Department of Electrical Engineering
Indian Institute of Technology Delhi

Project report on

Localization of sound source

Supervisor - **Prof. Arun Kumar**

By:
Ravi Parihar (2021EE10156)

Date of Project commencement - 06^{th} **May** 2024
Date of Conclusion/Presentation - 05^{th} **August** 2024

Abstract

This report presents a study on the localization of sound sources using Time Difference of Arrival (TDOA) based approaches. The main objectives were to estimate the TDOA using various techniques, implement localization algorithms, and analyze the performance under different conditions. The findings are based on extensive simulation and analysis using MATLAB.

Contents

List of Figures

1 Introduction

Localization of sound source using Time Difference of Arrival (TDOA) based methods involves capturing sound signals through multiple microphones and computing the time delay between signals. This time delay is then used to estimate the location of the sound source.

1.1 Background and Motivation

The localization of sound sources is a critical task in various fields such as surveillance, robotics, and communication systems. This project focuses on TDOA-based approaches due to their effectiveness in estimating the position of a sound source.

1.2 Objectives

The primary objectives of this project are:

- To estimate the TDOA using GCCPHAT methods.
- To implement localization techniques and analyze their performance.
- To evaluate the impact of factors like sampling frequency, noise, and microphone configuration on localization accuracy.

2 Theoretical viewpoint

Time Difference of Arrival (TDOA) based approaches are widely used in sound source localization. These methods rely on measuring the time differences at which a sound signal arrives at multiple spatially separated sensors. By calculating the difference in arrival times, the relative distance between the source and sensors can be estimated, allowing for triangulation of the sound source's location. TDOA is commonly used in applications like acoustic surveillance, robotics, and audio signal processing. The accuracy of these methods depends on factors such as the precision of time measurements, sensor placement, and environmental conditions like noise and reverberation.

3 Signal analysis

`Code for spectrum and waveform is given in file: ReadTDMS.m`

In this section, various signal properties were analyzed to understand their characteristics. The signal waveform was plotted to visualize the raw data over time. Spectral analysis was conducted to examine the frequency content, with spectrums plotted to reveal dominant frequencies. The power spectral density was estimated using the Welch method, providing insight into the signal's power distribution across frequencies. Additionally, spectrograms were generated to observe how the signal's frequency content evolves over time. These analyses form the foundation for accurately identifying and tracking drone acoustics.

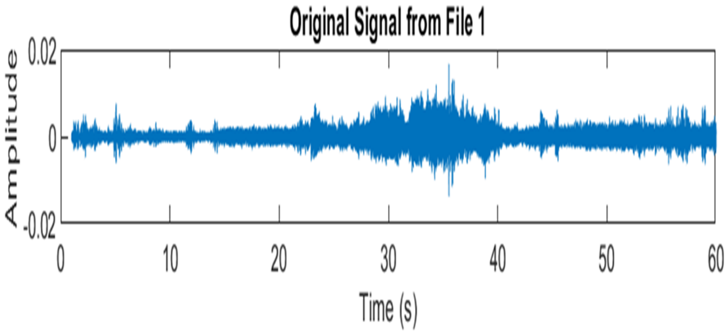

Figure 1: **Signal waveform**

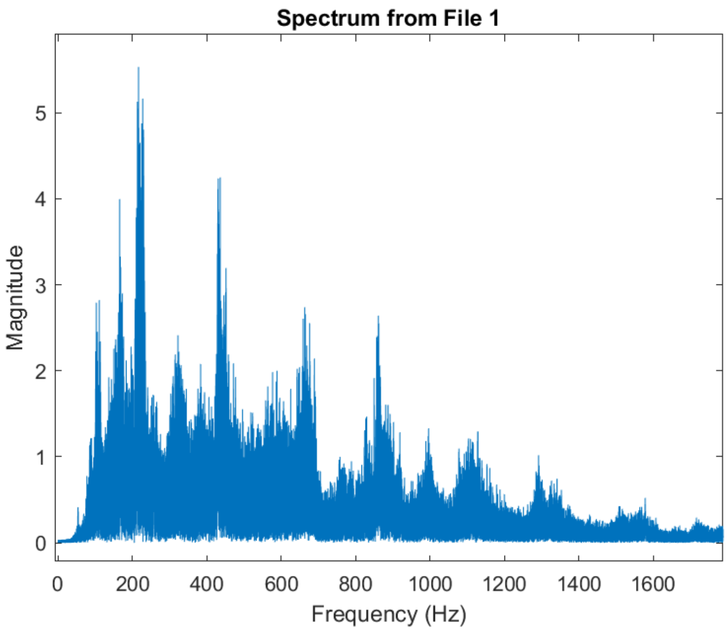

Figure 2: **Magnitude spectrum plot**

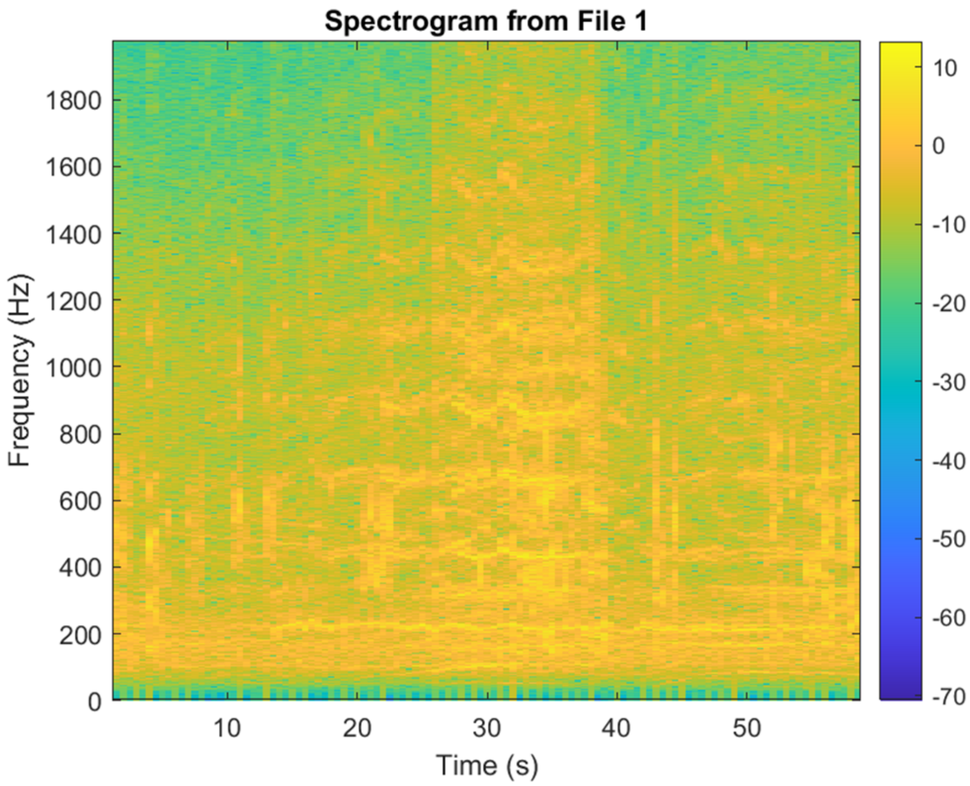

Figure 3: **Spactogram**

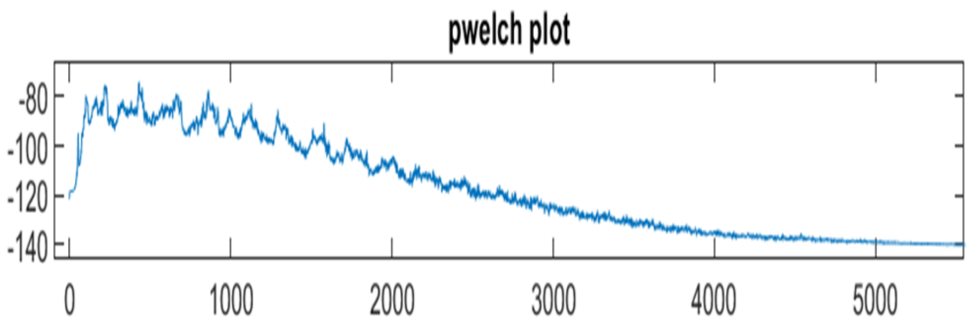

Figure 4: **pwelch plot**

4 Analysis the effect of sampling frequency on TDOA

Code file: TDOA_error_by_quantization.m

The sampling frequency significantly affects the accuracy of Time Difference of Arrival (TDOA) measurements for a sound signal recorded by two microphones. Higher sampling rates provide finer temporal resolution, enabling more precise determination of the time delay between the signals received by each microphone. This precision is crucial for accurate source localization. On the other hand, lower sampling frequencies can lead to aliasing and reduced accuracy in TDOA estimation, as the temporal details of the signal may be insufficiently captured. Choosing an appropriate sampling frequency is essential for reliable and accurate acoustic localization.

4.1 Analysis for 2 Microphones

Here two microphones are given in the y-axis at 1 m separation, and the source is varying its location from 5 m to 200 m on x axis.

- `mic1_position = [0, 0, 0];` % Mic 1 coordinates in m
- `mic2_position = [0, 1, 0];` % Mic 2 coordinate in m
- `Sampling frequency = 64000`

4.1.1 Up sampling the signal by a factor of 2

`fs = 64000*2 Hz;`

% varying the source position range (in m)

`source_min = 5;` % Minimum source distance in m (5 meters)

`source_max = 200;` % Maximum source distance in m (200 meters)

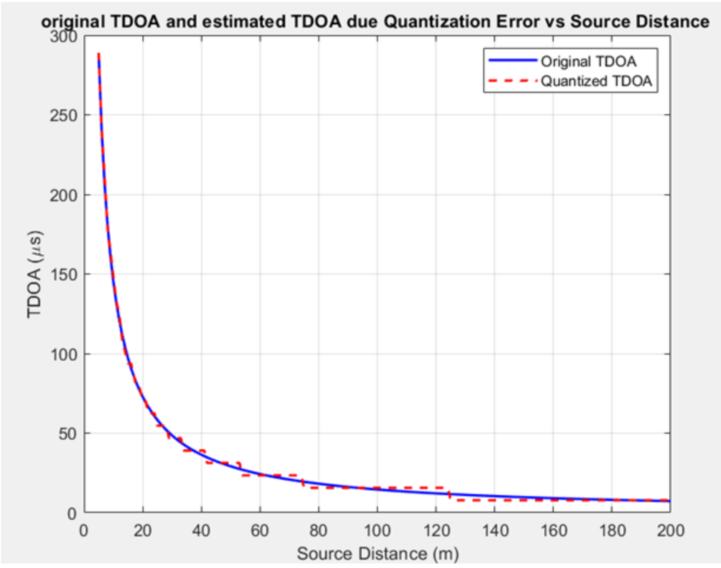

Figure 5: **Original TDOA and estimated TDOA due to upsampling the signal by factor 2 vs Source distance**

4.1.2 Upsampling by factor 4

`fs = 64000*4 Hz;`

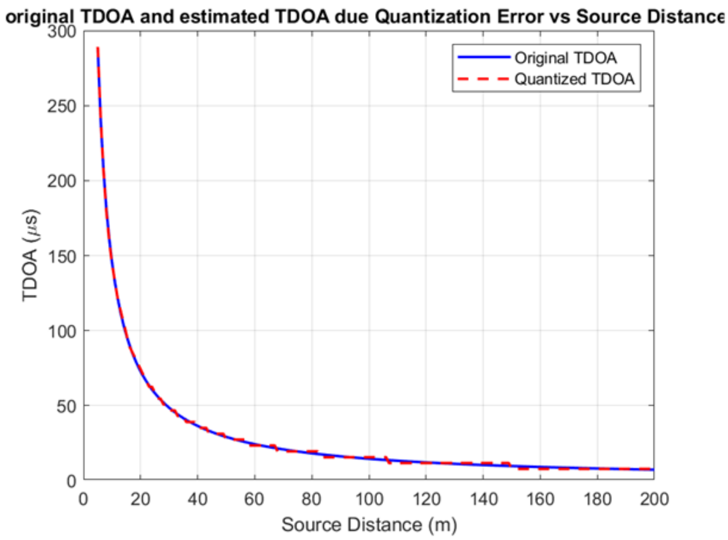

Figure 6: **Original TDOA and estimated TDOA due to upsampling the signal by factor 4 vs Source distance**

4.1.3 Upsampling by factor 8

```
fs = 64000*8 Hz;
```

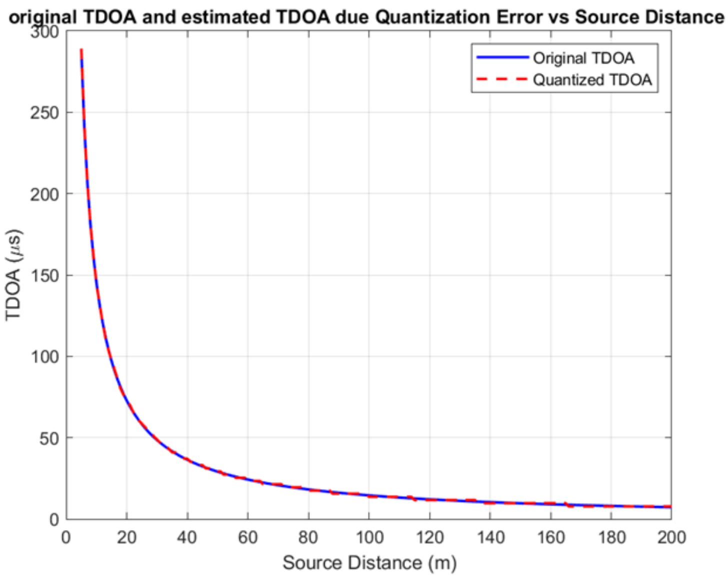

Figure 7: **Original TDOA and estimated TDOA due to upsampling the signal by factor 8 vs Source distance**

As sampling frequency increase(upsampling factor increase) error between the original TDOA and TDOA estimation from sampled signal (recorded signal) reduced.

4.2 Varying the distance between the microphones in 2 microphone system

`Upsampling by factor 8`
`fs = 64000*8 Hz;`
% varying the source position range (in m)
`source_min = 5;` % Minimum source distance in m (5 meters)
`source_max = 200;` % Maximum source distance in m (200 meters)

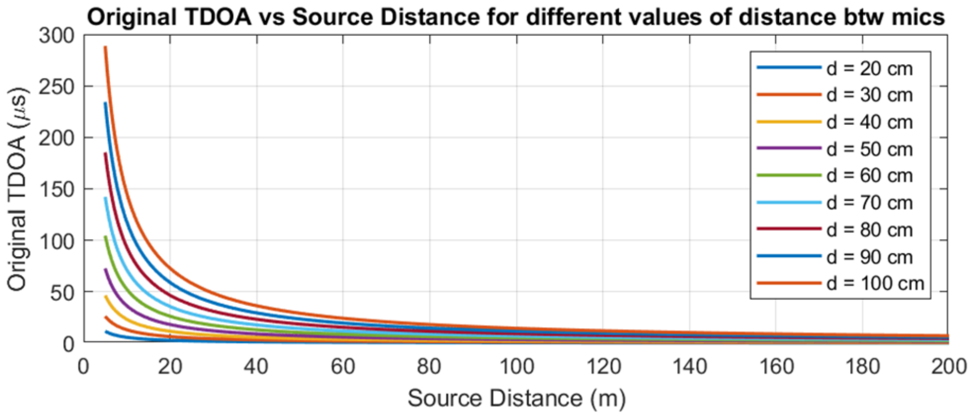

Figure 8: **Actual TDOA vs Source distance for different values of distance btw mics**

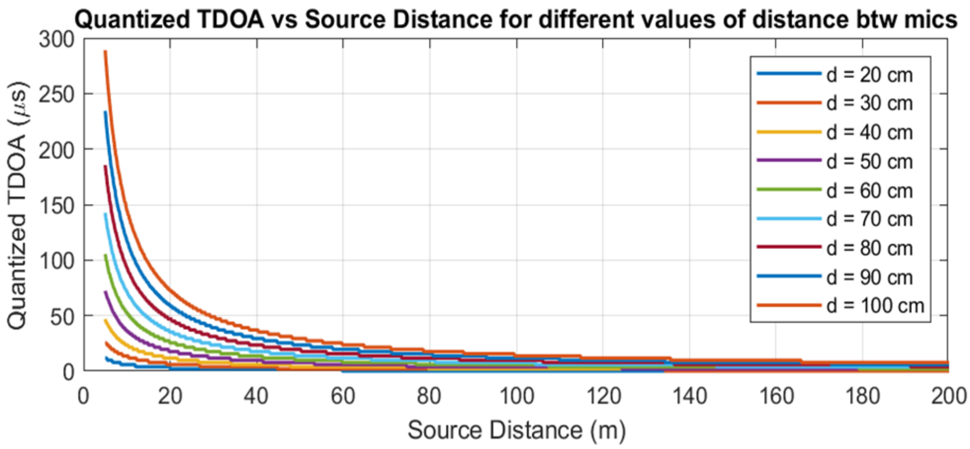

Figure 9: **Estimated TDOA vs Source distance for different values of distance btw mics**

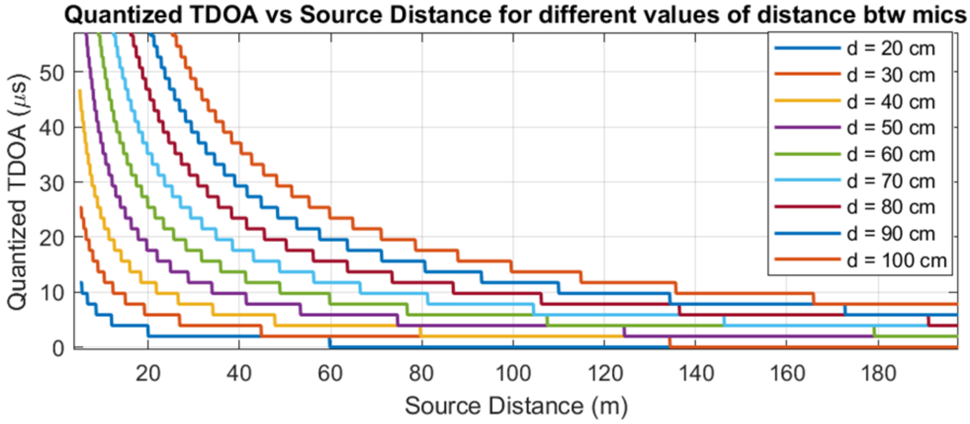

Figure 10: **Estimated TDOA vs Source distance for different values of distance btw mics**

5 Localization

5.1 implementation of a research paper

Code file: localization_as_given_in_paper.m

Paper link: A Surveillance System for Drone Localization and Tracking Using Acoustic Arrays - IEEE Conference Publication

first localization with 8 microphones (2 arrays, each array contains 4 microphones, there is one reference microphone considered from an array for each array)

```
% Define the microphone positions for both acoustic arrays
s1 = [0, 0, 0];
s2 = [0, 1, 0];
s3 = [1, 0, 0];
s4 = [0, 0, 1];
s5 = [0, 0, 5];
s6 = [0, 0, 6];
s7 = [-1, 0, 5];
s8 = [0, -1, 5];
% s1 is reference for first array, s5 is reference for second array

s = [40, 30, 110]; % original source location

Ts=1/(64000*8);

% Example TDOA values (replace these with actual TDOA calculations)
% TDOA = [t21, t31, t41, t65, t75, t85]; % Replace with actual TDOA dat
```

```
% Form the A matrix and B vector for least squares calculation

 A = 2 * [
     s2 - s1, c*t21, 0;
     s3 - s1, c*t31, 0;
     s4 - s1, c*t41, 0;
     s6 - s5, 0, c*t65;
     s7 - s5, 0, c*t75;
     s8 - s5, 0, c*t85
 ];


 B = [
     norm(s2)^2 - norm(s1)^2 - (c*t21)^2;
     norm(s3)^2 - norm(s1)^2 - (c*t31)^2;
     norm(s4)^2 - norm(s1)^2 - (c*t41)^2;
     norm(s6)^2 - norm(s5)^2 - (c*t65)^2;
     norm(s7)^2 - norm(s5)^2 - (c*t75)^2;
     norm(s8)^2 - norm(s5)^2 - (c*t85)^2
 ];

%  Solve for the drone's position using least squares
 XLos = (A' * A) \ (A' * B);

% Extract the position coordinates
% drone_position = XLos(1:3);
```

5.1.1 Observations for above microphones positions

for sampling frequency 64000*8 Hz

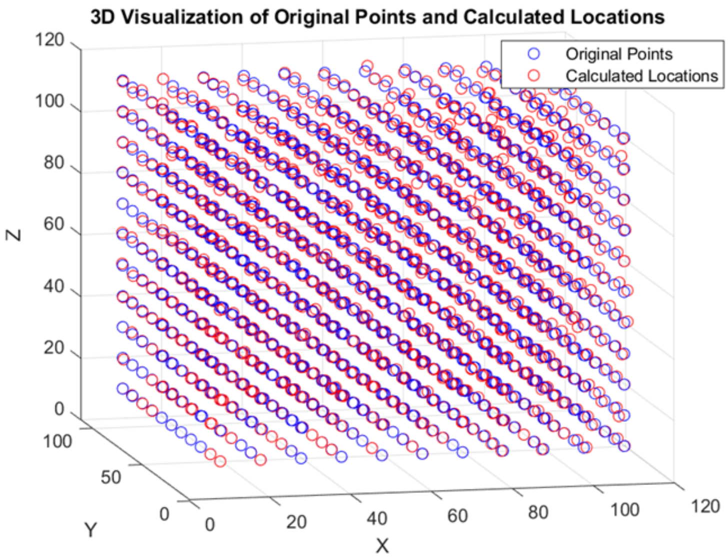

Figure 11: **3D visualization of original location of source vs estimated location**

`accuracy for x=linspace(10,50,41); y=linspace(10,50,41); z=40;`

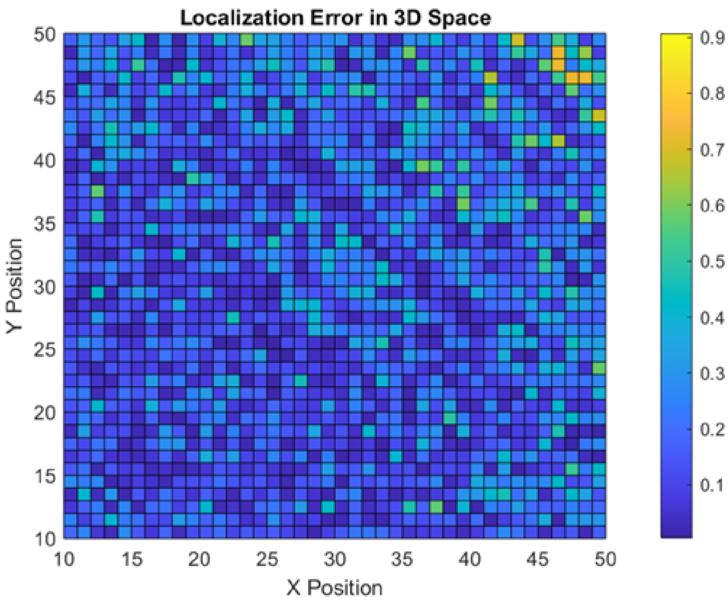

Figure 12: **Localization error in 3D space for** $x \in (10, 50)$**,** $y \in (10, 50)$**, and** $z = 40$

As we can see, the error in the estimated location from the actual location is less than 1 m for this model for source location coordinates $x \in (10, 50)$, $y \in (10, 50)$, and $z = 40$.

`Accuracy for x=linspace(10,50,41); y=linspace(10,50,41); z=40;`

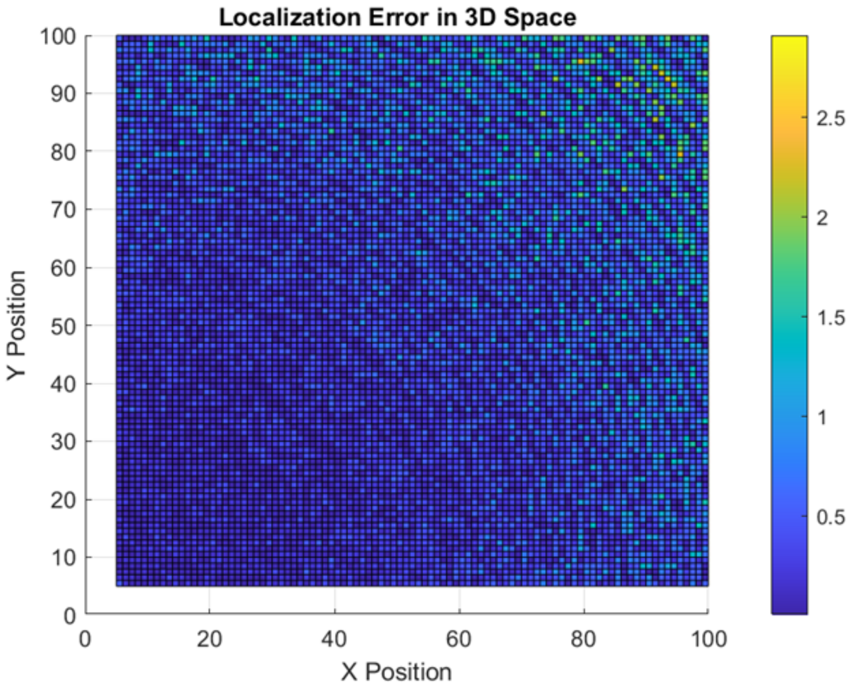

Figure 13: **Localization error in 3D space for $x \in (10, 100)$, $y \in (10, 100)$, and $z = 40$**

```
Accuracy for x=linspace(5,500,100); y=linspace(5,500,100); z=40;
```

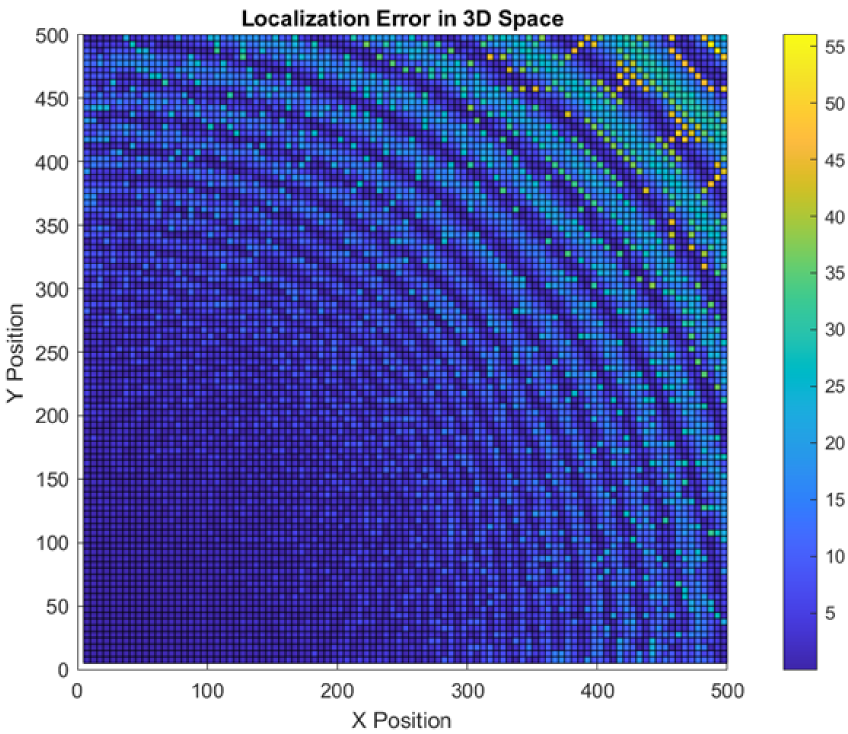

Figure 14: **Localization error in 3D space for $x \in (10, 500)$, $y \in (10, 500)$, and $z = 40$**

5.1.2 Observations for reducing size of each microphone array

```
% Define the microphone positions for both acoustic arrays
s1 = [0, 0, 0];
s2 = [0, 0.5, 0];
s3 = [0.5, 0, 0];
s4 = [0, 0, 0.5];
s5 = [0, 0, 5];
s6 = [0, 0, 5.5];
s7 = [-0.5, 0, 5];
s8 = [0, -0.5, 5];
```

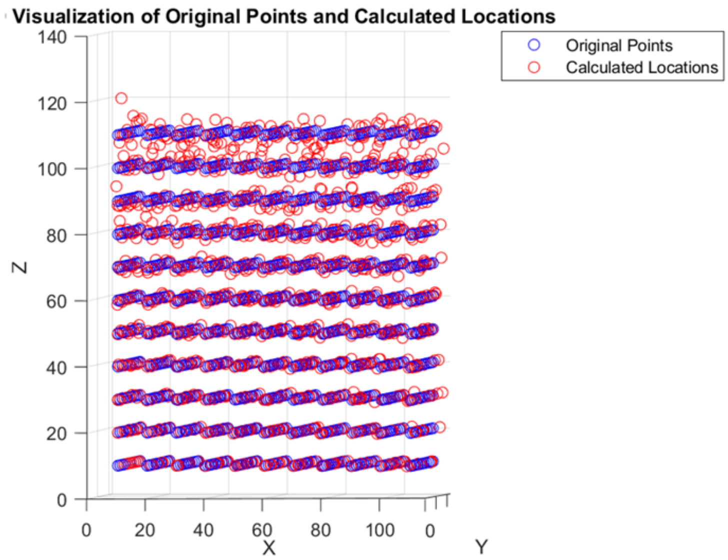

Figure 15: **3D visualization of original location of source vs estimated location**

```
accuracy for x=linspace(10,100,41); y=linspace(10,100,41); z=40;
```

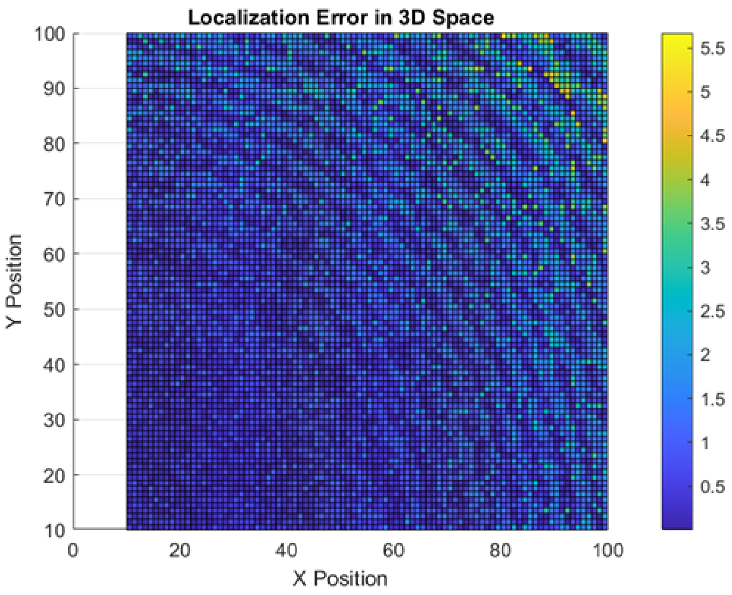

Figure 16: **Localization error in 3D space for $x \in (10, 100)$, $y \in (10, 100)$, and $z = 40$**

```
Accuracy for x=linspace(5,500,100); y=linspace(5,500,100); z=40;
```

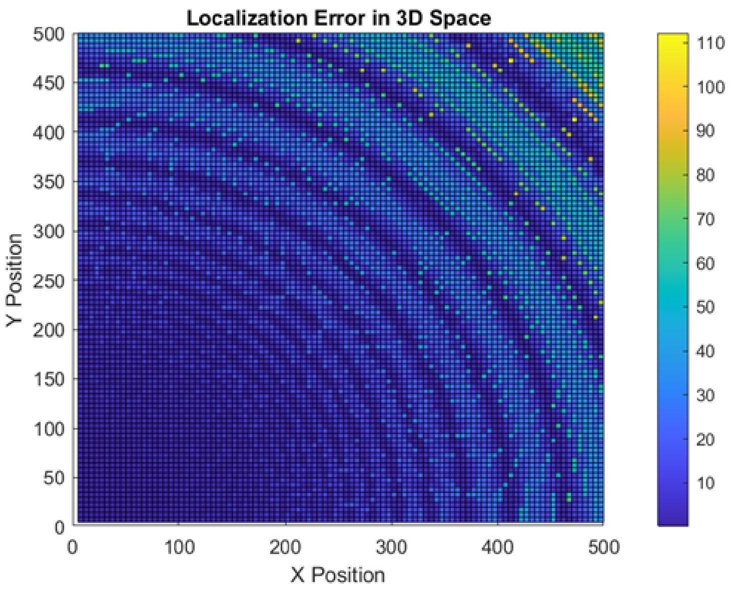

Figure 17: **Localization error in 3D space for $x \in (10, 500)$, $y \in (10, 500)$, and $z = 40$**

5.1.3 Observations for further reducing size of each microphone array

```
% Define the microphone positions for both acoustic arrays
s1 = [0, 0, 0];
```

```
s2 = [0, 0.3, 0];
s3 = [0.3, 0, 0];
s4 = [0, 0, 0.3];
s5 = [0, 0, 5];
s6 = [0, 0, 5.3];
s7 = [-0.3, 0, 5];
s8 = [0, -0.3, 5];
```

`accuracy for x=linspace(10,100,41); y=linspace(10,100,41); z=40;`

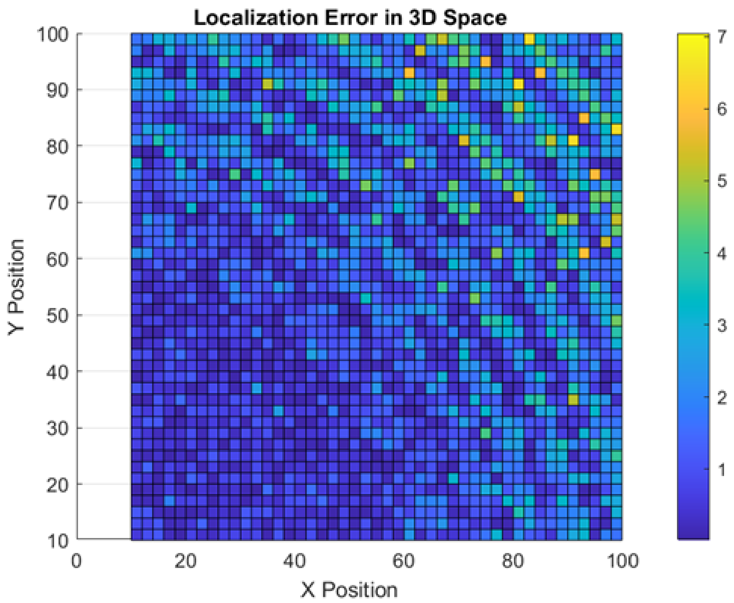

Figure 18: **Localization error in 3D space for $x \in (10, 100)$, $y \in (10, 100)$, and $z = 40$**

5.2 Solving least error with the help of optimization algorithms for lozalization

`code file: localization.m`

required toolbox: MATLAB optimization toolbox

$$t_i = \hat{t}_i + \epsilon_i, \quad i = 0, \ldots, N \tag{1}$$

Through multiplying these measurements by the propagation speed (c), the measured distance between the source and the ith sensor, D_i, can be defined as

$$D_i \equiv \|\mathbf{v} - \mathbf{s}_i\| + n_i = c(\hat{t}_i - t^*) + n_i, \quad n_i = c\epsilon_i, \quad i = 0, \ldots, N \tag{2}$$

$$d_{ij} \equiv D_i - D_j = \|\mathbf{y} - \mathbf{s}_i\| - \|\mathbf{y} - \mathbf{s}_j\| + n_{ij}, \tag{3}$$
$$n_{ij} = n_i - n_j$$

$$E = (\mathbf{e}^T)\mathbf{e}$$

Given the definition of e_i:

$$e_i = (\mathbf{y}^T\mathbf{s}_i) + d_i\|\mathbf{y}\| - 0.5(\|\mathbf{s}_i\|^2 - d_i^2)$$

This can be expressed in matrix form for all i:

$$\mathbf{e} = \mathbf{S}\mathbf{y} + \mathbf{d}\|\mathbf{y}\| - 0.5(\mathbf{c} - \mathbf{d}^2)$$

Where:

- $\mathbf{S}$ is a matrix with each row $\mathbf{s}_i^T$.
- $\mathbf{d}$ is a vector with elements d_i.
- $\mathbf{c}$ is a vector with elements $\|\mathbf{s}_i\|^2$.

The squared error E can be written as:

$$E = (\mathbf{S}\mathbf{y} + \mathbf{d}\|\mathbf{y}\| - 0.5(\mathbf{c} - \mathbf{d}^2))^T(\mathbf{S}\mathbf{y} + \mathbf{d}\|\mathbf{y}\| - 0.5(\mathbf{c} - \mathbf{d}^2))$$

```
% Define the coordinates of the ith microphone location (ai, bi, ci)
S = [
     1, 0, 0;
     0, 1, 0;
     0, 0, 1;
     1, 1, 1
];

s =[80,35.9,70.5]; % let actual  source location
% here reference microphone is placed at [0, 0, 0]

ts= 1/(64000*8);
v=343;
% Define vector d
d = [
    v*t10;
    v*t20;
```

```
    v*t30;
    v*t40
];

% Calculate vector c (squared norms of the rows of S)
c = sum(S.^2, 2); % N x 1 vector

% Define the objective function to minimize
objective = @(y) sum((S * y + d * norm(y) - 0.5 * (c - d.^2)).^2);

% Initial guess for y
y0 = rand(size(S, 2), 1);

% Use MATLAB’s optimization function to find y that minimizes
the objective function
options = optimoptions(’fminunc’, ’Algorithm’, ’quasi-newton’,
’Display’, ’iter’);
[y_opt, fval] = fminunc(objective, y0, options);
```

5.2.1 Observation for below microphones locations

```
% Now increasing the no. of microphones(si= [ai, bi, ci]):
microphones are equidistant in space
S = [
1, 0, 0;
0, 1, 0;
0, 0, 1;
-1, 0, 0;
0, -1, 0;
0, 0, -1
];

fs= 64000*8;
```

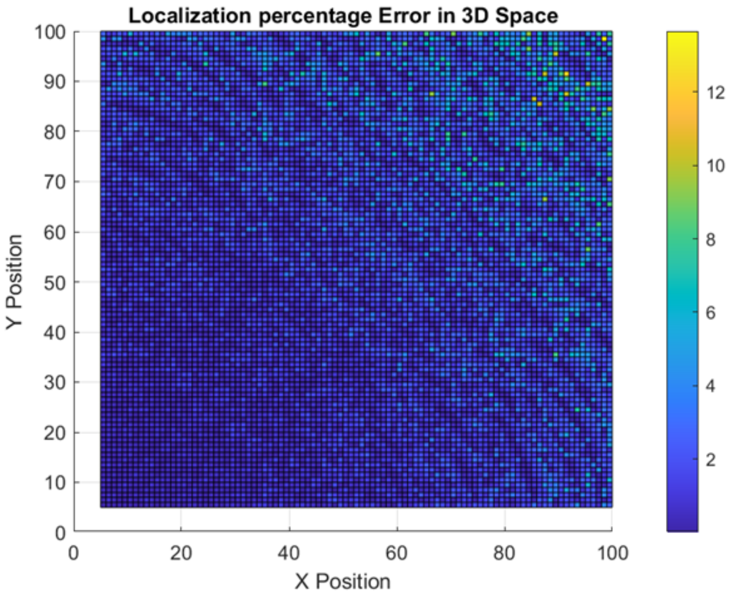

Figure 19: **Localization error in 3D space for** $x \in (10, 100)$**,** $y \in (10, 100)$**, and** $z = 35$

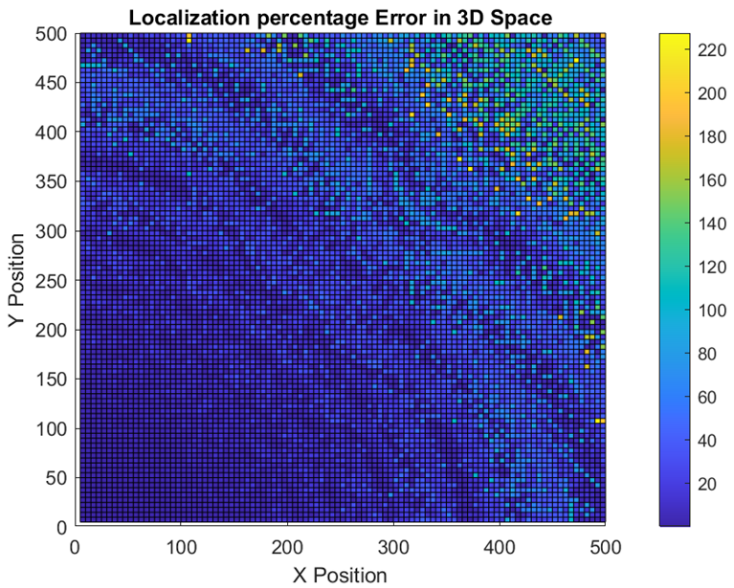

Figure 20: **Localization error in 3D space for** $x \in (10, 500)$**,** $y \in (10, 500)$**, and** $z = 35$

```
for fs= 64000*32;
```

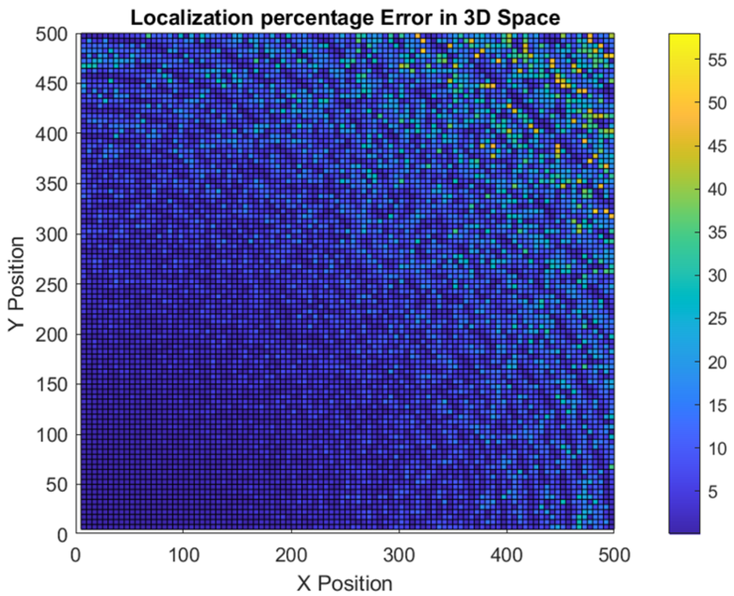

Figure 21: **Localization error in 3D space for $x \in (10, 500)$, $y \in (10, 500)$, and $z = 35$**

5.2.2 Observations for reducing the distance between the microphones

```
For microphones location:
S = [
0.7, 0, 0;
0, 0.7, 0;
0, 0, 0.7;
-0.7, 0, 0;
0, -0.7,0;
0, 0, -0.7];

for fs =64000*8 Hz
```

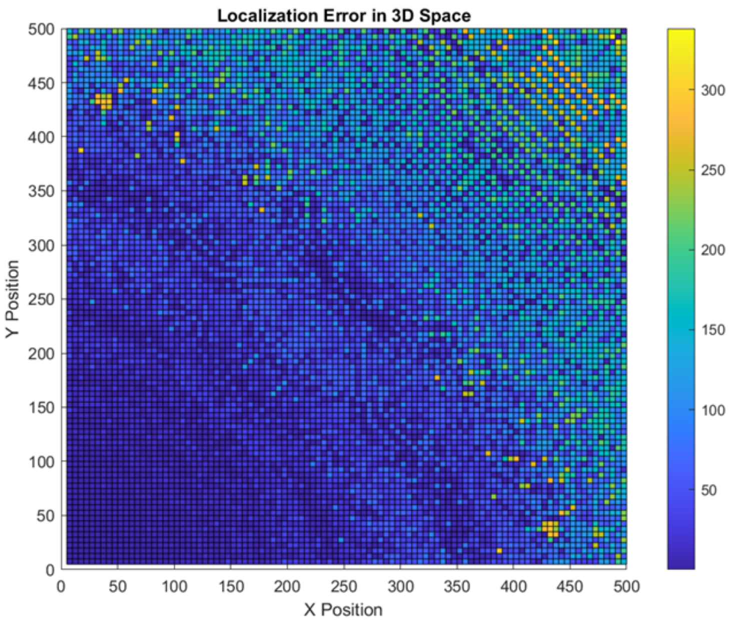

Figure 22: **Localization error in 3D space for** $x \in (10, 100)$**,** $y \in (10, 100)$**, and** $z = 35$

for fs =64000*32 Hz

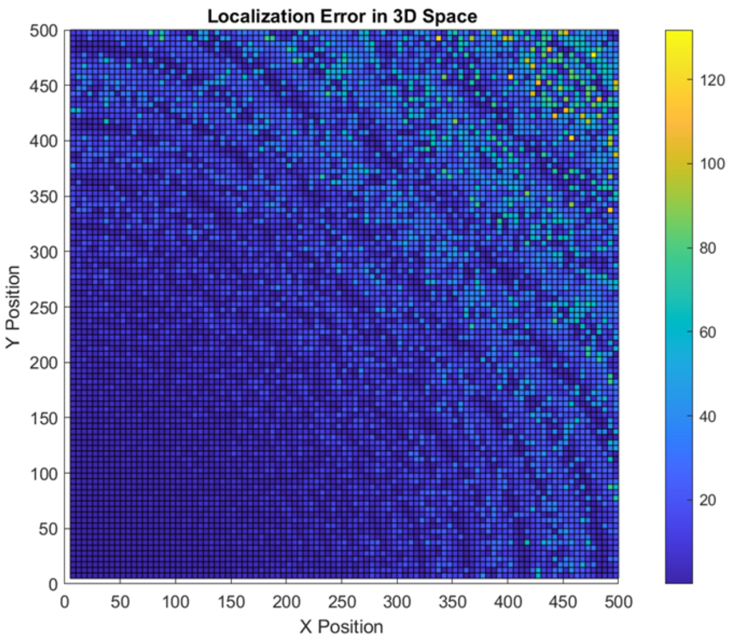

Figure 23: **Localization error in 3D space for** $x \in (10, 500)$**,** $y \in (10, 500)$**, and** $z = 35$

5.3 Localization with the help of Azimuthal and elevation angle calculated by Acoustic vector sensor

code file: localization.m

To find the localization of a sound source, we combined Time Difference of Arrival (TDOA) with angular information derived from an Acoustic Vector Sensor (AVS). The AVS provides the azimuthal angle (θ) and elevation angle (ϕ) of the sound source relative to the sensor array. These angles, combined with the TDOA between microphones located at the origin and 1 meter apart, enable the calculation of the source's distance from the origin.

The TDOA, Δt, is related to the difference in the sound source's distance from each microphone by:

$$\Delta t = \frac{d_2 - d_1}{c}$$

where d_1 and d_2 are the distances to the sound source from the two microphones, and c is the speed of sound.

Using the angles θ and ϕ obtained from the AVS, the position of the sound source in 3D space can be expressed as:

$$(x, y, z) = (r \cos\phi \cos\theta, r \cos\phi \sin\theta, r \sin\phi)$$

where r is the radial distance from the origin, calculated using the TDOA and the angular information.

5.3.1 Observation

```
fs = 64000*8;
for source range= linspace(10,100,91); theta = 45; phi=65;
```

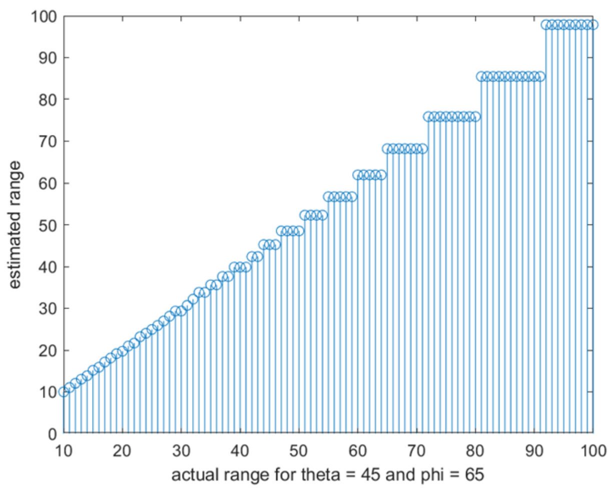

Figure 24: **Estimated range of source vs. Actual range of sound source for $\theta = 45°$ and $\phi = 65°$**

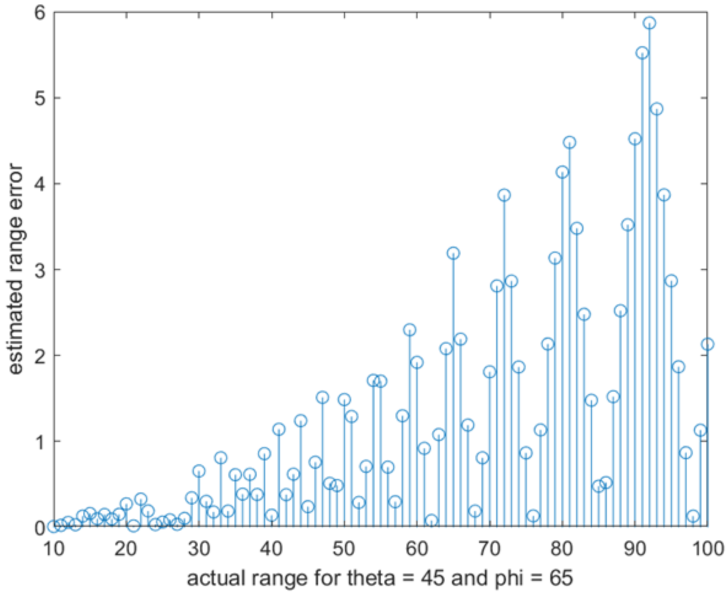

Figure 25: **Estimated range error vs. Actual range of sound source for $\theta = 45°$ and $\phi = 65°$**

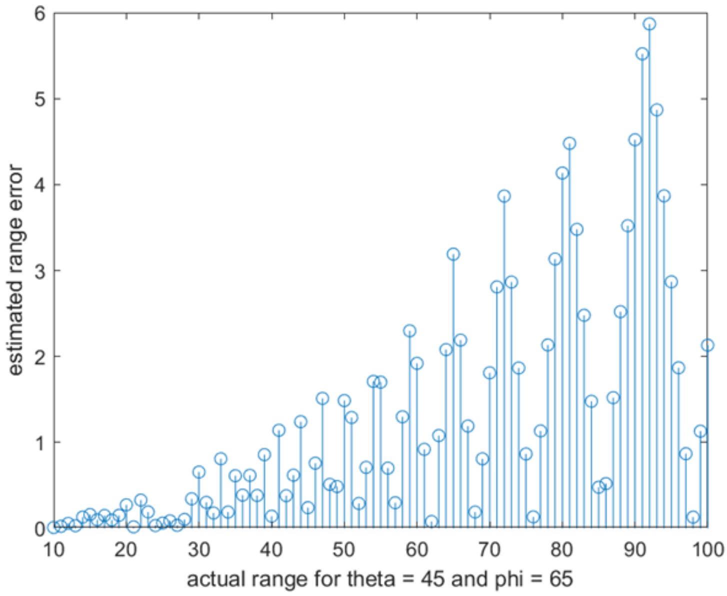

Figure 26: **Estimated range error vs. Actual range of sound source for $\theta = 45°$ and $\phi = 65°$**

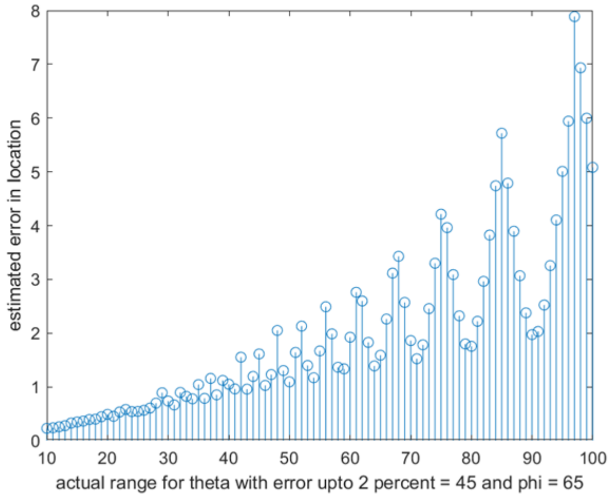

Figure 27: **Estimated range error vs. Actual range of sound source for $\theta = 45°$ and $\phi = 65°$ with random error up to 2 degree**

```
fs = 64000*8;
for source range= linspace(5,500,100); theta = 45; phi=65;
```

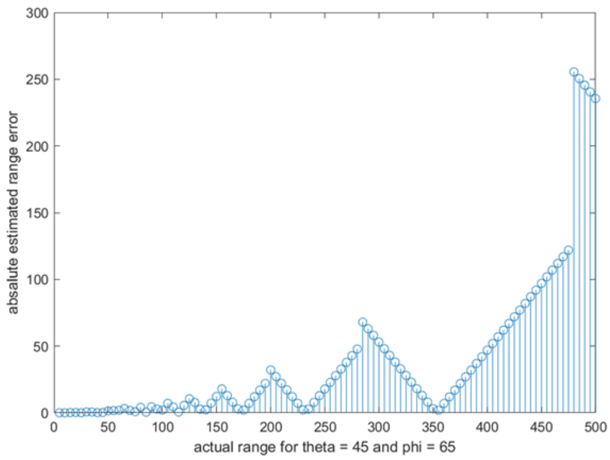

Figure 28: **Estimated range error vs. Actual range of sound source for $\theta = 45°$ and $\phi = 65°$ with random error up to 2 degree**

```
now fs = 64000*32 Hz;
for source range= linspace(5,500,100); theta = 45; phi=65;
```

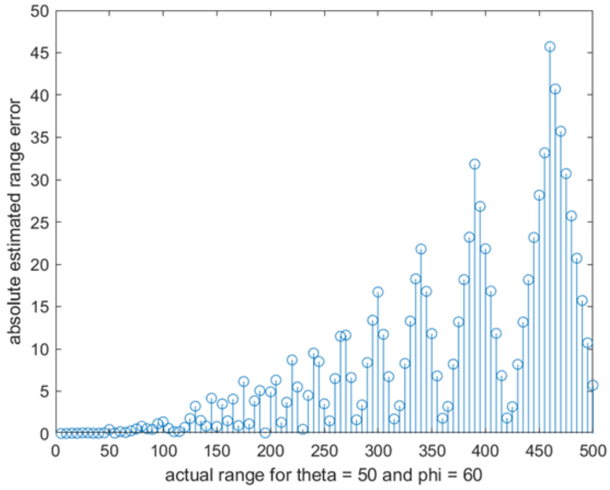

Figure 29: **Estimated range error vs. Actual range of sound source for $\theta = 45°$ and $\phi = 65°$**

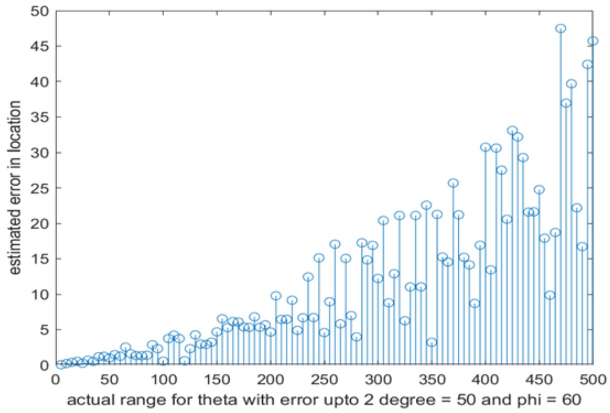

Figure 30: **Estimated range error vs. Actual range of sound source for $\theta = 45°$ and $\phi = 65°$ with random error up to 2 degree**

`now accuracy plot for x=linspace(10,100,46); y=linspace(10,100,46); z=35; ts= 1/(64000*8*4)`

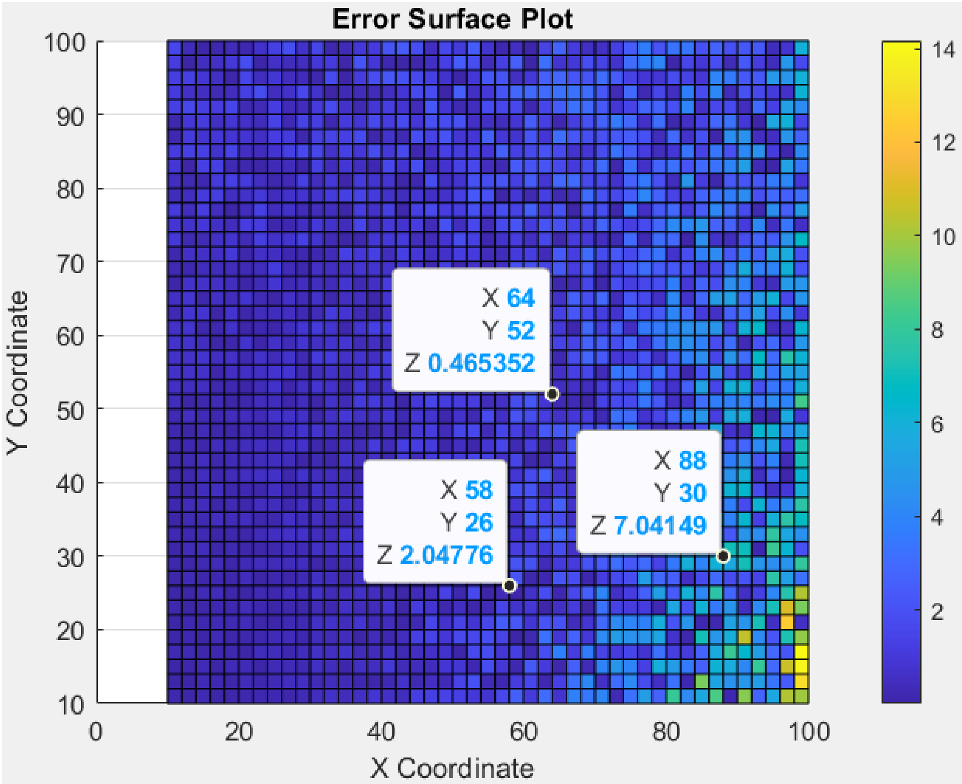

Figure 31: **Localization error in 3D space for** $x \in (10, 100)$**,** $y \in (10, 100)$**, and** $z = 35$

note: X,Y are coordinate of source, z coordinate of source is 35m. Z which is shown in plot is error in localization.

`now accuracy plot for x=linspace(10,100,46); y=linspace(10,100,46); z=35; ts= 1/(64000*8)`

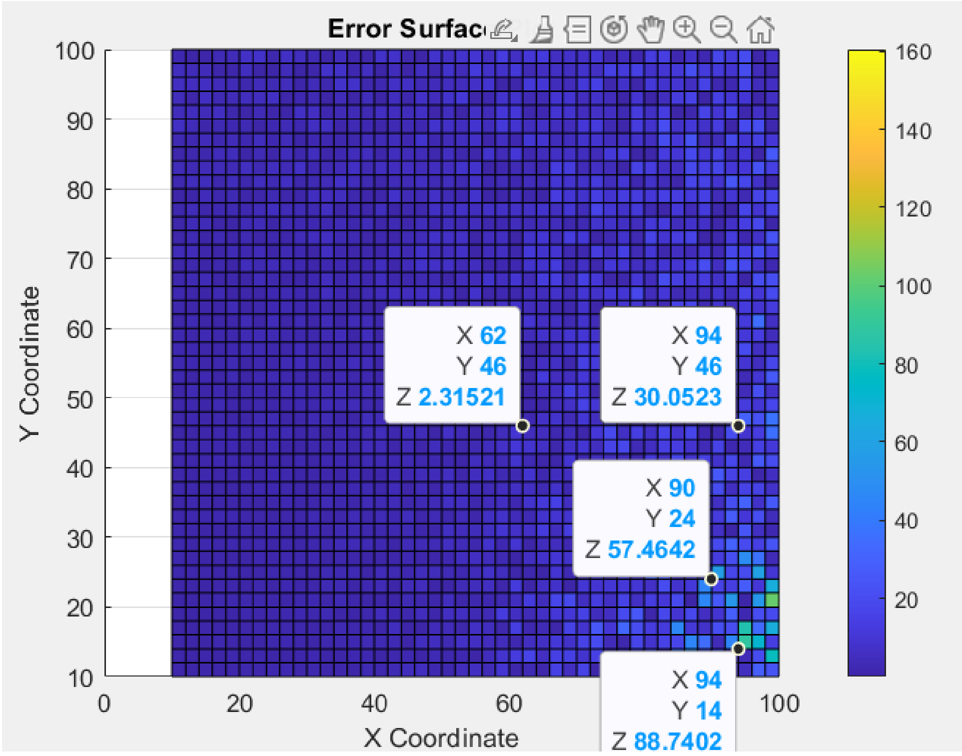

Figure 32: **Localization error in 3D space for** $x \in (10, 100)$**,** $y \in (10, 100)$**, and** $z = 35$

Results are not very good at the fs = 64000*8 Hz compared with the 6 microphone system of localization. At fs=64000*8*4 Hz its results are fine but not batter then 6 microphone system at same frequency for localization.

6 Time Difference of Arrival

`Code file: TDOA.m`

The Generalized Cross-Correlation with Phase Transform (GCC-PHAT) algorithm is a robust method for estimating the Time Difference of Arrival (TDOA) of a source signal between two sensors. By applying a phase transform to the cross-correlation function, GCC-PHAT enhances the peak corresponding to the TDOA, improving accuracy, especially in noisy environments. The method works by computing the inverse Fourier transform of the phase-corrected cross-power spectrum of the signals. This peak corresponds to the delay between the signals, which can then be used for precise localization of the sound source, making it ideal for real-time applications.

We have analyzed the performance of GCCPHAT to finding the TDOA of source sound on different SNR values. for this perpose we took drone sound signal recorded by one microphone, another signal is delayed versions of this signal. added noise according to required SNR value and then tested the GCCPHAT algorithm to find the accuracy of the estimated TDOA.

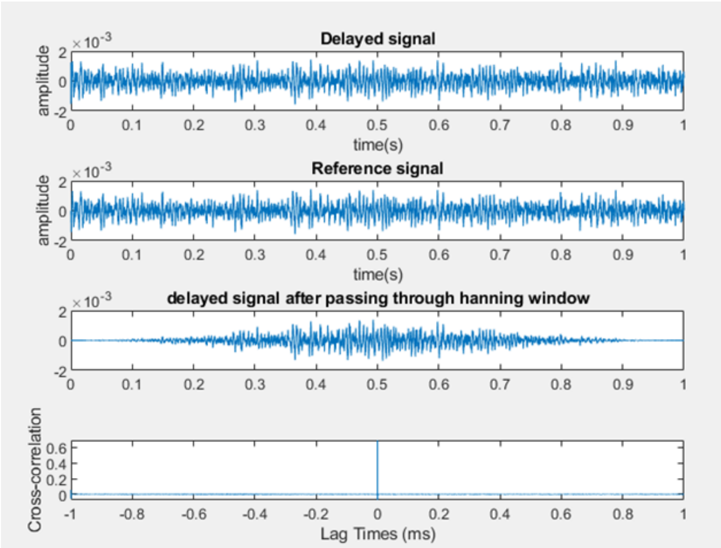

Figure 33: **Signal, reference signal and cross correlation plot**

For analysing the effect noise on finding TDOA, we added noise to signal on different SNR. For that perpose first generating the noise in it's frequency domain of range up to our required band limit then convert it into time domain using inverse Fourier transform. and then arrange according to our required SNR value and then add in signal.

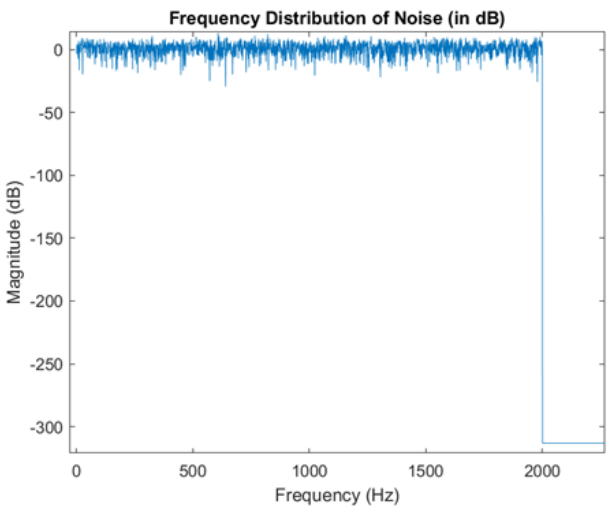

Figure 34: **frequency distribution of Noise**

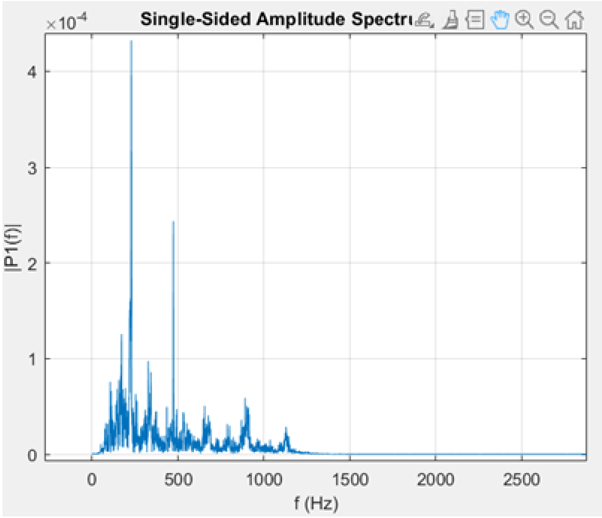

Figure 35: **frequency distribution of Signal**

now for a given delay of 15 samples from the reference signal to the delayed signal (our signal sample size is 512000 samples/ sample recorded in 1 second)

6.1 Observations

6.1.1 SNR is 10db for reference and delayed signals

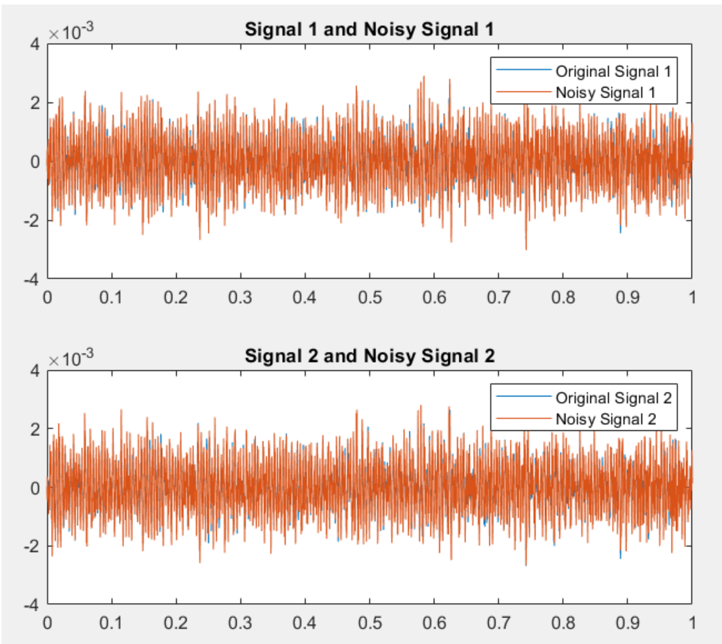

Figure 36: **signal 1 and signal 2**

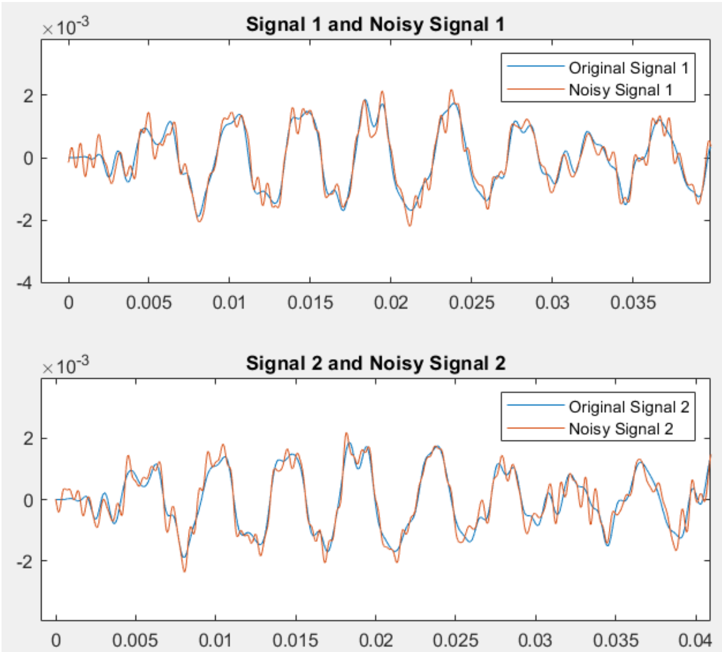

Figure 37: **signal 1 and signal 2 before and after addition of noise**

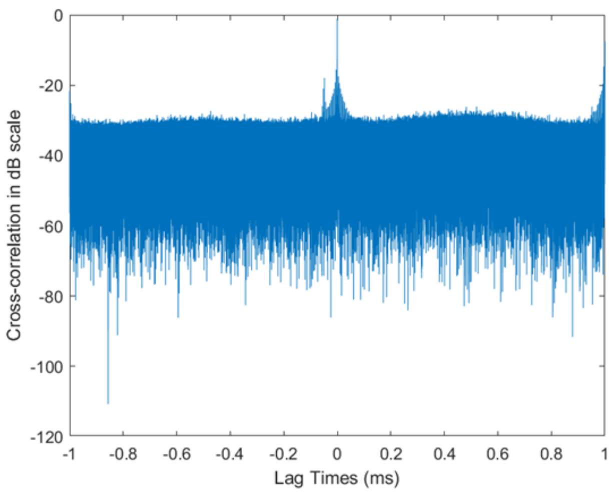

Figure 38: **Cross correlation plot in dB scale followed by GC-CPHAT**

6.1.2 SNR is -10db for reference and delayed signals

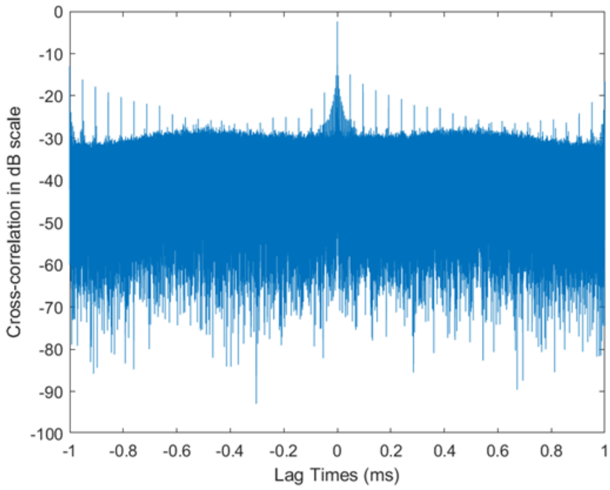

Figure 39: **Cross correlation plot in dB scale followed by GC-CPHAT**

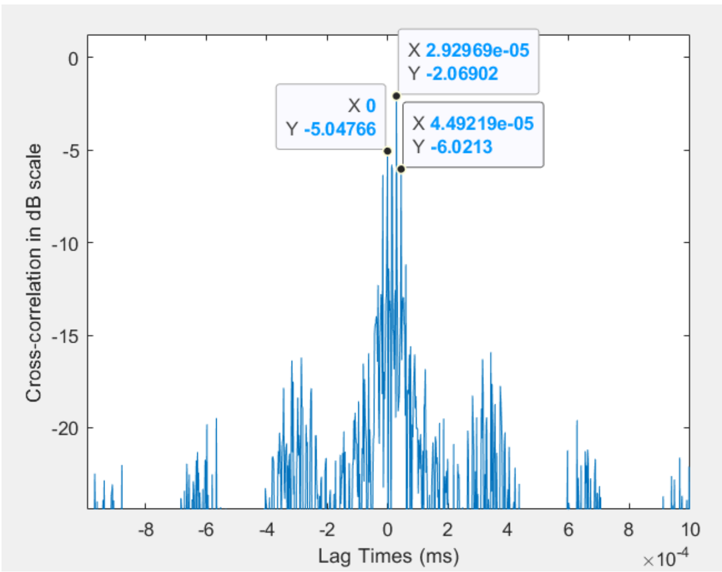

Figure 40: **Close observation of Cross-correlation plot**

6.2 Results

TDOA between the reference signal and delayed signal is calculated accurately up to -20 dB SNR values for 1 sec. sample of drone recording.
As the sample size is reduced, the estimated TDOA accuracy as we add noise will reduced. when 0.5 sec. of sample size is taken TDOA calculated accurately up to -5dB. it calculates the correct TDOA up to -10 dB but as we retest it, we sometimes get other values of TDOA. peaks other than

the actual main peak overwhelm at low SNR value when sample size is reduced.

7 Conclusions of Different Methods for Localization and Error Estimation

In this study, we explored various methods for localization and error estimation using different microphone configurations. Our findings indicate that the first method, based on the approach outlined in the referenced research paper, provided the most accurate results. This method utilizes two microphone arrays, each consisting of four microphones, with one microphone in each array designated as the reference.

In the second method, six microphones were arranged equidistantly in space from a central point. The accuracy of this method was highly dependent on both the sampling frequency and the separation distance between the microphones. The best results were achieved with a sampling frequency of fs=64,000×32 Hz, and when the microphones were positioned equidistantly in a space with a 1-meter radius.

Lastly, we evaluated localization using Acoustic Vector Sensors (AVS). when we checked at some angle of arrival and the source range varied, the results were satisfactory at a sampling frequency of fs=64,000×32 Hz. However, when the source location varied on the xy-plane with some elevation, localization errors became apparent, as depicted in Figure 31.

References

[1] X. Chang, C. Yang, J. Wu, X. Shi and Z. Shi, "A Surveillance System for Drone Localization and Tracking Using Acoustic Arrays," 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), Sheffield, UK, 2018, pp. 573-577, doi: 10.1109/SAM.2018.8448409.
paper link

[2] Vahid Heidari, Mohsen Amidzade, Khosrow Sadeghi, Amir Mansour Pezeshk "Exact solutions of time difference of arrival source localisation based on semi-definite programming and Lagrange multiplier: complexity and performance analysis"
paper link

[3] Zhe Dong and Ming Yu, ”Research on TDOA based microphone array acoustic localization,” 2015 12th IEEE International Conference on Electronic Measurement and Instruments (ICEMI), Qingdao, 2015, pp. 1077-1081, doi: 10.1109/ICEMI.2015.7494388.
paper link

[4] A. K. Zad Tehrani, B. Makkiabadi, A. Parsayan and S. H. Hozhabr, ”Sound Source Localization Using Time Differences of Arrival; Euclidean Distance Matrices Based Approach,” 2018 9th International Symposium on Telecommunications (IST), Tehran, Iran, 2018, pp. 91-95, doi: 10.1109/ISTEL.2018.8661037.
paper link