الجــــــــمــــهـــــوريـــــــة الجــزائـــريـــة الديمـــــقـــراطـيــــة الـــشــــعــبـيــــة

PEOPLE's DEMOCRATIC REPUBLIC OF ALGERIA

وزارة التــــــعـــــليــــــم العــــــــــــــالي والبــــــــــــحــث العــــــــــــلـــــــمــــــي

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

جـــــــــــــــــــــــــــــــامعة سعيدة – د. الطاهر مولاي –

UNIVERSITY OF SAIDA - Dr. MOULAY TAHAR

Faculty of Mathematics, Computer Science and Telecommunications

A Dissertation Submitted in Partial Fulfilment of the Requirements for a master's Degree in Telecommunications

Specialization: Telecommunications Systems

Presented by: HOCINI Kheira

GOT Israa

Sound Source Localization for Unmanned Aerial Vehicles (UAVs)

Defended on 17, June,2026 in front of the jury composed of:

Mr. HARZELLAOUI Abdelkader	MCB	President
Mr. TAMI Abdelkader	MCA	Supervisor
Mr. OUARDI Aissa	MCA	Examiner

2025 / 2026

Acknowledgments

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

Praise and thanks be to Allah, who granted us the strength, patience, and perseverance to accomplish this work.

We would like to express our deepest gratitude to our supervisor, Mr. Tami Abdelkader, for his valuable advice and continuous guidance throughout the months we spent preparing this dissertation.

We also extend our sincere thanks to all the faculty members of the Telecommunications Department for the quality education they provided during our academic journey.

Our heartfelt thanks also go to our families, Hocini and Got, for their support, love, and encouragement throughout our lives, especially during our years of study.

We would also like to thank our friends and classmates for the wonderful moments we shared together and for the beautiful memories that will remain with us forever.

To everyone who contributed, directly or indirectly, to the success of this work, we say: thank you from the bottom of our hearts.

Dedication

وَكَانَ فَضْلُ اللَّهِ عَلَيْكَ عَظِيمًا

Praise be to Allah who has blessed me with all the good things I have today. All praise and thanks are due to You, my Lord.

I dedicate this success to myself for all the hard work, perseverance, and determination, and for never giving up. I am proud of myself.

I also dedicate my success to my family, my support and source of strength.

To my father Ali: may Allah reward you with the best reward for your unconditional love and endless support. I love you so much, and words can never fully express my gratitude to you.

To my beloved mother: the dream has come true, but you are not here. Today, I dedicate my success to you. I deeply wished you were here so I could see the joy in your eyes. Thank you for your prayers that guided me to everything I have achieved. May Allah have mercy on you, my dear departed mother.

To my dear siblings: Abdelkader, Mohamed Toufik, Imad, and Asma,

thank you for your love and support.

To my partner Israa: thank you for your cooperation and commitment. You were truly the best companion throughout this journey. Thank you.

2026

Kheira

Dedication

وَكَانَ فَضْلُ اللَّهِ عَلَيْكَ عَظِيمًا

To my selff,

for never giving up and believing in my dreams despite every challenge and moment of doubt. I dedicate this success to my determination and perseverance.

To my dear father Ben Aissa,

my hero, thank you for your endless love, support, and sacrifices. Everything I am today is thanks to you.

To my beloved mother,

my heart and soul, thank you for your unconditional love and prayers that guided me throughout this journey. Words cannot express how much I love and appreciate you.

2026

To my sisters Hadjer and Marwa,

thank you for your kindness, encouragement, and constant presence by my side.

To my graduation companion Hocini Kheira,

thank you for sharing this academic journey with me, for your patience, cooperation, and mutual support during every difficult moment.

To my friends, especially Oualaa and Meriem,

thank you for your support, advice, and for standing by me during moments of doubt. Your friendship is one of the greatest gifts in my life.

Abstract

This dissertation aims to study and evaluate the acoustic localization of sound sources from unmanned aerial vehicles (UAVs), with a focus on methods based on Near-Field Microphone Arrays. The main objective is to accurately determine the three-dimensional position of the UAV by estimating the azimuth, elevation, and distance, while maintaining reliable performance in noisy environments.

Several localization algorithms were investigated in this work, including DAS, TDOA, and MUSIC. Simulations were carried out using MATLAB to assess and compare these methods in terms of localization accuracy, robustness to noise, and computational complexity.

The obtained results indicated that the DAS algorithm, although simple and easy to implement, provided the lowest localization accuracy. The TDOA method achieved better performance and offered a good compromise between estimation accuracy and execution speed. Finally, the MUSIC algorithm demonstrated the highest performance, providing superior accuracy and stronger resistance to noise compared with the other studied methods.

Keywords: Unmanned Aerial Vehicles (UAVs), Acoustic Localization, Microphone Array, DAS, MUSIC, TDOA

Résumé

Ce mémoire vise à étudier et à évaluer la localisation acoustique des sources sonores des drones, en mettant l'accent sur les méthodes basées sur les réseaux de microphones. L'objectif principal est de localiser avec précision un drone en trois dimensions en estimant l'azimut, l'élévation et la distance, tout en garantissant de bonnes performances dans des environnements bruités.

Plusieurs algorithmes de localisation ont été étudiés dans ce travail, à savoir DAS, TDOA et MUSIC. Des simulations ont été réalisées sous MATLAB afin d'évaluer et de comparer ces méthodes en termes de précision de localisation, de robustesse au bruit et de complexité de calcul.

Les résultats obtenus ont montré que l'algorithme DAS, bien que simple et facile à mettre en œuvre, présente la plus faible précision de localisation. La méthode TDOA offre de meilleures performances et constitue un bon compromis entre précision d'estimation et rapidité d'exécution. Enfin, l'algorithme MUSIC a démontré les meilleures performances, avec une précision supérieure et une meilleure résistance au bruit par rapport aux autres méthodes étudiées.

Mots-clés : drones, localisation acoustique, réseau de microphones, DAS, MUSIC, TDOA

ملخص :

يهدف هذا البحث إلى دراسة وتقييم الموقع الصوتي للمصادر الصوتية الخاصة بالطائرات بدون طيار مع التركيز على الطرق المعتمدة على مصفوفة ميكروفونات في المجال القريب (Near Field) ويتمثل الهدف الرئيسي في تحديد موقع الطائرة بدون طيار ثلاثي الأبعاد بدقة من خلال تقدير زاوية السمت والارتفاع والمسافة مع ضمان أداء جيد في البيئات المشوشة. تم دراسة عدة خوارزميات في هذا العمل من بينها DAS, TDOA ,MUSIC, كما تم تنفيذ محاكاة باستخدام MATLAB. بهدف تقييم أداء هذه الطرق ومقارنتها من حيث الدقة ومقاومة الضوضاء، والتعقيد الحسابي. أظهرت النتائج أن خوارزمية MUSIC حققت أفضل أداء من حيث الدقة ومقاومة الضوضاء بينما قدمت TDOA توازن جيد بين الأداء وسرعة التنفيذ في حين تميزت DAS بالبساطة وسهولة تطبيق رغم محدودية دقتها.

الكلمات الدالة: الطائرات بدون طيار، الموقع الصوتي، مصفوفة ميكروفونات DAS, MUSIC, TDOA

Table des matières

List of Abbreviations

AI	Artificial Intelligence
AOA	Angle of Arrival
AWGN	Additive White Gaussian Noise
CNNs	Convolutional Neural Networks
DAS	Delay And Sum
DOA	Direction of Arrival
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Techniques
FDOA	Frequency Difference of Arrival
GCC	Generalized Cross Correlation
GCC-PHAT	Generalized Cross-Correlation with Phase Transform
MSE	Mean Squared Error
MUSIC	MUltiple SIgnal Classification
MVDR	Minimum Variance Distortionless Response
RSS	Received Signal Strength
SED	Sound Event Detection
SNR	Signal-to-Noise Ratio
SSL	Sound Source Localization
TDOA	Time Difference of Arrival
TOA	Time of Arrival
UAVs	Unmanned Aerial Vehicles
UCA	Uniform Circular Array

List of Figures

List of Tables

General Introduction

General Introduction

Unmanned aerial vehicles (UAVs) have seen significant development in recent years thanks to advances in embedded computing, lightweight materials, and communication systems, which have expanded their range of applications in both civilian and military sectors. The effectiveness of these drones depends on the sensors they carry; Determining their location is no longer limited to cameras, which are affected by environmental conditions, but the use of acoustic sensors has become an effective option. Sound source localization is a fundamental technique that enables the drone to analyze acoustic signals to accurately determine the direction and location of the source using advanced algorithms. In this context, a system based on a regular circular array of microphones is used to estimate the three-dimensional position of the drone (azimuth, elevation, and distance) by comparing several algorithms such as sum-of-delays, the multi-signal classification algorithm, and the time-of-arrival difference method, all of which rely on the difference in the time of arrival of the sound wave at the microphones.

Unlike far-field models, where the source is assumed to be sufficiently distant and only the direction of arrival is estimated, the adopted near-field model also accounts for wavefront curvature, making accurate distance estimation possible in addition to azimuth and elevation estimation.

The audio signals are processed by calculating the time delays between microphones and adding noise to simulate a real-world environment, and then using the direction vector that represents the array's response to a specific sound source. Different methods exhibit varying characteristics: the delayed sum method is simple, the MUSIC algorithm offers high accuracy, and the TDOA method strikes a balance between accuracy and complexity. Despite the efficiency of these techniques, they face challenges such as noise from propellers and wind, reflection effects, and limited computational resources, requiring advanced filtering and performance optimization techniques. Thanks to this approach, an effective and low-cost system for locating drones based on their acoustic signals can be achieved, with the potential to improve accuracy in the future through the development of array designs and the integration of more advanced algorithms.

Chapter I

Overview of Sound Source Localization of Unmanned Aerial Vehicles (UAVs)

I.1 Introduction

Unmanned Aerial Vehicles (UAVs), commonly known as drones, have experienced rapid development over the last decade due to advances in embedded processors, lightweight materials, wireless communication systems, and intelligent control algorithms [1]. These technological improvements have enabled UAVs to perform a wide range of civilian and military applications, including environmental monitoring, infrastructure inspection, border surveillance, disaster management, and search-and-rescue missions [2].

The effectiveness of these applications depends greatly on onboard sensing and navigation systems. Conventional UAV platforms generally rely on cameras, Global Positioning System (GPS) receivers, and Inertial Measurement Units (IMUs) for navigation and target detection [3]. However, vision-based systems are highly affected by environmental conditions such as darkness, fog, smoke, and physical obstacles that may obstruct the line of sight. Consequently, acoustic sensing has emerged as a reliable complementary solution because it can operate independently of lighting conditions [4].

In the field of Sound Source Localization (SSL), acoustic sensors and microphone arrays are used to detect and estimate the position of a sound-emitting target. In this work, the objective is to localize an unmanned aerial vehicle using acoustic sensors by analyzing the sound generated by the drone’s propellers and motors. By processing the captured acoustic signals, the system can estimate important spatial parameters such as the direction of arrival and the relative position of the UAV [5]. This capability is particularly valuable in surveillance, security, and monitoring applications where rapid and accurate drone detection is required.

I.2. Sound Source Localization (SSL)

Sound Source Localization (SSL) refers to the process of determining the spatial position of a sound-emitting source using signals captured by one or more microphones [5]. In UAV acoustic applications, SSL techniques are employed to detect and localize drones by analyzing the acoustic signatures generated by their propellers and motors. Microphone arrays receive delayed and noisy versions of the emitted sound signal, where the propagation delay depends on the distance between the UAV and each microphone, as well as the speed of sound in air [6][7].

The main objective of SSL is to estimate the unknown position of the UAV from the recorded acoustic signals. This process is particularly important in surveillance, monitoring, and security applications where visual sensing systems may become unreliable due to poor lighting conditions, fog, or physical obstacles.

In mathematical terms, if a sound source is located at position:

$$\mathrm{s} = [x_s, y_s, z_s]^T \tag{I.1}$$

s: is the three-dimensional position vector of the sound source.

and the i-th microphone is positioned at:

$$\mathrm{m}_i = [x_i, y_i, z_i]^T \tag{I.2}$$

m_i: is the three-dimensional position vector of the i-th microphone.

then the received signal at microphone i can be modeled as:

$$x_i(t) = \alpha_i u(t - \tau_i) + n_i(t) \ \ ; \ \ \tau_i = \frac{\|\mathbf{s}-\mathbf{m}_i\|}{c} \tag{I.3}$$

$$\alpha_{\mathrm{i}} = \frac{1}{4\pi \mathrm{d}_{\mathrm{i}}} \ \ , d_i = \| \ s - mi \ \|$$

$$x_i(t) = \frac{1}{4\pi d_i} \ u(t - \frac{d_i}{c}) + n_i(t) \tag{I.4}$$

Where:

$x_i(t)$: Received acoustic signal at microphone $i$as a function of time.

$u(t\ )$: is the emitted acoustic signal.

α_i: Acoustic attenuation factor (or amplitude decay coefficient) for the i^{th}microphone. It represents the reduction of the signal amplitude due to propagation loss.

τ_i: is the propagation delay.

c: is the speed of sound.

d_i: Distance between the sound source and the i^{th}microphone.

$n_i(t)$: represents additive noise.

The main challenge is to recover the source position **s** from the observed signals $x_i(t)$.

In practical UAV acoustic localization systems, several factors can degrade localization accuracy. Since the UAV may be moving during operation, the received acoustic signals are often affected by Doppler effects, wind turbulence, environmental noise, and strong rotor-generated noise [8]. These disturbances can significantly reduce the accuracy of time-delay estimation between microphones. Therefore, robust signal processing techniques are required to improve localization performance under noisy and reverberant conditions.

Among the most widely used techniques is the Generalized Cross-Correlation with Phase Transform (GCC-PHAT) method [6], originally introduced by Charles Knapp and Gary Carter. GCC-PHAT enhances time-delay estimation by emphasizing phase information while reducing the influence of signal amplitude variations and noise, making it particularly suitable for UAV acoustic localization applications.

I .3. Detection and Localization

I.3.1 Sound Detection

In UAV acoustic localization systems, sound detection represents the first processing stage before localization can be performed.

Before performing localization, it is necessary to determine whether a significant acoustic event is present. Sound detection is typically formulated as a binary hypothesis test distinguishing between noise-only signals and signals containing both sound and [7]. Mathematically, this can be expressed as:

$$\begin{cases} H_0: x(t) = n(t) & \text{(noise only)} \\ H_1: x(t) = s(t) + n(t) & \text{(signal present)} \end{cases} \qquad \text{I.5}$$

Where:

$\mathcal{H}_0$: denotes the noise-only hypothesis.

$\mathcal{H}_1$: denotes the signal-plus-noise hypothesis.

$x(t)$: observed signal

t: time variable

$s(t)$: is the emitted acoustic signal.

$n(t)$: represents background noise.

A widely used approach for detection is the energy-based detector, where the energy of the received signal over an observation window of length N is computed as:

$$E = \sum_{k=1}^{N} | x(k) |^2 \qquad \text{I.6}$$

where:

E: measured signal energy.

k: sample index.

N: number of samples.

$x(k)$: sampled signal amplitude.

If the measured energy exceeds a predefined threshold γ, the presence of a sound source is declared: $E > \gamma \Rightarrow$ Sound detected.

Where γ: detection threshold.

Otherwise, the signal is considered to contain only noise. In UAV applications, this task becomes more difficult because the background noise is highly non-stationary due to propeller rotation, motor vibrations, and airflow turbulence [8]. Therefore, adaptive thresholding methods are often employed, where the threshold is continuously updated according to the current noise level in order to preserve reliable detection performance.

I.3.2 Sound Localization

Once a sound event has been detected, the next step is to estimate the direction or spatial position of the source. This stage is referred to as sound localization, and it aims to determine parameters such as azimuth angle, elevation angle, or source coordinates. Several techniques have been developed for this purpose; The most widely used localization techniques include:

Time Difference of Arrival (TDOA) [6][10]

Direction of Arrival (DOA) estimation [9]

Beamforming [9][11]

High-resolution subspace methods such as MUSIC [9][12]

These array-processing techniques have been widely studied in acoustic signal processing and are commonly used in UAV localization systems because of their robustness and localization accuracy [9].

I.4. Localization According to Spatial Dimensions

Spatial sound localization can be analyzed according to three principal dimensions: azimuth, elevation, and distance [9]. These parameters describe the direction and range of the acoustic source relative to the UAV or microphone array. In practical systems, accurate localization often requires estimating these three quantities simultaneously in order to obtain the complete three-dimensional position of the source.

I.4.1 Horizontal Localization (Azimuth)

In UAV acoustic localization systems, the sound generated by the drone propagates toward the microphone array and reaches each microphone at slightly different times depending on the relative position of the UAV. Azimuth estimation determines the horizontal direction of the sound source relative to the microphone array. This parameter indicates whether the UAV is located to the left or right of the reference axis [6].

$$\theta = \arctan\left(\frac{y}{x}\right) \qquad \text{I.7}$$

Alternatively:

$$\theta = \mathrm{atan2}(y, x) \qquad \text{I.8}$$

In practical implementations, the function atan2(y, x) is preferred because it correctly determines the angle quadrant.

where:

θ: azimuth angle

x: position along the x-axis

y: position along the y-axis

I.4.2 Vertical Localization (Elevation)

Elevation estimation determines whether the UAV sound source is located above or below the acoustic sensor array. Unlike azimuth estimation, this task requires microphones arranged in a three-dimensional geometry so that vertical spatial information can be captured [9]. Elevation cues are commonly extracted from phase differences, time delays, and amplitude variations between microphones positioned at different heights.

$$\phi = arctan\left(\frac{z}{\sqrt{x^2+y^2}}\right) \quad \text{I.9}$$

Where:

ϕ: elevation angle

z: height of the source

$\sqrt{x^2+y^2}$: horizontal distance from the origin

Accurate three-dimensional localization generally requires at least four spatially separated microphones so that the source position can be uniquely estimated [9]. In UAV applications, elevation estimation is particularly useful for detecting aerial objects flying above or below the platform.

I.4.3 Distance Estimation

Distance estimation determines the range between the UAV acoustic source and the acoustic sensor array. One classical approach relies on the attenuation of acoustic intensity with distance. Under ideal free-field propagation conditions, sound intensity decreases proportionally to the inverse square of distance [7], which can be expressed as:

$$I \propto \frac{1}{r^2} \quad \text{I.10}$$

Where:

I: acoustic intensity.

r: distance from source to receiver.

$\propto$: proportional to.

This means that as the source moves farther away, the received acoustic energy decreases significantly.

However, in UAV environments, reliable distance estimation is more challenging because reflections from surrounding surfaces, wind turbulence, and multipath propagation can distort the received signal amplitude [8]. For this reason, intensity-based methods are often combined with geometric triangulation or near-field localization techniques. When the source coordinates (x_s, y_s, z_s) and sensor position (x, y, z) are known, the distance can be expressed as:

$$\mathrm{r} = \sqrt{(x_s - x)^2 + (y_s - y)^2 + (z_s - z)^2} \qquad \text{I.11}$$

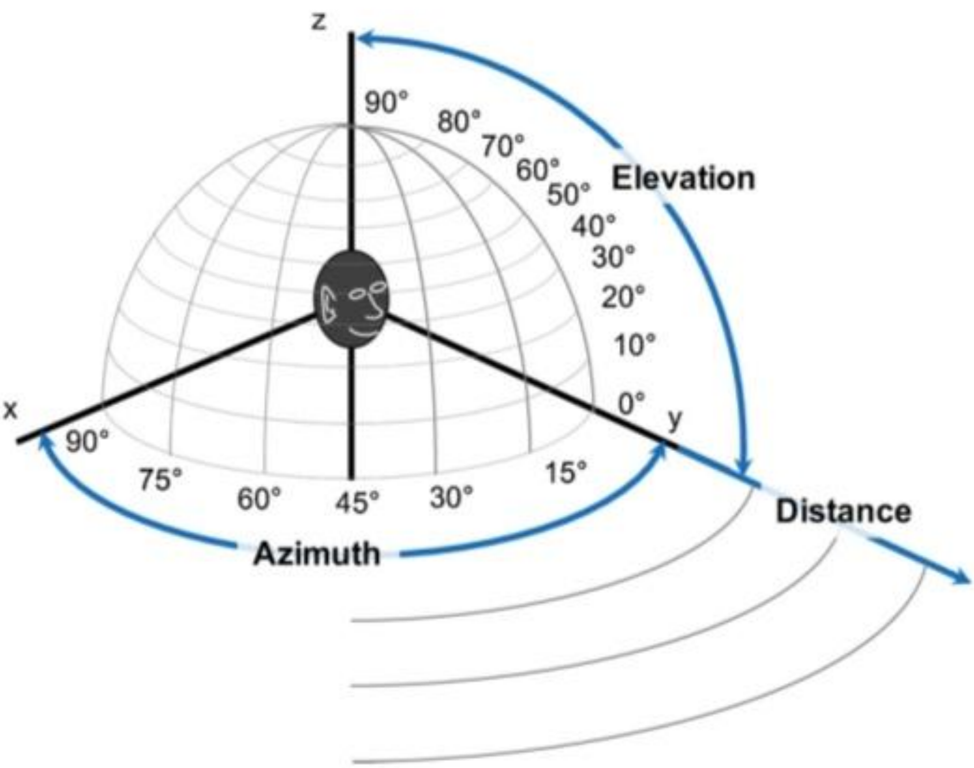
where:

r: distance between source and sensor.

x_s, y_s, z_s: source coordinates.

$\mathrm{x}, \mathrm{y}, \mathrm{z}$: sensor coordinates.

Such hybrid approaches generally provide more robust range estimation in realistic operating conditions.



FigureI.1: Polar coordinates System for Sound Source Localization [13].

I.5. Localization According to Number of Microphones

The performance and capability of sound source localization strongly depend on the number of microphones used in the system. In general, increasing the number of sensors improves spatial resolution and robustness, but it also increases system complexity and computational requirements. Therefore, localization methods can be classified according to whether they use a single microphone, two microphones, or a microphone array.

I.5.1 Single Microphone

When only a single microphone is available, true spatial localization is generally not possible because there is no spatial diversity to infer direction or position. In this case, localization cannot rely on time or phase differences between sensors. Instead, only limited information can be extracted from the signal, such as amplitude variations or spectral characteristics. As a result, only rough estimation of distance or event detection may be attempted, often with low reliability and strong sensitivity to noise and environmental changes [5].

I.5.2 Two Microphones

With two microphones, spatial information becomes available through the Time Difference of Arrival (TDOA) between the two received signals. This allows estimation of the azimuth angle of the sound source. The delay between the two microphones is defined as:

$$\Delta\tau = \frac{d\sin(\theta)}{c} \qquad \text{I.12}$$

Where:

d : is the distance between the microphones.

θ: is the azimuth angle.

c: is the speed of sound.

This configuration is computationally efficient and therefore well suited for lightweight UAV platforms [8]. However, it remains fundamentally limited to two-dimensional localization and cannot directly estimate elevation or full 3D position without additional assumptions or motion-based techniques.

I.5.3 Microphone Arrays

Microphone arrays consist of multiple spatially distributed sensors arranged in a specific geometry. These arrays enable more advanced spatial processing techniques such as beamforming and subspace-based methods, which significantly improve localization performance [9]. The main advantage of such systems is their ability to exploit spatial diversity to enhance signal quality and suppress interference from unwanted directions.

The beamforming output of an array with M microphones can be expressed as:

$$y(t) = \sum_{i=1}^{M} w_i \, x_i(t) \tag{I.13}$$

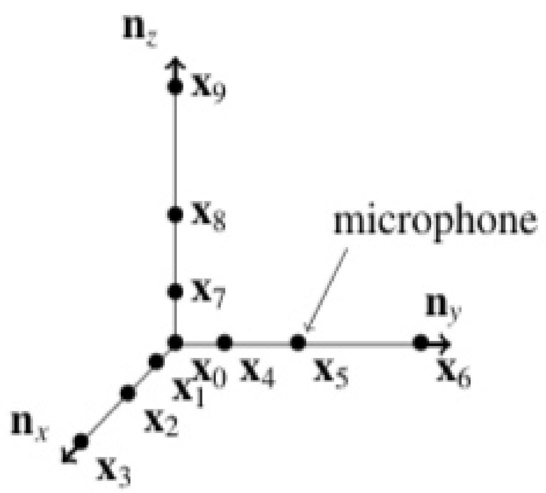
Where:

$y(t)$: beamformer output signal.

M: number of microphones.

w_i: weight coefficient of microphone i.

$x_i(t)$: signal received at microphone i.



FigureI.2: Geometry of a three-dimensional microphone array [14].

The sensors are arranged along three orthogonal axes n_x, n_y, and n_z. Positions X_0 to X_9 denote the spatial locations of the microphones, with a central reference point and additional sensors distributed along each direction to enable spatial sound source localization.

I.6. Comparison of Configurations

configuration	Spatial Capability	Accuracy	Computational Load	UAV Suitability
Single Microphone	\	Very Low	Very Low	Limited
Two Microphones	2D	Moderate	Low	Suitable for small UAVs
Microphone Array	2D/3D	High	High	Suitable for advanced UAVs

Table I.1:Performance Comparison of Microphone Configurations in Acoustic Localization.

While microphone arrays provide superior localization performance, simpler configurations may be preferred in energy-constrained UAV platforms due to lower computational and hardware requirements.

I.7. Applications of SSL in UAV Systems

Airport Protection: Microphones detect and locate unauthorized drones near airports to avoid collisions with aircraft.

Border and Urban Surveillance: Acoustic systems can monitor illegal drone activity in cities or border regions, especially when cameras fail because of fog, darkness, or obstacles.

Wildlife and Environmental Monitoring: Researchers can detect and track drones used in forest inspection or environmental missions without relying only on radar.

Early Warning Systems: SSL provides passive drone detection before the UAV becomes visible or detectable by other sensors.

Military Base Protection: Microphone arrays continuously monitor restricted airspace around military facilities.

Drone Classification and Identification: Different UAVs produce unique acoustic signatures that can be analyzed to identify drone types.

I.8. Challenges of SSL in UAV Platforms

Despite its advantages, implementing SSL on UAVs presents several challenges:

Propeller-induced self-noise

Wind turbulence

Limited payload capacity

Real-time processing constraints

Multipath propagation and reverberation

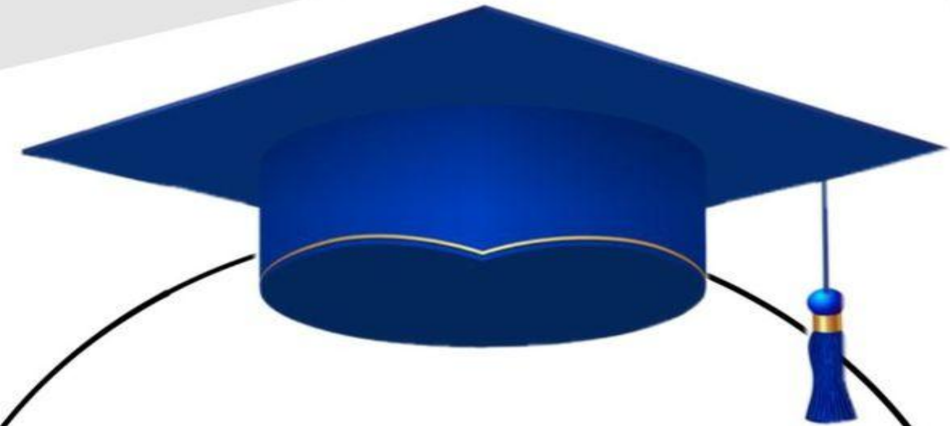
Rotor noise reduces the Signal-to-Noise Ratio (SNR), which directly affects localization accuracy [15]. To mitigate these effects, adaptive filtering, spatial filtering, and noise cancellation techniques are required.

Designing SSL systems for UAVs therefore requires balancing localization accuracy, computational complexity, and energy efficiency [8].

I.9.Conclusion

This chapter established the theoretical foundations of sound source localization in UAV systems. Acoustic propagation principles, detection strategies, and localization approaches were presented within a structured analytical framework.

Different spatial configurations and microphone setups were compared to identify the most suitable approach for aerial platforms. Furthermore, UAV-specific constraints such as self-noise, weight limitations, and processing capacity were examined. These considerations form the foundation for the algorithmic development and system implementation discussed in the subsequent chapters.



Chapter II

Survey of Sound Source Localization Methods

II.1 Introduction

The ability to identify and locate sound is a fundamental task in acoustic signal processing, known as Sound Event Detection (SED) and Sound Source Localization (SSL). This field has evolved through two main stages: Classical Methods, which rely on mathematical models and physics such as the Time Difference of Arrival (TDOA) and geometric triangulation and Modern Methods, which utilize advanced Deep Learning and Artificial Intelligence to process sound in complex, noisy environments.

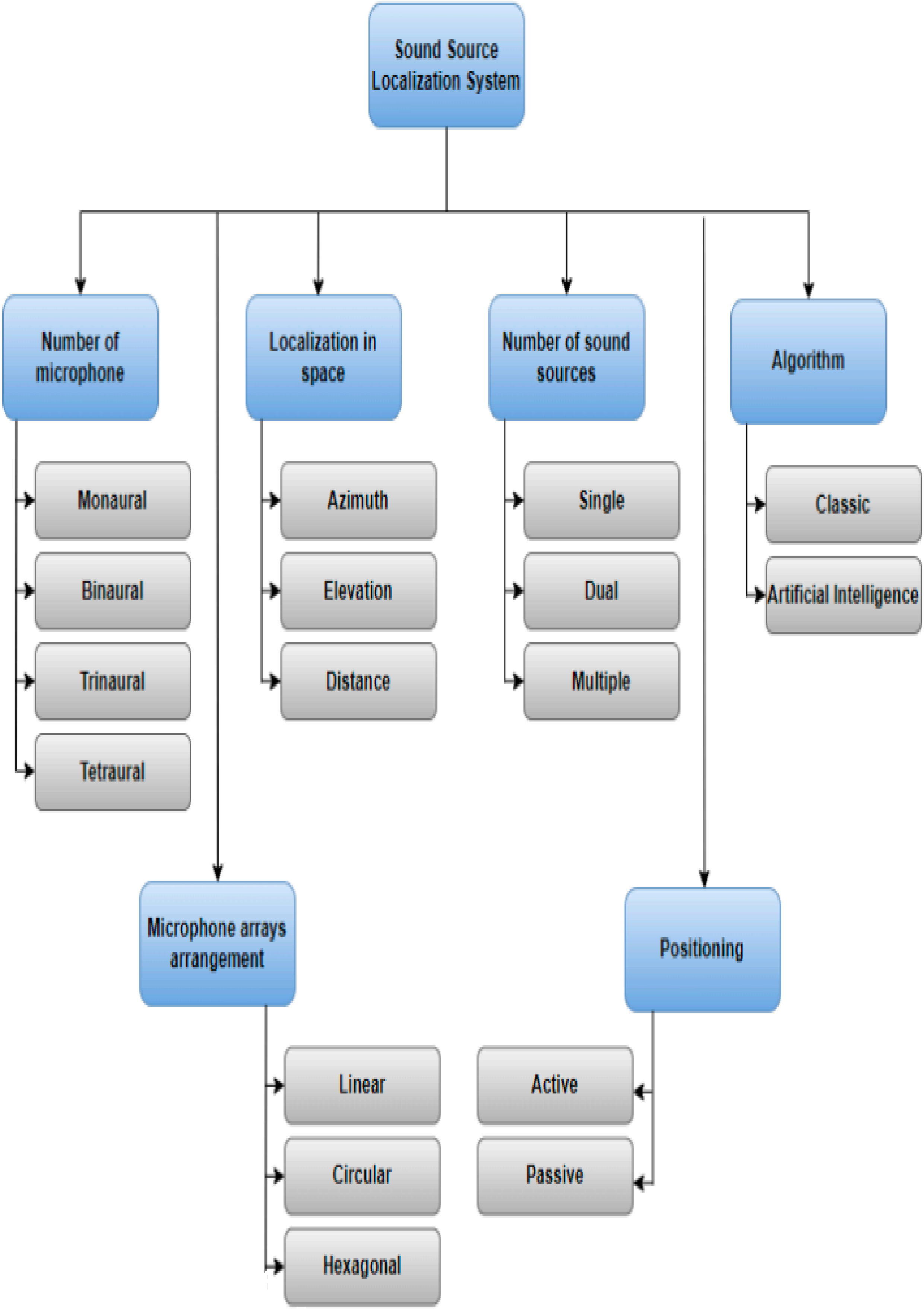
This chapter provides a comprehensive overview of how these evolving techniques are transforming our ability to sense and analyze the auditory world.

II.2 Classification Methods

Over the past decades, a wide range of acoustic detection and localization techniques has been developed, all of which share a fundamental requirement:

The use of microphones to capture sound signals. Beyond their primary role of converting acoustic waves into electrical signals, microphones may also perform additional functions such as attenuating ambient noise. In the scientific literature, several classification schemes have been proposed to organize sound source localization systems according to different criteria [19,20].

These schemes are synthesized and illustrated in Figure (II.1) [18].



FigureII.1: Sound source localization systems (SSL) classification [18].

II.3 Acoustic Source Localization Methods

Acoustic drone localization methods are generally divided into classical approaches, which are effective and diverse. In contrast, modern methods based on artificial intelligence leverage machine learning to improve robustness and accuracy in complex scenarios. Thus, all available approaches can be classified into these two categories:

II.3.1 Classical Methods

II.3.1.1 Measurement Techniques

A) Time of Arrival (TOA)

Time of Arrival (TOA) is a technique used to determine the precise location of an object by measuring the time it takes for a signal to travel between a transmitter and a receiver. The distance between the source and a microphone is calculated using the simple relationship between speed and time [19]. Since we know the speed of sound (v) in a given medium, we can find the distance (d) if we know the exact travel time (t) between the two points.

$$T_{TOA} = t_{received} - t_{transmitted} \tag{II.1}$$

Where:

$t_{transmitted}$: time when the signal is sent.

$t_{received}$: time when the signal is received.

Distance (Range) Equation

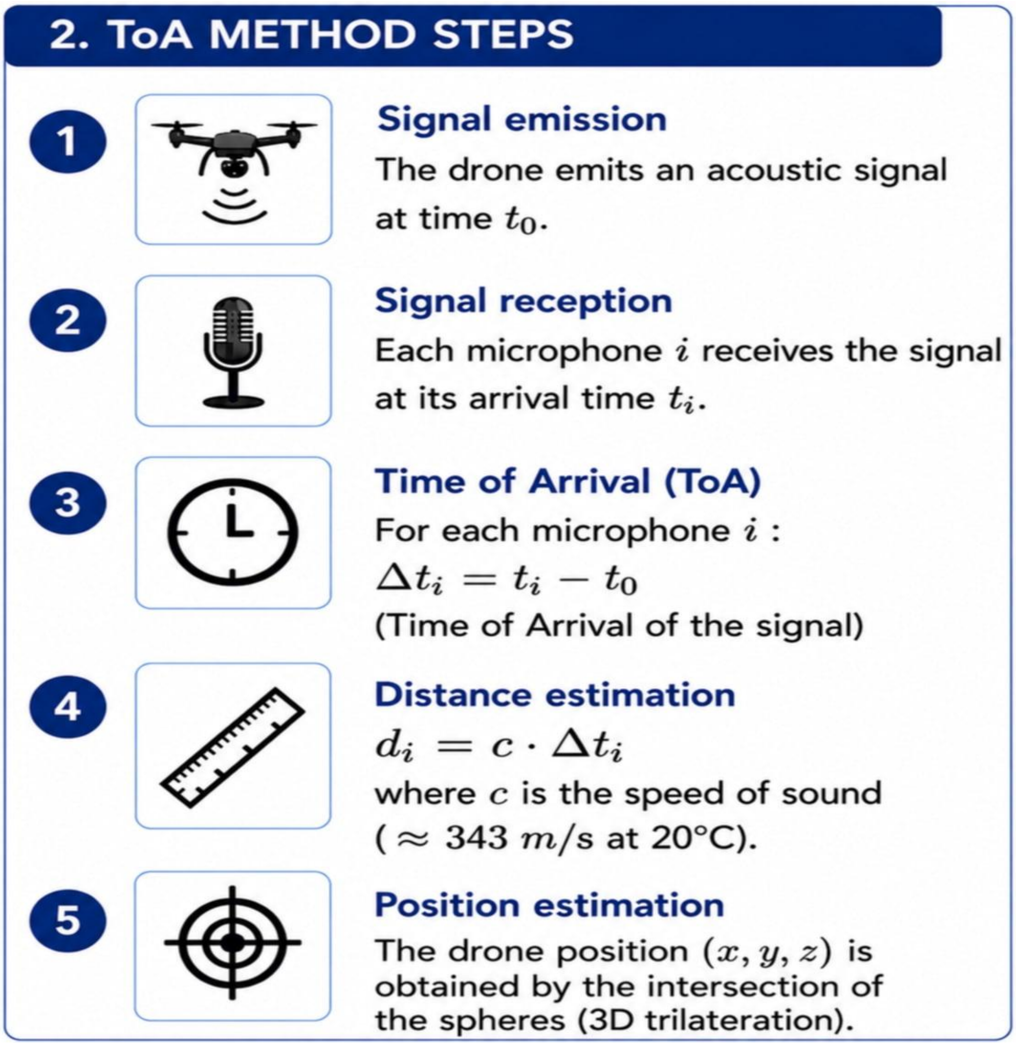
The distance is calculated using the propagation speed of the signal:

$$d = v \times T_{TOA} \tag{II.2}$$

Where:

d: distance between the drone and the receiver.

v: signal propagation speed (exp: Sound:343 m/s in air).



FigureII.2: Time of Arrival (TOA) Localization Principle.

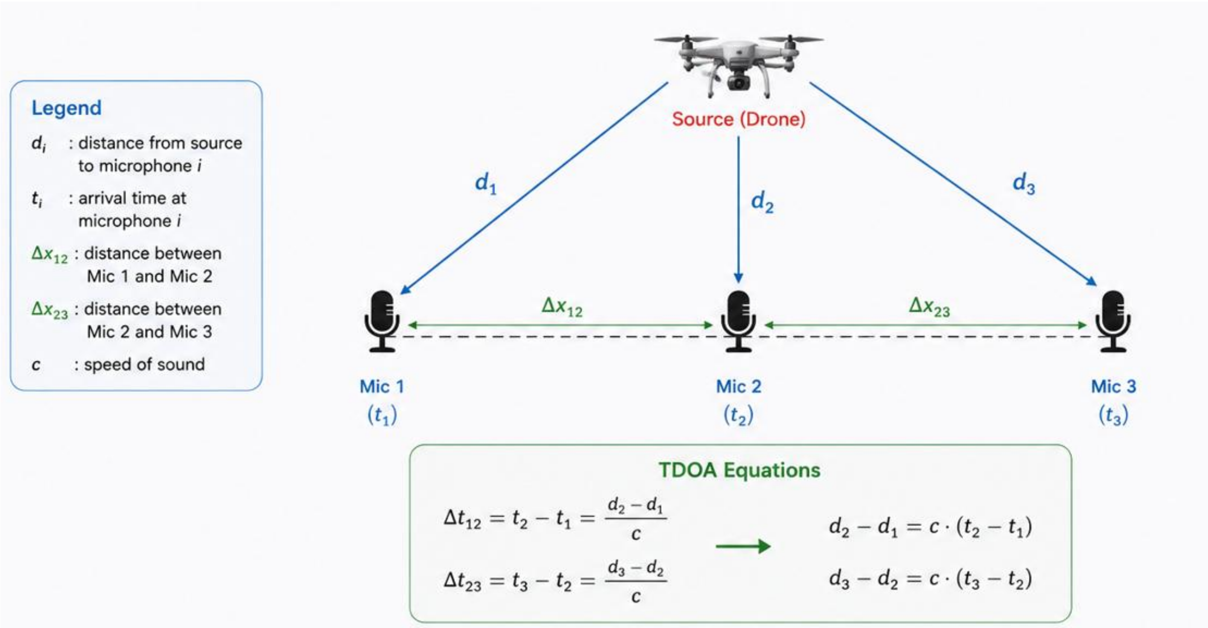
B) Time Difference of Arrival (TDOA)

Time Difference of Arrival (TDOA) technology is used to locate a signal source (such as a drone) by measuring the time differences in the signal's arrival at multiple receivers distributed at known locations. Since the signal travels at a constant speed the speed of sound it reaches each receiver at a different time depending on the distance. Instead of requiring knowledge of the transmission time, the method relies solely on calculating the differences between arrival times, making it suitable for asynchronous systems. These time differences are converted into distance differences; when we calculate the time difference between two receivers, we do not obtain a single point directly, but rather a set of possible points, all of which lie on this curve. When using three or more receivers, we obtain several curves (each pair of receivers yields a different hyperbolic section), and the true source location is the point of intersection of these curves. In two-dimensional space, three receivers are sufficient to obtain an intersection that determines the location, while in three-dimensional space, we typically need four receivers. It should be noted that the accuracy of the result depends heavily on the precision of the time measurement, the quality of the receiver distribution, and the effects of noise or signal reflection from the surrounding environment [20].

Calculating the time difference (TDOA):

$$\Delta t_{12} = t_2 - t_1 \quad \text{(II.3)}$$

$$\Delta t_{13} = t_3 - t_1 \quad \text{(II.4)}$$



FigureII.3: Principle of Time Difference of Arrival (TDOA) Localization.

C) Frequency Difference of Arrival (FDOA):

The FDOA localization process begins by measuring the frequency differences between the signals received at multiple sensors. These frequency shifts are then used to estimate the radial velocities between the source and the sensors based on the Doppler effect. Using the obtained FDOA measurements, a set of possible source locations is generated, and the source position is determined by identifying the intersection of the resulting curves or surfaces. Once the source coordinates have been estimated, the distance between the source and each sensor can be calculated using geometric relationships [21].

The received frequency at sensor i can be modeled as:

$$f_i = f_0\left(1 + \frac{\mathbf{v}_i \cdot (\mathbf{x} - \mathbf{s}_i)}{c \parallel \mathbf{x} - \mathbf{s}_i \parallel}\right) \tag{II.5}$$

Where:

f_i: The received frequency at sensor i.

f_0: is the transmitted frequency.

$\mathbf{x}$: is the source position.

$\mathbf{s}_i$ and $\mathbf{v}_i$: are the position and velocity of sensor i.

c: is the speed of sound.

By forming frequency differences between two sensors, the unknown transmitted frequency is eliminated:

$$\Delta f = f_i - f_j = \frac{f_0}{c}\left(\frac{\mathbf{v}_i \cdot (\mathbf{x} - \mathbf{s}_i)}{\parallel \mathbf{x} - \mathbf{s}_i \parallel} - \frac{\mathbf{v}_j \cdot (\mathbf{x} - \mathbf{s}_j)}{\parallel \mathbf{x} - \mathbf{s}_j \parallel}\right) \tag{II.6}$$

D) Received Signal Strength (RSS):

In localization, the Received Signal Strength (RSS) is used to estimate the location of a signal source, such as a drone or transmitter, by measuring the signal strength at multiple receivers with known locations. The basic idea is that signal strength decreases as the distance between the source and the receiver increases; therefore, the RSS value can be converted into a distance estimate using propagation models such as the log-distance path loss model. After obtaining approximate distances from multiple sensors, techniques such as multiliterate are used to determine the likely location of the source by finding the intersection point of the circles or spheres generated by these distances. However, positioning using RSS remains only an estimate, because the signal is affected by noise, reflections, and environmental obstacles, leading to errors in distance calculation and, consequently, in positioning [22][23].

$$\mathrm{P} = \frac{1}{N}\sum_{n=0}^{N-1} X[n]^2 \tag{II.7}$$

P: Mean Signal Power.

N: Number of samples.

n: Sample index.

x[n]: Signal amplitude at sample n.

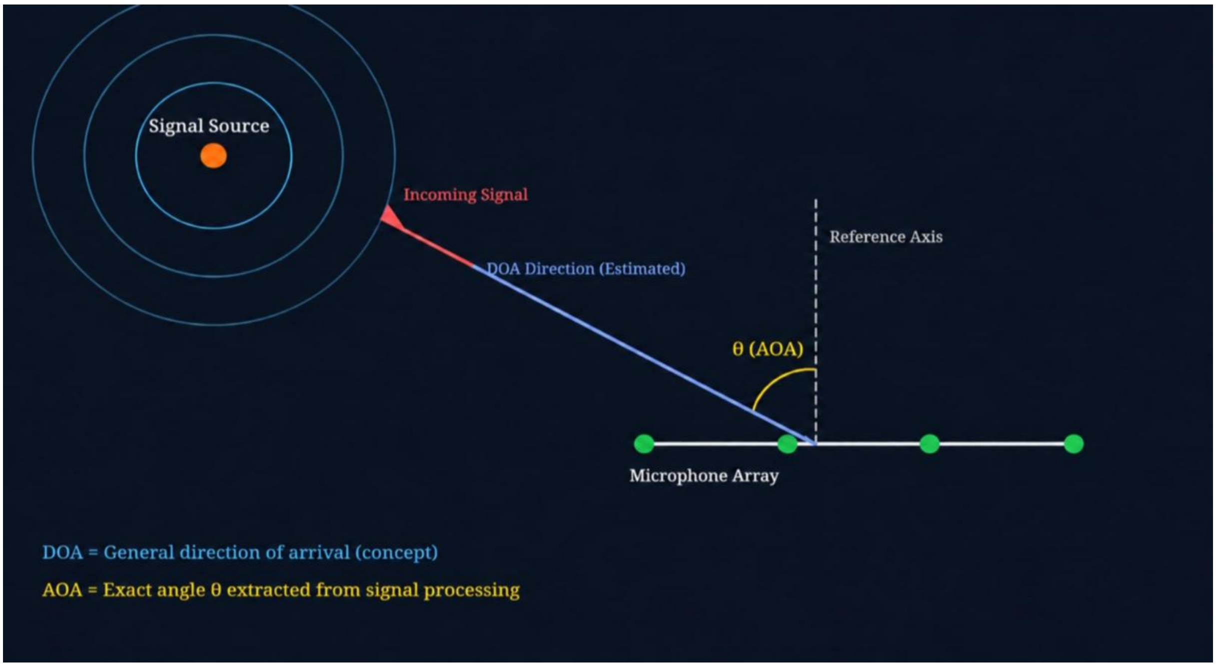
Log-distance path loss model is formally expressed as:

$$L = L_{\mathrm{Tx}} - L_{\mathrm{Rx}} = L_0 + 10\gamma \log_{10}\left(\frac{d}{d_0}\right) + X_{\mathrm{g}} \quad \text{(II.8)}$$

- L:is the total path_loss in decibels (dB).
- $L_{\mathrm{Tx}} = 10\log_{10}\frac{P_{\mathrm{Tx}}}{1\ \mathrm{mW}}$ dBm is the transmitted power level, and P_{Tx}: is the transmitted power.
- $L_{\mathrm{Rx}} = 10\log_{10}\frac{P_{\mathrm{Rx}}}{1\ \mathrm{mW}}$ dBm is the received power level where P_{Rx} is the received power.
- L_0:is the path loss in decibels (dB) at the reference distance d_0. This is based on either close-in measurements or calculated based on a free space assumption with the Friis free-space path loss model.
- $d$is the length of the path.
- d_0: is the reference distance, usually 1 km (or 1 mile) for a large cell and 1 m to 10 m for a microcell.
- γ: is the path loss exponent.
- X_g: is a normal (Gaussian) random variable with zero mean, reflecting the attenuation (in decibels) caused by flat fading. In the case of no fading, this variable is 0. In the case of only shadow fading or slow fading, this random variable may have Gaussian distribution with σstandard deviation in decibels.

E) Angle / Direction of Arrival (AOA / DOA)

The basic idea is that a sound source emits a wave that propagates in all directions; when it reaches a set of microphones at different times, the differences in arrival times are converted into path differences based on the speed of sound in air (approximately 343 m/s). Using this data, the AOA (Angle of Arrival) is calculated, representing the angle at which the wave reaches the microphone array. This angle is used directly to determine the DOA (Direction of Arrival), i.e., the true direction of the sound source relative to the drone. Thus, the DOA is determined by estimating the AOA, which ultimately allows for the creation of a line of bearing that precisely locates the sound source relative to the drone [24][18].



FigureII.4: Angle of Arrival (AOA) Estimation Geometry.

Converting to path difference:

$$d_i = v_{sound} \times \Delta t_i \tag{II.9}$$

Δt_i: time difference of Arrival $(\Delta t_i = t_i - t_j)$

Calculating the angle:

$$\theta_i = arcsin\left(\frac{d_i}{\Delta d}\right) \tag{II.10}$$

Δd: The distance between the microphones.

II.3.1.2 Feature Extraction Methods

A) Generalized Cross Correlation (GCC)

The Generalized Cross-Correlation (GCC) method is widely used to estimate the Time Difference of Arrival (TDOA) between pairs of microphones by performing cross-correlation in the frequency domain. For a known microphone array geometry, the TDOA values associated with each microphone pair are precomputed and mapped to their corresponding Direction of

Arrival (DOA). These pre-established lookup tables facilitate the estimation of the sound source direction.

First, the cross-correlation of the received signals is calculated in the frequency domain for every microphone pair. The obtained spectral data are then normalized, while frequency components below the microphone noise threshold are attenuated to reduce the influence of background noise. Afterward, zero-padding is applied to the frequency-domain signals before performing the Inverse Fast Fourier Transform (IFFT), which improves the resolution of the resulting time-domain correlation function. The generated cross-correlation signals exhibit peaks at positions corresponding to the TDOA of each microphone pair. Finally, the TDOA information from all microphone pairs is combined to determine the most accurate estimate of the DOA of the acoustic source [25].

A.1 The stages of GCC

1) **Signal Acquisition**

A microphone array composed of M microphones capture the acoustic signals:

$$x_i(t), i = 1,2, \dots, M$$

Where:

$x_i(t)$: signal received by microphone i.

2) **Pairwise Processing**

Microphone signals are grouped into pairs (i, j) for correlation analysis.

3) **Frequency Domain Transformation**

Each microphone signal is transformed into the frequency domain using the Fourier Transform:

$$X_i(f) = F\{x_i(t)\}$$
$$X_j(f) = F\{x_j(t)\}$$

where:

$X_i(f)$: Fourier transform of microphone i,

$X_j(f)$: Fourier transform of microphone j.

A.2 GCC-PHAT Formulation

1) Cross-Power Spectrum

For two microphone signals i and j, the cross-power spectrum is defined as:

$$G_{ij}(f) = X_i(f)X_j^*(f) \tag{II.11}$$

where:

$X_i(f)$: spectrum of microphone i,

$X_j^*(f)$: complex conjugate of microphone j.

2) PHAT Weighting

In the GCC-PHAT method, the cross-power spectrum is normalized using the PHAT (Phase Transform) weighting function:

$$\Psi(f) = \frac{1}{|\, G_{ij}(f) \,|} \tag{II.12}$$

This normalization suppresses magnitude information while preserving phase information, thereby improving the robustness of TDOA estimation in noisy and reverberant environments.

3) Generalized Cross-Correlation Function

The generalized cross-correlation function is expressed as:

$$R_{ij}(\tau) = \int_{-\infty}^{\infty} \Psi(f)\ G_{ij}(f)\, e^{j2\pi f\tau}\, df \tag{II.13}$$

In practical discrete implementations using FFT/IFFT:

$$R_{ij}(\tau) = IFFT\left(\frac{X_i(f)X_j^*(f)}{|\, X_i(f)X_j^*(f) \,|}\right) \tag{II.14}$$

where:

$R_{ij}(\tau)$: cross-correlation function between microphones i and j,

τ: time delay variable.

4) TDOA Estimation

The Time Difference of Arrival (TDOA) is estimated from the peak value of the cross-correlation function:

$$\hat{\tau}_{ij} = \arg \max_{\tau} R_{ij}(\tau) \tag{II.15}$$

where:

$\hat{\tau}_{ij}$: estimated TDOA between microphones i and j.

5) Relationship Between TDOA and DOA

For two microphones separated by a distance d, the relationship between TDOA and DOA is given by:

$$\tau_{ij} = \frac{d\sin\,\theta}{c} \tag{II.16}$$

where:

d: distance between microphones,

c: speed of sound,

θ: Direction of Arrival (DOA).

Thus, the DOA can be estimated as:

$$\theta = \sin^{-1}\left(\frac{c\tau_{ij}}{d}\right) \tag{II.17}$$

II.3.1.3 Direction Estimation Algorithms:

A) MUSIC (Multiple Signal Classification):

The MUSIC algorithm can be effectively applied to locate a drone using microphones, as it relies in this case on the acoustic signals emitted by the drone's propellers. Several microphones are arranged in a microphone array, and when sound reaches these microphones, a slight difference in phase and arrival time occurs between each microphone due to varying distances. The algorithm analyzes these signals by calculating the correlation matrix and then performing an eigenvalue analysis to separate the signal space from the noise space. Next, the

MUSIC Spectrum is calculated, which shows peaks at the angles from which the sound is coming, allowing the direction of arrival (DOA) of the drone to be determined with high accuracy. By using multiple microphone arrays at different locations, these directions can be intersected to determine the drone's actual location. [26]; More details in next chapter.

There is other direction estimation algorithm such as: ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) is a high-resolution direction-of-arrival (DOA) estimation algorithm based on the rotational invariance property of sensor subarrays. Unlike MUSIC, ESPRIT does not require spectral peak searching, which reduces computational complexity while providing accurate angle estimation [27].

II.3.1.4 Beamforming Techniques

A) Delay and Sum (DAS)

This is the simplest and most common method of beamforming. It works by introducing a slight delay to the signals received at each microphone [13], assuming a possible angle of arrival for the signal, The appropriate time delay for each microphone is calculated based on the array geometry and the speed of sound, so that these time differences are compensated for and the signals are aligned as if they arrived at the same moment. The signals are then multiplied by weighting coefficients to improve signal quality and reduce the effects of noise and side lobes. The resulting signals are then summed according to the delay-and-sum beamforming principle to obtain a single signal representing the estimated direction. The power of the resulting signal is then calculated; high power indicates that the signals were in phase, meaning the estimated angle is close to the true direction of the sound source, while low power indicates that the direction is incorrect. This process is repeated across a set of angles within a comprehensive angular scan, and ultimately the angle yielding the highest energy is selected as the estimate of the direction of arrival (DOA) [28].

$$z(t) = \sum_{m=0}^{M-1} w_m \, y_m(t - \Delta_m) \tag{II.18}$$

when:

$z(t)$: The output signal.

t: Time variable, indicates that signals are functions of time.

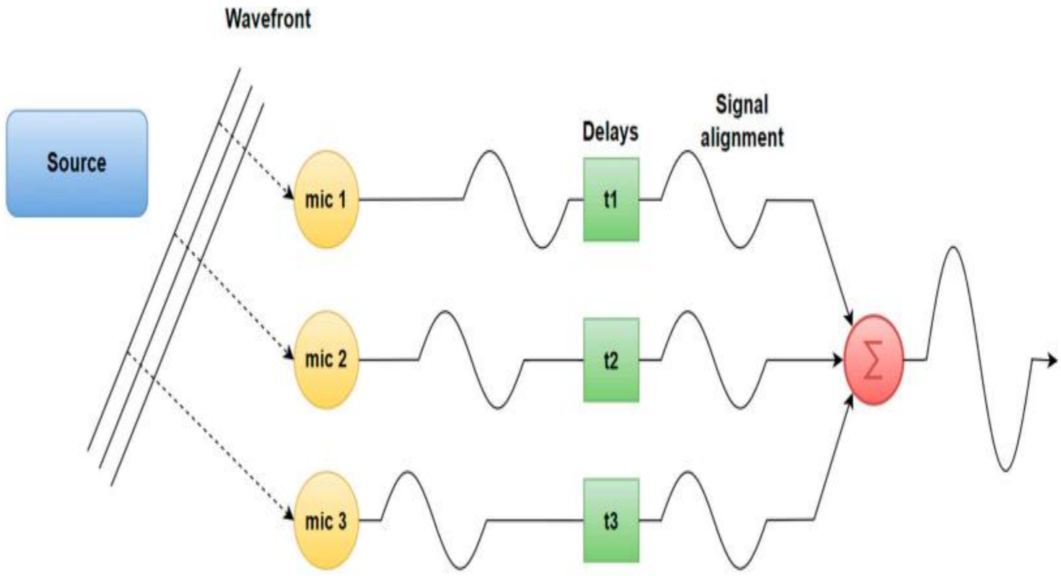
m: Identifies each microphone (or array element).

M: Total number of sensors (microphones) in the array.

y_m (t): The input signal from the m-th sensor.

w_m: Weight (scaling factor) for the m-th sensor.

Δ_m: Time delay applied to the m-th signal.



FigureII.5: Delay and sum basic idea [18].

B) Minimum Variance Distortion-less Response (MVDR)

MVDR is a widely used Beamforming technique for estimating the Direction of Arrival (DOA) of a signal, making it highly suitable for applications such as UAV localization. It relies on an array of sensors (e.g., microphones) that receive the same signal with slight time delays due to the angle of arrival., MVDR exploits them indirectly through phase differences [29]. The algorithm constructs a steering vector that models how a signal arriving from a given direction is observed across the array. It then computes an optimal weight vector that preserves the signal coming from the desired direction while minimizing the output power caused by noise and interference from other directions using the covariance matrix of the received signals [30]:

$$w_{MVDR} = \frac{R^{-1}a(\theta)}{a^H(\theta)R^{-1}a(\theta)} \quad \text{(II.19)}$$

w_{MVDR}:The optimal beamforming weight vector.

R:The spatial covariance matrix of the received signals.

$a(\theta)$:The steering vector corresponding to the direction θ.

The MVDR spatial spectrum is defined as:

$$P_{MVDR}(\theta) = \frac{1}{a^H(\theta)R^{-1}a(\theta)} \qquad \text{(II.20)}$$

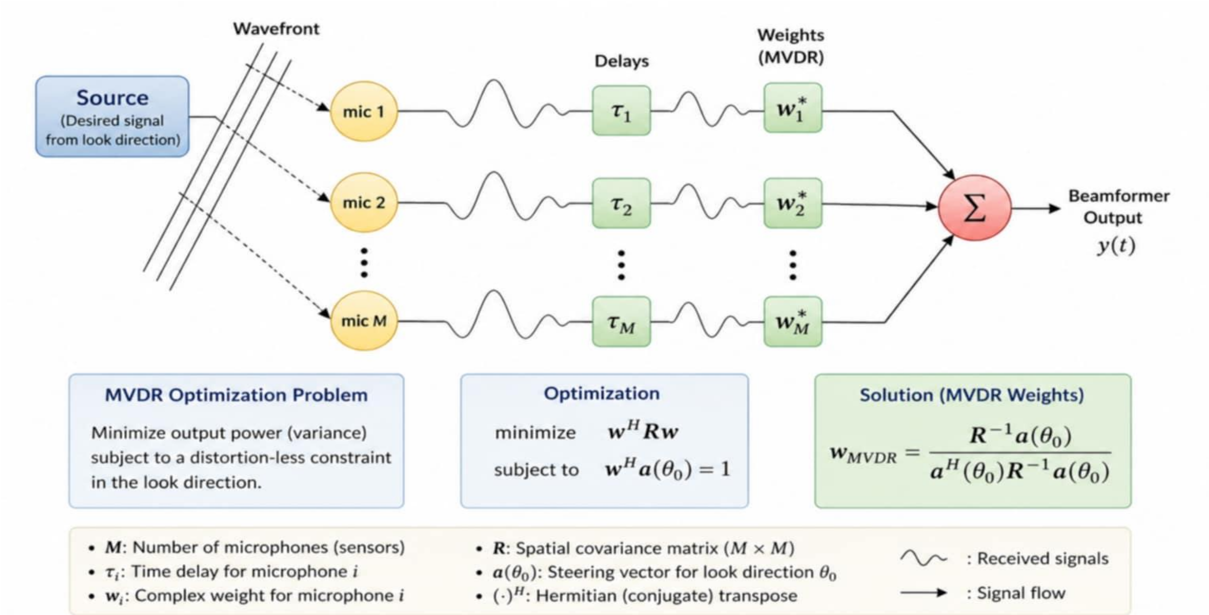
By scanning over all possible angles and evaluating the spatial power spectrum, the algorithm identifies the DOA as the angle corresponding to the maximum peak. However, MVDR provides only the direction, not the actual position. To estimate the UAV location, geometric methods are applied, either by intersecting the estimated direction with the ground plane when the altitude is known, or by combining multiple DOA estimates from different arrays through triangulation. Therefore, MVDR-based localization typically involves two main steps: DOA estimation using MVDR, followed by position estimation using geometric relationships, enabling accurate and robust localization even in the presence of interference.

The beamformer output can be expressed as follows:

$$y(t) = w^H x(t) \qquad \text{(II.21)}$$

w^H: the weight vector (Hermitan transpose).

$x(t)$:the vector of received signals from all microphones.

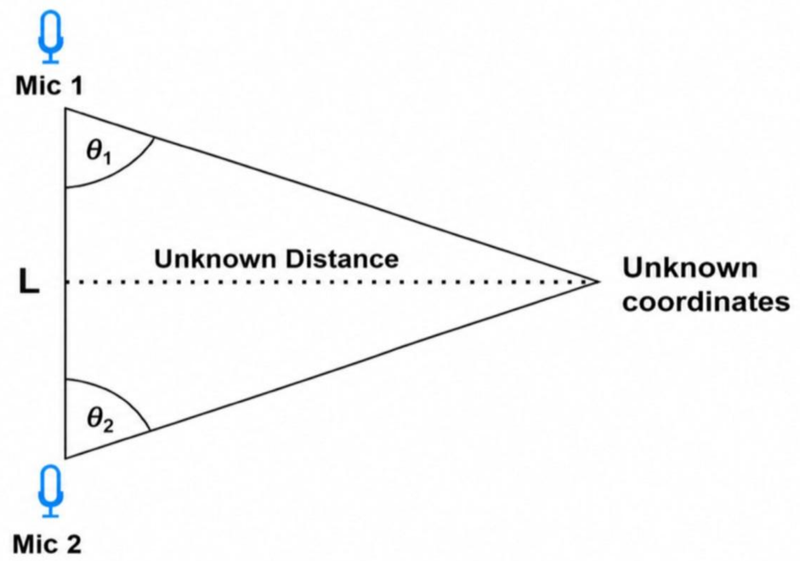


FigureII.6: Minimum Variance Distortion-less Response (MVDR) Beamformer.

II.3.1.5 Localization Algorithms

A) Triangulation (based on DOA/AOA)

Employs the geometric characteristics of triangles for localization determination. This approach calculates the angles at which acoustic signals arrive at the microphones Instead of calculating how far away a sound is, this method uses the Direction of Arrival (DOA) from two or more fixed points, Two receivers at known positions measure the angle of the incoming sound wave relative to a baseline by drawing a line from each receiver at the measured angle, the point where the lines intersect is the location of the source , Often used in radio direction finding and simple microphone arrays. To establish a two-dimensional localization, a minimum of two microphones is requisite. For precise spatial coordinates, a minimum of three microphones is indispensable. It is worth noting that increasing the number of microphones amplifies the method’s accuracy. Moreover, the choice of microphone significantly influences the precision of the triangulation. Employing directional microphones enhances the accuracy by precisely capturing the directional characteristics of sound [31].



FigureII.7: Two-dimensional triangulation scheme [18].

If the distance between two sensors is L, and the angles measured are the coordinates (x, y) are:

$$y = \frac{L}{\cot(\theta_1) + \cot(\theta_2)} \quad \text{(II.22)}$$

$$x = y \cot(\theta_1) \quad \text{(II.23)}$$

When:

y: The Altitude (or Depth). It represents the perpendicular distance from the baseline to the target point.

L: The Baseline. This is the known, fixed distance between your two sensors (e.g., two microphones or two observation stations).

θ_1: The first angle. The angle measured from the first sensor (usually at the origin (0,0)) toward the target.

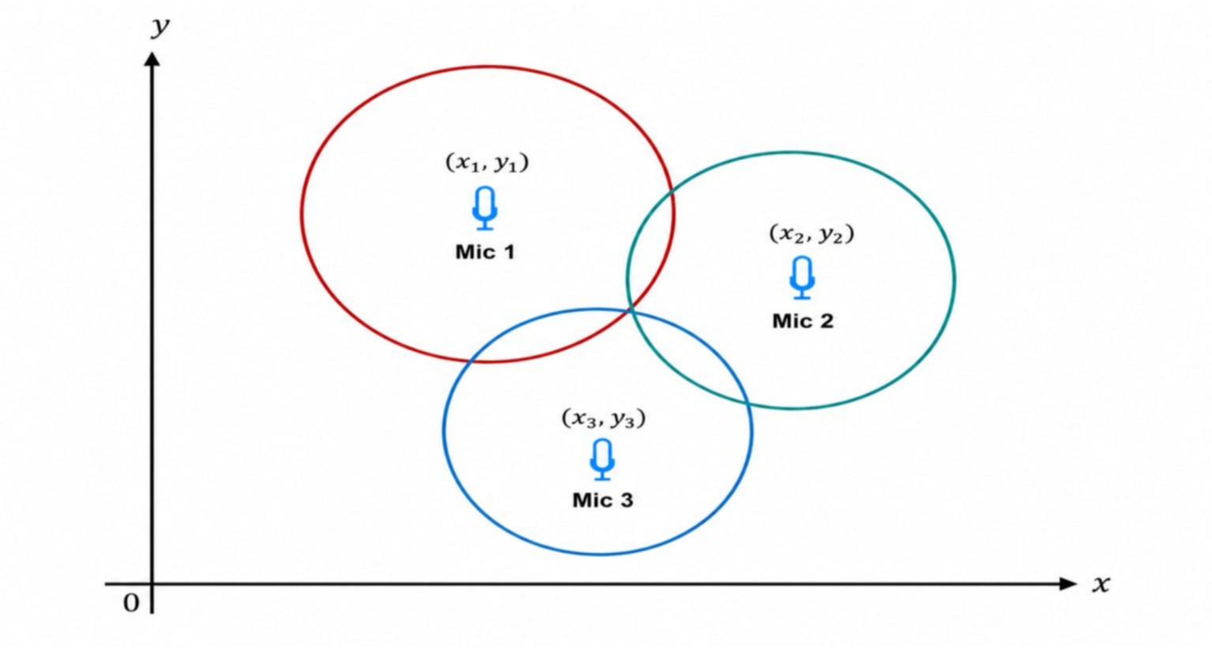
θ_2: The second angle. The angle measured from the second sensor toward the target.

cot: The Cotangent function

x: The Horizontal Coordinate. This tells you exactly where the point lies along the axis of the baseline

B) Trilateration (based on TOA/RSS)

It is used to determine localization based on the distance to three microphones (Figure II.7). Each microphone captures the acoustic signal at a different time, based on which the distance to the sound source is calculated. On this basis, the localization is determined by creating three circles with a radius corresponding to the distances from the microphones. The intersection point is the localization of the sound source. It is less dependent on the directional characteristics of the microphones, potentially providing more flexibility in microphone selection. sound source. It is less dependent on the directional. This is the basic principle behind how GPS receivers calculate your position relative to satellites [32].



FigureII.8: Two-dimensional trilateration scheme [18].

Trilateration determines a position by measuring the absolute distance (Range) from the source to at least three sensors.

For a sensor at (x_i, y_i, z_i) with a distance r_i to the source:

$$(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 = r_i^2 \tag{II.24}$$

To solve for (x, y), a system of three equations is typically linearized:

$$2x(x_2 - x_1) + 2y(y_2 - y_1) + 2z(z_2 - z_1) = r_1^2 - r_2^2 + x_2^2 - x_1^2 + y_2^2 - y_1^2 + z_2^2 - z_1^2 \tag{II.25}$$

When:

(x, y, z): These are the unknown coordinates of the source (the object or person you are trying to locate).

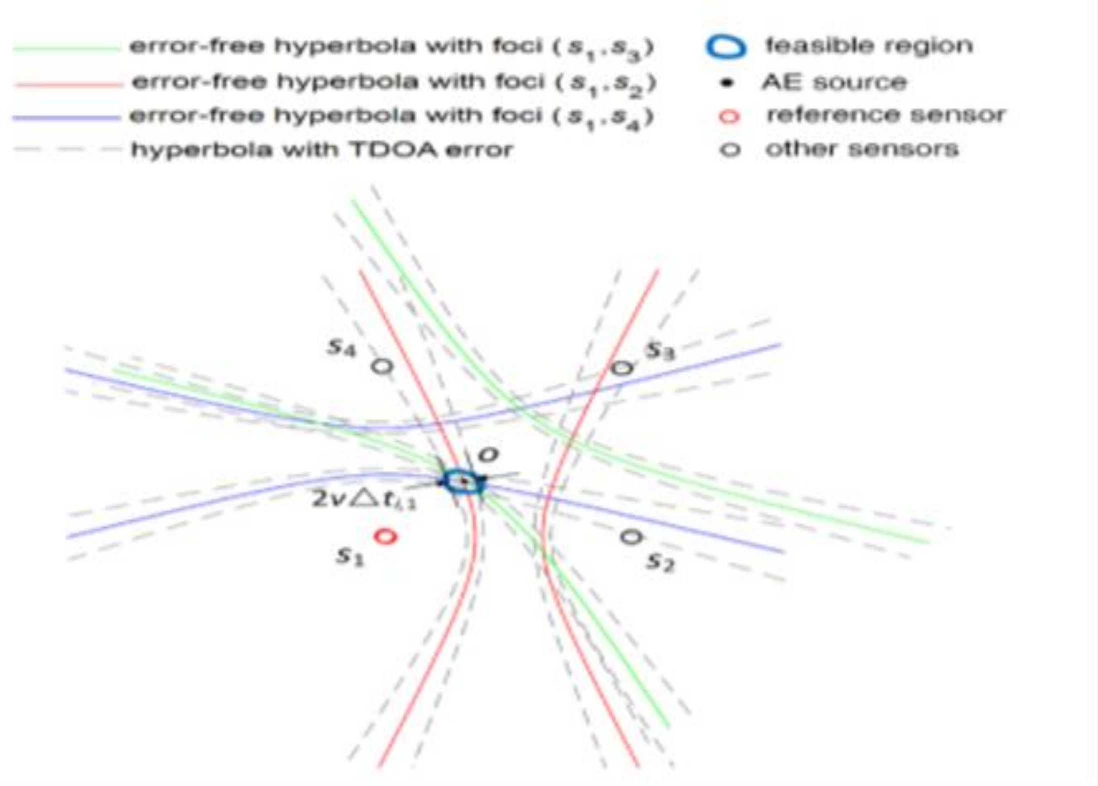
(x_i, y_i, z_i): These are the known coordinates of the i-th sensor (e.g., Sensor 1, Sensor 2, Sensor 3, etc.).

r_i:This is the Range (Distance) from the i-th sensor to the source.

i: This is an index representing which sensor is being used (e.g., if i=1, it refers to the first sensor).

c) Multilateration (based on TDOA)

It is used to determine the localization based on four or more microphones. The principle of operation is identical to trilateration. Using more reference points allows for a more accurate determination of the localization because, with their help, measurement errors can be compensated is the most common method for "blind" sound sources where the exact time the sound started is unknown, rather than measuring absolute distance, it measures the difference in time it takes for a sound to reach one sensor versus another. A constant time difference between two sensors defines a hyperbola. By using multiple pairs of sensors, you create multiple hyperbolic curves. The source is located at the intersection of these hyperbolas. However, this results in greater complexity and computational requirements. Despite this complexity intricacy, the accuracy and error mitigation benefits make multilateration a crucial technique in applications where precise localization determination is paramount [18].



FigureII.9: Hyperbolic Multilateration [33].

If (v) is the speed of sound and $(t_i - t_j)$ is the difference in arrival time between two sensors:

$$\sqrt{(x - x_i)^2 + (y - y_i)^2} - \sqrt{(x - x_j)^2 + (y - y_j)^2} = v \times (t_i - t_j) \tag{II.26}$$

When:

x, y: These are the unknown coordinates of the sound source that you are trying to detect.

x_i, y_i: The known coordinates of the first sensor (Sensor i).

x_j, y_j: The known coordinates of the second sensor (Sensor j).

v :The speed of sound in the given medium (e.g., air).

t_i: The arrival time of the signal at sensor i.

t_j: The arrival time of the signal at sensor j.

II.3.2 Artificial Intelligence Methods

Artificial Intelligence (AI) methods have become increasingly important in sound source localization (SSL), offering improved robustness and accuracy in complex and noisy environments. Unlike classical approaches that rely on explicit mathematical models, AI-based methods learn spatial and acoustic features directly from data, enabling better performance in reverberant and real-world scenarios. Deep learning architectures such as convolutional neural networks (CNNs) and hybrid models combining beamforming outputs with neural networks are widely used to enhance direction of arrival (DOA) estimation and localization accuracy. These methods are particularly effective in UAV and drone detection applications, where traditional techniques may struggle due to dynamic noise and interference [34], [35].

Conclusion:

This chapter has presented a classification of SSL systems, showing that most methods rely on microphone arrays as the primary sensing element. Classical localization techniques were detailed, including TOA, TDOA, FDOA, RSS, and DOA/AOA methods. These approaches are mainly based on physical and geometric principles, and they remain widely used due to their interpretability and solid mathematical foundations.

In addition, advanced signal processing techniques such as MUSIC and beamforming methods (DAS and MVDR) were presented. These methods improve direction estimation accuracy by exploiting spatial and spectral properties of the received signals. Finally, localization strategies such as triangulation, trilateration, and multilateration were described, showing how geometric relationships between sensors can be used to estimate the position of a sound source.

The chapter also highlighted the limitations of classical methods, particularly in noisy and complex environments, which has led to the increasing adoption of artificial intelligence and hybrid approaches in recent research.

Overall, this chapter demonstrates that while traditional methods remain essential for understanding the fundamentals of acoustic localization, the future of the field is moving toward more intelligent, adaptive, and data-driven systems that combine signal processing and machine learning to achieve higher accuracy and robustness.

Chapter III

3D Acoustic Localization Using UCA Array in Near-Field

III.1 Introduction

Acoustic source localization is widely used in surveillance, robotics, and drone detection systems. The objective is to estimate the spatial position of a sound source using multiple microphones arranged in a known geometry. In this work, a Uniform Circular Array (UCA) composed of multiple microphones is used to estimate the 3D position of a drone including azimuth, elevation, and distance.

Three localization algorithms are implemented and compared:

- Delay-and-Sum Beamforming (DAS)
- Multiple Signal Classification (MUSIC)
- Time Difference of Arrival (TDOA)

These methods rely on the propagation time differences of the acoustic wave arriving at different microphones in the array [9].

Unlike far-field localization where only direction is estimated, the proposed implementation considers Near-Field propagation, allowing the estimation of the distance between the drone and the array [36].

In this study, the source is assumed to be located in the Near-Field region, meaning that both direction and distance must be estimated.

All algorithms are implemented using real audio signals and evaluated under different Signal-to-Noise Ratio (SNR) conditions using Monte Carlo simulations. Performance is assessed using Mean Squared Error (MSE) and Normalized MSE (NMSE) [38].

III.2 System Model

III.2.1 Uniform Circular Array (UCA)

The microphones are arranged in a Uniform Circular Array with radius R. The position of the m^{th} microphone is defined as:

$$\begin{cases} x_m = R\cos(\theta_m) \\ y_m = R\sin(\theta_m) \\ z_m = 0 \end{cases} \quad \text{(III .1)}$$

Where:

x_m, y_m, z_m: Cartesian coordinates of microphone m (in meters).

R: Radius of the circular array (in meters).

θ_m: Angular position of microphone m (in radians).

Where:

$$\theta_m = \frac{2\pi(m-1)}{M}, m = 1, \ldots, M \quad \text{(III .2)}$$

m: Microphone index

M: Total number of microphones

θ_m: Angular position of microphone m

This geometry ensures uniform spatial sampling and is widely used in array signal processing applications [9].

As an example:

The array geometry used in our work consists of:

- $M = 8$ microphones.
- array radius $R = 1.72$ m

The circular configuration provides uniform angular coverage, which improves localization accuracy compared to linear arrays [37].

III.2.2 Source Model (Near-Field)

The 3D position of the source is expressed in spherical coordinates and converted into Cartesian form as follows:

$$P_s = \begin{bmatrix} rcos(\theta)cos(\phi) \\ rsin(\theta)\cos(\phi) \\ rsin(\phi) \end{bmatrix} \quad \text{(III .3)}$$

where:

P_s: Source position vector in 3D space.

r: distance between the source and microphone

ϕ: elevation angle

θ: azimuth angle

This representation allows direct computation of distances required for signal modeling [38].

III.2.3 Received Signal Model (Time Domain):

The received signal at each microphone is a delayed and noisy version of the emitted signal:

$$x_m(t) = s(t - \tau_m) + n_m(t) \quad \text{(III .4)}$$

Where:

$x_m(t)$: received signal at microphone m

$s(t)$: Source signal

τ_m: Propagation delay (s)

$n_m(t)$: Additive noise

The propagation delay is defined as:

$$\tau_m = \frac{d_m}{c} \quad \text{(III .5)}$$

$$d_m = \| ps - pm \| \qquad \text{(III .6)}$$

Where:

τ_m: Time delay at microphone m

d_m: Distance between source and microphone m(m)

c: Speed of sound (≈ 343 m/s) in air

p_s: Source position

p_m: Microphone position

Each microphone receives the same signal, but at a slightly different time due to different propagation distances. This time difference is the key information used for localization [9].

- The noise $n_{m(t)}$ is modeled as Additive White Gaussian Noise (AWGN).

III.2.4 Signal Model (Array Form)

The received signals can be written in vector form as:

$$\mathrm{x}(t) = \mathrm{a}(\theta, \phi, r)s(t) + \mathrm{n}(t) \qquad \text{(III .7)}$$

Where:

$\mathbf{x}(t)$: Received signal vector of size $M \times 1$

$\mathrm{a}(\theta, \phi, r)$: Steering vector of size $M \times 1$, dependent on azimuth θ, elevation ϕ, and distance r

$s(t)$: Source signal (scalar)

$\mathrm{n}(t)$: Noise vector of size $M \times 1$, modeled as Additive White Gaussian Noise (AWGN)

M: Number of microphones

For multiple samples:

$$\mathrm{X} = \mathrm{a}(\theta, \phi, r)\mathrm{s}^T + \mathrm{n} \qquad \text{(III .8)}$$

Where:

X: Received signal matrix of size $M \times N$

$\mathrm{a}(\theta, \phi, r)$: Steering vector of size $M \times 1$

s^T: Transpose of the source signal vector of size $1 \times N$

n: Noise matrix of size $M \times N$

N: Number of time samples

$(\cdot)^T$: Transpose operator

The covariance matrix is:

$$\mathrm{R}_{xx} = E[x(t)x^H(t)] \tag{III .9}$$

In practice, the covariance matrix is estimated using N temporal snapshots:

$$\mathrm{R}_{xx} = \frac{1}{N}\sum_{n=1}^{N} x[n]x^H[n] \tag{III .10}$$

This formulation is fundamental for subspace-based methods such as MUSIC [13].

III.3 Steering Vector for Near-Field UCA Localization

The steering vector represents the theoretical response of the microphone array to a signal arriving from a specific spatial position. It models the phase differences observed between microphones due to propagation delays of the acoustic wave [13].

III.3.1 Definition

The received signal vector can be written as:

$$\mathrm{x}(t) = \mathrm{a}(\theta, \phi, r)s(t) + \mathrm{n}(t) \tag{III .11}$$

$$\mathrm{a}(\theta, \phi, r) = \begin{bmatrix} e^{-j\frac{2\pi}{\lambda}d_1(\theta,\phi,r)} \\ e^{-j\frac{2\pi}{\lambda}d_2(\theta,\phi,r)} \\ \vdots \\ e^{-j\frac{2\pi}{\lambda}d_M(\theta,\phi,r)} \end{bmatrix} \tag{III .12}$$

With:

$$d_M(\theta,\phi,r) = \sqrt{(r\cos\,\theta\cos\,\phi - x_M)^2 + (r\sin\,\theta\cos\,\phi - y_M)^2 + (r\sin\,\phi - z_M} \quad \text{(III .13)}$$

$$\mathbf{a}(\theta,\phi,r) = \begin{bmatrix} e^{-j\frac{2\pi}{\lambda}\sqrt{(r\cos\theta\cos\,\phi - x_1)^2 + (r\sin\,\theta\cos\,\phi - y_1)^2 + (r\sin\,\phi - z_1)^2}} \\ e^{-j\frac{2\pi}{\lambda}\sqrt{(r\cos\theta\cos\,\phi - x_2)^2 + (r\sin\,\theta\cos\,\phi - y_2)^2 + (r\sin\,\phi - z_2)^2}} \\ \vdots \\ e^{-j\frac{2\pi}{\lambda}\sqrt{(r\cos\theta\cos\,\phi - x_M)^2 + (r\sin\,\theta\cos\,\phi - y_M)^2 + (r\sin\,\phi - z_M)^2}} \end{bmatrix} \quad \text{(III .14)}$$

where:

$x(t)$: received signal vector

$s(t)$: source signal

$a(\theta,\phi,r)$: steering vector

$n(t)$: noise vector

λ: Wavelength (m)

d_M: Distance from source to microphone m

This equation shows that the array output is a combination of the source signal scaled by the steering vector and noise.

Each element a_m corresponds to the phase delay at microphone m.

Each component represents the phase of the signal at one microphone. Together, they form a unique spatial signature of the source.

The steering vector plays a central role in all localization algorithms [39].

III.3.2 Near-Field Distance and Phase Delay

The distance between source and microphone (m) is:

$$d_m = \sqrt{(x_s - x_m)^2 + (y_s - y_m)^2 + (z_s - z_m)^2} \quad \text{(III .15)}$$

$$a_m = e^{-j\frac{2\pi}{\lambda}d_m} \quad \text{(III .16)}$$

$a_{m:}$ phase response at sensor (microphone) m

- The difference between the two values → This is what the algorithms exploit.

III.3.3 Steering Matrix in DAS, MUSIC, and TDOA Algorithms

Role of Steering Matrix is different from algorithm to another

- **DAS**: alignment of signals
- **MUSIC**: orthogonality test
- **TDOA**: distance/time relation

III.3.5 Near-Field vs Far-Field

- **Far-field**: Plane wave assumption; steering vector depends only on direction: $a(\theta, \phi)$
- **Near-field**: Spherical wave assumption; steering vector depends on distance: $a(\theta, \phi, r)$ Near-field modeling allows distance estimation, critical for drone localization [38].

III.4 DAS Localization Method

III.4.1 Definition

The Delay-and-Sum (DAS) algorithm is a classical beamforming technique that combines signals from multiple sensors after compensating for propagation delays in order to enhance signals coming from a specific direction [37].

III.4.2 Principle of DAS

The main idea of DAS is to scan the 3D space and compute the output power for each candidate position. The position that maximizes the power is considered as the estimated source location.

III.4.3 Covariance Matrix Decomposition

The covariance matrix is calculated as:

$$R_{xx} = \frac{1}{N}\sum_{n=1}^{N} x[n]x^{H}[n] \tag{III .17}$$

Where:

X: contains the signals received by the microphones.

The localization process consists of scanning a 3D grid of azimuth, elevation, and distance and selecting the point that maximizes the beamforming output power [13].

This expression represents the beamforming output power. High values indicate a strong match between the assumed position and the true source.

III.4.4 DAS Spectrum Formulation

$$P_{DAS}(p) = a^{H}R_{xx}a(p) \tag{III .18}$$

Where:

$P_{DAS}(p)$: beamformer output power (spatial spectrum) evaluated at position p

a: is the steering vector

R_{xx} : is the spatial covariance matrix

H: denotes the Hermitian transpose.

p : candidate source location, usually defined as: $p = (\theta, \phi, r)$

III.4.5 DAS Algorithm

Algorithm: Delay And Sum Beamforming

Input:
  X, grid(r,az,el)
Output:
  Estimated position ($\hat{r}$, $\hat{az}$, $\hat{el}$)
1: Compute covariance matrix Rxx
2: for each scan point p(r,az,el) do
3:   Compute steering vector a(θ, ϕ, r)
4:   Compute DAS spectrum: P(p) = a^H(p) Rxx a(p)
5: end for
6: Find maximum spectrum value
7: Return estimated position: $\hat{p}$ = argmax P(p)

III.4.6 Advantages and Limitations

III.4.6.1 Advantages

- Simple implementation
- Low computational cost

III.4.6.2 Limitations

- Low resolution
- Sensitive to noise

III.5 TDOA Localization Method

III.5.1 Definition

TDOA is a localization method that estimates the source position based on differences in arrival times between sensors [5].

III.5.2 Principle of TDOA

A reference microphone is selected, and the time delays between this reference and the other microphones are estimated using cross-correlation.

III.5.3 TDOA Signal Propagation Model

$$\tau_m = \frac{\parallel p_s - p_m \parallel}{c} - \frac{\parallel p_s - p_{ref} \parallel}{c} \quad \text{(III .19)}$$

where:

- τ_m: TDOA between microphone m and the reference microphone.
- c: speed of sound
- p_s: source position
- p_m: microphone position
- p_{ref}: position of the reference microphone

In practice, TDOA values are estimated using cross-correlation between microphone signals [6].

$$\tau_{ij} = \arg\ \max\ R_{x_i x_j} \quad \text{(III .20)}$$

Where:

τ_{ij}: is the time delay (TDOA) between signals x_i and x_j

$R_{x_i x_j}$: is the cross-correlation function between the two signals

The estimated delays are compared with theoretical delays for candidate positions in the 3D search grid.

III.5.4 Error Function

The optimal source position minimizes the error function:

$$J(p) = \sum_{m=2}^{M} \left(\Delta\tau_m^{meas} - \Delta\tau_m^{pred}\right)^2 \quad \text{(III .21)}$$

Where:

$J(p)$: Error function evaluated at source position p

M: number of microphones in the array

$\Delta\tau_m^{meas}$: measured time difference of arrival between microphone m and the reference microphone, estimated from the cross-correlation of received signals

$\Delta\tau_m^{pred}$: predicted time difference of arrival derived from the geometric distances between the source and microphones

- This function measures the error between measured and predicted delays.

The localization problem is formulated as a least squares optimization problem, where the best position minimizes the error.

III.5.5 TDOA Algorithm

Algorithm: Time Difference of Arrival
Input: X, grid Output: Estimated position (r̂, âz, êl) 1: Estimate τ_{meas} 2: For each scan point p(r,az,el) do 3: Compute τ_{pred} 4: J(p) = $\|\tau_{meas} - \tau_{pred}\|^2$ 5: end for 6: p̂ = argmin J(p)

III.5.6 Advantages and Limitations

III.5.6.1 Advantages

- Simple
- Practical implementation

III.5.6.2 Limitations

- Sensitive to delay estimation errors

III.6 MUSIC Localization Method

III.6.1 Definition

MUSIC is a high-resolution localization algorithm based on the decomposition of the signal space into signal and noise subspaces [39].

III.6.2 Principle of MUSIC

The algorithm exploits the orthogonality between the steering vector and the noise subspace. Peaks in the MUSIC spectrum correspond to possible source locations.

III.6.3 Covariance Matrix Decomposition

The covariance matrix is decomposed as:

$$R_{xx} = E_s \Lambda_S E_s^H + E_n \Lambda_n E_n^H \qquad \text{(III .22)}$$

where:

- E_s: signal subspace
- E_n: noise subspace
- E: eigenvectors
- Λ: eigenvalues

The steering vector corresponding to the true source direction is orthogonal to the noise subspace [38]:

$$\mathrm{a(p)} \perp \mathrm{E}_n \qquad \text{(III .23)}$$

This orthogonality property produces sharp spectral peaks and enables high-resolution localization.

III.6.4 MUSIC Spectrum

$$P(\boldsymbol{p}) = \frac{1}{a^H(p)E_n E_n^H a(\boldsymbol{p})} \qquad \text{(III .24)}$$

where:

$P(\boldsymbol{p})$: MUSIC pseudospectrum at location **p**

III.6.5 MUSIC Algorithm

Algorithm: Multiple Signal Classifier

Input:

X, grid(r,az,el)

Output:

Estimated position (r̂, âz, êl)

1: Compute covariance matrix Rxx

2: Eigen-decomposition → En

3: for each scan point p(r,az,el) do

4: Compute steering vector a(θ,φ,r)

5: Compute MUSIC spectrum: P(p) = $\frac{1}{a^H(p)E_n E_n^H a(p)}$

7: end for

8: Find maximum spectrum value

9: Return estimated position: p̂ = argmax P(p)

III.6.6 Advantages and Limitations

III.6.6.1 Advantages

- High resolution
- Ability to separate close sources

III.6.6.2 Limitations

- High computational complexity
- Requires knowledge of number of sources

III.7 Performance Evaluation Methodology

III.7.1 Evaluation Metrics

The Mean Square Error (MSE) is defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N} \| \hat{p}_i - p_s \|^2 \quad \text{(III .25)}$$

Where:

MSE: mean squared error (localization error metric)

N: number of trials / snapshots / simulations

$\hat{p}_i$: estimated position in the i^{th} snapshots

p_s: true source position

$\|\cdot\|^2$: squared Euclidean norm

The normalized version is:

$$NMSE = \frac{MSE}{\max(MSE)} \quad \text{(III .26)}$$

III.8 Conclusion

This chapter presented the complete theoretical framework for 3D acoustic source localization using a Uniform Circular Array. The steering vector was analyzed in detail, and three localization methods DAS, MUSIC, and TDOA were described with their mathematical models and physical interpretations.

The next chapter will focus on simulation results and a detailed comparison of these methods in terms of accuracy and robustness.

Chapter IV

Simulation & Results

IV.1. Introduction

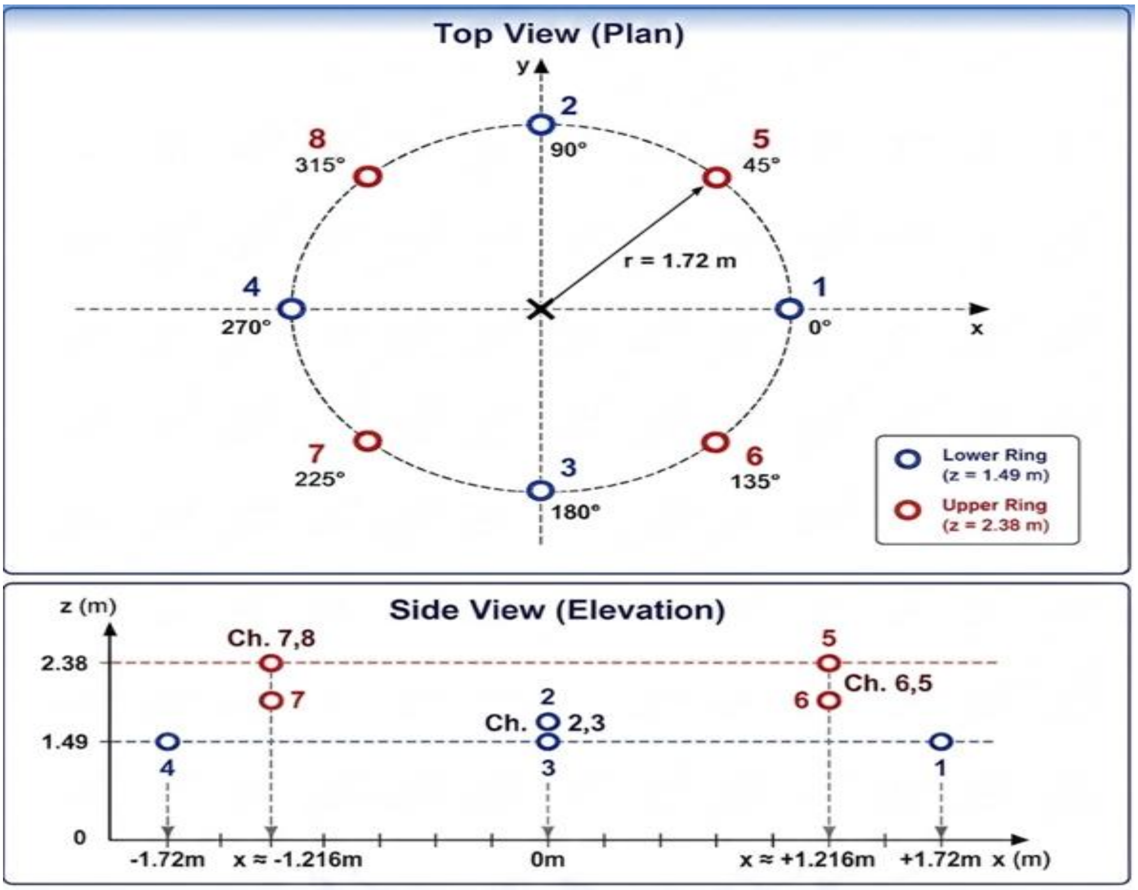
This chapter aims to determine the position of an unmanned aerial vehicle in three-dimensional space based on acoustic signals captured by an array of microphones. To achieve this, a uniform circular array (UCA) was used due to its ability to provide comprehensive angular coverage in the near field, where horizontal and vertical angles, as well as distance, are estimated. The study relied on three main methods for signal processing and localization: the Time Difference of Arrival (TDOA) method, the Delay and Sum (DAS) method, and the MUSIC algorithm, which is considered a high-precision method for signal analysis. Real-world audio recordings of drones taken from a dataset were used to ensure signal realism, and the system was simulated and evaluated using MATLAB, allowing for a comparison of the effectiveness of the three methods in near-field conditions and an assessment of their accuracy in estimating the sound source location and evaluating their accuracy through SNR versus MSE.

IV.2 Experimental Setup

The experimental setup used in this study is based on real acoustic drone recordings obtained from the UaVirBASE dataset [40] A segment containing 2000 samples is selected for analysis in order to reduce computational complexity while preserving the essential acoustic characteristics of the UAV signal. The recordings were acquired using a Uniform Circular Array (UCA) composed of eight microphones in order to achieve accurate three-dimensional sound source localization.

The target sound source corresponds to a DJI Mavic 3 Cine unmanned aerial vehicle (UAV), whose acoustic emissions were recorded under realistic outdoor conditions. The adopted microphone array geometry provides full azimuth coverage and improves localization performance in near-field environments.

Figure IV.1: DJI Mavic 3 Cine UAV used for acoustic signal acquisition.



Microphones Coordinates and Parameters									
Ch.	Ring	Az. (°)	Elev. (°)	Dist. (m)	Height (m)	Corr. X (m)	Corr. Y (m)	Lat.	Long. (°)
1	Lower	0	10	1.72	1.49	1.7200	0.0000	52.25533	22228067
2	Lower	90	10	1.72	1.49	0.0000	1.7200	52.25533	75777778
3	Lower	180	10	1.72	1.49	-1.7200	0.0000	52.25533	26462222
4	Lower	270	10	1.72	1.49	0.0000	-1.7200	52.25533	37777778
5	Upper	45	20	1.72	2.38	1.2162	1.2162	52.25533	55501004
6	Upper	135	20	1.72	2.38	-1.2162	1.2162	52.25533	55507004
7	Upper	225	20	1.72	2.38	-1.2162	-1.2162	52.25531	65507165
8	Upper	315	20	1.72	2.38	1.2162	-1.2162	52.25533	19371472

Figure IV.2: Uniform Circular Array (UCA) configuration used for the acoustic localization experiments.

IV.2.1. Simulation parameters:

The simulations were performed using acoustic drone signals extracted from the UaVirBASE dataset. The dataset provides multi-microphone recordings collected under different distances and angular positions for drone localization experiments [40].

The same acoustic data and simulation parameters were used for the three localization algorithms in order to ensure a fair comparison.

Simulation Parameters:

[x_audio, Fs] = audioread('20241115_102243\output.wav')

Drone type:" DJI Mavic 3 Cine"

Speed of drone 21 m/s

number of microphones= 8

microphone type:" NTG-2"

Array Geometry: Uniform Circular Array (UCA)

Signal-to-noise ratio SNR = -5,…,10 dB

azimuth angle = 180°

elevation angle = 30°

Distance = 20 m

array radius R = 1.72m

Speed of sound = 343 m/s

Number of Snapshots N = 2000

IV.2.2. Simulation results:

Each image displays the "spectrum" generated by the algorithm, which is a probability map where the peaks indicate the most likely location of the drone based on the azimuth and elevation angles.

IV.2.2.1. Result of the delay-and-sum (DAS) algorithm:

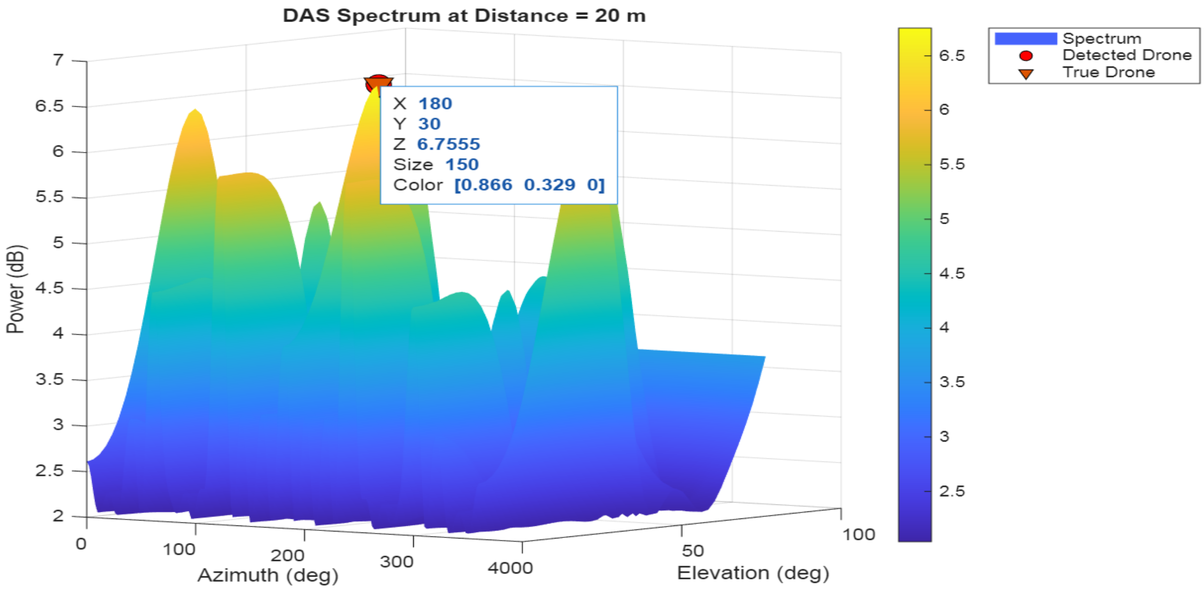


Figure IV.3: 3D Localization using DAS spectrum.

IV.2.2.2. Result of the Time Difference of Arrival (TDOA) algorithm:

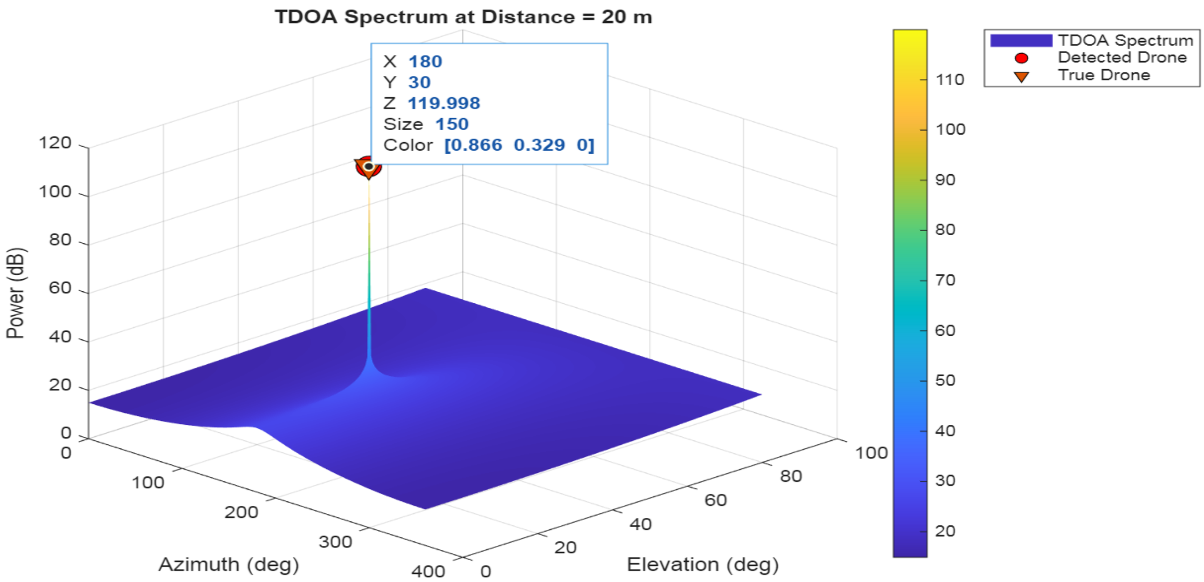


Figure IV.4: 3D Localization using TDOA spectrum.

IV.2.2.3. Result of the Multiple Signal Classification (MUSIC) algorithm:

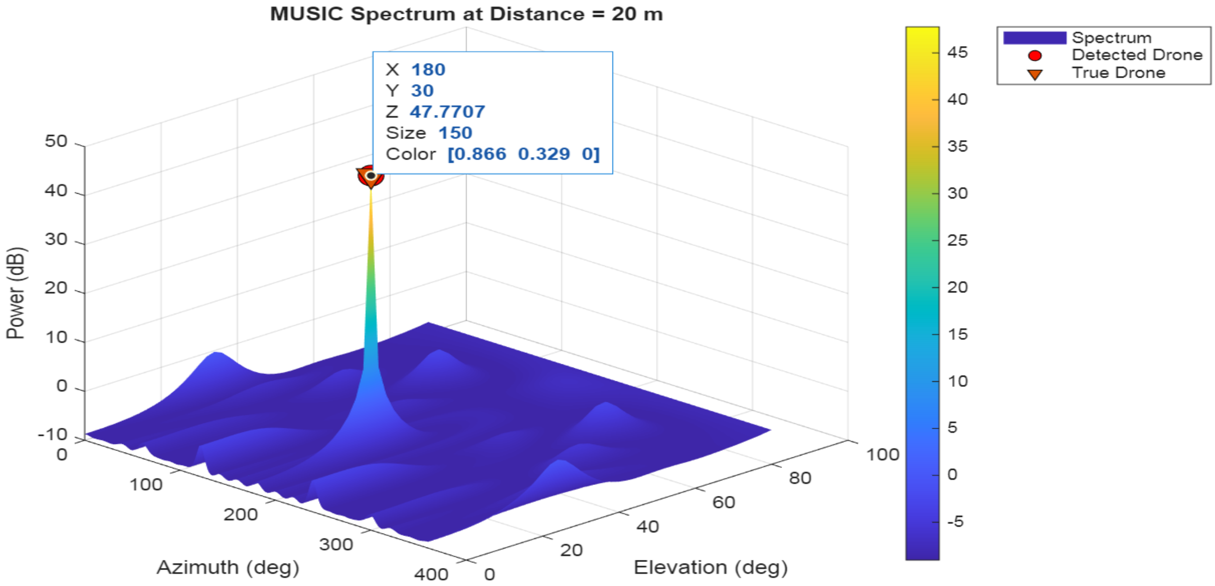


Figure IV.5:3D Localization using MUSIC spectrum.

IV.2.3. Comparison between algorithms:

- The DAS (Delay and Sum) algorithm is considered the weakest compared to MUSIC and TDOA because it relies on a simple principle of delaying signals and then summing them directly, without utilizing any advanced mathematical analysis or exploiting the statistical structure of the signal. This approach results in a spectrum characterized by broad, non-sharp peaks, which reduces directional accuracy and causes the appearance of multiple false peaks, making it difficult to identify the true source. In addition, DAS is highly sensitive to noise, as noise is collected along with the signal without any mechanism to separate it or reduce its impact, unlike the MUSIC algorithm.

- The TDOA (Time Difference of Arrival) algorithm estimates the time differences in signal arrival between sensors, giving it good source localization capabilities, particularly in terms of distance, as it excels at determining the geometric position (Distance); however, its accuracy in Azimuth and Elevation is lower than that of MUSIC. This algorithm produces a spectrum with a clear, concentrated peak, making it easier to distinguish the true source compared to simpler methods such as DAS. Its performance remains average because it relies heavily on the accuracy of time delay estimation, making it sensitive to noise and synchronization errors. Additionally, it does not utilize the full signal structure like MUSIC, particularly regarding precise angular information.

- The MUSIC (Multiple Signal Classification) algorithm is considered one of the most powerful algorithms because it divides the signal's mathematical space into two parts: the "signal space" and the "noise space." Furthermore, instead of searching for the strongest signal, MUSIC searches for angles at which the true signal is perfectly orthogonal to the noise space. This results in very sharp peaks, meaning very high angular accuracy. One of its advantages is that if two drones are very close to each other, MUSIC can distinguish between them, whereas DAS would see them as a single, indistinct target.

IV.3 Performance Evaluation Using SNR and MSE

This section evaluates the performance of the DAS, TDOA, and MUSIC algorithms using the Mean Squared Error (MSE) metric under different Signal-to-Noise Ratio (SNR) conditions.

IV.3.1. Simulation Parameters:

The drone acoustic recordings were obtained from the UaVirBASE

[x_audio, Fs] = audioread('20241115_102243\output.wav')

number of microphones= 8

Array Geometry: Uniform Circular Array (UCA)

Number of Snapshots N = 2000

Signal-to-noise ratio SNR = -5:2 :10 dB

Iteration (Itr) =40

azimuth angle = 180°

elevation angle = 30°

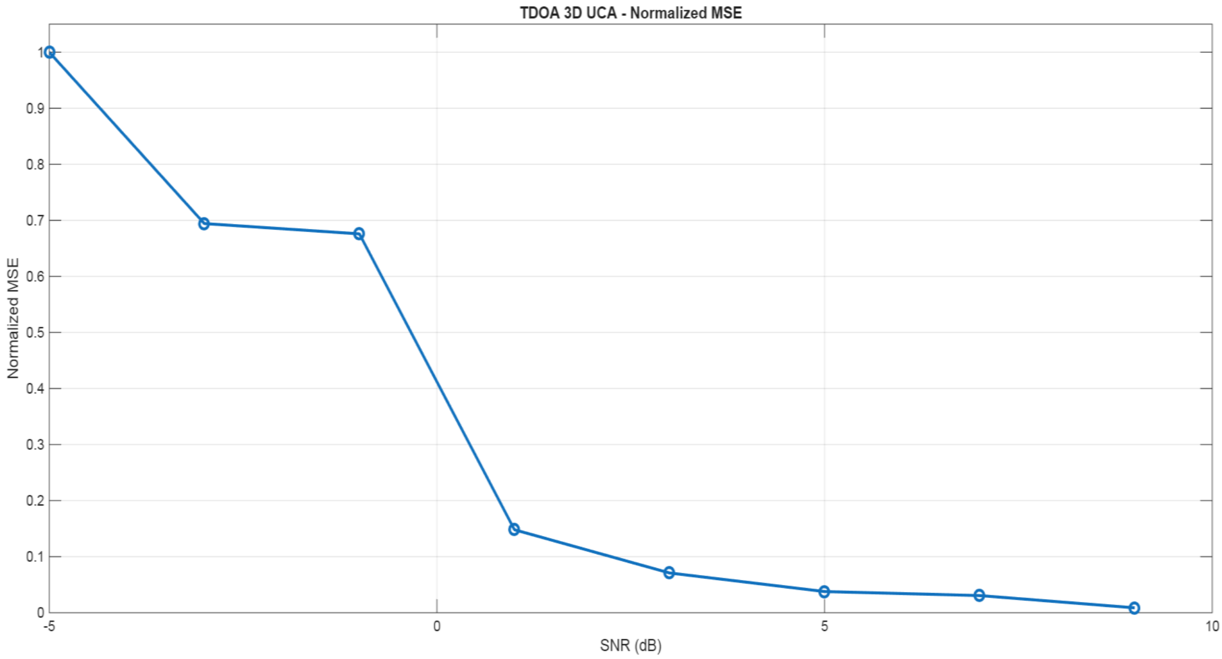
Distance = 20 m

array radius R = 1.72 m

Speed of sound = 343 m/s

IV.3.2. simulation results:

A) Comparative Performance Analysis between algorithms:



FigureIV.6: Performance Analysis of TDOA localization Using SNR and Normalized MSE

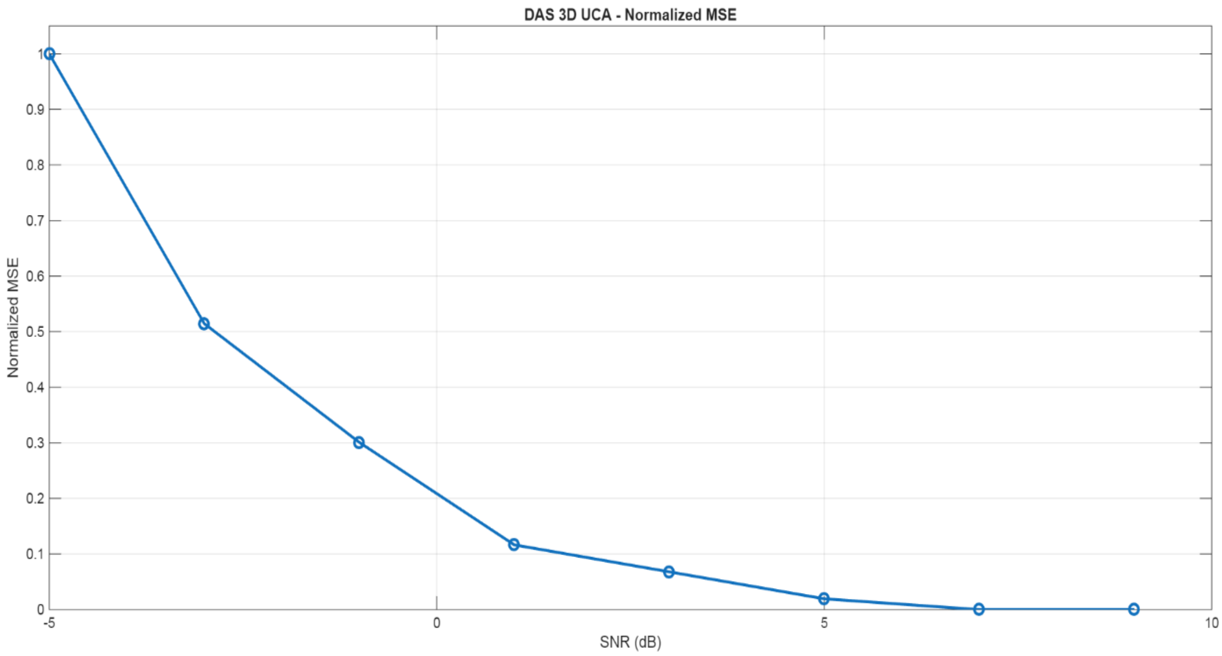


Figure IV.7: Performance Analysis of DAS localization Using SNR and Normalized MSE

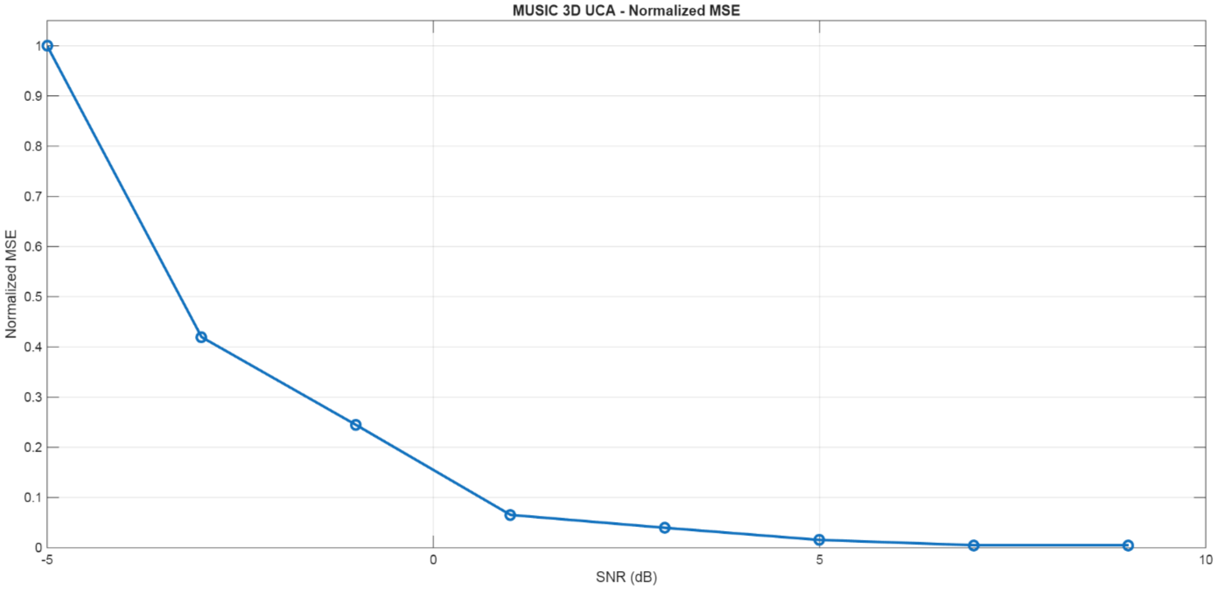


Figure IV.8: Performance Analysis of MUSIC localization Using SNR and Normalized MSE.

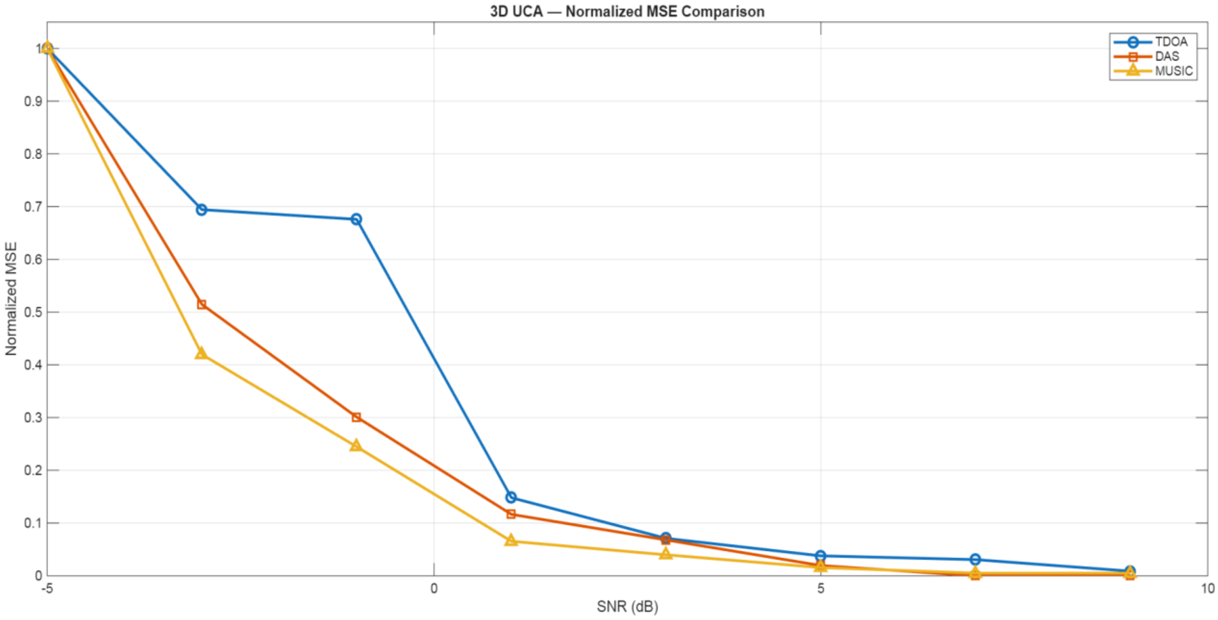


Figure IV.9: Performance Evaluation of DAS, MUSIC, and TDOA Algorithms Using Normalized MSE Across Different SNR Conditions

The figure compares the performance of the DAS, MUSIC, and TDOA algorithms based on the mean squared error (MSE) as a function of the signal-to-noise ratio (SNR). The TDOA algorithm records the highest error values, especially in noisy environments, reflecting its high sensitivity to noise and lower accuracy compared to the other algorithms, although its performance improves at high SNR values. The DAS algorithm exhibits moderate performance, with relatively high error values at low SNR, but it gradually improves as SNR increases, approaching the performance of MUSIC at high SNR levels. In contrast, the MUSIC algorithm consistently achieves the lowest error values across all SNR conditions, demonstrating superior accuracy and strong robustness against noise, particularly under low SNR conditions. Overall, these results indicate that MUSIC is the most accurate and reliable algorithm, DAS provides a balanced trade-off between performance and complexity, while TDOA remains the least efficient in high-noise environments.

IV.3.3 Runtime and localization accuracy comparison

Conditions:

SNR=1dB

Runs=5

RAM = 16GB

Processor = AMD Ryzen5 5600H with Radeon Graphics (3.30 GHz)

Algorithms	Average Run-time_sec	Average MSE	Average NMSE
TDOA	0.152865	0.242463	0.261546
DAS	1.029919	0.002122	0.047899
MUSIC	1.288727	0.002144	0.046233

Table IV.1: Comparative Evaluation of Runtime, MSE, and NMSE for TDOA, DAS, and MUSIC Algorithms

The runtime results show that TDOA is the fastest algorithm because it relies mainly on time-delay estimation rather than covariance matrix decomposition. DAS presents intermediate computational complexity due to beamforming operations over the scanning grid. In contrast, MUSIC requires eigenvalue decomposition and spectral scanning, which increases computational complexity and execution time.

IV.3.4 Comparative Table: DAS vs. MUSIC vs. TDOA:

Feature	Delay and Sum (DAS)	TDOA Algorithm	MUSIC Algorithm
Spatial Resolution	**Low**; produces a broad, blunt peak.	**High**; clear peak, though less sharp than MUSIC in dense environments.	**Very High**; produces a "super-resolution" sharp needle-like peak.
Accuracy (MSE)	**Moderate**; the error decreases steadily as SNR improves.	**Lower Efficiency**; requires a higher SNR to reach low error levels.	**Superior**; achieves the lowest Normalized MSE even at low SNR.
Noise Robustness	**Sensitive to noise**; peaks can easily get lost in background interference.	**Highly sensitive;** the error curve (Yellow) stays higher for longer as SNR increases.	**Highly Robust**; effectively separates the signal subspace from the noise subspace.
Run-time	**Moderate;** requires beamforming operations over the scanning grid.	**Fastest;** mainly relies on time-delay estimation.	**Slowest;** relies on eigenvalue decomposition and spectral scanning.
Computational Complexity	Moderate	Low	High
Suitable Applications	Systems requiring a balance between speed and accuracy.	Real-time applications requiring high speed.	Applications requiring very high accuracy despite longer execution time.

Table IV.2: Performance Comparison Between DAS, TDOA, and MUSIC Localization Algorithms.

IV.4. Conclusion:

This study conducted a comparative and in-depth analysis of three of the most prominent techniques for determining the direction of arrival (DOA) and positioning in radar and signal processing systems—namely, the DAS, TDOA, and MUSIC algorithms—and applied them to a scenario involving the detection of unmanned aerial vehicles (UAVs).

Based on simulation results obtained at a distance of 20 meters, the results demonstrated clear differences in the performance of each algorithm.

The Delay and Sum (DAS) algorithm proved to be the simplest in terms of software implementation; however, it suffers from low spatial accuracy and high sidelobes, making it less reliable in high-noise environments. In terms of computational performance, DAS achieved a moderate execution time compared to the other algorithms due to the beamforming operations performed over the scanning grid.

The Time Difference of Arrival (TDOA) technique delivered excellent results in terms of computational speed, achieving the shortest runtime among all tested algorithms. It also provided good distance estimation capability because it relies on propagation time differences between microphones. However, its efficiency remains dependent on a high signal-to-noise ratio, as the graphs showed greater sensitivity to noise compared to MUSIC.

The Multiple Signal Classification (MUSIC) algorithm demonstrated “super-resolution” superiority, as it successfully separated the signal domain from the noise domain, producing very sharp peaks and extreme accuracy in determining coordinates (azimuth and elevation). Additionally, the mean squared error (MSE) curves confirmed that it is the most stable and efficient even at low signal-to-noise ratios (SNR). Nevertheless, MUSIC required the longest execution time because of covariance matrix processing and eigenvalue decomposition operations.

In conclusion, the selection of the appropriate algorithm depends on the required balance between localization accuracy, computational complexity, execution time, and the characteristics of the operating environment.

General Conclusion

Generale Conclusion

This study addressed the problem of three-dimensional localization of a sound source from an unmanned aerial vehicle using an array of microphones. The work focused on the design, implementation, and evaluation of signal processing techniques capable of estimating the UAV's position in terms of azimuth, elevation, and distance under near-field conditions.

The dissertation began by outlining the fundamental principles of acoustic-based positioning, highlighting the importance of acoustic sensing as a complementary solution to traditional perception systems. It also addressed the key challenges associated with drone positioning, such as noise, platform motion, and the complexity of the surrounding environment. Subsequently, a comprehensive review of various methods used for detection and localization was presented, covering classical signal-processing-based methods as well as modern AI-based approaches. Methods such as Time Difference of Arrival (TDOA), beamforming, and the MUSIC algorithm were analyzed in terms of their theoretical foundations and practical applications.

In the next stage, a complete mathematical framework for acoustic localization in the near field was developed using a circular array of microphones. The signal model, steering vector, and contrast matrix processing were detailed, enabling the implementation of three main algorithms: Delay-and-Sum (DAS), MUSIC, and TDOA.

Simulation results showed that the MUSIC algorithm achieves the highest localization accuracy, thanks to its ability to separate the signal space from the noise space, resulting in sharp spectral peaks and low estimation error. The TDOA method also demonstrated a good balance between computational complexity and performance, particularly in distance estimation. In contrast, the DAS method, despite its simplicity and computational efficiency, suffers from lower accuracy and greater sensitivity to noise.

As the comparative study demonstrated, based on the mean squared error (MSE) as a function of the signal-to-noise ratio (SNR), the performance of all methods improves as noise decreases, with the MUSIC algorithm showing clear superiority, particularly in high-noise conditions.

Despite these encouraging results, challenges remain, including noise resistance in real-world environments, the impact of multipath reflections, and real-time implementation constraints on embedded platforms in unmanned aerial vehicles. Looking ahead, future work could explore the development of hybrid methods that combine traditional signal processing techniques with artificial intelligence approaches, with the aim of improving accuracy and reliability. Additionally, real-time system implementation and experimental validation using actual unmanned aerial vehicles (UAVs) represent important directions for future research.

In conclusion, this study demonstrates that acoustic localization using microphone arrays is an effective and robust solution for tracking and detecting UAVs, particularly in environments where traditional sensor systems such as radar, optical cameras, or infrared imaging are limited or ineffective.

References:

[1] A. Ollero and I. Maza, Multiple Heterogeneous Unmanned Aerial Vehicles. Springer, 2007.

[2] A. Waharte and N. Trigoni, “Supporting search and rescue operations with UAVs,” in IEEE Int. Conf. Emerging Security Technologies, 2010.

[3] R. Austin, Unmanned Aircraft Systems. Wiley, 2010.

[4] H. Liu, B. Sun, and M. Q.-H. Meng, “Sound source localization for UAVs,” IEEE Sensors Journal, 2018.

[5] J. Chen, J. Benesty, and Y. Huang, “Time delay estimation in room acoustic environments,” EURASIP Journal on Applied Signal Processing, 2006.

[6] C. Knapp and G. Carter, “The generalized correlation method for estimation of time delay,” IEEE Trans. Acoust., Speech, Signal Process., 1976.

[7] L. Kinsler et al., Fundamentals of Acoustics. Wiley, 2000.

[8] D. Cabric and R. W. Heath, “UAV signal processing challenges,” IEEE Communications Magazine, 2017.

[9] H. L. Van Trees, Optimum Array Processing. Wiley, 2002.

[10] Y. Zhang, J. Liu, and X. Chen, “Sound source localization system based on TDOA algorithm,” Advances in Transdisciplinary Engineering, 2024.

[11] B. D. Van Veen and K. M. Buckley, “Beamforming: A versatile approach to spatial filtering,” IEEE ASSP Magazine, vol. 5, no. 2, pp. 4–24, 1988.

[12] H. Wang and M. Kaveh, “Optimization of MUSIC algorithm for AOA estimation,” Journal of Systems Engineering and Electronics, 2013.

[13] M. Risoud et al., “Sound source localization,” Eur. Ann. Otorhinolaryngol. Head Neck Dis., 2018.

[14] T. Blanchard, J.-H. Thomas, and K. Raoof, “Acoustic localization estimation of an unmanned aerial vehicle using microphone array,” LAUM, Univ. du Mans, 2008.

[15] S. Grondin et al., “Microphone array processing for UAV noise reduction,” IEEE, 2017.

[16] H. Xing and X. Yang, “Sound source localization fusion algorithm and performance analysis of a three-plane five-element microphone array,” Applied Sciences, 2019.

[17] V. Pulkki, L. McCormack, and R. Gonzalez, “Superhuman spatial hearing technology for ultrasonic frequencies,” Scientific Reports, 2021.

[18] G. Jekaterynczuk and Z. Piotrowski, “A survey of sound source localization and detection methods and their applications,” Sensors, vol. 24, no. 1, 2024.

[19] M. Khalaf-Allah, “Time of arrival (TOA)-based direct location method,” Umm Al-Qura University.

[20] MathWorks, “Object tracking using time difference of arrival (TDOA),” MATLAB Documentation.

[21] K. C. Ho and Y. T. Chan, “Geolocation of a known altitude object from TDOA and FDOA measurements,” IEEE Trans. Aerosp. Electron. Syst., vol. 33, no. 3, pp. 770–783, 1997.

[22] J. Goldhirsh and W. J. Vogel, Handbook of Propagation Effects for Vehicular and Personal Mobile Satellite Systems. Wiley, 1998.

[23] T. S. Rappaport, Wireless Communications: Principles and Practice, 2nd ed. Prentice Hall, 2002.

[24] A. Dhabale, “Direction of arrival estimation using array signal processing,” M.S. thesis, UC Riverside, 2018.

[25] DSP Concepts, “Generalized cross-correlation direction of arrival,” Online.

[26] H. Tang, S. Nordebo, and P. Cijvat, “DOA estimation based on MUSIC algorithm,” 2014.

[27] R. Roy and T. Kailath, “ESPRIT—Estimation of signal parameters via rotational invariance techniques,” IEEE Trans. Acoust., Speech, Signal Process., 1989.

[28] I. Berrios, “Introduction to beamforming: DOA estimation,” Medium, 2024.

[29] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," Proc. IEEE, 1969.

[30] Y. Lu et al., "Improving DOA estimation using deep learning," Sensors, 2024.

[31] S. Y. Lee, J. Chang, and S. Lee, "Deep learning-enabled sound source localization," IEEE Trans. Instrum. Meas., 2022.

[32] R. Costa-Felix et al., CBEB 2018 Proceedings. Springer, 2019.

[33] Z. Zhou et al., "Algebraic solution for acoustic source localization without wave velocity," Sensors, 2021.

[34] P.-A. Grumiaux et al., "A survey of sound source localization with deep learning methods," J. Acoust. Soc. Am., 2022.

[35] K. Xu et al., "Deep learning-based sound source localization: A review," Applied Sciences, 2025.

[36] M. Li, "Near-field source localization using microphone arrays," IEEE Signal Processing Magazine, 2019.

[37] D. H. Johnson and D. E. Dudgeon, Array Signal Processing: Concepts and Techniques. Prentice Hall, 1993.

[38] P. Stoica and R. Moses, Spectral Analysis of Signals. Pearson, 2005.

[39] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag., 1986.

[40] G. Jekaterynczuk et al., "UaVirBASE dataset," Zenodo, 2025. https://zenodo.org/records/15391924