॥वसुधैव कुटुम्बकम्॥
SYMBIOSIS INSTITUTE OF TECHNOLOGY, NAGPUR

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

Anvesh Khode, 22070521021

VI, A

Arya Kashikar, 22070521036

VI, A

Feb 2025

Articulate problem statements and identify objectives

Sound localization is the process of determining the location of a sound source using multiple microphones. This is crucial in gunshot detection, surveillance, robotics, and smart acoustic systems. The problem involves accurately estimating the direction (DOA) and position (localization) of the sound source based on the time difference of arrival (TDOA) of sound waves at different microphones.

Objectives:

1. **Estimate Time Difference of Arrival (TDOA):**

- Measure the time delays between multiple microphones to estimate the DOA.

2. **Determine the Direction of Arrival (DOA):**

- Compute the angle of incidence of sound using the arcsin-based formula.

3. **Perform Localization:**

- Use mathematical models or machine learning-based approaches to compute (X, Y) coordinates of the sound source.

4. **Evaluate Performance:**

- Compare formula-based approaches with machine learning models.

5. **Optimize the Process:**

- Improve the localization accuracy using advanced techniques like GCC-PHAT (Generalized Cross-Correlation Phase Transform) or deep learning.

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

To solve the problem, we need engineering tools, algorithms, and variables for accurate sound localization.

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.
- Regression models for estimating TDOA and DOA.

3. **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. Time Difference of Arrival (TDOA) – Multilateration

- TDOA estimates the location of a gunshot by measuring the time delay between the arrival of the sound at different microphones. The time differences are used to compute hyperbolic curves, and the intersection of these curves determines the source position. This method is effective in controlled environments but is sensitive to noise and reverberation.

2. Steered Response Power (SRP) – Beamforming

- SRP uses a grid-based approach to scan possible source locations by maximizing the power of the received signal. It relies on phase alignment across microphone arrays to estimate the source direction. This method is robust in reverberant environments but is computationally expensive due to the exhaustive search over potential locations.

3. Multiple Signal Classification (MUSIC)

- MUSIC is a high-resolution spectral estimation technique that decomposes the received signal into signal and noise subspaces. The method identifies source directions by finding peaks in the pseudo-spectrum. It performs well in multipath environments but requires more microphones than sources to function effectively.

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

- ESPRIT is similar to MUSIC but computationally more efficient as it does not require exhaustive spectral search. It estimates the direction of arrival by leveraging the rotational invariance property of signal subspaces. While it provides accurate results, it requires prior knowledge of the microphone array geometry.

5. Feature-Based Machine Learning

- Feature-based machine learning methods extract acoustic features such as Mel Frequency Cepstral Coefficients (MFCCs), spectral centroid, and time delay values. These features are used as inputs for machine learning models such as Random Forest, SVM, or XGBoost to classify or regress the direction of arrival. This approach generalizes well in noisy conditions but requires a large labeled dataset for training.

6. Deep Learning with Convolutional Neural Networks (CNN)

- CNNs process spectrogram representations of gunshot sounds to learn spatial patterns in the signal. The network is trained to map input spectrograms to direction of arrival angles. This method is robust against environmental noise and does not require handcrafted features but needs a significant amount of labeled data.

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

- RNNs and LSTMs model the temporal dependencies in gunshot audio signals to predict the direction of arrival over time. They are effective in capturing sequential patterns in acoustic data but are computationally intensive and slower to train compared to CNNs.

8. Hybrid Approaches

- Combining traditional signal processing with machine learning enhances localization accuracy. For example, TDOA values can be extracted using GCC-PHAT and fed into a neural network for classification. Similarly, MUSIC pseudo-spectra can be used as input features for deep learning models. Hybrid approaches provide improved performance but require expertise in both domains for effective implementation.

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

A well-structured dataset is crucial for training an ML model. In this case, the dataset is synthetic and consists of gunshot recordings captured from multiple microphones. We decided to choose a synthetic dataset due to lack of real world datasets needed for localisation. Features extracted from these recordings include:

Input Features (X)

1. TDOA (Time Difference of Arrival) – The time delay of sound arrival at different microphones.

2. Zero-Crossing Rate – The rate at which the audio signal changes sign, indicating frequency content.

3. Spectral Features:

- Spectral Centroid – The "center of mass" of the spectrum, representing perceived pitch.
- Spectral Bandwidth – The range of frequencies present in the signal.
- Spectral Contrast – Difference between peaks and valleys in the spectrum.
- Spectral Rolloff – The frequency below which a certain percentage of the total energy is contained.

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)

2. True_Y (Y-coordinate of the gunshot location)

Out of all the above mentioned approaches we decided to use 2 Approches :

a) ML-Based Approach for Gunshot Sound Direction of Arrival (DOA) and Localization

1. **Problem Definition**
   This approach estimates the Direction of Arrival (DOA) and spatial location (X, Y coordinates) of a gunshot sound using machine learning models trained on acoustic features extracted from multiple microphones.

2. **Dataset and Features**
   A dataset of gunshot recordings is used, with features including Time Difference of Arrival (TDOA), spectral features (spectral centroid, bandwidth, contrast, rolloff), energy-based features (entropy energy, short-time energy), and Mel-Frequency Cepstral Coefficients (MFCCs). The target variables are the true X and Y coordinates of the gunshot.

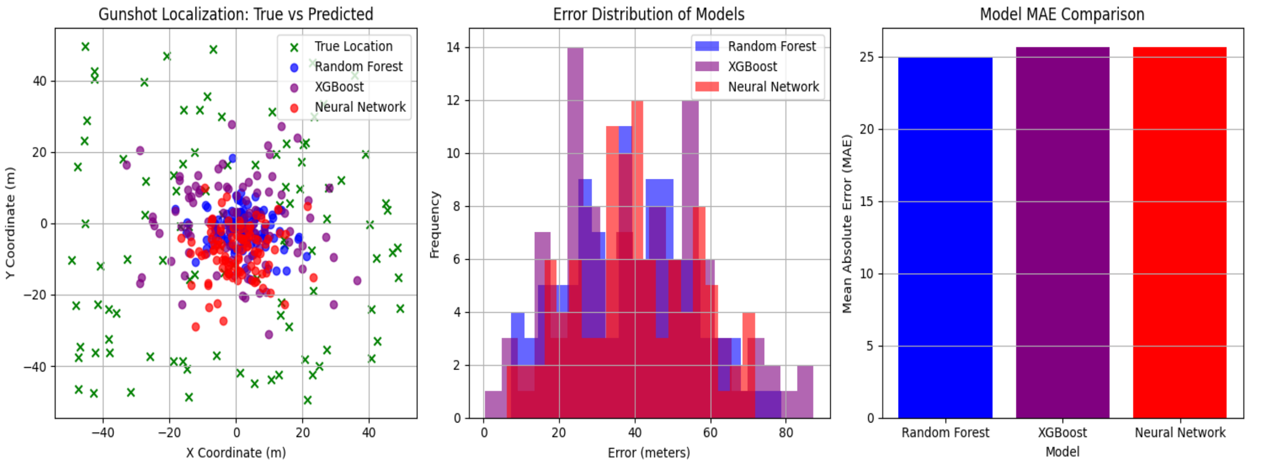
3. **ML Models for Localization**
   Three machine learning models—Random Forest, XGBoost, and a Neural Network (MLP Regressor)—are trained to predict the gunshot’s location. These models handle non-linearity in data and generalize well for real-world scenarios.

4. **Model Training and Evaluation**
The dataset is split into training (80%) and testing (20%) sets. Each model is trained on extracted features and evaluated using Mean Absolute Error (MAE) to measure localization accuracy. Error distributions are analyzed to assess model performance.

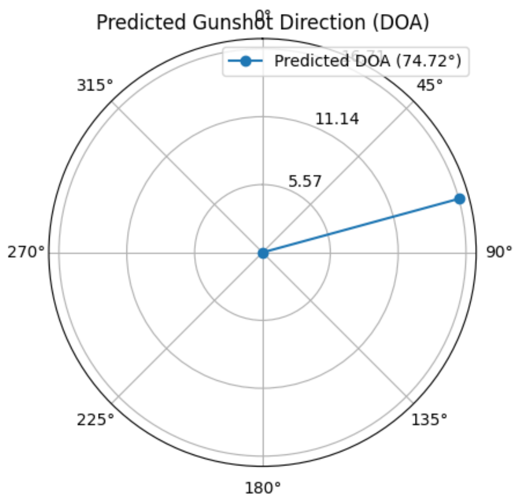
```
Training Random Forest...
Training XGBoost...
Training Neural Network...
Random Forest MAE: 24.9482
XGBoost MAE: 25.6441
Neural Network MAE: 25.6618

Model Performance:
Random Forest: MAE = 24.9482
XGBoost: MAE = 25.6441
Neural Network: MAE = 25.6618
```

5. **Visualization of Predictions**
Scatter plots compare true and predicted gunshot locations, while histograms illustrate error distributions. This helps in understanding the accuracy of different models in predicting gunshot locations.



6. **Direction of Arrival (DOA) Estimation**
DOA is computed using the Euclidean norm and arctan2 function to determine the angle relative to the microphone. The predicted direction is visualized using polar plots, aiding in understanding the gunshot's orientation.



Predicted Gunshot Coordinates: X = 4.40, Y = 16.12
Estimated DOA: 74.72°

7. **Geolocation Mapping**
   The predicted (X, Y) coordinates are converted to real-world latitude and longitude by considering the microphone's GPS position. This enables practical deployment for law enforcement and public safety applications.

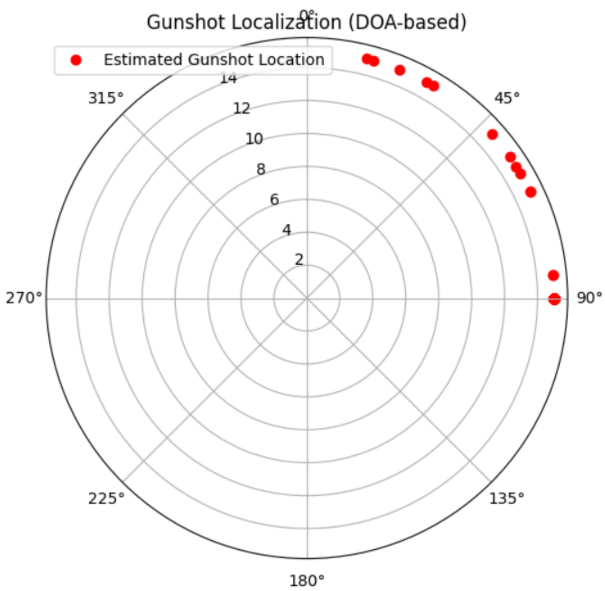
```
Predicted Gunshot Location (X, Y): (-2.899, -5.885)
Estimated Gunshot Geolocation: Latitude = 37.774847, Longitude = -122.419433
```

b) Formula-Based Approach for Gunshot Sound Direction of Arrival (DOA) and Localization

1. **Problem Definition**
   This approach estimates the Direction of Arrival (DOA) and the spatial location (X, Y coordinates) of a gunshot sound using mathematical formulas derived from acoustics and geometry.

2. **Time Difference of Arrival (TDOA) and DOA Computation**
   The Time Difference of Arrival (TDOA) between microphones is used to determine the angle of arrival of the gunshot.
   The DOA is computed using the formula:

$$\sin(\theta) = \frac{\text{TDOA} \times \text{Speed of Sound}}{\text{Microphone Distance}}$$




```
Unique DOA 1: 14.03 degrees
Unique DOA 2: 15.53 degrees
Unique DOA 3: 22.11 degrees
Unique DOA 4: 28.91 degrees
Unique DOA 5: 30.62 degrees
Unique DOA 6: 48.39 degrees
Unique DOA 7: 55.17 degrees
Unique DOA 8: 57.81 degrees
Unique DOA 9: 59.81 degrees
Unique DOA 10: 64.43 degrees
Unique DOA 11: 64.43 degrees
Unique DOA 12: 84.50 degrees
Unique DOA 13: 84.62 degrees
Unique DOA 14: 90.00 degrees
```

3. **Conversion to Cartesian Coordinates**
   Assuming the gunshot occurred within a fixed distance R, the estimated X, Y coordinates are computed as:

$$X = R\cos(\theta), \quad Y = R\sin(\theta)$$

```
Gunshot 1: Estimated Coordinates -> X: 48.51, Y: 12.12
Gunshot 2: Estimated Coordinates -> X: 48.18, Y: 13.38
Gunshot 3: Estimated Coordinates -> X: 46.32, Y: 18.82
Gunshot 4: Estimated Coordinates -> X: 43.77, Y: 24.17
Gunshot 5: Estimated Coordinates -> X: 43.03, Y: 25.47
Gunshot 6: Estimated Coordinates -> X: 33.20, Y: 37.38
Gunshot 7: Estimated Coordinates -> X: 28.56, Y: 41.04
Gunshot 8: Estimated Coordinates -> X: 26.64, Y: 42.31
Gunshot 9: Estimated Coordinates -> X: 25.14, Y: 43.22
Gunshot 10: Estimated Coordinates -> X: 21.58, Y: 45.10
Gunshot 11: Estimated Coordinates -> X: 21.58, Y: 45.10
Gunshot 12: Estimated Coordinates -> X: 4.79, Y: 49.77
Gunshot 13: Estimated Coordinates -> X: 4.69, Y: 49.78
Gunshot 14: Estimated Coordinates -> X: 0.00, Y: 50.00
```

These coordinates represent the estimated position of the gunshot relative to the microphone.

4. **Visualization of DOA and Localization**
- Polar plots are used to visualize the gunshot's estimated direction.
- Scatter plots compare estimated and actual locations of gunshots.

5. **Geolocation Mapping**
The computed (X, Y) coordinates are converted to real-world latitude and longitude using the microphone's known GPS location:

$$\text{Latitude} = \text{Ref Latitude} + \frac{Y}{\text{Meters per degree latitude}}$$

$$\text{Longitude} = \text{Ref Longitude} + \frac{X}{\text{Meters per degree longitude}}$$

The meters per degree longitude is adjusted based on the reference latitude

$$\text{Meters per degree longitude} = 111320 \times \cos(\text{Ref Latitude})$$

```
Gunshot 1: Estimated Geolocation -> Latitude: 37.775009, Longitude: -122.418849
Gunshot 2: Estimated Geolocation -> Latitude: 37.775020, Longitude: -122.418852
Gunshot 3: Estimated Geolocation -> Latitude: 37.775069, Longitude: -122.418874
Gunshot 4: Estimated Geolocation -> Latitude: 37.775117, Longitude: -122.418903
Gunshot 5: Estimated Geolocation -> Latitude: 37.775129, Longitude: -122.418911
Gunshot 6: Estimated Geolocation -> Latitude: 37.775236, Longitude: -122.419023
Gunshot 7: Estimated Geolocation -> Latitude: 37.775269, Longitude: -122.419075
Gunshot 8: Estimated Geolocation -> Latitude: 37.775280, Longitude: -122.419097
Gunshot 9: Estimated Geolocation -> Latitude: 37.775288, Longitude: -122.419114
Gunshot 10: Estimated Geolocation -> Latitude: 37.775305, Longitude: -122.419155
Gunshot 11: Estimated Geolocation -> Latitude: 37.775305, Longitude: -122.419155
Gunshot 12: Estimated Geolocation -> Latitude: 37.775347, Longitude: -122.419346
Gunshot 13: Estimated Geolocation -> Latitude: 37.775347, Longitude: -122.419347
Gunshot 14: Estimated Geolocation -> Latitude: 37.775349, Longitude: -122.419400
```

Advantages of Both Approaches:

1. Formula-Based Approach

This approach relies on mathematical formulas derived from acoustics and geometry for estimating the **Direction of Arrival (DOA)** and **localization** of gunshots.

Advantages:

- **Computational Efficiency:** Requires simple mathematical calculations, making it fast and lightweight, suitable for real-time applications.
- **No Training Required:** Unlike ML-based approaches, this method does not need a large dataset for training, reducing dependency on data collection.

- **Theoretical Accuracy:** Based on established physics and trigonometry, ensuring precise DOA estimation when input values (TDOA, microphone spacing) are accurate.
- **Low Hardware Requirements:** Can be deployed on low-power devices as it does not require high computational resources.
- **Deterministic Output:** Provides consistent results based on input values without the risk of overfitting or underfitting seen in ML models.
- **Easier Debugging and Interpretation:** Since the results come from direct calculations, it is easier to identify and fix errors.

2. **Machine Learning-Based Approach**

   This approach uses ML models (Random Forest, XGBoost, Neural Networks) trained on acoustic features to predict the gunshot's location.

   Advantages:

   - **Handles Complex Acoustic Environments:** Can model non-linear relationships between microphone inputs and gunshot locations, improving accuracy in urban or noisy settings.
   - **More Robust to Noise and Variability:** Unlike the formula-based approach, ML models can learn from real-world data and adapt to different environments.
   - **Higher Accuracy in Large-Scale Deployment:** Trained ML models can generalize better across different conditions, reducing errors from microphone misalignment or reflections.
   - **Integrates Multiple Features:** Uses a combination of spectral, energy-based, and MFCC features, making it more robust in real-world conditions.
   - **Can Improve Over Time:** With more training data, ML models continuously improve their accuracy, unlike the formula-based approach, which remains static.
   - **Enables Advanced Predictive Analytics:** Can be combined with deep learning models (e.g., CNNs, LSTMs) to detect and classify different types of gunshots.

Comparison and Selection

- If fast, real-time gunshot localization with minimal resources is required, the formula-based approach is more suitable.
- If higher accuracy and robustness in complex environments are needed, the ML-based approach is preferable.
- A hybrid approach combining both methods can offer high speed with adaptive learning, making it a practical real-world solution.

Table 3: Summary of Formulae Used:

Formula	Purpose
$\Delta t_{ij} = \frac{d_{ij}}{c}$	Time Difference of Arrival (TDOA)
$d_i = \sqrt{(X - x_i)^2 + (Y - y_i)^2}$	Distance from microphones
$\theta = \tan^{-1}\left(\frac{Y - y_{mic}}{X - x_{mic}}\right)$	Direction of Arrival (DOA)
$r = \sqrt{X^2 + Y^2}$	Radial distance
$\theta = \tan^{-1}\left(\frac{Y}{X}\right)$	DOA in radians
$\Delta lat = \frac{Y}{R} \times \frac{180}{\pi}$	Latitude offset for geolocation
$\Delta lon = \frac{X}{R\cos(lat)} \times \frac{180}{\pi}$	Longitude offset for geolocation

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

Technical information

Technical information in sound localization involves mathematical modeling and machine learning techniques for estimating the Direction of Arrival (DOA) and position (X, Y) of a gunshot.

- Formula-Based Approach: Uses TDOA and DOA equations to calculate the sound source location based on time delays between microphones.
- Machine Learning-Based Approach: Extracts acoustic features (MFCCs, spectral features, energy-based features) and trains models like Random Forest, XGBoost, and Neural Networks to predict gunshot location.
- Performance Metrics: Evaluates accuracy using Mean Absolute Error (MAE) and scatter plots to compare predicted vs. actual locations.

Non- Technical information

Real-World Constraints Affecting Sound Localization

1. Environmental Noise:
    - Background noise affects the reliability of TDOA calculations.
    - Solutions: GCC-PHAT improves noise robustness.
2. Reverberation and Echoes:
    - Indoor environments cause sound reflections, leading to false TDOA estimates.

- Solutions: Beamforming techniques and machine learning models improve accuracy.

3. Microphone Synchronization Issues:
   - Time drift between microphones leads to incorrect time delay measurements.
   - Solutions: GPS or high-precision clock synchronization ensures accuracy.
4. Deployment in Real-World Scenarios:
   - Law enforcement applications require real-time processing.
   - Solutions: Formula-based methods are faster, while ML models provide higher accuracy.

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.