

**Gunshot Sound Direction of Arival Classification and Localisation**

PBL Review Report – I

Submitted To:

Dr Sudhanshu Maurya

Associate Professor

Submitted By:  
Aditya Joshi, 22070521016

VI, A

Feb 2025

**Articulate problem statements and identify objectives**

The process of determining sound source location involves using multiple microphones for the purpose of sound localization. Multiple microphone systems require this function to function properly in detection systems, robotic control and automated sound processing. The problem requires a determination of sound source position and direction through analyzing time differences between sound waves detected at different microphones.

Objectives:

1. A system must calculate Time Difference of Arrival (TDOA) for the purpose.

The estimation of Direction of Arrival depends on measuring the time delays existing between multiple microphones.

2. Measurements must estimate Direction of Arrival (DOA).

The sound incidence angle will be determined through an arcsin formula computation.

3. Perform Localization:

The system uses mathematical formulas together with machine learning models to generate (X, Y) coordinates of the sound source.

4. Evaluate Performance:

The performance of both formula-based approaches and machine learning models should be compared during the evaluation process.

5. Optimize the Process:

The localization precision can be enhanced through advanced approaches which include GCC-PHAT (Generalized Cross-Correlation Phase Transform) or deep learning.

**Identify engineering systems/ tools, variables, and parameters to solve the problems**

The solution requires engineering tools in combination with algorithms to measure variables that lead to accurate sound localization abilities.

**Engineering Systems/Tools:**

1. Mathematical Models:

• Formula-based TDOA and DOA estimation.

• Hyperbolic equations for source localization.

2. Machine Learning-Based Approach:

• Neural Networks (CNN, LSTM, CRNN) for sound classification and localization.The system implements regression models to calculate both time difference of arrival and direction of arrival estimates.

**Hardware & Software Tools:**

• Microphone Arrays (for recording sound waves).

• Python Libraries: numpy, scipy, librosa, matplotlib, pandas, tensorflow/sklearn.

• DSP Algorithms: Fourier Transforms, Cross-Correlation, GCC-PHAT.

**Table 1: Variables and Parameters**

|  |  |
| --- | --- |
| **Variable** | **Description** |
| **TDOA (Δt)** | Time difference between microphones. |
| **DOA (θ)** | Estimated angle of arrival (degrees). |
| **Speed of Sound (c)** | 343 m/s (assumed in air). |
| **Microphone Distance (d)** | Known spacing between microphones. |
| **Waveform Features** | Amplitude, frequency, entropy, zero-crossing rate. |
| **X, Y Coordinates** | Estimated position of the source. |

**Identify existing processes/ solution methods for solving the problem**

1. The Time Difference of Arrival (TDOA)

The technique uses multilateration to calculate the gunshot location. The position of a gunshot is determined by TDOA which measures the time duration it takes sound to reach different microphones. The position of the source is calculated through the intersection of hyperbolic curves that result from time difference measurements. The technique functions well within structured settings yet its operation is negatively influenced by sound disturbances and room echoes.

1. When applying Steered Response Power (SRP)

Beamforming the system activates directional beams that search for signal power maximums to identify the source location. The SRP system performs a grid search across specific locations while detecting maximum power signals to establish its source position. The processing method depends on the continuous alignment of microphone array phases to determine source positions. The search over all possible locations in this robust method makes it computationally costly when operating in reverberant rooms.

1. Multiple Signal Classification (MUSIC)

MUSIC allows high-resolution spectral estimation by separating the received signal between signal and noise subspaces. The identification of source directions in the method occurs through detecting peaks in its pseudo-spectrum. The system needs more microphones than the number of sources to work well in multipath environments.

1. Estimation of Signal Parameters via Rotational Invariance (ESPRIT)

The computational efficiency of ESPRIT outmatches MUSIC since its spectral search process is non-exhaustive. The directional arrival detection uses signal subspace properties to determine angle measurements because of rotational invariance. The method produces precise results although it needs clear information about microphone array geometry during execution.

1. Feature-Based Machine Learning

Acoustic feature extraction using machine learning methods includes the capability to obtain characteristics like Mel Frequency Cepstral Coefficients (MFCCs), spectral centroid and time delay values. The extracted features serve as input data for Random Forest or SVM or XGBoost models to perform direction-of-arrival classification or regression analysis. Taking this method yields good performance under noisy conditions while needing extensive labeled data for model training.

6. Deep Learning with Convolutional Neural Networks (CNN)

The CNN analyzes spectrogram data of gunshot sounds to recognize spatial patterns existing in the signal. Direction of arrival angles form the output of a neural network that processes input spectrograms. With this method you can achieve robust performance in noisy conditions without needing human-designed features yet it needs vast amounts of labeled gunshot audio data.

7. Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM)

Time-based data dependencies in gunshot audio signals are captured by RNNs and LSTMs for continuous direction of arrival predictions. The analysis of sequential acoustic patterns succeeds through these models yet their training phase takes longer relative to CNNs while using substantial computational resources.

8. Hybrid Approaches

Through the unison of traditional signal processing techniques with machine learning methods one can achieve superior localization precision. The determined TDOA values can be processed through GCC-PHAT before being entered into a neural network for classification. The use of MUSIC pseudo-spectra serves as valuable features when implementing deep learning models. The implementation of hybrid approaches needs domain experts for both fields to achieve their enhanced performance.

**Table 2: Comparison of Different existing Methods**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Accuracy** | **Computational Cost** | **Robustness to Noise** | **Scalability** |
| **TDOA (Formula-Based)** | Medium | Low | Low | Low |
| **DOA (Formula-Based)** | Medium | Low | Medium | Medium |
| **Hyperbolic Localization** | High | Medium | Medium | Medium |
| **CNN-Based DOA** | High | High | High | High |
| **CRNN-Based TDOA** | Very High | Very High | Very High | Medium |
| **GCC-PHAT + Deep Learning** | Very High | High | High | High |

**Compare and contrast alternative solution processes to select the best process**

**Dataset and Features**

The training process of an ML model requires proper organization of its dataset. A synthetic dataset makes up the recording collection where different microphones have documented gunshot sounds. A synthetic dataset became our selection because we didn't have the necessary real world dataset to perform localisation. The recording features consist of:

**Input Features (X)**

1. The time delay as sound travels between different recording devices serves as TDOA (Time Difference of Arrival).

2. Audio zero crossing rate describes how frequently the signal changes from positive to negative values while indicating frequencies present within the signal.

3. Spectral Features:

The Spectral Centroid functions as the spectrum's center weight which indicates heard pitch levels. The signal contains frequencies within the spectrum bandwidth which defines its range of present frequencies.

* Spectral Contrast – Difference between peaks and valleys in the spectrum.
* Spectral Rolloff quantifies the area of lower frequency where the signal maintains a certain amount of energy content.

4. Energy-Based Features:

* Entropy Energy – A measure of the distribution of energy in the frequency domain.
* Short-Time Energy – The sum of squared signal amplitudes over short time windows.

5. MFCCs (Mel-Frequency Cepstral Coefficients) [MFCC\_1, MFCC\_2, ..., MFCC\_5] Compact representation of spectral properties.

**Target Variables (y)**

1. True\_X (X-coordinate of the gunshot location defines the target value.

2. The Y-coordinate measurement of the gunshot location is True\_Y.

The chosen approaches from all possible solutions include 2 specific methods.Out of all the above mentioned approaches we decided to use 2 Approches :

1. **ML-Based Approach for Gunshot Sound Direction of Arrival (DOA) and Localization**
2. **Problem Definition**

The solution detects Direction of Arrival as well as X and Y spatial coordinates of gunshot sounds by combining machine learning models with acoustic features derived from multiple microphone arrays.

1. **Dataset and Features**

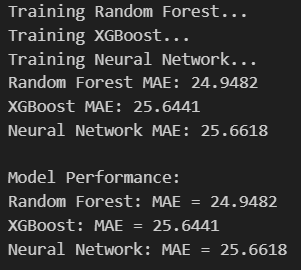
A gunshot database consists of TDOA measurements together with spectral features (spectral centroid, bandwidth, contrast, rolloff) and energy-based features (entropy energy, short-time energy) and MFCCs features. The prediction targets include the physical X and Y positions of the gunshot.

1. **ML Models for Localization**

Three predictive models comprise of Random Forest and XGBoost and a Neural Network (MLP Regressor) used to forecast gunshot positions. These models effectively deal with data non-linearities along with providing accurate predictions for real-world applications.

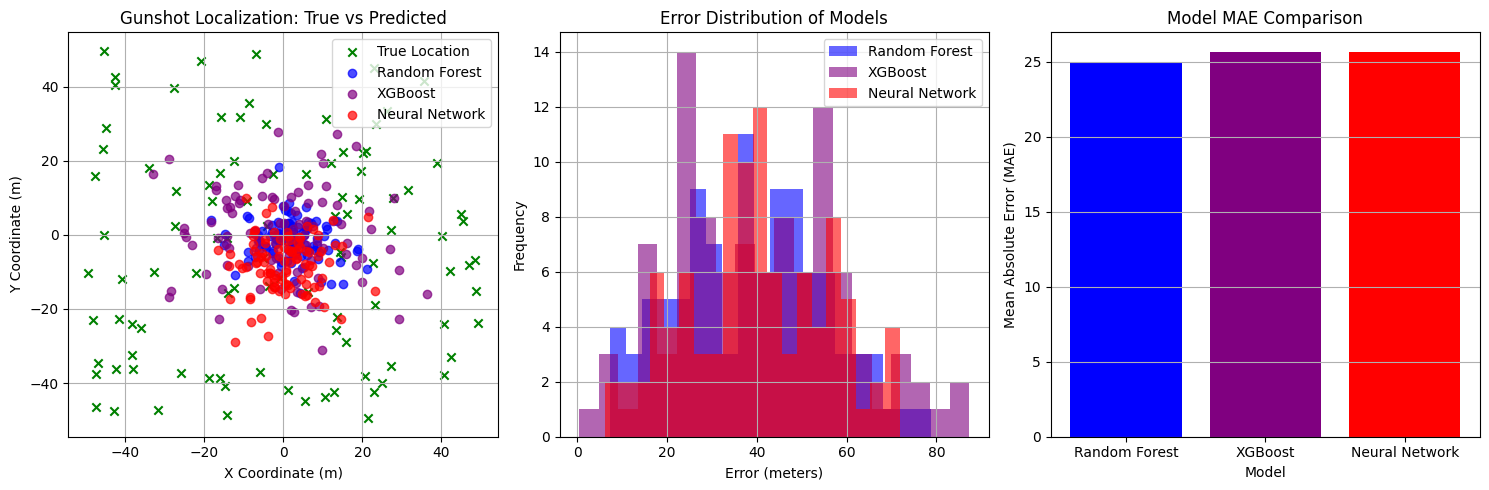
1. **Model Training and Evaluation**

To prepare the dataset it was divided into a training set (80%) together with a separate testing set (20%). The prediction models receive quantified features during training before Mean Absolute Error (MAE) evaluates their localization precision. Statistical error patterns serve to evaluate how well the models function.



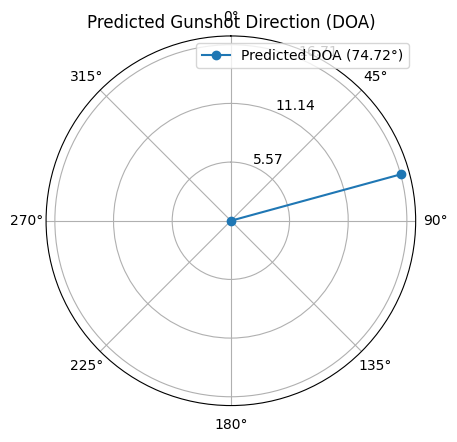
1. **Visualization of Predictions**

True and predicted gunshot positions get displayed through scatter plots along with error information displayed using histograms. The accuracy level of different models in predicting gunshot locations becomes clearer through this method.



1. **Direction of Arrival (DOA) Estimation**

The Euclidean norm and arctan2 function calculate DOA by generating an angle measurement relative to the microphone position. Polar plots display the direction prediction while providing shooters valuable information about the orientation of the gunshots.



Predicted Gunshot Coordinates: X = 4.40, Y = 16.12

Estimated DOA: 74.72°

1. **Geolocation Mapping**

The (X, Y) coordinates from prediction go through a conversion process to become real-world latitude and longitude values following the GPS position of the microphone. The system allows deployment for practical use by law enforcement and public safety agencies.

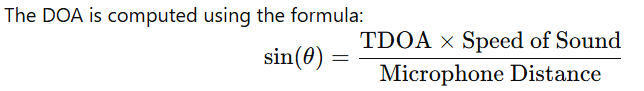


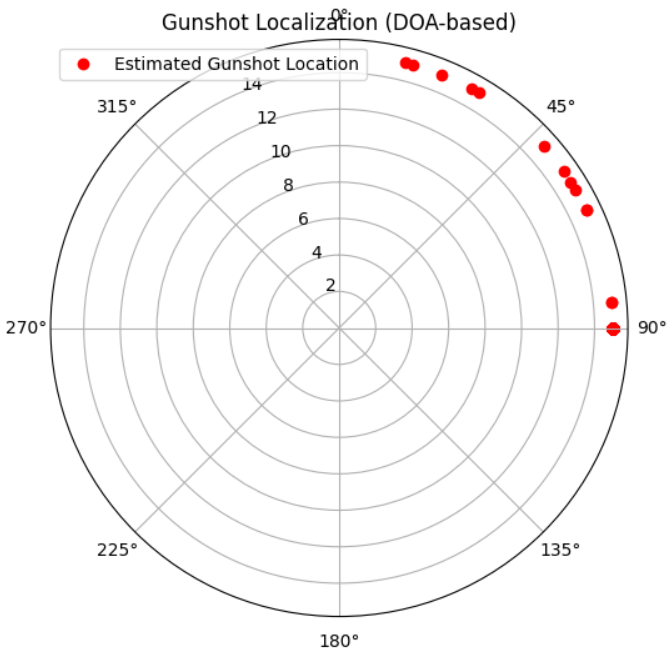
1. **Formula-Based Approach for Gunshot Sound Direction of Arrival (DOA) and Localization**
2. **Problem Definition**

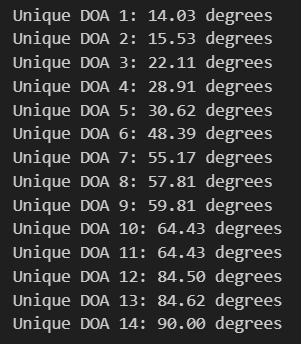
The formula-based method obtains Direction of Arrival (DOA) data alongside spatial X,Y location data for gunshot sounds using mathematical acoustic and geometric procedures.

1. **Time Difference of Arrival (TDOA) and DOA Computation**

The system uses Time Difference of Arrival measurements to compute Direction of Arrival values. The Time Difference of Arrival calculation between adjacent microphones enables detection of the gunshot angle.

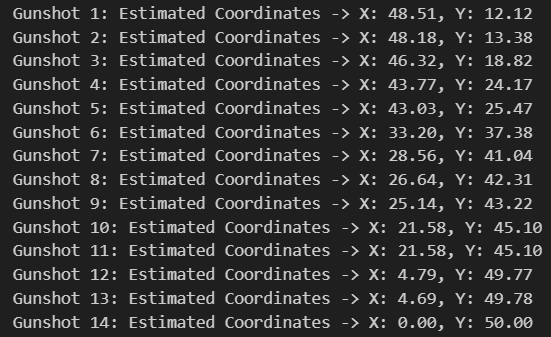






1. **Conversion to Cartesian Coordinates**

The estimated X and Y coordinates can be calculated by applying the formula when the sound came from a distance R within a specified area. 



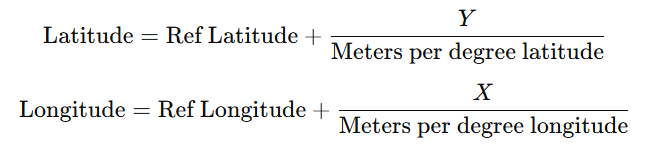
These computational values determine where the microphone detects the gunshot to be positioned from its location.

1. **Visualization of DOA and Localization**

* Graphic displays based on polar coordinates show how the gunshots travel in direction.
* The estimated and actual gunshot positions appear as points in scatter plots.

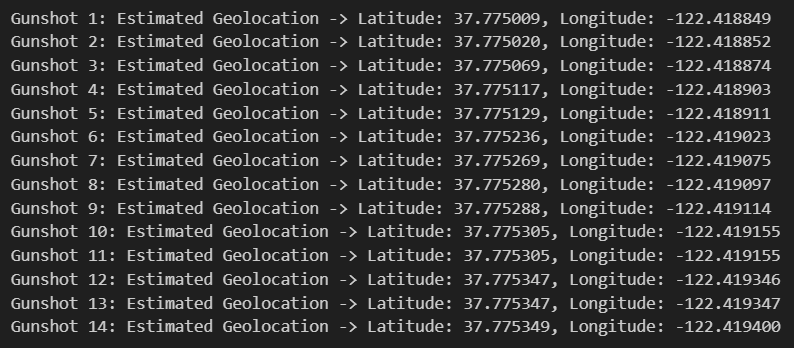
1. **Geolocation Mapping**

The program converts (X, Y) values into real-world location data through the GPS coordinates of the microphone unit.



The adjustment of distance measurements to degree longitude depends specifically on the reference latitude value.





**Advantages of Both Approaches:**

**1. Formula-Based Approach**

The system depends on acoustical and geometrical mathematical expressions to calculate the Direction of Arrival (DOA) and location of gunshot data.

**Advantages:**

* The system uses basic mathematical operations which give it quick processing capabilities and real-time operation potential.
* The approach needs no training because it requires no large dataset for its operation which decreases data collection needs.
* A theoretical framework founded on physics and trigonometry produces exact DOA estimation provided that precise TDOA and microphone spacing input values are given.
* This approach requires few hardware needs because it works efficiently even on devices which have small computational power.
* This method delivers exact outcome results from input parameters while protecting against two prevalent problems (overfitting or underfitting) which occur in ML model operations.
* Debugging processes become straightforward and result interpretations become simpler because computed output directly generates results from calculations.

1. **Machine Learning-Based Approach**

The method depends on ML models consisting of Random Forest and XGBoost and Neural Networks that detect gunshot location by processing acoustic features.

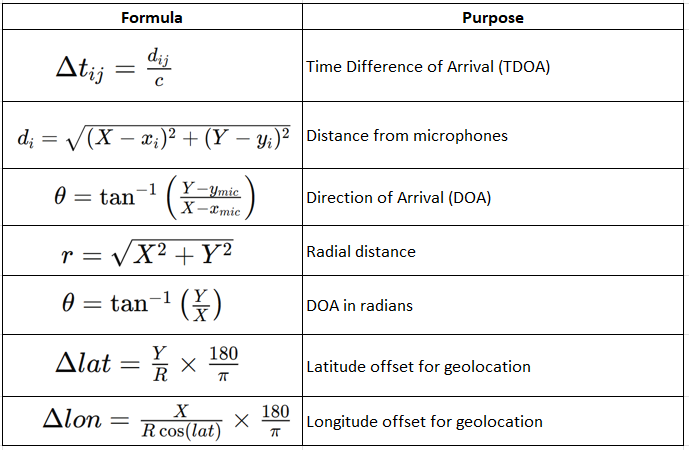
**Advantages:**

* This model can process complicated acoustic settings because it establishes non-linear associations between microphone inputs and gunfire positions thus achieving better accuracy in noisy urban environments.
* The model learns from actual world data and adapts to different environments because it uses ML models rather than formula-based approaches.
* ML models demonstrate improved performance for large-scale implementations because their trained algorithms maintain accuracy across various conditions even when microphones do not align properly or when acoustic reflections occur.
* The system leverages multiple characteristics which unite spectral energetic features with MFCC features to handle real-world operational environments better.
* The system learns to grow better when it uses expanded datasets because ML models advance steadily when compared with formula-based methods that have no dynamic capabilities.
* The technology can integrate with deep learning models including CNNs or LSTMs because of its capabilities to detect and classify multiple gunshot types.

**Comparison and Selection**

* Instances requiring rapid shot localization involving limited resources should implement the formula-based method instead of alternative systems.
* An ML-based system would be the best selection for achieving high accuracy levels and robust performance in complex conditions but it needs greater computational resources.
* A combination of these methods applied together creates a solution which provides fast results alongside learning adaptation capabilities for more effective practical gun position detection applications.

**Table 3:** Summary of Formulae Used:



**Read, understand, and interpret technical and non-technical information**

**Technical information**

* Sound localization requires technical information that uses mathematical models together with machine learning to calculate gunshot Direction of Arrival (DOA) and determine X and Y positions.
* Applying TDOA and DOA equations enables the calculation of sound source positions through microphone delay time comparisons.
* A machine learning framework extracts acoustic features (MFCCs, spectral features, energy-based features) to train models such as Random Forest, XGBoost and Neural Networks in order to determine gunshot position.
* Performance evaluation measures the accuracy through Mean Absolute Error (MAE) and it uses scatter plots for comparing predicted target locations against actual positions.

**Non- Technical information**

1. Environmental Noise:

* Background noise produces unpredictable results which decrease the accuracy of Time Difference of Arrival calculations.
* Solutions: GCC-PHAT improves noise robustness.

2. Reverberation and Echoes:

* The characteristics of indoor areas produce sound reflections that result in invalid TDOA calculations.
* Two approaches enable enhancement of accuracy in the localization system through beamforming applications combined with machine learning technology.

3. Microphone Synchronization Issues:

* A time lag between microphone devices produces inaccurate measurements of time delays.

GPS or high-precision clock synchronization achieves accurate measurements.

4. Deployment in Real-World Scenarios:

* Audio processing for law enforcement must operate simultaneously with real-time scanning demands.
* The speed advantage belongs to formula-based solutions yet ML detection models achieve better precision levels.

**Table 4: Approch Comparison**

|  |  |  |
| --- | --- | --- |
| **Aspect** | **Formula-Based Approach** | **Machine Learning-Based Approach** |
| **Computation** | Low, uses direct equations. | High, requires GPU for training. |
| **Data Requirement** | No dataset needed. | Requires large labeled datasets. |
| **Noise Robustness** | Low, affected by noise. | High, adapts to different environments. |
| **Scalability** | Limited to fixed mic arrays. | Can generalize to different setups. |

**References (If Any):**

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, 2nd ed. Cambridge, MA, USA: MIT Press, 1997.

[2] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3048–3058, Nov. 1998.

[3] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[4] M. Cobos, J. J. Lopez, and S. Spors, "A Survey of Sound Source Localization Methods in Wireless Acoustic Sensor Networks," *Appl. Acoust.*, vol. 139, pp. 64–83, 2018.

[5] "Acoustic Beamforming and Localization," *ScienceDirect*, available at: <https://www.sciencedirect.com/science/article/pii/S0003682X20300492> (accessed Feb. 2025).

[6] "BGG Gunshot Dataset," *Papers With Code*, available at: <https://paperswithcode.com/dataset/bgg-dataset> (accessed Feb. 2025).

[7] "UrbanSound8K Dataset," *Urban Sound Dataset*, available at: https://urbansounddataset.weebly.com/ (accessed Feb. 2025).

[8] Park, Junwoo & Cho, Youngwoo & Sim, Gyuhyeon & Lee, Hojoon & Choo, Jaegul. (2022). Enemy Spotted: in-game gun sound dataset for gunshot classification and localization. 10.48550/arXiv.2210.05917.

[9] J. Svatos and J. Holub, "Impulse Acoustic Event Detection, Classification, and Localization System," in *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-15, 2023, Art no. 6501515, doi: 10.1109/TIM.2023.3252631.

[10] Irungu, John & Ancel, Jamelia & Mahmoud, Wagdy & Denis, Max. (2023). Gunshot detection from audio excerpts of urban sounds using transfer learning. 045003. 10.1121/2.0001783.