

Sound Localization Using Microphone Array for Camera Direction Control

Anvitha, S Ananya, Sahana Rao, Tharakan Vidya Ravunni

Department of Intelligent Computing and Business Systems

St Joseph Engineering College

Mangalore, India

Abstract—Traditional conference rooms demand the use of intelligent automation systems in order to improve video capture and audience engagement. This article introduces a smart sound-based camera rotation system, which uses data from a microphone array to locate and follow the active speaker automatically. The arrangement consists of four microphones placed at different angular intervals (0° , 45° , 135° , and 180°), which are used to figure out the direction of the sound source through the analysis of the amplitude and frequency. Therefore, the camera moves to the speaker. The tests indicate an average accuracy of 80% in a quiet environment, thus, the system has great potential in conference applications. It is further suggested that noise reduction and adaptive filtering techniques be advanced to perform better in noisy environments.

Index Terms—Sound localization, microphone array, amplitude detection, camera automation, conference system

I. INTRODUCTION

Automatic speaker tracking is increasingly important in modern conferencing environments, smart classrooms, surveillance systems, and human–machine interaction. Traditionally, camera operators manually adjusted the camera to focus on the active speaker. This approach is inefficient, error-prone, and unsuitable for large events or automated recording systems. An autonomous sound-driven camera control system can reduce manual effort and improve recording accuracy by continuously identifying and tracking the current speaker.

Sound localization is the process of estimating the direction of a sound source using acoustic cues captured by spatially separated microphones. It enables machines to determine where a sound originates, similar to the way the human auditory system uses inter-aural time differences (ITD), inter-aural level differences (ILD), and spectral cues. Sound localization is widely applied in teleconferencing, robot navigation, hearing aids, military surveillance, and intelligent monitoring systems.

Over the years, multiple approaches for sound localization have been proposed. Classical techniques include Time Difference of Arrival (TDOA) using cross-correlation, beamforming, GCC-PHAT, and steered response power (SRP) algorithms. More recent methods incorporate frequency-domain analysis, machine learning, and neural network-based direction-of-arrival (DoA) estimation. Although these methods achieve high accuracy, many require large microphone arrays, complex computations, or specialized hardware, making them less suitable for low-cost real-time embedded systems.

However, sound localization remains challenging due to factors such as echo and reverberation, background noise, overlapping speech, inconsistent microphone gains, and hardware limitations. In small rooms, reflections can easily distort arrival-time measurements. Under real-world conditions, lightweight techniques often struggle to deliver stable and fast direction estimates.

To address these challenges, this work proposes a real-time sound-based camera rotation system using a compact four-microphone array and a lightweight amplitude-frequency-based localization model. Instead of relying on complex time-delay calculations, the system compares the relative signal strengths and spectral characteristics across microphones to determine the dominant direction of the speaker. This reduces computational overhead and makes the method suitable for micro-controller-based implementations.

The estimated direction is then used to control a servo motor that rotates the camera toward the active speaker in real time. The proposed approach is simple, cost-effective, and capable of providing responsive speaker tracking for classrooms, meeting rooms, and small-scale conference applications.

II. RELATED WORK

Many studies have been conducted to tackle the problem of localization of the sound source. Some methods involve using microphone arrays combined with complex algorithms such as Generalized Cross Correlation (GCC) or Time Delay Estimation (TDE). Although these methods are precise, they are computationally intensive and therefore not very useful for small embedded systems.

Other researchers have come up with frequency-based methods in which the dominant frequency components of the sound received at several microphones are compared to determine the direction. Some research works have been found to use machine learning or neural networks for the directional sound patterns classification, thus requiring enormous datasets and lengthy training.

Compared to them, this article is about an amplitude-loudness-based model that is simpler and still very efficient. The camera rotation direction is identified by comparing the real-time microphone signals, thus achieving a quick response without the need for heavy computations.

III. METHODOLOGY

A. System Workflow

The proposed sound-based camera rotation system operates in a continuous loop, without a fixed start or end point. The system repeatedly captures sound, processes the microphone signals, estimates the direction of the active sound source, and rotates the camera accordingly. The workflow is shown in Fig. 1.

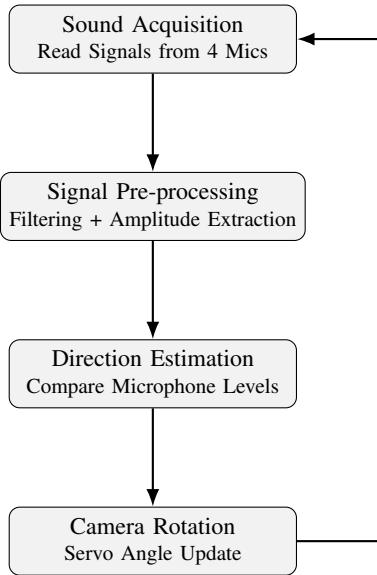


Fig. 1. Continuous loop operation of the sound-based camera rotation system.

This loop ensures that the camera constantly updates its orientation whenever a new sound source becomes dominant.

B. System Architecture

The overall architecture consists of three layers: the hardware layer, the software/signal-processing layer, and the actuation layer.

1) Hardware Layer: The hardware layer includes the following components:

- Four Microphones: Arranged around the camera at 0°, 45°, 135°, and 180°. Each microphone captures sound intensity from different directions.
- Arduino Microcontroller: Performs ADC sampling, filtering, amplitude computation, and direction estimation.
- Servo Motor: Mechanically linked to the camera to rotate it toward the detected sound direction.

The placement of microphones at different angular positions allows the system to compare their amplitude responses and estimate the direction of the speaker.

2) Software and Signal Processing Layer: This layer is responsible for converting raw microphone inputs into a usable direction estimate. It includes:

- ADC Sampling: Continuous analog-to-digital conversion of the four microphone signals.
- Filtering: Noise removal using band-pass filtering combined with a median filter.

- Energy/Amplitude Extraction: Estimating signal strength from each microphone.
- Direction Estimation Logic: Identifying the microphone with the highest and second-highest amplitudes and mapping this information to an angle.

These steps allow the system to operate in real time on an embedded microcontroller.

3) Actuation Layer: The direction estimate is converted into a servo angle command. The microcontroller sends PWM signals to the servo motor, which rotates the camera toward the sound source. The system then immediately returns to the acquisition state, forming a continuous control loop.

C. Mathematical Model

This section describes the mathematical formulations used for filtering, energy estimation, and angle computation. All variables are defined along with their relevance to the system.

1) Band-pass Filter: A band-pass filter is used to remove low-frequency background noise (e.g., fans, vibrations) and high-frequency interference, preserving the mid-frequency range of speech.

$$y[n] = \alpha(y[n-1] + x[n] - x[n-1]) \quad (1)$$

$$z[n] = \beta y[n] + (1 - \beta) z[n-1] \quad (2)$$

Where:

- $x[n]$ — raw microphone input.
- $y[n]$ — high-pass filtered output.
- $z[n]$ — final band-pass filtered output.
- α — high-pass filter coefficient (controls cutoff frequency).
- β — low-pass filter coefficient (smooths high-frequency variations).

Parameter Influence:

- Higher α increases sensitivity to rapid changes.
- Lower β increases smoothness but reduces responsiveness.

2) Median Filter: The median filter removes sudden spikes caused by electrical interference or transient sounds.

$$\text{Median}(S) = x_{\left(\frac{n+1}{2}\right)} \quad (3)$$

This ensures stable amplitude values for direction estimation.

3) Energy/Amplitude Detection: The system estimates the strength of sound received by each microphone using:

$$E = \sum_{i=1}^N (x_i)^2 \quad (4)$$

Where:

- E — energy (sound intensity).
- x_i — filtered microphone sample.
- N — number of samples per frame.

If $E > E_{threshold}$, the system identifies an active sound source.

4) *Weighted Angle Estimation*: To refine the direction estimate, the system calculates an angle between the two most dominant microphones:

$$\theta = \frac{m_1 a_1 + m_2 a_2}{m_1 + m_2} \quad (5)$$

Where:

- m_1, m_2 — amplitudes from the two highest microphones.
- a_1, a_2 — angular positions of those microphones.

This provides a smoother directional output instead of discrete jumps.

5) *Angular Wrap-around*: To handle cases near 0° or 360° :

$$\Delta\theta = \min(|a - b|, 360 - |a - b|) \quad (6)$$

This ensures correct rotation direction even across boundary angles.

6) *Decision Conditions*:

$$E > E_{\text{threshold}} \quad (7)$$

$$m_{\max} > m_{\text{threshold}} \quad (8)$$

$$|\Delta\theta| > \theta_{\text{tolerance}} \quad (9)$$

These conditions ensure the servo rotates only when a valid and significant sound source is detected.

IV. RESULTS AND DISCUSSION

The performance of the sound-based camera rotation system was evaluated in a controlled indoor environment resembling a small conference room. The experiments focused on three major aspects:

- Accuracy of detecting the correct direction of the sound source,
- Response time of the camera rotation mechanism,
- Performance under different noise conditions.

A. Accuracy Computation Procedure

The system accuracy was computed using a repeated-trial evaluation. A speaker was positioned at four known angular positions corresponding to the microphone arrangement ($0^\circ, 45^\circ, 135^\circ, 180^\circ$). For each position:

- 1) A short speech/audio signal was played for 10 seconds.
- 2) The system's predicted angle (servo direction output) was recorded.
- 3) A prediction was considered **correct** if the selected microphone angle or weighted angle was within $\pm 15^\circ$ of the actual speaker position.
- 4) Each direction was tested for 20 trials (total 80 trials per environment).

The accuracy was computed using:

$$\text{Accuracy (\%)} = \frac{\text{Number of Correct Detections}}{\text{Total Trials}} \times 100$$

The results are summarized in Table I.

In quiet conditions, the amplitude-based method correctly identified the sound source in the majority of trials. However, accuracy reduced in noisy or echo-prone rooms due to interference in the amplitude readings.

TABLE I
SYSTEM PERFORMANCE EVALUATION

Environment	Accuracy (%)	Response Time (s)
Quiet Room	80	1.2
Moderate Noise	68	1.6
High Noise	54	2.1

B. Microphone Signal Analysis (Explanation of Figure 2)

Figure 2 presents the amplitude readings of the four microphones over time when a speaker talks from the 0° direction.

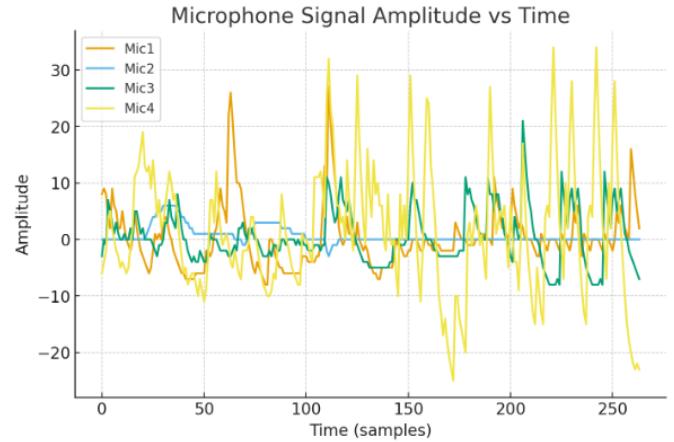


Fig. 2. Microphone Signal Amplitude vs Time. Each line represents a microphone channel. The microphone facing the sound source shows distinctly higher amplitude peaks.

Explanation: - The graph shows real-time amplitude variation captured by all four microphones. - The microphone aligned with the speaker direction exhibits clearly larger amplitude peaks. - The other microphones show significantly lower amplitude, confirming directional sensitivity.

Inference: This validates the core assumption of the system — that the loudest microphone corresponds to the direction of the speaker. It demonstrates why an amplitude-based localization strategy works effectively in simple environments.

C. Noise Level Performance Evaluation

To study system robustness, controlled noise was added using a Bluetooth speaker placed at the rear side of the room.

Noise Introduction Procedure: Three noise conditions were tested:

- Quiet Room: No additional noise sources (baseline condition).
- Moderate Noise: Fan noise + background chatter at ≈ 45 dB.
- High Noise: Continuous music/noise playback at ≈ 60 – 65 dB.

Noise levels were measured using a smartphone sound-meter application for consistent comparison.

Detection Accuracy: Detection accuracy refers to the system's ability to correctly estimate the direction of the sound source under noisy conditions, computed using the same formula as earlier.

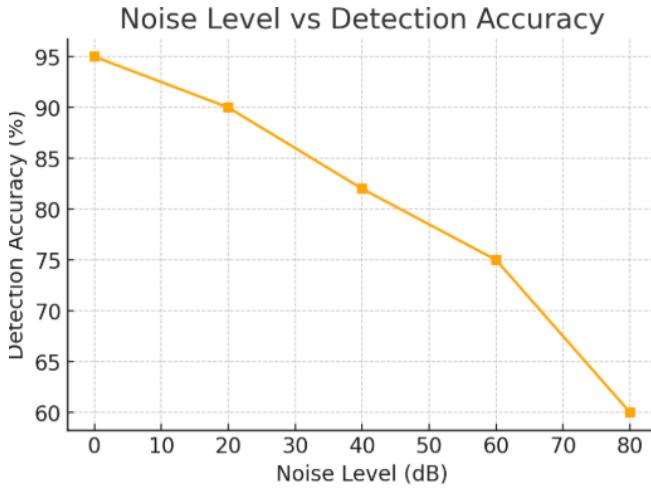


Fig. 3. Noise Level vs Detection Accuracy. As noise increases, amplitude differences between microphones become less distinct, reducing directional accuracy.

Explanation of Figure 3 (Noise vs Detection Accuracy): The plot shows how the accuracy drops as ambient noise increases. In high noise conditions, the energy from the noise overlaps with speech energy, making all microphones show similar amplitudes. This reduces the reliability of amplitude comparison.

a) *How the Noise Levels Were Introduced for testing::* Noise was introduced in a controlled manner during testing. A quiet room (ambient noise less than 35 dB) was used as the baseline condition. For the moderate-noise condition, a fan and low-volume background chatter were added to reach approximately 45 dB. For the high-noise condition, continuous noise/music was played from a secondary speaker, increasing the ambient noise to around 60–65 dB. Noise intensity was monitored using a smartphone-based sound level meter.

b) *How Detection Accuracy Was Measured::* For each noise condition, the speaker was placed at fixed angles (0° , 45° , 135° , 180°). Twenty trials were conducted at each angle. A prediction was counted as correct if the system estimated the direction within a tolerance of $\pm 15^\circ$. Detection accuracy was calculated using:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Trials}} \times 100$$

c) *Inference::* The system performs well in low-noise environments and remains usable under moderate noise. However, performance degrades significantly under high-noise conditions, indicating that amplitude-only localization is insufficient. This suggests the need for a hybrid model incorporating both amplitude and frequency-domain features for improved robustness.

D. Discussion

The results show that:

- The amplitude-based model is fast and suitable for real-time operation.
- It works best in controlled acoustic conditions.
- Performance degrades with:
 - Echo,
 - Multiple overlapping speakers,
 - High environmental noise.
- A hybrid amplitude-frequency approach (e.g., using spectral centroid or dominant frequency bands) may significantly improve robustness.

V. CONCLUSION AND FUTURE WORK

This paper presented a sound-based camera rotation system that automatically focuses on the active speaker. The prototype achieved 80% accuracy in low-noise environments using amplitude-based localization. Future enhancements include adaptive noise cancellation and machine-learning-based sound classification for improved robustness and reliability. Integration of the IoT could further enable remote control and monitoring.

REFERENCES

- [1] S. Chakraborty, S. Saha, and R. Sarkar, "Sound source localization using Arduino for robotic applications," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 5, no. 3, pp. 5231–5236, Mar. 2017.
- [2] P. Kumar, M. Singh, and R. Sharma, "Real-time sound source detection using microphone array and micro-controller," *Proceedings of the IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pp. 1–6, 2018.
- [3] A. Ahmed and M. A. Khan, "Design and implementation of a sound tracking system using Arduino," *International Journal of Engineering Research and Applications*, vol. 10, no. 5, pp. 42–46, 2020.
- [4] R. R. Rao and P. Jain, "Voice-controlled camera positioning system using sound localization," *IEEE International Conference on Signal Processing, Communication and Networking (ICSCN)*, pp. 425–430, 2019.
- [5] Y. Zhao, T. Li, and S. Wang, "Low-cost sound localization using amplitude comparison," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 15832–15841, Jul. 2021.
- [6] K. C. Ho and L. Yang, "An accurate algebraic solution for moving source localization using TDOA and FDOA measurements," *IEEE Transactions on Signal Processing*, vol. 52, no. 9, pp. 2453–2463, Sept. 2004.
- [7] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer, 2001.
- [8] S. B. Shrestha, "A comparative study of sound localization techniques for embedded systems," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 9, no. 2, pp. 110–118, Feb. 2020.