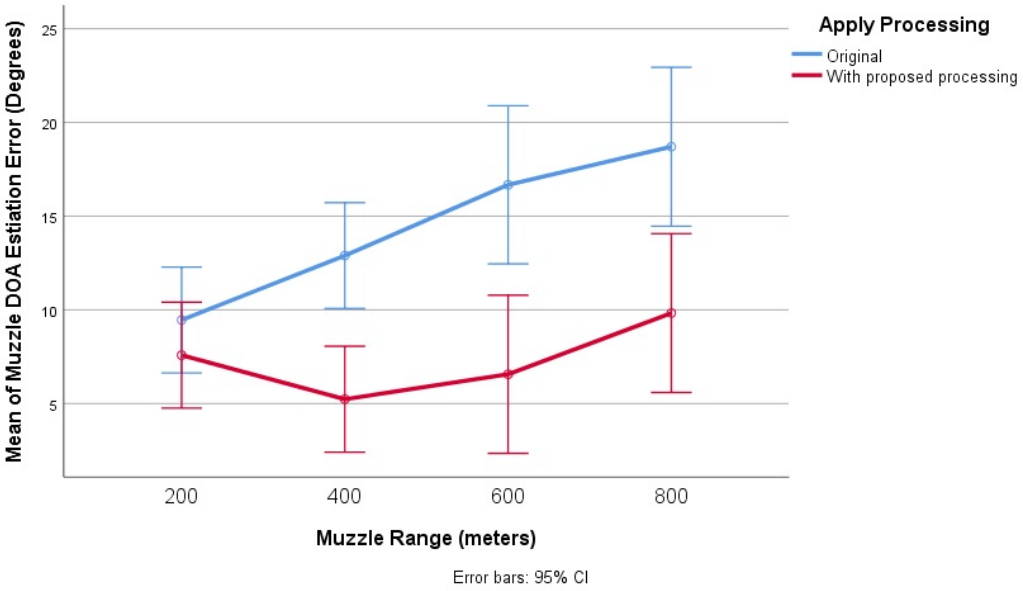


Figure 26: DOA Estimation Performance Improvement by the Proposed Processing wrt Muzzle Range

in Table 3, the SPL levels (at 3 m) of the 308r, AR15, and colt45 are 145 dB, 144 dB, and 137 dB, respectively. This means that performances without adaptive processing are directly related to the range and the gun's SPL. Checking the performance for different gun types with the proposed processing, there is a slight difference in ordering the gun types. However, the AR15's DOA estimation error drop is smaller than the other two. This results in a change in the ordering of the error magnitudes between gun types. Still, it should be noted that all three gun types' DOA estimation errors are reduced. Similarly, all three gun types' performance vs. range characteristics are enhanced to be nearly-range-independent.

6.5 Direction of Arrival

To examine the effects of the incidence angle of detection and DOA estimation performance, the dataset includes four different angles of arrival, as given in Table 13. As for the detection performance of the proposed method, Chapter 5.3 shows that detection performance is not affected by the direction of arrival of the muzzle blast component. On the other hand, in the DOA estimation case, some level of effect is expected. This is due to the angle-dependent nature of DOA algorithms rather than the nature of the proposed detection or the pre-processing method. Examining Table 14, for the muzzle DOA parameter, we can see that for both the original array signal and the signal cleaned with the proposed processing, the DOA error performance is affected by the muzzle DOA.

The mean DOA error against DOA of the gunshot signals and performance comparison before vs. after applying the proposed method is given in Figure 29. Figure 29 reveals the mean DOA estimation error for the four angles: 45, 60, 75 and 90 degrees. The DOA error is again defined as the differences between the 3-D vectors of the actual direction of arrival for the muzzle blast component (truth table)

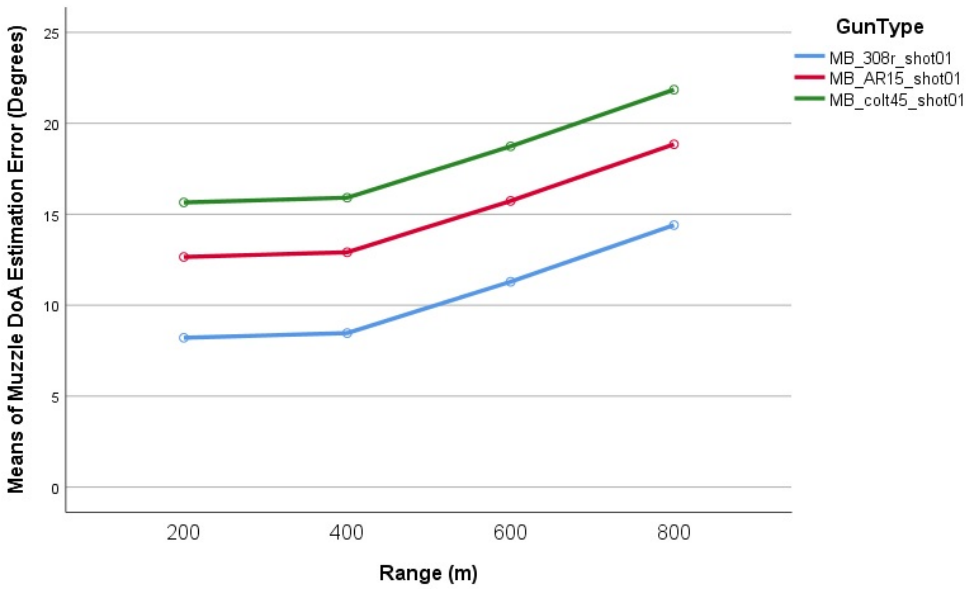


Figure 27: Mean Muzzle DOA Estimation Errors wrt Range per Gun Type with Original Array Data

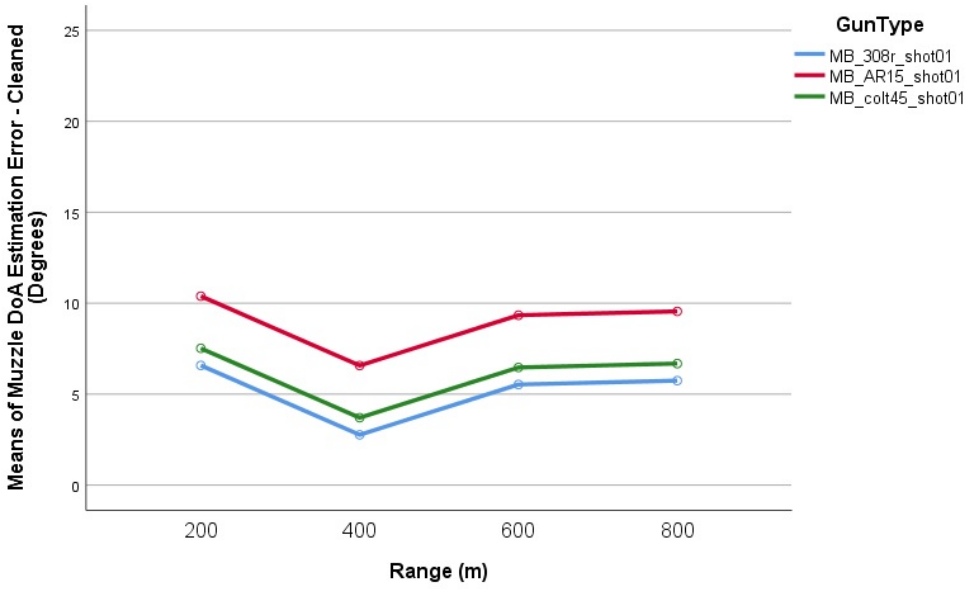


Figure 28: Mean Muzzle DOA Estimation Errors wrt Range per Gun Type with Proposed Processing

in azimuth and elevation and the azimuth-elevation output of the DOA algorithm. The error values are between 10 to 25 degrees. We also see that the error with the proposed pre-processing drops to 4 to 11 degrees, an improvement over the original array (no pre-processing) case. This improvement can be attributed to increased SNR.

Moreover, the effect of different angles should be investigated. We can see from Figure 29 that the angle at 90 degrees is a particular angle for both cases. This is as expected since the way DOA is geometrically defined, 90 degrees correspond to the directly opposite direction of the array. In this case, the inter-arrival times of the muzzle blast signal to array microphones are the smallest. That is, geometrically speaking, the distances between the sound source and the microphone number N are closest to each other. That is why the DOA performance is expected to decrease for cases around DOA:90 degrees.

An interesting point would be to exclude the DOA:90 degrees scenarios while examining DOA error performances. Removing such rows from the list of all scenarios, we obtain a significance value of 0.405 for the original array data. This means, initially, the DOA dependence of the deployed DOA methods is dominant around 90 degrees, and for other incidence angles, little effect is expected. Still, the significance value for the cleaned case is 0.009, meaning even when DOA around 90 degrees is removed, we can not expect angle-independent DOA estimation performance from the proposed sequence of methods. Compare this to the detection performance with respect to the angle of incidence, which was not affected significantly by the incidence angle.

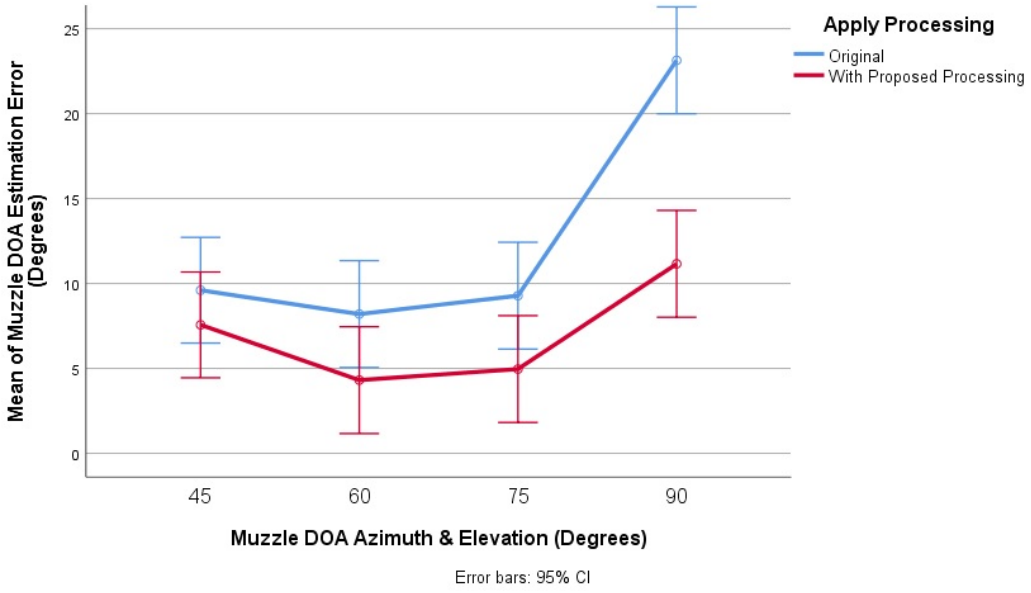


Figure 29: DOA Estimation Performance Improvement by the Proposed Processing wrt DOA

6.6 DOA Method

This study's emphasis is not on any specific DOA estimation algorithm. Instead, the claim is on the novel detection algorithm that signal processing methods, following this detection, will perform better thanks to having precise detection instant and an adaptive scale (frequency) information at hand. Therefore, two conventional DOA estimation algorithms are deployed to compare the overall DOA estimation performance with and without the proposed pre-processing. Since the following signal processing problem is DOA estimation, the DOA method selection is one of the parameters. As explained

in Section 6.1, DOA performance is examined over two DOA estimation methods, namely MUSIC and MVDR.

With the conventional reference DOA methods, the intention is not to make conclusions about the DOA algorithms' effects. However, since the number of scenarios is large, it is possible to comment on the DOA method results statistically. As summarized in Table 13, 193 trials for the MUSIC algorithm and 193 for the MVDR. The significance value of 0.962 in Table 14, asserts that the DOA error performance is not affected by the DOA method used when the proposed method is used. For the original array data, on the other hand, the DOA algorithm is expected to affect the performance, which is the default case and is expected.

The situation is supported by Figure 30. Before the proposed method, there was a distinct performance difference between the DOA estimation methods in Figure 30. The difference is eliminated with the method, that is, for the cleaned signal. Other than comparing the DOA methods, it is also possible to assert that the amount of error is reduced in the latter case compared to that of the original array data. The DOA estimation error falls from 11-18 degrees to around 7 degrees.

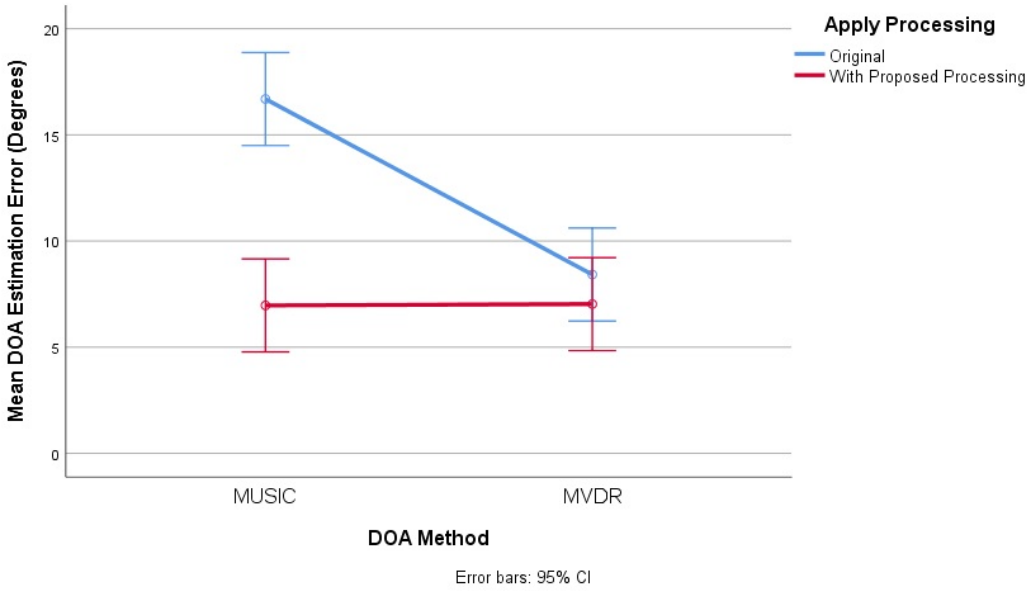


Figure 30: DOA Estimation Performance Improvement for Different DOA Algorithms

The results for the two DOA methods are also visualized on top of the Range and Angle of Arrival parameters to examine the effects of DOA methods together with other parameters. First, let us look at the effect of the range together with the DOA method. Figure 31 and Figure 32 reveal both Range and DOA Method performance. Comparing the two figures of the same scale, the improvement in the error is clear with the proposed processing. Also, we can say that the MVDR outperforms the MUSIC for all range values. Finally, the expected trend of DOA error increasing with the Range is observed for both DOA methods.

Similarly, Figure 33 and Figure 34 compare performances of MVDR and MUSIC against all four DOA scenarios. Firstly, the proposed pre-processing decreases the DOA estimation error for all the angles

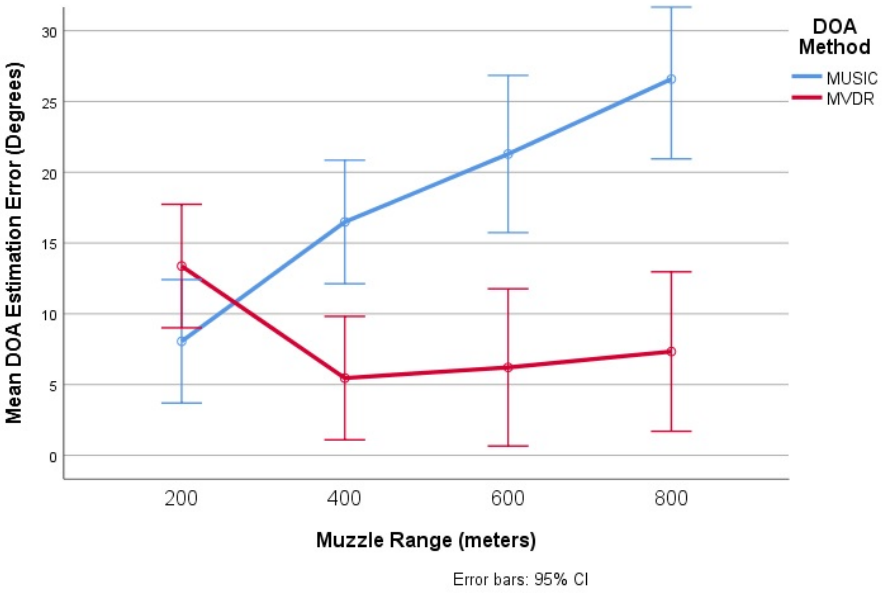


Figure 31: Mean DOA Estimation Errors for Different DOA Methods Against Range - with Original Array Data

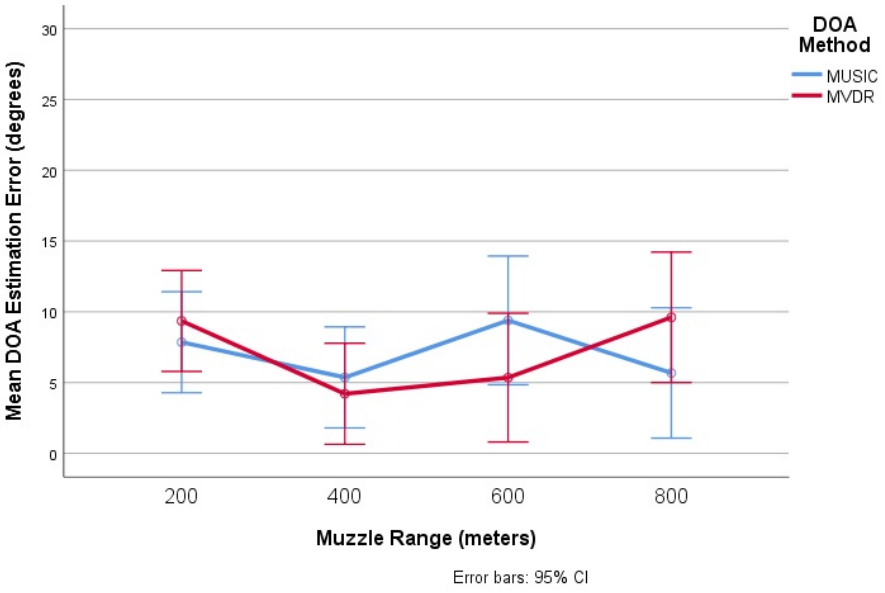


Figure 32: Mean DOA Estimation Errors for Different DOA Methods Against Range - with Proposed Processing

and both DOA methods. Also, the trend of both algorithms against DOA angles is similar, including the reduction of performance around the DOA of 90 degrees. Lastly, like the Range parameter's case, MVDR seems superior to the MUSIC algorithm regarding DOA estimation error performance.

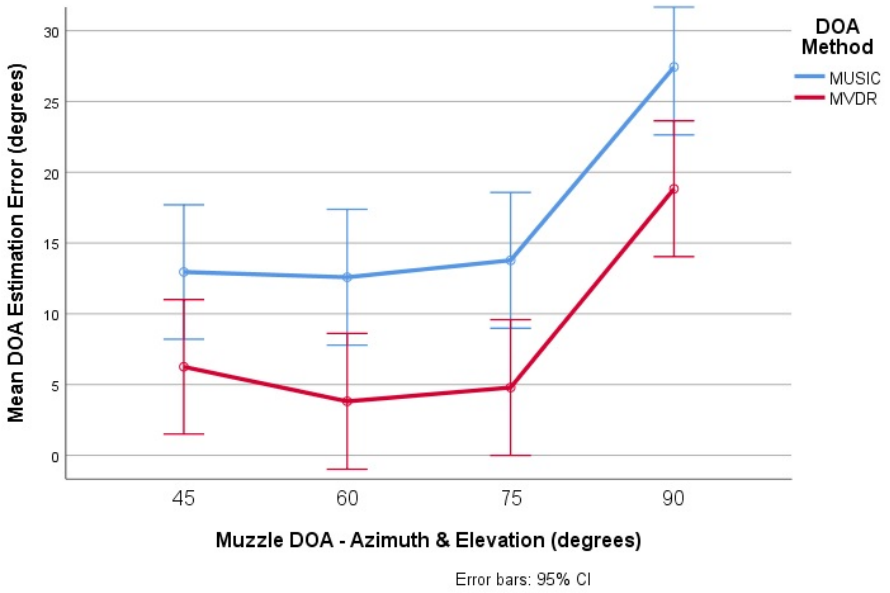


Figure 33: Mean DOA Estimation Errors for Different DOA Methods Against Angle of Arrivals - with Original Array Data

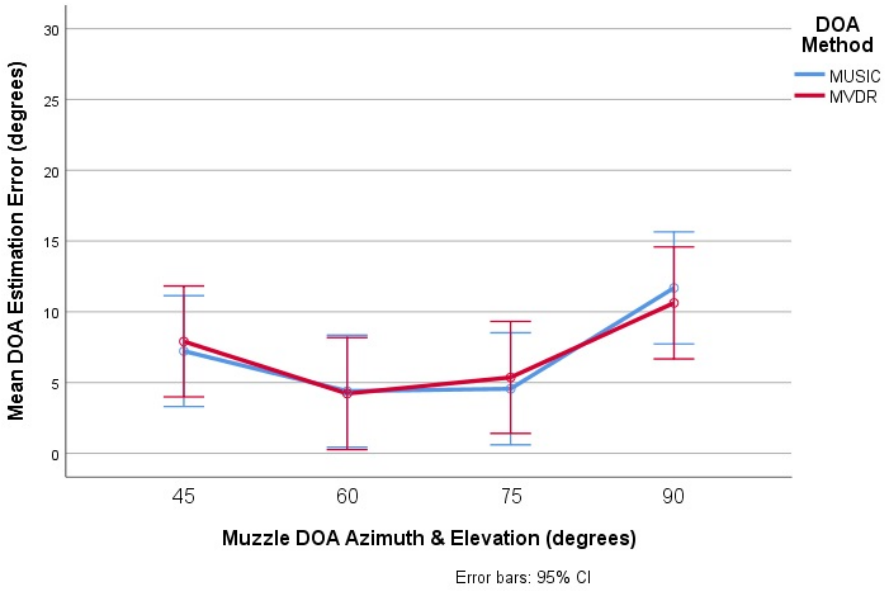


Figure 34: Mean DOA Estimation Errors for Different DOA Methods Against Angle of Arrivals - with Proposed Processing

While examining all figures in this section, the interesting point is that the DOA error values are different from the original array data and nearly the same once the proposed method is used. Knowing which scales (frequencies) to focus on eliminates the difference between different DOA methods'

performance. This conclusion on DOA algorithm performance is promising since the method's performances were originally different for the same set of scenarios.

6.7 Gun Type

Examining Table 14, the significance value for both Gun-Type in the case of the original (not cleaned) array signal indicates a strong correlation with DOA performance, as expected. Different gunshot signals are expected to result in different performances considering sound characteristics differences. Here, it should be remembered that muzzle blast SPL levels of varying gun types are different, a major natural component of SNR differences. For this parameter, there are two row entries in Table 14, one for the original array data and one for the signal cleaned by the proposed method. Since the increase in the significance value is not large enough to conclude by itself (not > 0.05), let us examine the effect of gun type for both cases in Figure 35.

In Figure 35, the mean DOA estimation error for three different gun types is revealed before and after applying the proposed method. The error is defined as the differences between the 3-D vectors of the actual direction of arrival for the muzzle blast component (truth table) in azimuth and elevation and the azimuth-elevation output of the DOA algorithm. The mean error value of all scenarios, obtained with original array data, varies between 10 to 20 degrees. The same vectorial DOA error difference with the proposed pre-processing is also given. The mean error values range between 4.5 to 9 degrees after the proposed processing, a significant improvement over the original array (no pre-processing) case.

The DOA estimation error performance for different gun types was visualized on top of the Range characteristics in Section 6.4. Since some level of difference is statistically expected when the scenarios are grouped by gun type, it is not possible to claim to be independent of gun type. Still, since the number of trials (scenarios) is large, it can be asserted that the more minor error differences mean improvement in different gun types' performance. The effect of gun types is also visualized on top of other parameters in the following sections.

Judging by Figure 35, we can claim that the improvement in DOA estimation performance is valid for different gun types.

6.8 Drone (Noise) Type

Examining Table 14, the significance value of 0.009 for Noise Type (drone signal) in the case of the original (not cleaned) array signal indicates a strong correlation with DOA performance, as expected. Reminding the drone's non-stationary characteristics, it is impossible to cover all possible noise space for this problem domain. On top of the large number of drone movement types in the dataset, selecting any instant in drone noise yields a different noise signal. Still, it is at least possible to keep the noise type constant and gather the effect of different parameters. Also, since the number of scenarios is large, it is possible to gather if there is a statistically significant difference between sample noise selections.

For this parameter, there are two-row entries in Table 14, one for the original array data and one for the signal cleaned by the proposed method. The significance value for the row with the "cleaned" entry is 0.204 (>0.05), depicting that there is not a significant effect of Noise Type on DOA estimation error

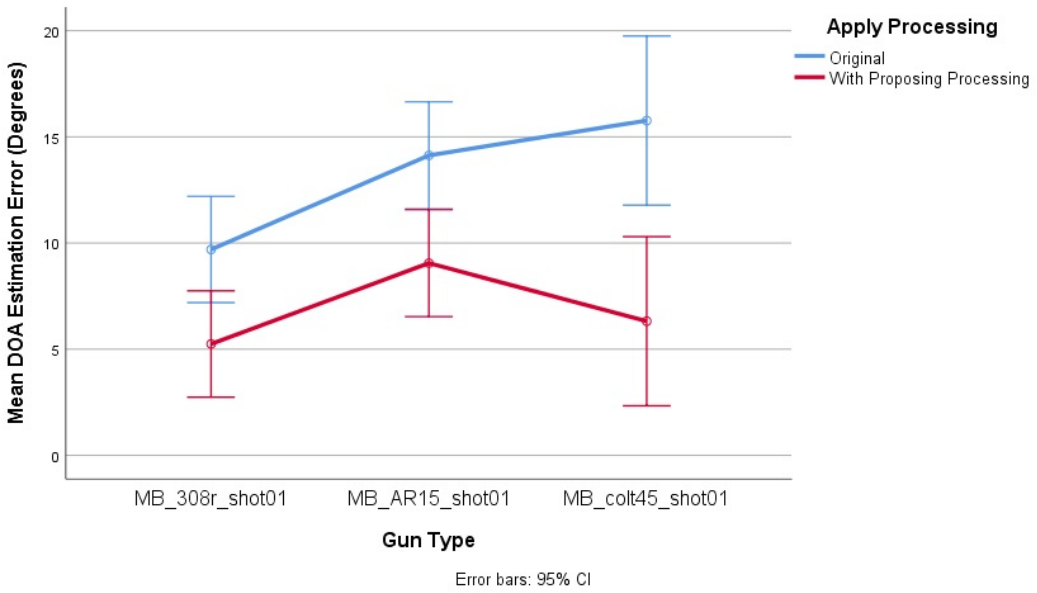


Figure 35: DOA Performance Improvement for Different Gun Types

with the proposed method. The mean error and performance comparison details can be seen in Figure 36.

The DOA estimation error performance for different noise types was already visualized with the Range characteristics in Section 6.4. Moreover, Figure 36, gives the mean DOA estimation error separately for two different noise signals before and after the method is applied. The DOA error is again defined as the differences between the 3-D vectors of the actual direction of arrival for the muzzle blast component (truth table) in azimuth and elevation and the azimuth-elevation output of the DOA algorithm. As given by the scenario summary Table 13, the two noise parameters are a free flight portion and a hovering portion of the flight. The mean error value of all scenarios, obtained with original array data, varies between 12 to 17 degrees. On the other hand, the same vectorial DOA error difference with the proposed pre-processing varies between 6 to 7.5 degrees, which is an improvement over the original array (no pre-processing) case.

Judging by Figure 36, we can claim that the improvement in DOA estimation performance is valid for different noise instants.

6.9 DoA Estimation of the Shock Wave Component

Section 6.3 revealed DOA estimation performance starting with a reference problem geometry. Similarly, Figure 37 also gives the geometry and parameter definitions for the shock wave component. The same dataset is also used since two of the three rifles have supersonic projectiles.

The shock wave DOA estimation performance parameters can be listed as miss distance, DOA algorithm, gun type, noise of the specific instant, and shock DOA. Since the shock wave component's

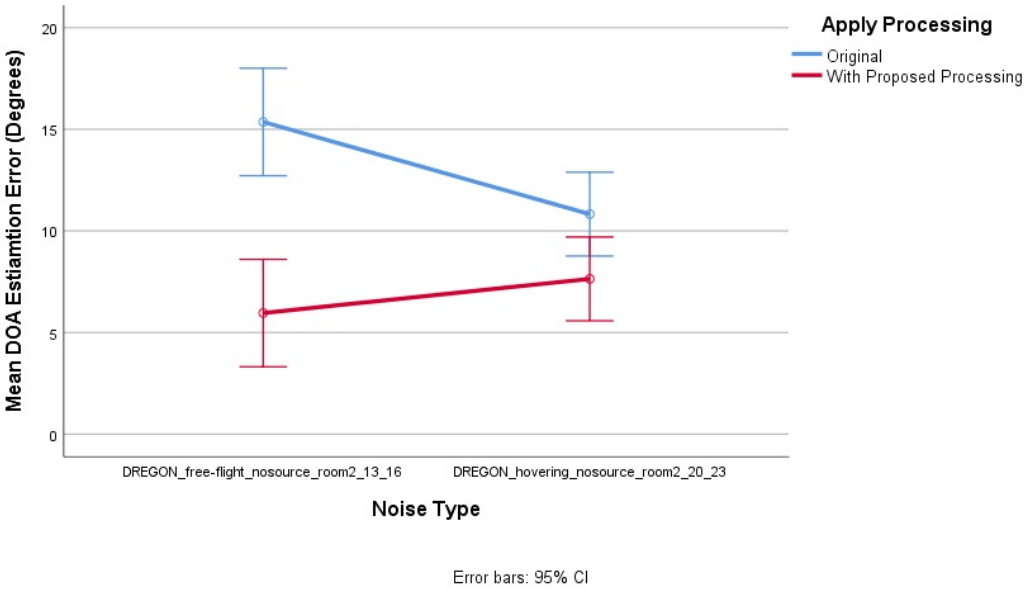


Figure 36: DOA Performance Improvement for Different Noise Samples

Table 15: Multivariate Analysis Table for Shock Wave DOA Estimation

Tests of Between-Subjects Effects

Source		Type III Sum of Sq.	df	Mean Sq.	F	Sig.
DOA Method	DOA Err Orig.	6068.105	1	6068.105	116.987	0.000
	DOA Err Cleaned	0.051	1	0.051	0.004	0.952
Miss Distance	DOA Err Orig.	711.538	1	711.538	13.718	0.000
	DOA Err Cleaned	781.689	1	781.689	55.233	0.000
Gun Type	DOA Err Orig.	512.849	1	512.849	9.887	0.002
	DOA Err Cleaned	1853.199	1	1853.199	130.945	0.000

presence depends on the gun type and the geometry of the particular gunshot occurrence (that is, the shock wave will not always be there), and since the signal SNR is high in case the miss distance is small, the more challenging and crucial part of this problem is deemed as the muzzle blast component. A smaller set of parameters are still examined, especially those specific to shock wave components: the miss distance, gun type, and DOA method.

As an introduction, multivariate general linear analysis F and significance values by SPSS can be examined in Table 15. Similar to Table 14 of the muzzle DOA estimation case in Section 6.3, Table 15 also lists two cases, with and without the proposed processing. For each scenario, the table has two rows corresponding to the original array data and the signal "cleaned" by the auto-tuned CWT scales of this study. In other words, before performing DOA estimation, the tuned-scales knowledge obtained automatically at the detection phase is used as a pre-processing. If this pre-processing reduces the effects of parameters on muzzle DOA estimation, then this means an extra benefit at the DOA estimation phase.

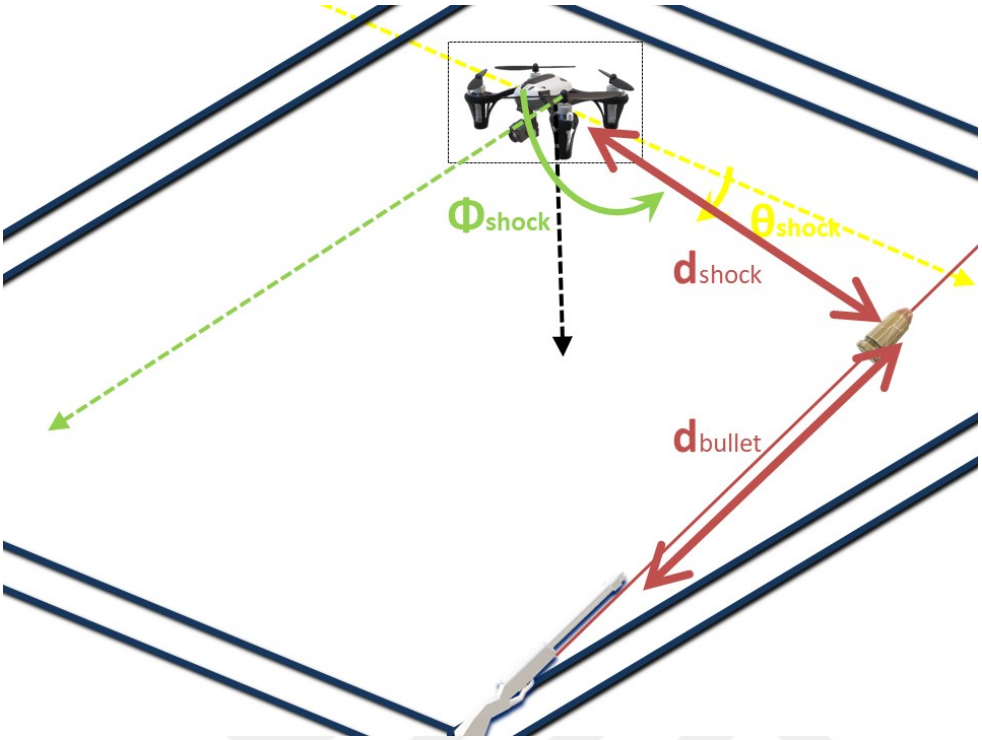


Figure 37: Arrival path and geometric parameters of the shock wave component.

The significance values of the ANOVA analysis in Table 15 reveal that the DOA estimation error for the shock component varies with miss distance and gun type. This is as expected; indeed, the miss distance directly affects the SNR of the shock wave, and the performance is likely to depend on it. Moreover, it should be noted that this dependence is not removed with the proposed method, either.

For the DOA method, the table reveals that the proposed method removes performance dependency on the DOA method selection, at least for the reference DOA algorithms of this thesis. More detailed results are given in the following figures of this section.

Similar to the Range parameter of the muzzle blast case, miss distance is the major contributor to the performance of the shock wave signal's DOA estimation. Once the shock wave originates from the bullet's trajectory, it starts traveling with the speed of sound and with power attenuation. The longer the miss distance, the more attenuation, so this is directly related to the resulting SNR. To reveal this expected effect, Figure 38 gives DOA estimation error performance for two different miss distances. Although the simulation framework makes trying a broader range of miss distances possible, only two alternatives are included here. The reason why the muzzle blast component is examined in more detail was already discussed above. The intention here is to show the concept instead of making claims in a broader miss distance range.

The DOA error slightly increases for a miss distance of 24 m, compared to a miss distance of 12 m. The performance is acceptable in both cases. Note that the shock wave is a strong impulsive component and is still strong for two cases, as examples in Chapter 3 revealed. This figure also compares two DOA algorithms' outcomes. As with the muzzle blast component case, the mean DOA error with MVDR is less than that of the MUSIC algorithm.

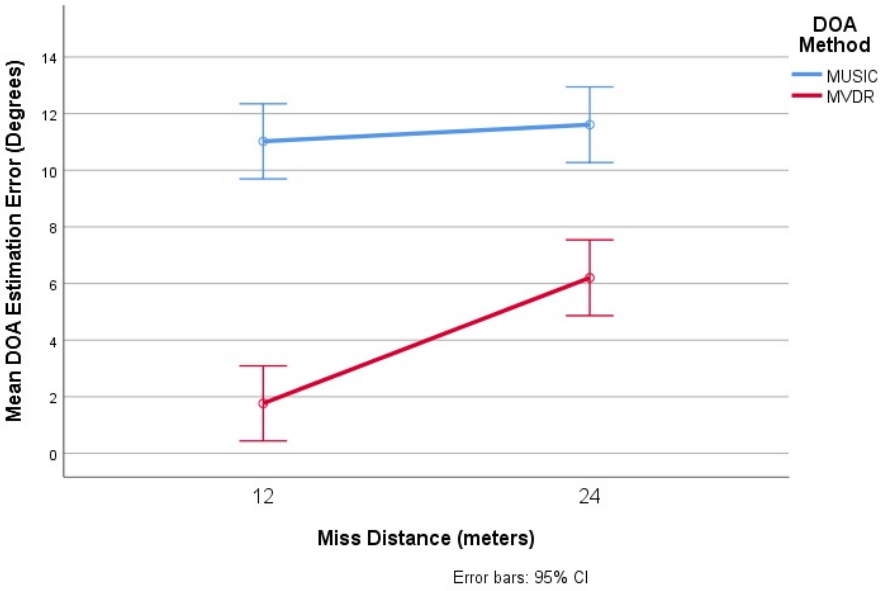


Figure 38: DOA estimation error performances for different DOA methods

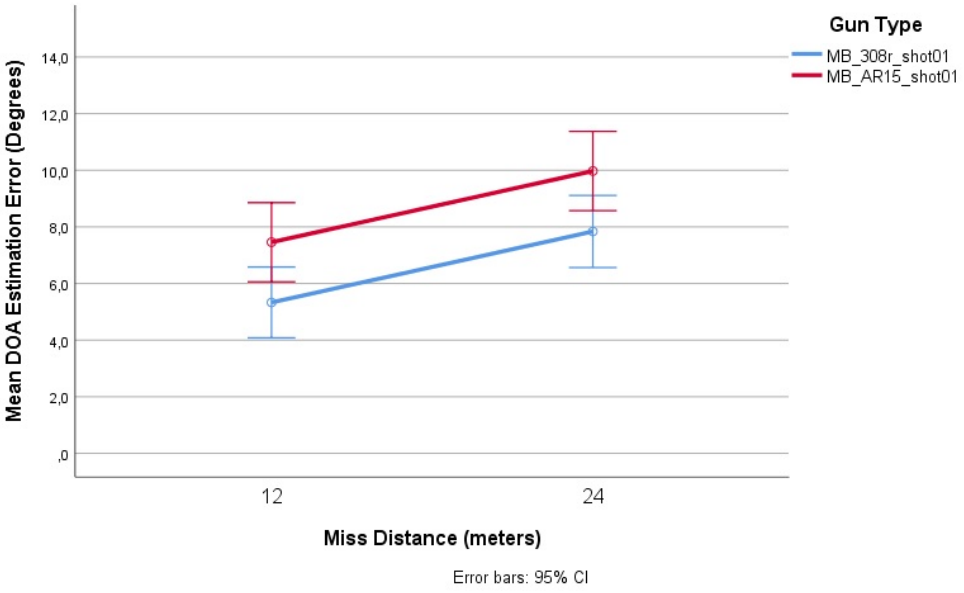


Figure 39: DOA estimation error performances for different miss distances and gun types

Similarly, Figure 39 gives the effect of different gun types on top of the DOA estimation error vs. miss distance graph. The error amounts reveal similar behavior as the distance changes from 12 m to 24 m. The same tilt in the curves can be attributed to the sound attenuation calculation of the simulation.

Effect of the methodology proposed by this thesis is examined in the following comparative Figure 40, Figure 41 and Figure 42. Especially the miss distance and gun type comparisons reveal that we can not claim that the method brings improvement independent of these parameters. It may be because all shock wave scenarios covered here are high SNR ones, and the method does not bring a distinctive advantage for these SNRs.

Figure 41 gives the DOA method case a different characteristic. For both DOA algorithms, the before-after perspective reveals that the DOA estimation performance is improved by the proposed processing. Moreover, DOA errors for the two algorithms are very similar when the proposed method is applied, although it was significantly different before the technique was applied. This characteristic is similar to the investigation for the muzzle blast component's DOA estimation performance in Figure 30 and is an interesting note.

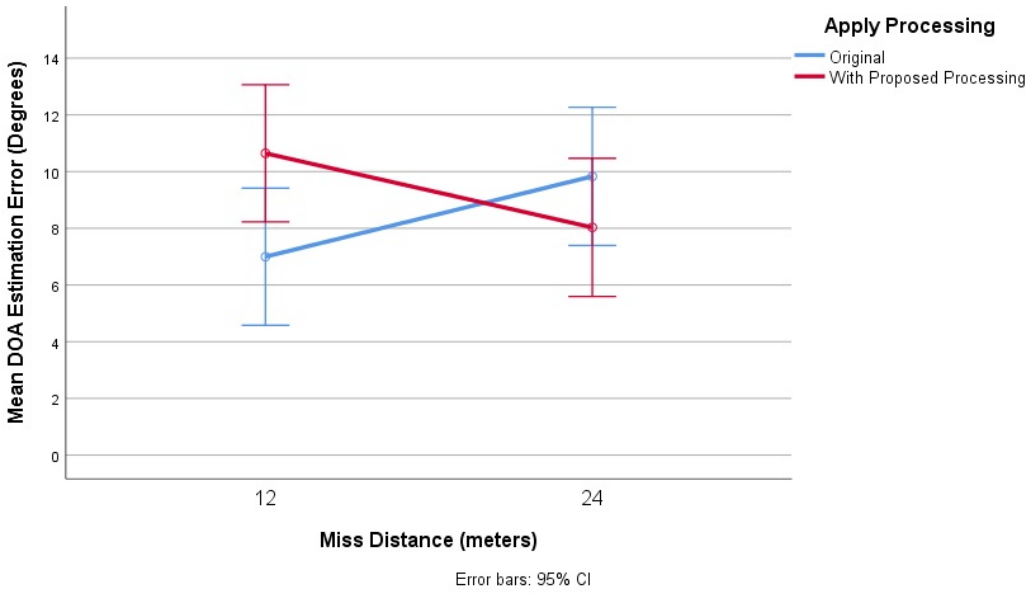


Figure 40: DOA Estimation Performance Comparison with the Proposed Method with respect to the Miss Distance

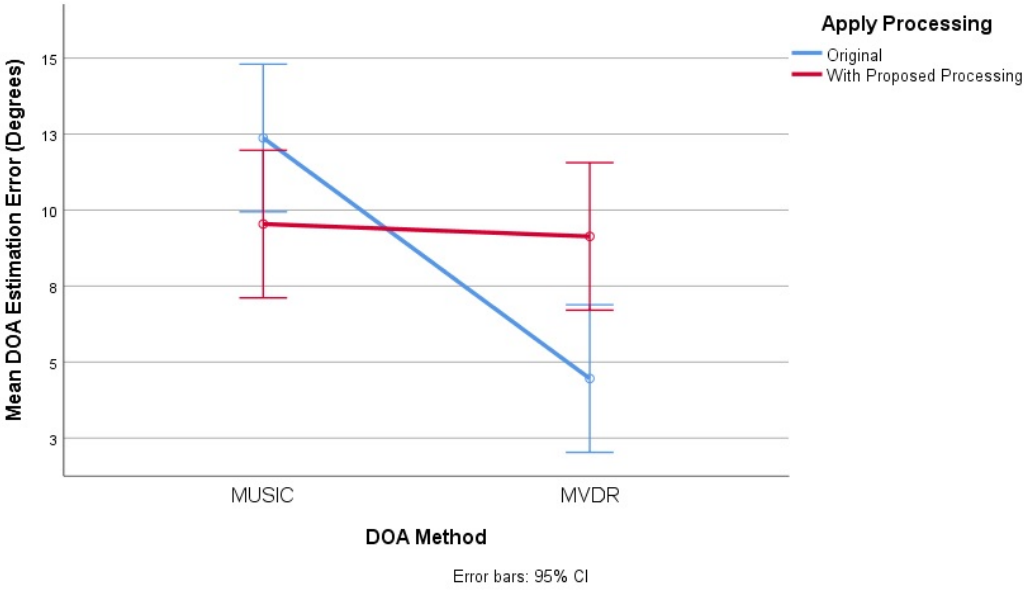


Figure 41: DOA Estimation Performance Comparison with the Proposed Method with respect to the DOA Method

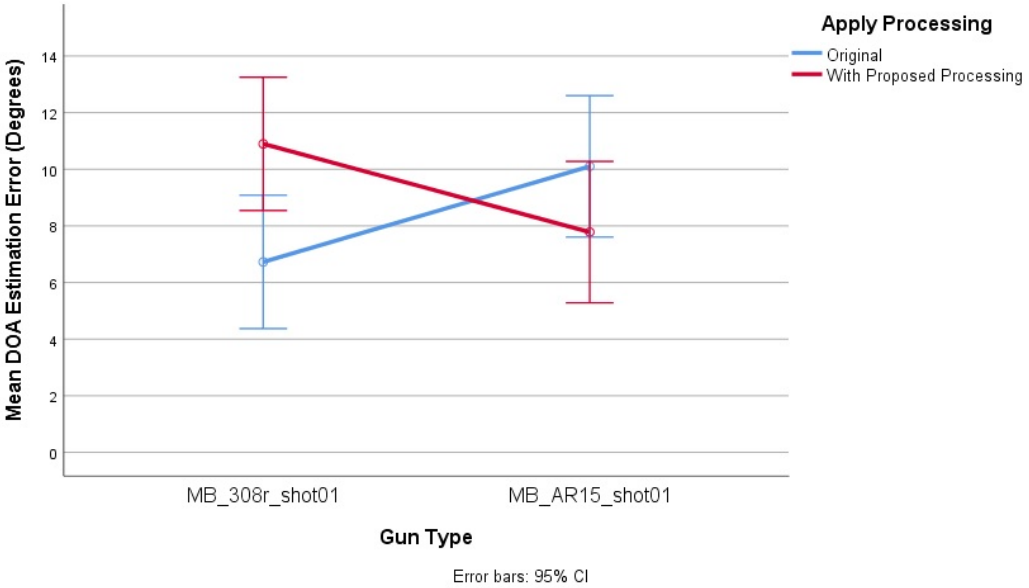


Figure 42: DOA Estimation Performance Comparison with the Proposed Method with respect to the Gun Type

CHAPTER 7

DISCUSSION

7.1 Discussions on Muzzle and Shock Wave Detection

The novel method proposed by this thesis improves SNR in challenging sound field scenarios in an adaptive manner. The useful output of such a system should enhance the performance of signal-processing tasks. Chapter 4 revealed such improvements for SNR level as a product of high-resolution precision at detecting the impulsive signal instant and adapting to frequencies of interest. More focused on the gunshot problem of this study Chapter 5 reveals detection performance with F-score, error rate, and ROC curves.

At first glance, it should be noted that the SNR improvement is evident by the calculations and analysis of Chapter 4. Moreover, the detection method focuses on extracting directional and impulsive signals, even with very low SNR. Thanks to these two aspects, the offered method has promising detection performance for both muzzle blast and shock wave components.

Concerning a possible implementation of the method for drones, the following aspects can be noted. First, judging by the geometric aspects of the problem domain, the simulated muzzle ranges are expected to bring operational advantages up to several hundred meters. This range is larger than any onboard sound detection or DOA estimation application since most studies in the literature are directed at sound sources at several meters, as listed in Chapter 2. There are only a few studies with ranges in this scale, which are continuous narrow-band signals, taking advantage of a specific frequency and continuous signal. Moreover, such examples can address only fixed-wing UAVs with smaller noise and a smaller range of applications.

The ROC curves and AUC results, presented in Chapter 5, should be reminded here, proofs of the proposed method's detection performance. Firstly, the threshold at detection decision is selected with a sound method, with the help of ROC analysis. Still more importantly, the curves make it evident that the proposed method acts like a decent binary classifier for the problem at hand.

Concerning shock wave detection, the miss distances used in the dataset scenarios are purposefully limited. As discussed in Section 2.1, the shock wave component is present only if the microphone array is inside the cone of shock wave travel, meaning that the solution should not depend on its presence. Moreover, once there is a shock wave, such as a scenario where the gunshot fire was aimed at the drone, the shock wave's SPL is much larger than the muzzle blast's. Examples in Chapter 3 visualize this larger SNR and clearance at shock wave signals.

Another conclusion can be made on the detection accuracy regarding time resolution. As stated in Chapter 5, a sample is accounted as correct detection if the detection instant is in the 50 ms vicinity of the known truth table. This definition is in line with examples from literature and is also necessary since there has to be a threshold for concluding. Also, judging by problem geometry, such a threshold is in line with the requirements of the problem field. In this judgment, remember that:

- the dataset sound samples of this study are 1 s long
- for the largest range (800 m), the arrival time difference between muzzle blast and shock wave components dictates 3 s long simulations
- the minor frame for detection algorithm windowing is 5 ms, and the major frame for peak-in-neighboring-windows decision is 45 ms
- 50 ms resolution at the system's output is more than a human user can react

More precisely than this minimal threshold, detection results reveal less than 5 ms between the detection output and the known truth table value for all muzzle blast and shock wave scenarios. This error interval of 50 ms includes the latency brought by the nine peak-in-neighboring-windows and is much smaller for the actual detection algorithm in a singular processing window. To sum up, it is considered that such precision and latency are more than required for the muzzle blast and shock wave detection problem.

As for the DOA estimation part of the problem, the generic conclusion can be noted as the DOA estimation error has decreased with the proposed method for different DOA algorithms, ranges, incidence angles, gun types, and noise types. Note that this improvement comes as an outcome of the SNR enhancement. That is, the method does not utilize means specifically improving DOA estimation accuracy. It is rather focused on an ultra-performance detector. Still, there is a clear improvement in DOA estimation performance. As illustrated in Chapter 3 and Chapter 6, this enhancement comes from two different aspects: the increase at signal SNR thanks to CWT scales' adaptive selection and the benefit of ultra-resolution insight about signal detection instant.

As a result, the DOA estimation errors are reported as lower than 15 degrees even for several hundreds of meters range, up to 800 m. Instead of focusing on converting this angle error to distance error for the known range, the following aspects should be considered.

First, concluding in terms of angles is more reasonable in terms of functional scenarios. Camera usage on drones is already very common, as stated in Chapter 1. Such an angle resolution is more than adequate for steering the camera to frames of interest. Secondly, only two DOA estimation algorithms were utilized in this study as references, which is a very effective factor for DOA error performance. This study's method concentrates on knowing which exact moment to focus so that, instead of several degrees of improvement, DOA estimations that are normally impossible are made possible. Thirdly, note that instead of only DOA estimation, localization of sound sources is a broader question, also adding the location of the sound source in 3-D. There are many methods for such a broader problem, and this method serves as a signal improvement stage for any possible implementation.

Although related studies from literature differ from the scenarios of this thesis in one aspect or another, it is still possible to compare their performances. It is impossible to compare the result metrics directly;

Table 16: Comparison of Results and Performance with Related Studies

RELATED STUDY	Results/Performance	Range / SNR
Strauss et al., 2018 (IEEE SPCUP 2019 Challenge Results)	DOA estimation error below 5 degrees is achieved.	Less than 10 m in the free flight task
Basiri, M., et al., 2012	DOA estimation error is less than 10 degrees. Detection range of 150 meters is achieved.	Up to 150 m
Go, Y.J., Choi, J.S., 2021	DOA estimation error is less than 8 degrees	For ranges up to 100 m. At 150 m, the error increases to more than 12 degrees.
Manamperi et al., 2022	DOA estimation error is less than 9.43 degrees	For SNRs between -15 dB and -30 dB
Method of this Study (Help of precise detection, only)	DOA estimation error is less than 18 degrees	Up to 800 m range.
Method of this Study (Reconstructed-cleaned signal)	DOA estimation error is less than 10 degrees	Up to 800 m range.

however, similarities and differences can be commented on. Table 16 summarizes such a comparison from the studies of [38], [15], [41], [42] and the method proposed by this study. The details of the related work were already given in Section 2.5.

7.2 Notes for a Real-Life Drone Implementation

When implementing this concept on drones, the following topics from the problem field and real-life aspects can be considered. First, for a drone in flight or landing position, since the geometry of the DREGON dataset's microphone array has a 3-D, directionless custom geometry, reasonable performance is expected for gunshots in any direction. Apart from being theoretically obvious since the direction information is not utilized in the solution or the array type, the independence of different directions is also illustrated by different DOA scenarios in this thesis. Therefore, the application of the proposed method is not limited to this thesis's range and direction scenarios.

Another conclusion can be made about positioning the microphone array on the drone. The performance of such a system depends very much on the array positioning on the drone. Moreover, the microphone selection and microphone geometry (spacing, distance from the drone, types of microphones, etc.) can all be included in this concern. This study does not examine all these parameters one by one. Instead, as mentioned above, the relation between the method and these parameters is commented on. Moreover, this study can offer the same geometry and positioning of the array on the drone at the least.

This brings up another aspect of how the datasets and scenarios were handled. Besides the microphone array geometry and the positioning on the drone, the interaction between the two is expected to bring

up additional problems for real-life implementations. One such aspect is the practical difference between noise components for different array elements. This electrical and physical phenomenon is out of the scope of any simulations. Even more challenging is the airflow around the drone. The microphones will record this as noise components since drones create a challenging air flow around them, as discussed in Section 2.2. Modeling such airflow would be subject to another challenging study, still not making certain that the simulation of the microphone array is realistic enough to conclude that the method will work in real-life conditions. Instead, this study uses drone (noise) samples as whole-microphone-array recordings, meaning that the drone noise utilized is not a simulation of N-channels; it is an actual array recording. Apart from electrical and simulation concerns, this thesis's methodology is realistic in terms of one of the most challenging factors of this problem: the noise due to the airflow around the drone.

7.3 On Possible Challenges Not Covered by the Dataset Scenarios

Simultaneous gunshot cases were not considered in the dataset and results of this study. That is, in real-life scenarios, it is possible that there are simultaneous gunshots from different guns and different DOAs. Since the processing window length is small, the probability of two gunshots happening in the same window is expected to be small. As illustrated by the ROC analysis for different window lengths in Chapter 5, 5-ms window length brings better performance than 10-ms window length for detection. That is as expected, due to the advantage when the window length is comparable to muzzle blast signal length. If the simultaneous gunshots fit into even the 5 ms windows, it is still possible to try smaller processing windows up to approximately 1 ms, which is the typical muzzle blast length.

Moreover, remember from Chapter 3 that the final detection is decided depending on an N-neighboring window (which is nine windows for the specific trials of the thesis). If the gunshot signals are so close in time that they coincide even with small windows, there will be a detection decision that can not distinguish the two shots. Still, the detection decision is correct. Successful detection with possible false DOA or other signal processing results may be expected in this case.

As another challenging scenario, there are no sound components in the dataset other than the noise and the gunshot signals. The method favors the directional signals over the noise coming from all directions and the impulsive signals over the continuous noise. In this case, a third component in the sound field may mislead the algorithm if it is directional and transient yet is not a gunshot signal. It is possible to include such signals in the dataset mixes and assess the performance; however, it was not implemented in the scope of this study.

Firstly, note that other transient signals, such as fireworks, clapping, or sudden crashes, may appear as impulsive, but the situation can be remedied since their duration is longer than the gunshot signals'. Also, one can change the window length parameter in case such signals are causing false positives. Moreover, remember that, by design, the method has a focus on good rejection performance for false positives. This was so, since the airflow, wind gusts, and shifting motor RPMs create transients in the array data. The initial onboard gunshot recordings reveal such false positives in Chapter 3, while the ROC analysis of Chapter 5 shows that the method successfully removes such false positives. All directional, transient, and non-gunshot signals are false positives, and the method has a fair performance for rejecting such signals. Lastly, being a binary classifier problem, it is always possible to select the threshold of the classifier to trade off between false positives and false negatives, in case false positives

are critical for the specific implementation. Conversely, many drone-mounted applications can accept false positives if it is affordable. One such example would be to steer the camera to the angle interval of interest if there is a suspicious detection decision.

7.4 On Generalizability of the Proposed Method to Similar Problems

The method offered by this study was worked out specifically for gunshot detection and DOA estimation on drones. Starting with importance of a sound detection step, and since it is not practical to account for the whole noise subspace, an adaptive means for detection was sought.

Concerning generalizability, let us first consider the generalizability of the solution for different parameters of this problem. That is, many kinds of drones, gunshots, array geometries, flight maneuvers, gunshot occurrence azimuth and elevation, characteristics of the bullet (in case the projectile was supersonic)... There are many parameters in place, as discussed in Chapter 2. Even weather conditions such as temperature and humidity may be counted as affecting factors. Despite its impracticality to check all parameters and impossible to claim success under all conditions, the following characteristics can be depended on.

First, the solution does not depend on a specific array geometry or sound field geometry, provided that any 3-D symmetrical geometries are used. Secondly, although specific examples of gun and drone sounds are used, the signals' frequency characteristics are not dependent on. That is, the method selects advantageous frequency regions in terms of signal vs. noise dominance. It does not manually select specific bands or eliminate others. More soundly, there is no frequency-specific action in the algorithm. This feature was a primary design decision in the first place, considering real-life feasibility and nonstationary noise characteristics.

The following lists the characteristic challenges of the problem:

- Very low SNR
- Extra-ordinarily varying noise characteristics
- Noise sources close to microphone array, signal of interest is at a large range
- Both noise and signal have wide-band frequency characteristics, making simple band-selection style denoising not useful

While working on the solution, the following domain characteristics were important in making this solution accurate:

- Impulsive nature of the signal of interest
- Signal of interest being directional
- Having many microphones brings advantages of array processing; using microphones' coherence in this case, brings advantage

- Although not completely separated, it is possible to select advantageous frequency regions in terms of signal vs. noise dominance

The methodology may be useful for other signal-processing tasks with similar challenges and similar advantageous characteristics. A signal processing problem that needs to favor the instantaneous signal above the continuous noise may find the offered detection method useful. Similarly, where the signal is far and directional with respect to the microphone array, while noise sources are close at hand and in all directions with respect to the microphone array, the method may bring an advantage. Therefore, applications other than gunshot forensics may successfully apply the technique, too. In other words, the methodology of this study may be helpful in tasks other than gunshot forensics.

The method may find usage in radar and sonar systems when considering array processing applications other than sound processing. To extract low SNR events in radar applications, it may be useful for especially instantaneous signals of interest. Medical applications are another example; since multiple sensors are used, detecting patterns while eliminating false positives is usually the concern. Another possible field is array processing for astronomy applications, where directional signal transmission or reception is essential. That is, an updated version of the proposed method may be useful to combine elements in a phased array such that some signals experience destructive inference and others experience constructive inference. Applications of ultrasound imaging, smart antennas, and underwater acoustic imaging systems may also apply a similar approach, considering the autonomous decisive nature of the proposed solution.

The method can also be viewed as a generic, adaptive SNR enhancer. Despite being focused on detection and DOA estimation, the gain from the proposed solution is actually through adaptively enhancing the input signal. In other words, the benefit is expected not to be specific to the detection and DOA estimation, but is a way to obtain an SNR-improved signal focused on impulsive directional signal components. In other words, the method may be used as an SNR-improving pre-processing phase for any problem. Knowing which instant to concentrate with a high time accuracy and which frequency to tune adaptively, the signal obtained after this pre-processing phase will improve following signal processing tasks. In cases where the frequency characteristics distortion caused by the CWT and then scale-selective inverse-CWT is unacceptable, using only the precise detection part (without scale selection) can still bring significant advantages.

One step further than a generic SNR enhancer, the approach can be packaged as a transform for impulsive signals of interest, taking a noisy signal as input and outputting the instant to focus and frequencies (or scales) to tune. For a sample that is known to include a signal of interest, such a transform may be used for extracting insight into this information or enhancing SNR. More importantly, the transform can be used on unknown samples to include a signal of interest. In such a case, this transform may yield a good rejection ratio for negative samples; that is, it should be good at eliminating false positives. The initial ROC analysis in Chapter 5 reveals promising results.

CHAPTER 8

CONCLUSIONS

This study introduces a novel method for detecting muzzle blast and shock wave components of gunshot sound with a microphone array mounted on a drone. A method is proposed that brings precise detection in time and SNR improvement. Simply put, the Array Correlation Map (ACM) used in this way distinguishes directional and transient event candidates if there is coherence in the ACM, and selection on CWT scales selects/tunes the candidate that matches with frequency of interest. The following sections comment on SNR improvement, detection results, DOA estimation results, novel features of the proposed methods, the generalizability of the solution to other problems and domains, and possible future work.

8.1 Key Features of the Proposed Method

The solution addresses challenges of highly non-stationary ego-noise, ultra-low SNR, and vicinity of the sensors to noise sources and varying nature of drone and gunshot sounds. The performance of the proposed solution in terms of SNR improvement, signal detection, and signal DOA estimation are given in the three results chapters of this thesis. The proposed method has the following novel features, making these results possible.

Firstly, a way of using the multi-channel array's output as a whole, namely the Array Correlation Map (ACM), is offered. The idea is to obtain an N x N state from the whole array's reading and use this information to suit the needs of the specific problem. For this directional detection problem, the post-step with the ACM favors directional signal over ego-noise in the vicinity of the sensor array, though one may use that map differently for other problem geometries.

Secondly, the method takes advantage of the wavelet transform, with its high resolution in time, but still uses frequency information (CWT scales). On top of regular usage of scales' info, such as applying filter banks or hard-coding specific band selection, an adaptive manner is used for selecting the CWT scales. For this specific problem, this means being adaptive to different gunshots, drones, and different instants of the flight. This adaptive feature is crucial for the non-stationary ego noise of such a carrier device. For other signal processing problems that require low SNR and adaptive nature, the same method may bring similar results.

The third one is a tuning feature. The known strength of CWT is that it has a variable time-frequency resolution that yields a sharp time-frequency representation for nonstationary signals. This strength is made sharper with ACM usage. Despite the strong drone noise, the method can "tune to" muzzle

and shock components. Typically, these two signal components dominate in different frequencies; however, without prior knowledge of the specific gunshot, the method automatically catches which scales have sharp peaks and uses those scales.

Another key feature is the extra precision in time. For this specific problem, the processing window's length is 5 ms, considering the typical muzzle blast signal length of 1 to 3 ms. A 9-window major frame is used for finding sharp occurrences in neighboring windows. Therefore, the method outputs its decisions within 45 ms. Also note that the windowing is applied in a sliding window manner, meaning signals placed closer than 45 ms can still be processed correctly. This is the case even with the shock wave and muzzle blast components of the same shot, which are positioned closely in time. Also, checking the detection results against the known truth tables, it is observed that the final decision is true in the vicinity of 1 to 2 ms to truth table values. This precision at the detection step brings advantages in the following steps: DOA estimation, localization, or classification tasks.

Moreover, it should be noted that the method does not depend on the direction of arrival of the signal of interest. The array geometry information is not used by the detection step at all. Therefore, it is clear that the ACM + adaptive CWT detection does not depend on the signal's arrival direction. The same claim is valid for any application interested in the SNR improvement part of the solution. As for the DOA estimation part, the results reported in Chapter 6, for most DOA angles, the DOA estimation performance does not change significantly with arrival angle, either. It should be noted that the specific angles around 90 degrees yield less performance; however, this correlation with DOA comes from the DOA method itself rather than the methods proposed by this study.

Lastly, the method does not require previous knowledge of noise signals. The simulations and results of this study use only a limited number of noise samples and noise signal instants. Considering the generalizability of the solution to many drones and the non-stationary nature of the drone noise in time, it is impossible to cover a representative set for possible noise cases. Instead, an adaptive solution was sought in this study, not using any apriori information on the specific drone type or flight instant other than typical drone noise characteristics. The few noise samples are reported in the result chapters of this thesis to reveal the concepts.

8.2 Future Work

From a performance enhancement point of view, variations of Array Correlation Map (ACM) usage should be noted first. As discussed in Chapter 3, thanks to the ACM, whole-array data is used to emphasize the directional impulsive signal. Other combinations may improve detection or SNR performance or yield a better ROC curve and AUC characteristics. The specific ACM usage reported in this thesis visualizes the concept and yields reasonable performance. Moreover, once the SNR is improved, other than detection and DOA estimation, complete sound source localization in 3-D can be applied, which was not covered in the scope of this thesis.

Regarding the scope of this study, the effect of wind was not accounted for in this thesis. There are methods for modeling wind noise and its impact on sensors of a microphone array, which can be added to the sound field simulator. Still, it should be noted that the airflow around the drone is a larger contributor to noise, which was accounted for since the multi-channel noise data is not synthesized and are real drone recordings. Another aspect not covered in this thesis is the degradation performance of the method upon losing one or more sensors. Since the technique depends on the coherence of multiple

sensors regarding detection decisions, the performance is expected to be still fair, even when several microphones are lost.

Regarding the generalizability of the method, applying the solution to other challenging array processing tasks would be interesting. Trials and performance comparisons with specific low-SNR detection problems for radar, smart antenna, or image-processing tasks may bring promising results. Finally, for the applicability of the method as a generic transform or SNR enhancer, a similar approach is necessary, using known datasets and performance comparisons against known solutions.

REFERENCES

[1] A. K. Bandi, M. Rizkalla, and P. Salama, "A novel approach for the detection of gunshot events using sound source localization techniques," in *2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 494–497, IEEE, 2012.

[2] J. Sallai, P. Volgyesi, K. Pence, and A. Ledeczi, "Fusing distributed muzzle blast and shockwave detections," in *14th International Conference on Information Fusion*, pp. 1–8, IEEE, 2011.

[3] R. C. Maher, "Modeling and signal processing of acoustic gunshot recordings," in *2006 IEEE 12th Digital Signal Processing Workshop & 4th IEEE Signal Processing Education Workshop*, pp. 257–261, IEEE, 2006.

[4] R. C. Maher and S. R. Shaw, "Deciphering gunshot recordings," in *Audio Engineering Society Conference: 33rd International Conference: Audio Forensics-Theory and Practice*, Audio Engineering Society, 2008.

[5] M. Yılmaz and B. G. Kılıç, "Evaluation of acoustic gunshot localization methods on helicopters with environmental sound simulations," in *Cybernetics Perspectives in Systems: Proceedings of 11th Computer Science On-line Conference 2022, Vol. 3*, pp. 366–385, Springer, 2022.

[6] M. Yılmaz, "A comprehensive simulation of factors that affect performance of array based acoustic gunshot localization on helicopters," Master's thesis, Middle East Technical University, 2016.

[7] R. C. Maher, "Acoustical characterization of gunshots," in *2007 IEEE Workshop on Signal Processing Applications for Public Security and Forensics*, pp. 1–5, IEEE, 2007.

[8] A. Satué-Villar and J. Fernández-Rubio, "Environment simulator for audio signals," in *2005 13th European Signal Processing Conference*, pp. 1–4, IEEE, 2005.

[9] L. Gerosa, G. Valenzise, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and gunshot detection in noisy environments," in *2007 15th European Signal Processing Conference*, pp. 1216–1220, IEEE, 2007.

[10] S. D. Beck, H. Nakasone, and K. W. Marr, "An introduction to forensic gunshot acoustics," in *Proceedings of the 162nd Meeting of the Acoustical Society of America, San Diego, CA, USA*, vol. 31, 2011.

[11] A. L. Ramos, S. Holm, S. Gudvangen, and R. Otterlei, "A spectral subtraction based algorithm for real-time noise cancellation with application to gunshot acoustics," *International Journal of Electronics and Telecommunications*, vol. 59, no. 1, pp. 93–98, 2013.

[12] F. Bronuzzia, L. Monaib, and M. Patrucccob, "Correct and effective characterization of firearms noise: a basic aspect to provide reliable input data for the reduction of emitted noise from shooting ranges in urbanized area," *Chem. Eng*, vol. 26, pp. 507–512, 2012.

[13] A. Chacon-Rodriguez, P. Julian, L. Castro, P. Alvarado, and N. Hernández, "Evaluation of gunshot detection algorithms," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 2, pp. 363–373, 2010.

[14] O. Sugiyama, S. Uemura, A. Nagamine, R. Kojima, K. Nakamura, and K. Nakadai, "Outdoor acoustic event identification with dnn using a quadrotor-embedded microphone array," *Journal of Robotics and Mechatronics*, vol. 29, no. 1, pp. 188–197, 2017.

[15] M. Basiri, F. Schill, P. U. Lima, and D. Floreano, "Robust acoustic source localization of emergency signals from micro air vehicles," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4737–4742, IEEE, 2012.

[16] L. Wang and A. Cavallaro, "Ear in the sky: Ego-noise reduction for auditory micro aerial vehicles," in *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 152–158, IEEE, 2016.

[17] L. Wang and A. Cavallaro, "Time-frequency processing for sound source localization from a micro aerial vehicle," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 496–500, IEEE, 2017.

[18] G. Ince, K. Nakadai, and K. Nakamura, "Online learning for template-based multi-channel ego noise estimation," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3282–3287, IEEE, 2012.

[19] L. Wang and A. Cavallaro, "Microphone-array ego-noise reduction algorithms for auditory micro aerial vehicles," *IEEE Sensors Journal*, vol. 17, no. 8, pp. 2447–2455, 2017.

[20] M. Strauss, P. Mordel, V. Miguet, and A. Deleforge, "Dregon: Dataset and methods for uav-embedded sound source localization," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–8, IEEE, 2018.

[21] B. Günel and H. Hacihabiboglu, "Sound source localization: Conventional methods and intensity vector direction exploitation," in *Machine Audition: Principles, Algorithms and Systems*, pp. 126–161, IGI Global, 2011.

[22] G. L. Duckworth, D. C. Gilbert, and J. E. Barger, "Acoustic counter-sniper system," in *Command, control, communications, and intelligence systems for law enforcement*, vol. 2938, pp. 262–275, SPIE, 1997.

[23] S. Khan, A. Divakaran, and H. S. Sawhney, "Weapon identification using hierarchical classification of acoustic signatures," in *Sensors, and command, control, communications, and intelligence (C3I) technologies for homeland security and homeland defense VIII*, vol. 7305, pp. 230–234, SPIE, 2009.

[24] P. Suman, S. Karan, V. Singh, and R. Maringanti, "Algorithm for gunshot detection using mel-frequency cepstrum coefficients (mfcc)," in *Proceedings of Ninth International Conference on Wireless Communication and Sensor Networks: WCSN 2013*, pp. 155–166, Springer, 2014.

[25] K. Furukawa, K. Okutani, K. Nagira, T. Otsuka, K. Itoyama, K. Nakadai, and H. G. Okuno, "Noise correlation matrix estimation for improving sound source localization by multirotor uav," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3943–3948, IEEE, 2013.

[26] E. Ghaderpour, S. D. Pagiatakis, and Q. K. Hassan, "A survey on change detection and time series analysis with applications," *Applied Sciences*, vol. 11, no. 13, p. 6141, 2021.

[27] A. Graps, "An introduction to wavelets," *IEEE computational science and engineering*, vol. 2, no. 2, pp. 50–61, 1995.

[28] A. Grossmann and J. Morlet, "Decomposition of hardy functions into square integrable wavelets of constant shape," *SIAM journal on mathematical analysis*, vol. 15, no. 4, pp. 723–736, 1984.

[29] E. Alickovic, J. Kevric, and A. Subasi, "Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction," *Biomedical signal processing and control*, vol. 39, pp. 94–102, 2018.

[30] A. Bhattacharyya, M. Sharma, R. B. Pachori, P. Sircar, and U. R. Acharya, "A novel approach for automated detection of focal eeg signals using empirical wavelet transform," *Neural Computing and Applications*, vol. 29, pp. 47–57, 2018.

[31] J. Fernández-Macho, "Wavelet multiple correlation and cross-correlation: A multiscale analysis of eurozone stock markets," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 4, pp. 1097–1104, 2012.

[32] Q. Nguyen and J. Choi, "Selection of the closest sound source for robot auditory attention in multi-source scenarios," *Journal of Intelligent & Robotic Systems*, vol. 83, pp. 239–251, 2016.

[33] M. Basiri, F. Schill, P. U. Lima, and D. Floreano, "Localization of emergency acoustic sources by micro aerial vehicles," *Journal of Field Robotics*, vol. 35, no. 2, pp. 187–201, 2018.

[34] K. Okutani, T. Yoshida, K. Nakamura, and K. Nakadai, "Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3288–3293, IEEE, 2012.

[35] F. Ruffier, S. Benacchio, F. Expert, and E. Ogam, "A tiny directional sound sensor inspired by crickets designed for micro-air vehicles," in *SENSORS, 2011 IEEE*, pp. 970–973, IEEE, 2011.

[36] H. de Bree, I. J. Wind, I. E. Druyvesteyn, and H. te Kulve, "Multi purpose acoustic vector sensors for battlefield acoustics," in *Proceedings of the DAMA Conference*, 2010.

[37] K. Hoshiba, K. Washizaki, M. Wakabayashi, T. Ishiki, M. Kumon, Y. Bando, D. Gabriel, K. Nakadai, and H. G. Okuno, "Design of uav-embedded microphone array system for sound source localization in outdoor environments," *Sensors*, vol. 17, no. 11, p. 2535, 2017.

[38] A. Deleforge, D. Di Carlo, M. Strauss, R. Serizel, and L. Marcenaro, "Audio-based search and rescue with a drone: highlights from the ieee signal processing cup 2019 student competition [sp competitions]," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 138–144, 2019.

[39] B. Yen and Y. Hioka, "Noise power spectral density scaled snr response estimation with restricted range search for sound source localisation using unmanned aerial vehicles," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2020, pp. 1–26, 2020.

[40] B. Dahlan, W. Mansoor, M. Abbasi, and P. Honarbakhsh, "Sound source localization for automatic camera steering," in *The 7th International Conference on Networked Computing and Advanced Information Management*, pp. 20–25, IEEE, 2011.

[41] Y.-J. Go and J.-S. Choi, "An acoustic source localization method using a drone-mounted phased microphone array," *Drones*, vol. 5, no. 3, p. 75, 2021.

[42] W. Manamperi, T. D. Abhayapala, J. Zhang, and P. N. Samarasinghe, "Drone audition: Sound source localization using on-board microphones," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 508–519, 2022.

[43] M. Yılmaz and B. Günel, "Investigation of acoustic source localization algorithm performances for gunshot localization on helicopters," in *2016 24th Signal Processing and Communication Application Conference (SIU)*, pp. 1885–1888, IEEE, 2016.

[44] H.-E. De Bree, "Acoustic vector sensors increasing uav's situational awareness," tech. rep., SAE Technical Paper, 2009.

[45] H.-E. De Bree and G. De Croon, "Acoustic vector sensors on small unmanned air vehicles," *the SMi Unmanned Aircraft Systems*, 2011.

[46] I. 9613-1, "Attenuation of sound during propagation outdoors, part 1: Atmospheric absorption," 1993.

[47] C. M. Harris, "Absorption of sound in air versus humidity and temperature," *The Journal of the Acoustical Society of America*, vol. 40, no. 1, pp. 148–159, 1966.

[48] T. K. Routh *et al.*, *Observing variation of acoustical characteristics of several common firearms in a quasi anechoic environment at a high sampling rate*. PhD thesis, Montana State University-Bozeman, College of Engineering, 2016.

[49] T. K. Routh and R. C. Maher, "Recording anechoic gunshot waveforms of several firearms at 500 kilohertz sampling rate," in *Proceedings of Meetings on Acoustics 171ASA*, vol. 26, p. 030001, Acoustical Society of America, 2016.

[50] A. Hero and S. Schwartz, "Alternatives to the generalized cross correlater for time delay estimation," in *ICASSP'84. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 9, pp. 624–627, IEEE, 1984.

[51] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE transactions on acoustics, speech, and signal processing*, vol. 24, no. 4, pp. 320–327, 1976.

[52] A. Mesaros, S. Adavanne, A. Politis, T. Heittola, and T. Virtanen, "Joint measurement of localization and detection of sound events," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 333–337, IEEE, 2019.

[53] H. Lim, J.-S. Park, and Y. Han, "Rare sound event detection using 1d convolutional recurrent neural networks.," in *DCASE*, pp. 80–84, 2017.

[54] "Dcase2017 challenge web page." `https://dcase.community/challenge2017/index`. Accessed: 2023-02-19.

[55] G. E. Poliner and D. P. Ellis, "A discriminative model for polyphonic piano transcription," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–9, 2006.

[56] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[57] K. A. Spackman, "Signal detection theory: Valuable tools for evaluating inductive learning," in *Proceedings of the sixth international workshop on Machine learning*, pp. 160–163, Elsevier, 1989.

[58] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[59] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *ICASSP'83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 8, pp. 336–339, IEEE, 1983.

[60] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 7, pp. 984–995, 1989.

[61] Y. Jia, M. Jia, L. Li, J. Gao, and S. Yang, "Multiple speech source separation by using mvdr for b-format recordings," in *Proceedings of the 2020 9th International Conference on Computing and Pattern Recognition*, pp. 400–404, 2020.

[62] L. Tzafri and A. J. Weiss, "High-resolution direct position determination using mvdr," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6449–6461, 2016.

[63] J. S. Garofolo, "Timit acoustic phonetic continuous speech corpus," *Linguistic Data Consortium, 1993*, 1993.

APPENDIX A

ROC THRESHOLD PERFORMANCES

Table 17: Muzzle Detection Strength ROC Threshold Performances for Window Length of 256 Samples

Threshold peak/peak (unitless)	Sensitivity	1 - Specificity
-0.652	1.000	1.000
0.393	1.000	0.997
0.439	1.000	0.994
0.472	1.000	0.992
0.519	1.000	0.989
0.548	1.000	0.986
0.579	1.000	0.983
0.601	1.000	0.961
0.631	1.000	0.917
0.668	1.000	0.906
0.700	1.000	0.883
0.735	1.000	0.872
0.759	1.000	0.850
0.768	0.967	0.850
0.770	0.967	0.817
0.776	0.967	0.808
0.799	0.967	0.775
0.819	0.967	0.742
0.825	0.967	0.708
0.831	0.967	0.697
0.839	0.967	0.675
0.845	0.967	0.642
0.851	0.967	0.619
0.863	0.967	0.597
0.896	0.967	0.553
0.922	0.933	0.553
0.961	0.933	0.519
0.997	0.933	0.517

1.012	0.933	0.514
1.035	0.933	0.503
1.075	0.933	0.469
1.116	0.933	0.436
1.130	0.933	0.414
1.143	0.933	0.369
1.153	0.933	0.367
1.165	0.933	0.364
1.218	0.933	0.342
1.341	0.933	0.319
1.504	0.933	0.297
1.612	0.933	0.264
2.531	0.933	0.242
3.425	0.933	0.239
3.425	0.933	0.236
3.425	0.933	0.233
3.426	0.933	0.231
3.430	0.933	0.064
3.463	0.917	0.064
3.504	0.900	0.064
3.593	0.883	0.064
4.010	0.867	0.064
4.424	0.850	0.064
4.524	0.833	0.064
4.662	0.817	0.064
5.565	0.800	0.064
6.365	0.792	0.064
6.442	0.783	0.064
6.647	0.775	0.064
6.857	0.767	0.064
6.944	0.750	0.064
7.049	0.733	0.064
7.188	0.717	0.064
7.791	0.700	0.064
8.415	0.692	0.064
8.489	0.683	0.064
8.594	0.675	0.064
8.722	0.675	0.058
8.747	0.667	0.058
8.754	0.667	0.053
8.782	0.650	0.053
8.834	0.650	0.047
8.885	0.650	0.042
8.953	0.633	0.042
9.020	0.617	0.042
9.076	0.617	0.036

9.180	0.617	0.031
9.247	0.600	0.031
9.396	0.600	0.025
13.591	0.600	0.019
17.791	0.583	0.019
17.951	0.567	0.019
18.179	0.550	0.019
20.956	0.533	0.019
23.572	0.533	0.017
23.720	0.533	0.014
24.005	0.517	0.014
24.439	0.500	0.014
24.931	0.483	0.014
25.244	0.475	0.014
25.661	0.458	0.014
26.195	0.450	0.014
26.449	0.442	0.014
26.571	0.433	0.014
26.724	0.425	0.014
26.855	0.417	0.014
27.155	0.408	0.014
28.624	0.400	0.014
30.116	0.383	0.014
30.740	0.367	0.014
31.367	0.350	0.014
33.526	0.333	0.014
35.542	0.317	0.014
35.851	0.300	0.014
36.667	0.283	0.014
47.993	0.267	0.014
59.198	0.250	0.014
59.767	0.233	0.014
60.879	0.217	0.014
65.646	0.200	0.014
69.331	0.183	0.014
69.397	0.167	0.014
69.484	0.150	0.014
69.566	0.133	0.014
69.792	0.117	0.014
70.114	0.100	0.014
70.615	0.083	0.014
71.098	0.067	0.014
71.479	0.050	0.014
71.935	0.033	0.014
72.318	0.017	0.014
113.747	0.000	0.014

613.672	0.000	0.011
1078.824	0.000	0.008
1220.687	0.000	0.006
1378.428	0.000	0.003
1401.784	0.000	0.000

APPENDIX B

CV

Murat YILMAZ

Electrical & Electronics Engineer

EDUCATION

2016 – 2023 **Middle East Technical University**
Informatics Institute, Information Systems (Ph.D.)

2012 – 2016 **Middle East Technical University**
Informatics Institute, Information Systems (M.S.)

2006 - 2011 **Middle East Technical University**
Faculty of Engineering, Electrical and Electronics Engineering

2002 - 2006 **Bursa Anatolian High School**

WORK EXPERIENCE

2022- **Turkish Aerospace, Inc.**
Software Engineering Manager
DO-178C DAL-A flight control SW development, verification, certification
Four reporting groups, ~10 SW products
Created a DEVOPS team, lead CI transition of projects
Creation of new program: DO-178C DAL-A mission computer software

2019- 2022 **Turkish Aerospace, Inc.**
SW Development Chief Engineer
DO-178C DAL-A Flight control SW development & certification
Coded & managed Avionics, Flight Control and utility SW projects
Both Model Based Design and manual coding
Managed three avionics SW projects in parallel through delegation

Airborne Electronic Hardware CVE for DO-254 compliance
HW compliance & airworthiness checking
Auditing & reporting of 30+ equipment from 15+ different suppliers

2011- 2019 **Turkish Aerospace, Inc.**
Design Engineer
Flight Control Software Development
Integrity & Vxworks RTOS Integration
SW-HW integration & I/O management SW design
Real-time Software Development and Integration
YARASA+ Project - Avionic Software Development

2010 - 2011 **Turkish Aerospace, Inc.**
Part-time Design Engineer
YARASA Project-Software Regression Testing

PROFESSIONAL PROJECTS

- **T625 Autopilot SW**
  - DO-178C DAL-A Flight Control Software project. I have led a team of twelve developers, developing the first DO-178C certified DAL-A autopilot SW developed in Turkey.
  - First transition to formal SW lifecycle, formal releases and continuous integration. First (and following) releases of the program. I have led 20 SW engineers through this transition.
  - Code & design reviews, SCA, OCA, WCET

- **GOKBEY SW Development Tools & Infrastructure - SW Development Chief Engineer,** 2019-Ongoing

  - ~20 SW tools that serve in different phases of real-time system design and SW development
  - Digital Twin of Integration Laboratory equipment

- **GOKBEY Helicopter - Airborne Electronic Hardware CVE activities,** 2016-Ongoing

  - DO-254 compliance checking and auditing for all HW units of the helicopter
  - ~30 hardware units, worked with >15 different HW suppliers, both local & international
  - Performed / led SOI audits with delegation from certification authority
  - Auditing and SOI reporting in accordance with the FAA Job Aid – "Conducting Airborne Electronic Hardware Reviews"
  - Built a HW compliance team, mentoring four new CVEs

- **T929 Avionics Operational Flight Program – Mission Computer SW**

  - Pilot Interface GUI & Logic SW. Using MBD, built from scratch with SCADE Suite and SCADE Display. Completely new SW architecture with modular design
  - Flight Controller SW. Reuse of GOKBEY Flight Controller SW architecture with major functionality additions

- **Flight Control Computer – Real-time SW developer**

  - Real-time system & SW design
  - SW planning, requirement, coding processes for Flight Control Computer (FCC)
  - Integration of both VxWorks 653 and GHS Integrity-178 tuMP RTOS
  - Integration to target HW, both PPC7448 and PPC2041 architectures
  - All FCC interfaces: ARINC429, Resolver, LVDT, PWM, discrete & analog drives, CPU-synchronization
  - Hardware requirements and technical subcontractor issues for Flight Control Computer HW

- **Manned Helicopter Autopilot Project**

  - Real-time SW design & development for Flight Control Computer (FCC) with VxWorks RTOS running on NAI's SIU3 hardware.
  - HIL and aircraft level integration of FCC with aircraft systems sensors and actuators.
  - Preparation of the Software Design Document (SDD)

- **YARASA+ Project**

  - Multi-threaded SW design & POSIX compliant code
  - Porting of legacy Black Hawk Operational Flight Program (3.2M SLOC)
    - from VxWorks to Green Hills Integrity RTOS
    - to two different avionics computers (AITECH and GE FANUC)
  - Data Concentrator Unit integration t avionics OFP
  - VAPS integration (OpenGL porting to new HW)

- **Embedded Data Collection & Onboard Signal Processing Platform**
  - Real-time multi-threaded design & development
  - Audio & array signal processing design on MATLAB
  - MATLAB Embedded code generation for Raspberry Pi target
  - Cross-compiling & integration to LINUX environment
  - Live data-stream over UDP & accompanying PC-based GUI SW

SKILLS / ABILITIES

Programming Languages:

C / C++	(project experience)
C #	(project experience)
Python	(intermediate)
Groovy	(intermediate)
MATLAB	(project experience)
SCADE	(project experience)
SIMULINK	(intermediate)
VAPS	(Intermediate)

Operating Systems:

Green Hills Integrity-178 tuMP
Green Hills Integrity 5.x
Wind River VxWorks 653
Wind River VxWorks 6.x
Linux

IDE:

Green Hills MULTI & AdaMULTI
Wind River Workbench
Visual Studio
STM32CubeIDE
MPLAB X

Languages:

Turkish	(Native)
English	(Fluent)

Standards Literacy:

DO-178C
DO-254
ARINC 653 & APEX
ARINC661
FACE 2.1
POSIX
UML 2.0
IEEE-12207

Additional Interests & Experiences:

Real-time Design	(Proj. Exp.)
Multi-thread Design	(Proj. Exp.)
Keras	(intermediate)
Tensorflow	(intermediate)
Agile Development	(Proj. Exp.)
AWS	(intermediate)
Databases	(intermediate)
CUDA GPGPU	(Beginner)

Academics & Research Areas:

Signal Detection
Signal Denoising
Sound Source Localization
Gunshot Acoustics
Sound Field Simulation
Neural Networks

TRAININGS

- **Leadership Base - Manager Training Program**, presented by Turkish Aerospace, 2021, Ankara, Turkey
- **DEOS RTOS Training,** presented by DDC-I Inc, 2019, Ankara, Turkey
- **DO-254 Training Course,** presented by RTCA, 2017, Washington D.C., USA
- **VxWorks653 Platform 2.5 Essentials and Advanced Topics Training,** presented by Windriver, 2017, Ankara, Turkey
- **Integrity RTOS and MULTI IDE Training,** presented by Green Hills Software, 2013, Santa Barbara, CA, USA

PARTICIPATIONS & AWARDS

- SW Development Chief through **first flight of GÖKBEY AFCS SW** (autopilot), first domestic / indigenous DAL-A autopilot project developed in Turkey, 03/2021
- SW Development Chief through **GÖKBEY AFCS SW SOI#3**, SW Verification Audit in the scope of DO-178C, 03/2020
- SW Development Chief through **GÖKBEY AFCS SW SOI#2**, SW Development Audit in the scope of DO-178C, 03/2019
- Alternative means for certification of Real Time Operating Systems, **Cost-reducing Suggestion Award**, Turkish Aerospace Idea Management System, 2020
- **Optimum Cost/Performance Award**
  A Microcontroller Based Robot Moving in a Grid According to Pre-defined Traffic Rules, METU Electrical and Electronics Engineering Graduate Projects Exhibition-2011

PUBLICATIONS & IP

Journal of Intelligent & Robotic Systems – *"Introducing Array Correlation Map with Adaptive CWT for Gunshot Detection on Drones"*, Publishing in Progress.

Demirer, S., Duran, S., & Yilmaz, M. (2023). *"Development of an AFCS SIL Twin for Autopilot Software Design and Integration."* In AIAA AVIATION 2023 Forum (p. 3598).

Yılmaz, M., & Kılıç, B. G. (2022, April). *"Evaluation of Acoustic Gunshot Localization Methods on Helicopters with Environmental Sound Simulations."* In Computer Science On-line Conference (pp. 366-385). Cham: Springer International Publishing.

Yılmaz, M. (2016). *"A Comprehensive simulation of factors that affect performance of array based acoustic gunshot localization on helicopters"* (Master's thesis, Middle East Technical University)

Yılmaz M., Gunel B., *"Investigation of Acoustic Source Localization Algorithm Performances for Gunshot Localization on Helicopters",* 2016, 24th Signal Processing and Communication Applications Conference, Zonguldak, pp. 1885-1888.

PATENTS & IP

Günel Kılıç, B., Yılmaz, M., *"An Impulsive Event Detection Method for Acoustic Multi-channel Data on Noisy Carrier Vehicles"*, Patent in Progress.

Yılmaz, M., Bayrak, Y. (2021), Registered Trademark: *"APSWARE™, Autopilot SW by Turkish Aerospace Inc."*

INTERESTS

Playing baglama, books on world history, playing football, skiing

REFERENCES