International Journal on Advanced Technology, Engineering, and Information System
https://ojs.transpublika.com/index.php/IJATEIS
Online ISSN 2828-5425

https://doi.org/10.55047/ijateis.v4i2.1741

Original Article

Sound Source Localization: A Survey

Hatice Tombul Çalışkan[1,3*], Hacer Karacan[2]

[1,2]Gazi University, Graduate School of Natural and Applied Sciences, Department of Computer Engineering, Turkey
[3]TÜBİTAK SAGE (TÜBİTAK Defense Industries Research and Development Institute), Turkey
Email: [1)] hatice.tombul@tubitak.gov.tr, [2)] hkaracan@gazi.edu.tr

Received : 21 April - 2025
Accepted : 25 May - 2025
Published online : 30 May - 2025

Abstract

In modern defense systems, there is a growing demand for technologies that detect and track threats without revealing the observer's position. Sound Source Localization (SSL) fulfills this requirement by passively estimating the position of sound-emitting targets using spatially distributed microphone arrays. Unlike active sensing systems, SSL operates solely on incoming acoustic signals, extracting location information from time delays, amplitude differences, or phase shifts. This survey provides a structured review of recent studies, covering both classical SSL methods (e.g., TDOA, GCC-PHAT) and artificial intelligence (AI)-based models (e.g., CNNs, RNNs). Classical techniques offer low computational complexity and reliable spatial resolution under ideal conditions but often degrade in noisy or reverberant environments. In contrast, AI-based approaches exhibit higher adaptability and robustness to environmental variability, though they require substantial labeled data and computational resources. Moreover, the performance of SSL systems is closely tied to microphone array geometry: while linear arrays are simple and widely used, circular, spherical, and irregular configurations provide better angular coverage and enable 3D localization. The review concludes that SSL performance is highly application-dependent, and no single method is universally superior. Hybrid approaches that combine signal processing with machine learning, as well as adaptive array designs, emerge as promising directions for improving SSL accuracy, robustness, and scalability in real-world scenarios. The comparative analysis result also underscores that optimal SSL design hinges on a trade-off between algorithmic complexity, environmental conditions, and array geometry, with hybrid methods offering a viable path forward.

Keywords: Acoustic Source Localization, Microphone Array, Passive Detection, Sound Source Localization, Time Difference of Arrival.

1. Introduction

Numerous tools, materials and machines produce sounds that surround us in daily life. These sounds are perceived through pressure changes at the ear, which travel as waves through the middle ear. The human brain possesses a remarkable ability to identify the location of sound sources in the environment. For example, when driving a vehicle, we can detect the direction of an approaching fire engine and pull over accordingly. In neuroscience, this capacity is known as auditory spatial localization, and it has inspired efforts to replicate similar capabilities in machines. SSL is the process of determining the position of an acoustic source using the information captured by spatially distributed microphones. SSL systems typically rely on signal processing techniques such as Time Difference of Arrival (TDOA), Generalized Cross-Correlation (GCC), or modern artificial intelligence methods. However, accurate localization remains challenging due to factors such as environmental noise, reverberation, and the geometry of the microphone array. For instance, reverberation causes reflections that can distort timing cues, while poor array geometry may limit spatial resolution or introduce ambiguity in source direction. Recent advancements in machine learning have also introduced AI-based SSL methods, such as convolutional and recurrent neural networks (CNNs, RNNs), which offer improved performance in complex environments but require large labeled datasets and high computational power. Recent advances in computer science, communication and electronics have increased interest in SSL across many disciplines (Papez, 2013; Tuma et al., 2012; Kunin, 2011).

Microphone array based SSL has diverse applications in healthcare, robotics, navigation technologies, RADAR/SONAR systems, and defense operations including artillery detection and gunshot localization. Particularly in the context of national security, accurate sound localization plays a crucial role in enhancing defense systems.

This survey aims to provide a comprehensive review of SSL methods, examining their technical foundations, strengths and limitations, and the types of microphone configurations used. The key research questions guiding this review are:

1) What are the main strengths and weaknesses of classical signal processing and AI-based SSL algorithms?
2) What are the most commonly used microphone array geometries in classical and AI-based SSL approaches?
3) How do different approaches perform in terms of accuracy, robustness, and computational efficiency?

By addressing these questions, this paper seeks to help researchers and engineers select suitable SSL techniques and system designs for their specific applications.

The structure of the paper is as follows: Section 2 presents a literature review; Section 3 outlines the major sound localization techniques discussed in previous studies; Section 4 includes results and discussion; and finally, Section 5 concludes the paper.

2. Literature Review

Sound source localization has been extensively studied in both biological and technological contexts. This section explores both the biological basis of human sound localization and the engineering techniques inspired by it.

2.1. Biological Foundations of Sound Localization

Humans possess an innate capacity to localize acoustic sources with remarkable accuracy. In neuroscience, this perceptual ability is referred to as sound localization and relies on several physiological cues. The human auditory system primarily uses three types of information: Interaural Time Difference (ITD), Interaural Level Difference (ILD), and the Head-Related Transfer Function (HRTF). ITD represents the time lag between sound arrival at the two ears, while ILD accounts for differences in sound intensity reaching each ear. These two are collectively known as binaural cues. In addition, HRTF provides monaural spectral cues shaped by the listener's head, torso, shoulders, and especially the pinna of each ear. The brain interprets these modified signals to determine the three-dimensional location of the sound source. Subtle head movements allow listeners to dynamically optimize ITD and ILD and enhance perception based on HRTF (Kapralos, 2003; Erkal, 2007; Jin et al., 1999; Song et al., 2016; Xiong, 2013).

2.2. Engineering Approaches to Sound Localization

Inspired by this biological mechanism, researchers have developed numerous computational methods to equip machines with similar localization capabilities. In engineering, sound source localization is defined as the process of determining the position of an acoustic source using two or more spatially separated microphones or sensors. These systems estimate parameters such as time delay, phase shift, or signal energy differences between channels to identify the direction or location of the source (Dhull, 2015; Pradip Kashid, 2015; Song et al., 2016). A crucial component in these systems is the microphone array, which consists of multiple microphones arranged in a specific geometric configuration. Microphone arrays enable the collection of spatial information and improve the signal-to-noise ratio by combining signals from multiple microphones. As a result, the reconstructed signal is typically stronger and more reliable than one captured by a single microphone. Microphone arrays can be classified by their geometry. For instance, linear arrays consist of microphones aligned in a straight line, offering accurate localization in the horizontal plane. In contrast, circular arrays distribute microphones around a circular boundary, enabling effective omnidirectional localization. In literature different microphone arrays are used for various studies such direction of arrival estimation, source localization estimation and tracking (Birinci & Leblebicioglu, 2006; Çontar, 2008; Jekaterynczuk, 2023; Liaquat, 2021). A variety of algorithms have been developed to estimate the location of sound sources based on the data acquired from microphone arrays. These methods will be discussed in detail in the following section.

3. Methods

Acoustic systems localize sound sources by analyzing signals captured from multiple microphones. These methods rely on spatial differences in time, phase, or intensity. Performance is influenced by the environment and signal processing strategies (Dhull, 2015; Khan et al., 2025; Yılmaz & Günel, 2016).

3.1. Time Difference of Arrival (TDOA)

TDOA is widely used in microphone arrays to localize sound sources by estimating the time delay between signals captured at spatially separated microphones. This delay can be calculated as in Equation (1) and is often estimated using the GCC method (Jekaterynczuk, 2023; Jiang, 2012; Paulose, 2013; Song et al., 2016).

$$\mathrm{TDOA}_{12} = \frac{(\|\mathrm{m}_1 - \mathrm{s}\| - \|\mathrm{m}_2 - \mathrm{s}\|)}{\mathrm{v}} \tag{1}$$

where s is the source location, v is the speed of sound and m_1, m_2 are microphone positions. TDOA-based methods offer high accuracy, low computational cost, and easy implementation under line-of-sight (LOS) conditions but degrade in reverberant or multi-source environments (Al-Sheikh et al., 2013; Chen et al., 2011; Papez, 2013; Ranjkesh, 2015; Kashid, 2015; Qinqin & Linghua, 2015; Liu, 2012; Li, 2016).

Many configurations exist beyond basic two-microphone setups. For instance, Mahmud et al. (2023) used linear arrays of 2–3 microphones, while Heydari et al. (2023) implemented distributed 3D localization using two rhombus-shaped arrays composed of seven and nine microphones. Wang et al. (2022) combined TDOA and ILD using only three microphones arranged in an equilateral triangle. Lellouch et al. (2025) proposed a low-cost, TDOA-based method to track animal vocalizations in natural environments, utilizing four omnidirectional microphones arranged in a tetrahedral formation.

3.2. Time of Arrival (TOA)

TOA methods estimate the signal travel time to each microphone and are common in ultrasonic indoor localization under LOS. The model is given by Equation (2) (Jekaterynczuk, 2023; Xu et al., 2011).

$$t_i = \frac{1}{v}\|x_i - y\| + t_0 + n_i \tag{2}$$

Here, v is sound speed, t_o is unknown emission time and n_i is noise. A major limitation is the requirement for tight synchronization and knowledge of t_o, often unavailable in practice. Nevertheless, TOA offers high accuracy when synchronization and LOS are maintained and forms the basis for deriving TDOA (Amundson & Koutsoukos, 2009; Dhull, 2015; Jekaterynczuk, 2023; Xiao et al., 2016; Li, 2016; Xu et al., 2011).

Studies using TOA include Saric et al. (2010), who evaluated TOA, AOA, and RSS in sensor nodes contained four microphones positioned at the vertices of a regular tetrahedron. Zou et al. (2020) used a circular eight-microphone array and showed improved accuracy over earlier methods.

3.3. Cross Correlation (CC)

CC is the earliest TDE method and calculates delay by identifying the peak of the correlation between two signals. Given signals $x_1(t)$ and $x_2(t)$, the CC function is defined as shown in Equation (3) (Chen et al., 2006; Dhull, 2010; Liu, 2012).

$$R_{x_1x_2}(\tau) = \lim_{\tau\to\alpha}\int_{-\frac{T}{2}}^{\frac{T}{2}} x_1(t)x_2{}^*(t-\tau)dt \tag{3}$$

Here, τ is delay parameter and delay parameter defines the maximum correlation between the two signals and this can be found as shown in Equation (4).

$$\tau_{CC} = \arg\max[R_{x_1x_2}(\tau)] \tag{4}$$

While CC offers high spatial resolution, it suffers in noisy or reverberant conditions. To mitigate this, the GCC method was introduced (Crocco, 2016; Qinqin & Linghua, 2015; Liu, 2012).

3.4. Generalized Cross Correlation (GCC)

GCC improves upon CC by applying frequency-domain weighting to enhance time-delay estimation. In reverberant or noisy environments, however, localization accuracy tends to decrease. GCC calculates the cross power spectrum between signals from two microphones and applies frequency-domain weighting functions before transforming the result back into the time domain. The output is a correlation function similar to CC, where the peak indicates the estimated time delay. (Chen et al., 2006; Chen et al., 2011; Khaddour, 2011; Liu, 2012; Yılmaz & Günel, 2016). The mathematical representation of GCC is given by Equation (5) and (6).

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} \psi(f) G_{x_1x_2}(f) e^{j2\pi f\tau}\, df \tag{5}$$

$$\tau_{CC} = \arg\max\left[R_{x_1x_2}(\tau)\right] \tag{6}$$

where $\psi(f)$ is the weighting function and $G_{x1x2}(f)$ is the cross spectrum of the two signals. GCC is accurate, computationally efficient, and relatively simple to implement. Various weighting functions have been developed to improve performance under different conditions. These can be seen in Table 1 (Dhull, 2015; Yang, 2023; Imran et al., 2016; Knapp & Carter, 1976; Marmaroli, 2012; Zu, 2017; Li, 2016).

Table 1.Weighting Functions

Weighting Function	Value
CC (Cross Correlation)	$\psi_{CC}(f) = 1$
PHAT (Phase Transform)	$\psi_{PHAT}(f) = \frac{1}{\lvert G_{x1x2}(f)\rvert}$
ML (Maximum Likelihood)	$\psi_{ML}(f) = \frac{1}{\lvert G_{x1x2}(f)\rvert} \frac{\lvert \gamma_{x1x2}(f)\rvert^2}{1 - \lvert \gamma_{x1x2}(f)\rvert^2}$ $\lvert \gamma_{x1x2}(f)\rvert^2 = \frac{\lvert G_{x1x2}(f)\rvert^2}{G_{x1x1}(f) G_{x2x2}(f)}$
ROTH (Roth Processor)	$\psi_{Roth}(f) = \frac{1}{G_{x1x1}(f)}$ or $\psi_{Roth}(f) = \frac{1}{G_{x2x2}(f)}$
SCOT (Smoothed Coherence Transform)	$\psi_{scot}(f) = \frac{1}{\sqrt{G_{x1x1}(f) G_{x2x2}(f)}}$
Eckart Filter	$\psi_{Eckart}(f) = \frac{G_{S1S2}(f)}{G_{N1N1}(f) G_{N2N2}(f)}$

In Table 1, $\lvert\gamma_{x1x2}(f)\rvert^2$ is the magnitude coherency function, $G_{S1S2}(f)$ is the signal cross spectrum, $G_{N1N1}(f)$ and $G_{N2N2}(f)$ are the noise power spectral densities.

CC is a special case of GCC because of its weighting filter equal to 1 and it is sensitive to external noise. ML considers the effect of background noise during TDE. PHAT produces sharp peaks, performs well in reverberant conditions, and is commonly used for multi-source detection. ML estimator aims to increase the accuracy of estimated delay with the help of attenuating signal. ROTH suppresses noisy frequency bands but may broaden the peak. SCOT estimator take into account the background noise effect in delay estimation as ML. SCOT tries to minimize the influence of the sound source signal power. Eckart suppresses frequency bands dominated by noise (Chen et al., 2011; Aguilar, 2012; Dhull, 2015; Jiang, 2012; Salvati, 2012; Kumar, 2013; Zu, 2017).

GCC has been widely used in sound source localization studies. Chen et al. (2011) evaluated CC, ROTH, SCOT and PHAT using three microphones in a triangular configuration. Lee et al. (2020) presented a method based on GCC PHAT using two linear microphones.

Firoozabadi et al. (2022) proposed an algorithm based on generalized eigenvalue decomposition and adaptive GCC-PHAT/ML methods. They used thirty-eight microphones which six T-shaped microphone arrays containing five microphones and a circular microphones array containing eight microphones. They provided accurate locations' estimations with low computational complexity but installation cost is high because of number of microphones.

3.5. Adaptive Eigenvalue Decomposition (AED)

AED estimates time delays by modeling impulse responses between a single source and microphones in linear time-invariant systems. It is robust to reverberation and suitable for real-time use (Benesty, 1999; Dhull, 2015; Jiang, 2012; Huang et al., 1999; Kashid, 2015). The algorithm extracts impulse responses from source to microphones, isolates direct paths, and calculates TDOA. It identifies the minimum eigenvalue's eigenvector in a covariance matrix using a Least Mean Square (LMS)-based iterative approach.

The AED method has computational complexity when it compared to another time delay method such as GCC-PHAT and also the AED method accuracy is lower at very noisy environments (Brutti, 2008). Shin et al. (2021) explored enhancements in TDE within reverberant environments by applying the AED algorithm. Their study, conducted using a two-microphone setup, demonstrated that the approach remains effective even in highly reflective rooms, provided the ambient noise level is moderate. Firoozabadi et al. (2023) have proposed a subband system for speaker counting using DOA estimation with AED. A centralized star-shaped microphone array, consisting of twelve microphones, was employed alongside four additional T-shaped peripheral arrays, each containing four microphones. This distributed configuration of microphone arrays contributes to enhanced algorithmic accuracy by improving spatial resolution and coverage.

Average Square Difference Function (ASDF)

ASDF estimates delay by minimizing the mean squared error between two signals as shown in Equation (7) and (8) (Aguilar, 2012; Dhull, 2015; Jacovitti & Scarano, 1993; Kashid, 2015).

$$R_{ASDF}(\tau) = \frac{1}{N}\sum_{n=0}^{N-1}[x_1(n) - x_2(n+\tau)]^2 \tag{7}$$

$$\tau_{ASDF} = argmin[R_{ASDF}(\tau)] \tag{8}$$

It is computationally simple and avoids spectral analysis, but its accuracy is lower than GCC-based methods (Aguilar, 2012; Zhang, 2005). Despite limited adoption in practical applications, ASDF has been explored in some specialized studies. Hochradel et al. (2019) employed it alongside visual 3D tracking to monitor bat trajectories using a cross-shaped sensor array comprising eight microphones. Ameran et al. (2015) also explored a hybrid approach combining ASDF and cross-correlation to improve flow velocity measurements using two sensors.

3.6. Direction of Arrival (DOA)

DOA estimates the incoming angle of a sound wave using time differences between microphones and known array geometry. It typically assumes a line-of-sight condition and can be implemented via TDOA or beamforming-based methods (Imran et al., 2016; Mahdinejad, 2015; Kunin et al., 2011; Liaquat, 2021; Paulose, 2013; Li, 2016).

The delay-angle relation is given by Equation (9) and (10) where *c* is the speed of sound.

$$c\tau_{01} = d_1 \cos(\theta) \tag{9}$$

$$\theta = cos^{-1}\left(\frac{c\tau_{01}}{d_1}\right) \tag{10}$$

To understand an environment acoustically, the most fundamental thing is to determine the sound source location, such as direction of arrival (DOA). Farmani et al. (2015) conducted a study to estimate target sound source DOA for Hearing Aid System. They proposed TDOA based DOA method for two microphones linearly mounted on the ears. Gunjal et al. (2020) have proposed a method for estimation of the desired direction using the MUSIC algorithm. They have used uniform linear array with varying number of microphones.

3.7. Angle of Arrival (AOA)

AOA estimates the direction from which sound signals arrive and is often used interchangeably with DOA. It is commonly derived from inter-sensor time delays and estimated using algorithms such as MUSIC, ESPRIT, or TDE techniques. For example, Koppula et al. (2021) improved indoor localization using AOA, focusing on the MUSIC algorithm with a four-microphone square array. Li et al. (2023) achieved accurate 3D localization by combining AOA and TDOA from LOS/echo paths using a four-microphone uniform linear array.

AOA-based methods offer practical benefits, including the elimination of strict time synchronization between sensors and reduced computational requirements. However, their performance is sensitive to line-of-sight conditions and the number of microphones used (Jekaterynczuk, 2023; Mahdinejad, 2015; Xiao et al., 2016; Li, 2016)

3.8. Beamforming

Beamforming estimates sound source direction by aligning and summing delayed microphone signals. In Delay-and-Sum (DAS) beamforming, signals are shifted based on assumed directions and summed. The direction producing the highest energy is chosen as the estimated source location. Unlike TDOA-based methods, beamforming does not require exact time-of-arrival estimates, offering robustness in Gaussian noise environments. However, it is computationally expensive (Weinstein, 2007; Kundu, 2014; Papez, 2013; Tuma et al., 2012; Veen & Buckley, 1988; Yılmaz & Günel, 2016). The beamforming output is described in Equation (11).

$$y(t) = \sum_{i}^{N} w_i \cdot x_i(t - \tau_i(\kappa)) \tag{11}$$

where x is a sound signals from microphones, w represents weighting function, τ denotes time delay, κ indicates unit direction vector, N is the number of microphones.

Recent studies have extended DAS with circular arrays or spherical harmonics for improved localization. Zhanh et al. (2021) demonstrated an optimized sound localization method building upon conventional DAS beamforming. A dual-ring uniform circular array was employed, with eight microphones each in the inner and outer concentric circles They showed effectiveness of proposed algorithm using the microphone array with experimental results. Szwajcowski et al. (2023) proposed an iterative algorithm for localizing a single sound source. They implemented two beamformers one is based on DAS beamforming and the second is spherical harmonic beamforming. They used spherical microphone array with 32 microphones. For the delay and sum beamforming the optimal results were obtained by using all thirty-two microphones in every iteration.

3.9. Multiple Signal Classification (MUSIC)

MUSIC is a high-resolution subspace algorithm for DOA estimation that relies on the orthogonality between signal and noise subspaces derived from the covariance matrix of microphone signals. Although accurate, it is computationally intensive due to exhaustive spatial scanning and requires additional processing for tracking multiple sources in real time. The array output is modeled as shown in Equation (12) (Huthaifa Obeidat, 2021; Hacıvelioğlu; Imran et al., 2016; Schmidt, 1986; Orul, 2012; Yılmaz & Günel, 2016).

$$u(t) = \sum_{i=0}^{D-1} a(\phi_i) S_i(t) + n(t) \tag{12}$$

where *s(t)* is received signal vector, *n(t)* is noise vector, $a(\phi_i)$ is the steering vector. The input covariance matrix $\hat{R}_{uu}$ is calculated shown in Equation (13). Eigenvalue decomposition of $\hat{R}_{uu}$ yields eigenvalues $\{\lambda_0, \ldots, \lambda_{M-1}\}$ and eigenvectors $[q_0\ q_1 \ldots\ q_{M-1}]$. The number of sources D is estimated as shown in Equation (14). The MUSIC spectrum is computed by Equation (15) (Desai, 2022; Hacıvelioğlu; Schmidt, 1986).

$$\hat{R}_{uu} = \frac{1}{K} \sum_{k=0}^{K-1} u_k u_k^H \tag{13}$$

$$D = M - K \tag{14}$$

$$\hat{P}_{MUSIC}(\phi) = \frac{a^H(\phi) a(\phi)}{a^H(\phi) V_n V_n^H a(\phi)},\ V_n = [q_D\ \ q_{D+1} \ldots\ q_{M-1}] \tag{15}$$

D peak values are detected in P ˆ_MUSIC (ϕ) and DOA for received signals are obtained. MUSIC has been applied across diverse microphone array configurations. Sasaki et al. (2018) employed a sixteen -microphone spherical array, demonstrating MUSIC's effectiveness for directional sound detection. Hoshiba et al. (2017) used two different MUSIC methods for sound source localization in outdoor environments. They used hexagonal and spherical microphone arrays which include respectively 16 and 12 microphones. Results of simulation in this work show that hexagonal geometry more accurate for outdoor localization. Hogg et al. (2021) proposed an extended MUSIC variant for noisy/reverberant environments, validated on an eight-microphone uniform circular array, outperforming conventional MUSIC. Kumari et al. (2025) reformulated MUSIC and MVDR in the spherical harmonics domain, testing on a twenty-microphone uniform circular array, with MUSIC surpassing MVDR in multi-source scenarios.

3.10. Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT)

ESPRIT is a high-resolution subspace method for DOA estimation that operates on narrowband signals. It supports multiple source localization and has lower computational complexity than MUSIC. The input covariance matrix $\hat{R}_{uu}$ is computed from the array inputs as shown in Equation (16), followed by eigen decomposition to separate the signal and noise subspaces (Obeidat, 2021; Hacıvelioğlu; Roy & Kailath, 1989; Orul, 2012).

$$\hat{R}_{uu} = V \Lambda V \tag{16}$$

Here, $\Lambda = \mathrm{diag}(\lambda_0, \lambda_1, .., \lambda_{M-1})$ shows Eigenvalues and $V = [q_0 \quad q_1 \quad \dots \quad q_{M-1}]$ shows Eigen vectors. D number of signal is calculated as shown in Equation (17) using minimum Eigenvalues. The signal subspace $\hat{V}_s$ is then used for further processing (Hacıvelioğlu; Orul, 2012).

$$D = M - K \tag{17}$$

By partitioning $\hat{V}_s$ into shifted subarrays, and applying eigenvalue decomposition, the matrix $\Psi = -V_{12}V_{22}^{-1}$ is formed. The eigenvalues of Ψ are used to compute DOA angles $\widehat{\Phi}_k$ as shown in Equation (18) and (19) (Hacıvelioğlu; Orul, 2012).

$$\widehat{\Phi}_k = Eigenvalues(-V_{12}V_{22}^{-1}), \forall k = 0, \dots, D-1 \tag{18}$$

$$\widehat{\Phi}_k = cos^{-1}\left[c\frac{\left(\arg(\widehat{\Phi}_k)\right)}{\beta\Delta x}\right] \tag{19}$$

The MUSIC algorithm gives more accurate results and gives better resolution than the ESPRIT algorithm (Hong, 2012; Sattar). However, there are studies using the ESPRIT method instead of MUSIC. Choi et al. (2022) implemented an Eigenbeam-ESPRIT technique using four spherical microphone arrays arranged in a tetrahedral configuration, with each array containing thirty-two omnidirectional microphones. Similarly, Zhou et al. (2023) employed a thirty-two-channel spherical microphone array with uniform surface distribution to address parameter pairing issues in ESPRIT-based localization, achieving performance superior to conventional methods.

3.11. Steered Power Response (SPR)

The SPR method estimates sound source location by steering a microphone array over candidate positions and identifying the location with maximum output power. It maximizes the function $W(x) \in \mathbb{R}$ where x=$(x, y, z) \in \mathbb{R}^3$ represents a possible source position, as shown in Equation (20) (Lima, 2015; Nunes et al., 2014).

$$W(x) \triangleq \sum_{p=1}^{P} \emptyset_p[\zeta_p(x)] \tag{20}$$

Here, P is the number of microphone pairs and $\zeta_p(x)$ denotes the TDOA for a position x. These are defined in Equation (21) and (22), where M is the number of microphones (Lima, 2015; Nunes et al., 2014).

$$P \triangleq \frac{M(M-1)}{2} \tag{21}$$

$$\zeta_p(x) \triangleq round\left\{\frac{\|m_{p,2} - x\| - \|m_{p,1} - x\|}{c} f_s\right\} \tag{22}$$

f_s is the sampling rate and c is the speed of sound. For microphone signals $s_{p,1}[n]$ and $s_{p,2}[n]$, the CC function $\phi_p[\zeta]$ is given by Equation (23) (Lima, 2015; Nunes et al., 2014).

$$\emptyset_p[\zeta] \triangleq \sum_{n\in\mathbb{Z}} s_{p,1}[n]s_{p,2}[n-\zeta] \tag{23}$$

Although robust in noisy and reverberant environments, SPR is computationally intensive and unsuitable for real-time use (Imran et al., 2016; Lima, 2015; Ranjkesh, 2015).

Since its introduction, SPR approach has been widely adopted for sound source localization across numerous applications. Nie et al. (2022) developed an SRP variant with a coherence constraint to suppress local extrema, validated using a five-microphone cross-shaped array against 20 randomized acoustic sources. Their method maintained localization accuracy while reducing spurious peaks. Lai et al. (2024) enhanced SRP's precision through a novel planar array processing scheme. Their sixteen-microphone planar array demonstrated superior performance compared to conventional SRP in experimental tests.

3.12. Received Signal Strength (RSS)

RSS estimates the source-microphone distance based on signal attenuation. RSS is simple, low-cost, and does not require time synchronization. However, its accuracy is limited due to sensitivity to distance, environment, and multipath reflections. Under LOS conditions, the RSSI value is given by Equation (24) (Jekaterynczuk, 2023; Taraktaş, 2010; Tu-Ly, 2016; Shin, 2008; Yang, 2021; Li, 2016).

$$RSSI = -(10n\log_{10} d + A) \tag{24}$$

where n is the propagation constant, d is distance and A is a reference value. While RSS is low cost and time synchronization free, its accuracy suffers due to environmental sensitivity and multipath effects (Ekim, 2013; Li, 2016).

When looking at RSS-based localization studies, Gong et al. (2023) enhanced RSS localization accuracy using Kalman filtering, demonstrating through simulations with 4 to 20 uniformly and randomly distributed base stations in a rectangular area that positioning error decreases gradually with increasing station count. Williams et al. (2023) combined RSS metrics with beamforming, achieving improved localization performance at medium-to-high SNR using a four-microphone square array, with particular gains in reverberant environments.

3.13. Energy Based Method

Energy-based source localization relies on energy measurements from microphones. Assuming energy decays with distance, the received energy at microphone i, can be expressed as Equation (25) (Dhull, 2015; Khaddour, 2011; Kashid, 2015; Sheng, 2003).

$$y_i(t) = \frac{1}{T}\sum_{t-\frac{T}{2}}^{t+\frac{T}{2}} x_i^2(t) \tag{25}$$

Here, t, T, $x_i(t)$ denote the time instant, window and received signal respectively. Although less accurate and sensitive to reverberation, it requires lower sampling rates and computational resources (Crocco, 2016). Pandey et al. (2022) used acoustic energy in a regression model and achieved 76% accuracy using only two microphones.

3.14. Feed Forward Neural Networks (FFNN)

FFNN is a basis type of artificial neural network consisting of input layer, one or more hidden layers and output layer with unidirectional data flow. Correia et al. (2021) have proposed method for solving the energy based sound source localization based on deep FFNN. They used three, six, nine, twelve, fifteen microphones arranged on a circle centered and deep FFNN was trained with measures taken from the microphones. Simulation results showed that proposed method performance matched the performance of its counterparts. FFNN is a successful method in noisy or reverberation environments. Artificial neural networks require a large training dataset to make accurate predictions. They may not generalize well to unseen situations, and there is a high risk of overfitting as well as significant computational complexity (Jekaterynczuk, 2023; Correia, 2021).

3.15. Convolutional Neural Networks (CNNs)

CNNs, originally designed for image processing, have been applied to sound source localization by converting audio data into suitable formats. Jo et al. (2025) used GCC-PHAT features and a 1D CNN with three microphones arranged in a triangle, achieving 8.75% higher accuracy than traditional methods. Other studies have employed CNNs with various array geometries, including an irregular quadrilateral array with four microphones (Boztas, 2022), a tetrahedral array composed of four elements (Sakavičius et al., 2022) and a hexagonal array of six omnidirectional microphones (Peng et al., 2022). CNNs offer high accuracy and robustness to noise and reverberant but require large datasets, are computationally intensive, and may struggle with generalization (Jekaterynczuk, 2023; Tan, 2021).

3.16. Recurrent Neural Networks (RNNs)

RNNs are designed for sequential data processing and maintain memory of past inputs through feedback loops. Though not direct localization tools, RNNs are useful in learning temporal patterns in audio signals. Nguyen et al. (2021) proposed an RNN-based system to align sound event detection with DOA outputs, using a four-microphone tetrahedral array. RNNs are robust to noise and effective for dynamic environments with moving sound sources, but demand large training datasets and computationally expensive (Guérin, 2022).

4. Results and Discussion

4.1. Research Results

Sound source localization using microphone arrays has become a popular topic in many areas. In this paper, the broad review of sound source localization has presented and a variety of methods have been examined. As a result of this paper the advantages and disadvantages of source localization methods and microphone array geometry and number of microphones used in these methods are summarized accordingly Table 2 and Table 3 bellow.

Table 2.The Pros and Cons of Localization Techniques

Measurement Techniques	Algorithm (Classic/AI)	Advantages	Disadvantages
TDOA	Classic	High accuracy, easy to achieve, less computation	LOS is required
TOA	Classic	High accuracy	Time Synchronization(between source and all microphones) and LOS is required, costly, signal transmit time should be known
CC	Classic	High spatial resolution, low computational cost, easy to implement	Reverberation or noise weakens the maximum peak value of CC
GCC	Classic	High accuracy, low computational cost, easy to implement, high spatial resolution, use weighting functions to improve the TDE performance	Low Accuracy when the reverberation time and noise is high.
ASDF	Classic	Not require multiplication, no require knowledge of input spectra, low computational cost	Low accuracy
DOA	Classic	Simple to implement, low bandwith usage	LOS is required for accuracy
AED	Classic	A robust method in reverberant environments, easy to implement	Computationally expensive, low accuracy when the noise is high.
AOA	Classic	No need to time synchronization, sufficient accuracy	LOS is required
Beamforming	Classic	Not require precise time-of-arrival information, remains effective in noisy signals if noise is Gaussian white noise	Computationally expensive
SPR	Classic	Strong approach for determining sound source location in a noisy and reverberant environment	Time consuming
RSS	Classic	Simple, low cost, no need to time synchronization	Low accuracy
MUSIC	Classic	Able to detect multiple sources	Computationally expensive
ESPRIT	Classic	Able to detect multiple sources, less computational cost than MUSIC	Lower accuracy when it compared to MUSIC
Energy Based	Classic	Not require high sampling rates	Generally low accuracy, sensitivity to reverberation
FFNN	Artificial Intelligence	High accuracy in noisy and reverberant environment	Large training dataset need, high computational complexity, high risk of overfitting
CNN	Artificial Intelligence	Robust to noisy environment, adaptable for various types of data	Large training dataset need, complex architecture, computationally expensive
RNN	Artificial Intelligence	Robust to noisy environment, suitable for dynamic environments	Large training dataset need, computationally expensive

Table 3. Microphone Array Configurations and Number of Microphones

Author	Measurement Techniques	Microphone Array Configuration	Number of Microphones
Mahmud et al. (2023)	TDOA	Linear x 2	2 and 3
Heydari et al. (2023)	TDOA	Rhombuses shaped x 2	7 and 9
Wang et al. (2022)	TDOA	Equilateral triangle	3
Saric et al. (2010)	TOA	Regular tetrahedron x 3	3 x 4
Zou et al. (2020)	TOA	Circular	8
Chen et al. (2011)	GCC-ROTH/SCOT/PHAT and CC)	Triangle shaped	3
Lee et al. (2020)	GCC PHAT	Linear	2
Firoozabadi et al. (2022)	GCC-	T-shaped x 6 and Circular	6 x 5 and 8
Shin et al. (2021)	AED	Linear	2
Firoozabadi et al. (2023)	AED	Star and T shaped x 4	12 and 4 x 4
Hochradel et al. (2019)	ASDF	Cross shaped	8
Ameran et al. (2015)	ASDF	Linear	2
Farmani et al. (2015)	DOA	Linear	2
Gunjal et al. (2020)	DOA	Uniform linear	Varying number of
Koppula et al. (2021)	AOA	Square shaped	4
Li et al. (2023)	AOA	Uniform linear	4
Zhanh et al. (2021)	Beamforming	Uniform concentric circular array (inner circle and outer circle)	8 and 8
Szwajcowski et al. (2023)	Beamforming	Spherical	32
Sasaki et al. (2018)	MUSIC	Ball shape	16
Hoshiba et al. (2017)	MUSIC	Hexagonal and Spherical	16 and 12
Hogg et al. (2021)	MUSIC	Uniform circular	8
Kumari et al. (2025)	MUSIC	Uniform circular	20
Choi et al. (2022)	ESPRIT	Spherical arrays positioned at vertices of a tetrahedron x 4	4 x 32
Zhou et al. (2023)	ESPRIT	Spherical	32
Nie et al. (2022)	SPR	Cross shaped	5
Lai et al. (2024)	SPR	Planar	16
Gong et al. (2023)	RSS	Rectangular shaped	4 to 20
Williams et al. (2023)	RSS	Square shaped	4
Pandey et al. (2022)	Energy Based	Linear	2
Correia et al. (2021)	FFNN	Circular	3, 6, 9, 12 and 15
Jo et al. (2025)	CNN	Triangle shaped	3
Boztas (2022)	CNN	Irregular quadrilateral	4
Sakavičius et al. (2022)	CNN	Tetrahedral	4
Peng et al. (2022)	CNN	Regular hexagon	6
Nguyen et al. (2021)	RNN	Tetrahedral	4

Each method listed in Table 2 offers unique advantages, challenges, and constraints. The selection of an appropriate technique largely depends on the specific application, the required localization accuracy, and the environmental conditions in which the acoustic source localization is carried out. The problem of positioning a target has been very important in many fields such as RADAR, SONAR. Many different techniques, like the ones mentioned in this study, have been developed to determine the position of source. Analyzes of the signals emitted by a sound source and received by microphones can be performed to develop more efficient algorithms for sound source localization. When using methods given in Table 2, microphones can be of various geometries and in various numbers as Table 3. The configuration of the microphone array and the number of microphones directly affect the success of the studies. For this reason, these parameters should be determined carefully, taking into account the area of the study, the expectation of accuracy and the cost. This paper will form the basis for us to choose the method and microphone array configuration most suitable for usage in studies.

4.2. Discussion

The comparative analysis of localization techniques in Table 2 highlights the particularly in terms of computational efficiency, accuracy, and robustness to environmental. Classic approaches such as TDOA, TOA, and GCC remain widely adopted due to their relatively low computational costs and ease of implementation. The recent integration of Artificial intelligence based models such as FFNN, CNN, and RNN demonstrate a shift toward more robust solutions in dynamic and noisy conditions. These models exhibit impressive performance in adverse environments, but their high computational requirements, complex architectures, and dependence on large annotated datasets remain key limitations. The increased risk of overfitting also poses challenges in generalization, especially in real-world, unseen scenarios.

Table 3 presents an overview of the microphone array configurations and the number of microphones used across a diverse set of studies. It reveals a clear trend toward utilizing geometrically diverse and often nonlinear array structures to optimize spatial resolution and source localization performance. While linear and circular arrays are still used due to their simplicity and ease of implementation, more complex configurations such as tetrahedral, spherical, and concentric circular arrays are being adopted to enhance three-dimensional localization capabilities. Additionally, the number of microphones employed varies widely, from minimal setups using only two microphones to highly sophisticated arrays with over 30 sensors. This variation reflects the trade-off between system complexity and localization accuracy.

5. Conclusion

In conclusion, the growing demand for passive sensing in defense and surveillance underscores the critical role of sound source localization (SSL) in enabling real-time situational awareness without exposing the observer. This survey highlights the wide spectrum of SSL algorithms—from traditional time-delay and direction-based methods to advanced AI-driven models—each tailored to specific environmental and computational constraints. System performance is strongly influenced by microphone array design, including geometry, number of microphones, and spatial configuration. While AI-based approaches (e.g., FFNN, CNN, RNN) demonstrate robustness in noisy or reverberant conditions, their dependence on large-scale annotated datasets remains a key limitation.

In addition to technical accuracy, sound source localization (SSL) technologies have wider technological and societal impacts. These systems are increasingly used in smart cities to detect unusual sounds, in autonomous vehicles to recognize sirens or horns, and in public areas for security monitoring. While these applications offer many benefits, they also raise concerns about privacy, data ownership, and ethical use. For instance, continuous audio monitoring may lead to unauthorized recording of people's conversations. Moreover, the way sound data is stored and who can access it needs careful regulation. Therefore, before SSL systems are widely deployed, these issues must be considered to ensure responsible and transparent use.

Future research should focus on: (1) hybrid methods that combine model-based and data-driven techniques to enhance adaptability, (2) scalable solutions for low-resource settings with limited labeled data, and (3) adaptive array configurations capable of handling dynamic acoustic environments. Addressing these challenges will further improve the reliability and operational utility of SSL systems. This survey provides a foundation for method selection and encourages targeted innovation in the field.

6. References

Al-Sheikh, B., Elshebli, A., Al-Assaf, A., Almashaqbeh, S., Alrawabdeh, W., Al-Tahat, A., Baker, A. A., Batayneh, J., & Jabra, F. (2013). Sound source direction estimation in horizontal plane using microphone array. 2013 IEEE Jordan Conference on AEECT.

Ameran, H. L. M., Jumaidin, E. J. M., Rahim, R. A., Rashid, W. N. A., Mohamad, M. M., Hashim, H., Zakaria, Z., Shaib, M. F., & Marwah, O. M. F. (2015). Velocity measurement simulative study of twin plane ECT using advanced cross correlation technique. ARPN Journal of Engineering and Applied Sciences, 10(20), 9556–9561.

Amundson, I., & Koutsoukos, X. D. (2009). A survey on localization for mobile wireless sensor networks. In Proceedings of the 2nd International Conference on Mobile Entity Localization and Tracking in GPS-less Environments, Orlando, FL, USA.

Benesty, J. (1999). Adaptive eigenvalue decomposition algorithm for passive acoustic source localization. The Journal of the Acoustical Society of America, 107(1), 384–391. https://doi.org/10.1121/1.428436

Birinci, I. Y., & Leblebicioğlu, K. (2006). Optimum placement of microphone array for sound capture using genetic algorithms. In Proceedings of the 2006 IEEE 14th Signal Processing and Communications Applications Conference

Boztas, G. (2022). Sound source localization for auditory perception of a humanoid robot using deep neural networks. Neural Computing and Applications.

Brutti, A., Omologo, M., & Svaizer, P. (2008). Comparison between different sound source localization techniques based on a real data collection. 2008 Hands-Free Speech Communication and Microphone Arrays, Trento, Italy.

Chen, J., Benesty, J., & Huang, Y. (2006). Time delay estimation in room acoustic environments: An overview. EURASIP Journal on Applied Signal Processing, 2006, 170. https://doi.org/10.1155/asp/2006/26503

Chen, L., Liu, Y., Kong, F., & He, N. (2011). Acoustic source localization based on generalized cross-correlation time-delay estimation. Procedia Engineering, 15, 4912–4919. https://doi.org/10.1016/j.proeng.2011.08.915

Chen, P., & Chen, N. (2022). Sound source localization based on convolutional neural network. In Proceedings of the CCET 2022

Choi, J.-W., Zhang, F., Jo, B., & Yoo, J.-H. (2022). Multiarray eigenbeam-ESPRIT for 3D sound source localization with multiple spherical microphone arrays. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30, 1483–1495.

https://doi.org/10.1109/TASLP.2022.3178824

Correia, S. D., Tavares, S., & Beko, M. (2021). A feed-forward neural network approach for energy-based acoustic source localization. Journal of Sensor and Actuator Networks, 10(2), 29. https://doi.org/10.3390/jsan10020029

Crocco, M., Cristani, M., Trucco, A., & Murino, V. (2016). Audio surveillance: A systematic review. ACM Computing Surveys (CSUR), 48(4), 1–46. https://doi.org/10.1145/2871183

Çontar, E., & Bayrak, H. S. (2008). Sound source localization in microphone arrays using genetic algorithm. Bilişim Teknolojileri Dergisi, 1, 23–33.

Dehghan Firoozabadi, P. A., Zabala-Blanco, D., Palacios-Játiva, P., & Azurdia-Meza, C. (2023). Speaker counting by scattered microphone array based on DOA and eigenvalue estimations in adverse environments. IEEE Xplore.

Dehghan Firoozabadi, P. I., Adasme, P., Zabala-Blanco, D., Palacios Játiva, P., & Azurdia-Meza, C. (2022). 3D multiple sound source localization by proposed T-shaped circular distributed microphone arrays in combination with GEVD and adaptive GCC-PHAT/ML algorithms. Sensors, 22(3), 1011. https://doi.org/10.3390/s22031011

Desai, D., & Mistry, N. (2022). A Review on Sound Source Localization Systems. Archives of Computational Methods in Engineering (2022), 29:4631–4642. https://doi.org/10.1007/s11831-022-09747-2

Dhull, R., & Singh, S. K. (2015). Review on Acoustic Source Localization Techniques [Review Article]. European Journal of Advances in Engineering and Technology, 72-77.

Dhull, S., Aggarwal, D. S., & Sahu, O. P. (2010). Comparison of time-delay estimation techniques in acoustic environment. International Journal of Computer Applications, 8(6), 1–6. https://doi.org/10.5120/1234-1634

Ekim, P. O., Gomes, J., & Oliveira, P. (2013). RSS-based cooperative sensor network localization with unknown transmit power. In Proceedings IEEE ICC (pp. 2765–2770). https://doi.org/10.1109/ICC.2013.6654905

Farmani, M., Sedighizadeh, M., Tan, Z.-H., & Jensen, J. (2015). Informed TDoA-based direction of arrival estimation for hearing aid applications. In Proceedings IEEE GlobalSIP (pp. 1055–1059). https://doi.org/10.1109/GlobalSIP.2015.7418313

Gong, Z., Zhang, J., & Zhou, F. (2023). Accuracy-enhanced position-sensing via Kalman filter based on received signal strength. In 7th International Conference on Communication and Information Systems.

Guérin, P.-A., Gannot, S., Kowalczyk, K., Laufer-Goldshtein, B., & Ahrens, J. (2022). A survey of sound source localization with deep learning methods. The Journal of the Acoustical Society of America, 151(5), 2801–2819. https://doi.org/10.1121/10.0011809

Gunjal, M. M., & Bhosale, A. A. R. (2020). Improved direction of arrival estimation using modified MUSIC algorithm. In Proceedings of the Fifth International Conference ICCES (pp. 1086–1090). IEEE. https://doi.org/10.1109/ICCES48766.2020.9137904

Hacıvelioğlu, İ., & Demir, H. (t.y.). Akıllı anten sistemlerinde işaret geliş açısı kestirim yöntemleri. TMMOB Elektrik Mühendisleri Odası.

Heydari, Z., & Movahedi, A. (2023). Scalable real-time sound source localization method based on TDOA. Multimedia Tools and Applications, 82, 23333–23372.

Hochradel, K., Hohler, T., Becher, A., Wildermann, S., & Sutor, A. (2019). Three-dimensional localization of bats: Visual and acoustical. IEEE Sensors Journal, 19(24), 11931–11940. https://doi.org/10.1109/JSEN.2019.2937877

Hogg, A. O. T., Neo, V. W., Weiss, S., Evers, C., & Naylor, P. A. (2021). A polynomial eigenvalue decomposition MUSIC approach for broadband sound source localization. In 2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.

Hong, J.-G., Park, C.-S., & Seo, B.-S. (2012). Comparison of MUSIC and ESPRIT for direction of arrival estimation of jamming signal. In 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings (pp. 1047–1050). IEEE.

https://doi.org/10.1109/I2MTC.2012.6229339

Hoshiba, K., Wakabayashi, M., Ishiki, T., Kumon, M., Bando, Y., Gabriel, D., Nakadai, K., & Okuno, H. G. (2017). Design of UAV-embedded microphone array system for sound source localization in outdoor environments. Sensors, 17(11), 2535. https://doi.org/10.3390/s17112535

Huang, Y., Benesty, J., & Elko, G. W. (1999). Adaptive eigenvalue decomposition algorithm for real time acoustic source localization system. In 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Vol. 2, pp. 917–920). IEEE. https://doi.org/10.1109/ICASSP.1999.758098

Imran, M., Hussain, A., Qazi, N. M., & Sadiq, M. (2016). A methodology for sound source localization and tracking: Development of 3D microphone array for near-field and far-field applications. In 2016 13th IBCAST (pp. 103–108). IEEE. https://doi.org/10.1109/IBCAST.2016.7429859

Jacovitti, G., & Scarano, G. (1993). Discrete time techniques for time delay estimation. IEEE Transactions on Signal Processing, 41(2), 525–533. https://doi.org/10.1109/78.193195

Jekaterynczuk, G., & Pliszka, Z. (2023). A survey of sound source localization and detection methods and their applications. Sensors, 24(1), 68. https://doi.org/10.3390/s24010068

Jiang, H., Mishra, B., & Willett, P. (2012). Sound localization using compressive sensing. In Proceedings of the 1st International Conference on Sensor Networks (pp. 159–166).

Jo, H.-M., Kim, T.-W., & Kwak, K.-C. (2025). Sound source localization using deep learning for human–robot interaction under intelligent robot environments. Electronics, 14(5), 1113. https://doi.org/10.3390/electronics14051113

Kashid, P., Chaudhari, K., Shah, S. K., & Agrawal, S. S. (2015). A review: Sound source localization and audio beam formation. In Proceedings of the 46th IRF (pp. 34–38)

Khaddour, H. (2011). A comparison of algorithms of sound source localization based on time delay estimation. Electrotechnics Magazine, 2(1).

Khan, A., Waqar, A., Kim, B., & Park, D. (2025). A review on recent advances in sound source localization techniques, challenges, and applications. Sensors and Actuators Reports, 9, 100313. https://doi.org/10.1016/j.snr.2025.100313

Knapp, C., & Carter, G. (1976). The generalized correlation method for estimation of time delay. IEEE Transactions on Acoustics, Speech, and Signal Processing, 24(4), 320–327. https://doi.org/10.1109/TASSP.1976.1162830

Koppula, S. (2021). Using angle of arrival for improving indoor localization. arXiv. https://doi.org/10.48550/arXiv.2101.09904

Kumar, S., & Garg, D. R. (2013). Estimating time delay using GCC for speech source localisation. International Journal of Application or Innovation in Engineering & Management (IJAIEM), 2(5), 24–30.

Kumari, D., & Kumar, L. (2025). Spherical sector harmonics based acoustic source localization using circular array. In Proceedings of the ICASSP 2025.

Kundu, T. (2014). Acoustic source localization. Ultrasonics, 54(1), 25–38. https://doi.org/10.1016/j.ultras.2013.06.009

Kunin, V., Jia, W., Turqueti, M., Saniie, J., & Oruklu, E. (2011). 3D direction of arrival estimation and localization using ultrasonic sensors in an anechoic chamber. In 2011 IEEE International Ultrasonics Symposium (pp. 1024–1027). IEEE. https://doi.org/10.1109/ULTSYM.2011.0276

Kunin, V., Saniie, J., & Oruklu, E. (2011). Direction of arrival estimation and localization using acoustic sensor arrays. Journal of Sensor Technology, 71–80. https://doi.org/10.4236/jst.2011.13010

Lai, W.-T., Chen, X., Bastine, A., Abhayapala, T. D., & Samarasinghe, P. N. (2024). Source localization by multidimensional steered response power mapping with sparse Bayesian learning. arXiv.

Lee, R., Kim, M.-S., Kim, B.-H., Park, K.-H., Lee, S. Q., & Park, H.-M. (2020). Sound source localization based on GCC-PHAT with diffuseness mask in noisy and reverberant environments. IEEE Access, 8, 207611–207623. https://doi.org/10.1109/ACCESS.2020.3037715

Lellouch, L., Haupert, S., & Sueur, J. (2025). Sound source localization in a natural soundscape with autonomous recorder units based on a new time-difference-of-arrival algorithm. Applied Acoustics, 235, 110648. https://doi.org/10.1016/j.apacoust.2025.110648

Li, Z., Zhang, Y., Li, C., Zang, H., & Wang, W. (2023). MicEye: Locating sound source in 3D with small-size microphone array. In ICCC 2023.

Liaquat, M. U., Munawar, H. S., Qadir, Z., Kouzani, A. Z., & Mahmud, M. A. P. (2021). Localization of sound sources: A systematic review. Energies, 14(13), 3910. https://doi.org/10.3390/en14133910

Lima, M. V. S., Martins, W. A., Nunes, L. O., Biscainho, L. W. P., Ferreira, T. N., Costa, M. V. M., & Lee, B. (2015). Efficient steered-response power methods for sound source localization using microphone arrays. IEEE Signal Processing Letters, 22(8), 1229–1233. https://doi.org/10.1109/LSP.2014.2385864

Liu, S., Zhang, C., & Huang, Y. (2012). Research on acoustic source localization using time difference of arrival measurements. In 2012 International Conference on MIC (pp. 777–780). IEEE.

Li, X., Dinh, Z. D., Rauchenstein, L. T., & Carlson, T. J. (2016). Contributed review: Source-localization algorithms and applications using time of arrival and time difference of arrival measurements. Review of Scientific Instruments, 87. https://doi.org/10.1063/1.4947001

Llerena Aguilar, C., & Alvarez Perez, L. (2012). A comparative study of time-delay estimation techniques for convolutive speech mixtures. Advances in Computer Science.

Mahdinejad, K., Bayat, S., & Ziaee Seghaleh, M. (2015). Implementation of TDOA positioning with different arrays of microphone using GCC-PHAT and Chan algorithm. Cumhuriyet University Faculty of Science Science Journal (CSJ), 36(4), 1–10.

Marmaroli, P., Feng, X., & Lissek, H. (2012). A comparative study of time delay estimation techniques for road vehicle tracking. In Acoustics 2012 Nantes Conference Proceedings (pp. 1–6). Nantes, France.

Mohammud@Mahmud, N. S. S., Haron, I. S. A., Hassan, S. L. M., & Abdullah, W. F. H. (2023). Design and performance analysis of sound source localization using time difference of arrival estimation. Journal of Advanced Research in Applied Mechanics, 106(1), 14–26

Nguyen, T. N. T., Nguyen, N. K., Phan, H., Pham, L., Ooi, K., Jones, D. L., & Gan, W.-S. (2021). A general network architecture for sound event localization and detection using transfer learning and recurrent neural network. In Proceedings of ICASSP 2021.

Nie, P., Liu, B., Chen, P., & Han, Y. (2022). Coherence-weighted steered response power for acoustic source localization. Acoustics Australia, 50, 71–77. https://doi.org/10.1007/s40857-021-00238-y

Nunes, L. O., Martins, W. A., Lima, M. V. S., Biscainho, L. W. P., Costa, M. V. M., Gonçalves, F. M., Said, A., & Lee, B. (2014). A steered-response power algorithm employing hierarchical search for acoustic source localization using microphone arrays. IEEE Transactions on Signal Processing, 62(19), 5171–5183. https://doi.org/10.1109/TSP.2014.2336636

Obeidat, H., Shurman, W., Obeidat, O., & Abd-Alhameed, R. (2021). A review of indoor localization techniques and wireless technologies. Wireless Personal Communications, 119, 289–327. https://doi.org/10.1007/s11277-021-08209-5

Orul, T. (2012). İşaret geliş açısı kestirim algoritmaları [Estimation of direction of arrival algorithms]. In Proceedings of ELECO 2012.

Pandey, S. P., Singh, S., Mishra, S. P., & Kumar, M. U. (2022). Sound localisation of an acoustic source using time delay and distance estimation. In 2022 International Conference on SMART GENCON (pp. 1–5).

Papez, M., & Valtr, K. (2013). Acoustic source localization based on beamforming. In Recent Advances in Systems Science (pp. 325–330).

Paulose, S., & Paul, B. (2013). Acoustic source localization. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2(2), 1084–1089

Qinqin, Z., & Linghua, Z. (2015). Study of delay estimation in acoustic source localization based on microphone array. In 2015 IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (pp. 1563–1566). IEEE. https://doi.org/10.1109/IAEAC.2015.7428676

Ranjkesh, M., & Hashemi, R. (2015). A fast and accurate sound source localization method using the optimal combination of SRP and TDOA methodologies. Journal of Information Systems and Telecommunication, 3(3), 209–217.

Roy, R., & Kailath, T. (1989). ESPRIT-estimation of signal parameters via rotational invariance techniques. IEEE Transactions on Acoustics, Speech, and Signal Processing, 37(7), 984–995. https://doi.org/10.1109/29.32276

Sakavičius, S., Striūgas, A., & Abromavičius, V. (2022). Multiple sound source localization in three dimensions using convolutional neural networks and clustering-based post-processing. Sensors, 22(4), 1586. https://doi.org/10.3390/s22041586

Salvati, D. (2012). Acoustic source localization using microphone arrays [Master's thesis, University of Udine].

Saric, Z. M., Krstanovic, D. D., & Teslic, N. D. (2010). Acoustic source localization in wireless sensor network. Circuits, Systems, and Signal Processing, 29, 837–856. https://doi.org/10.1007/s00034-010-9187-3

Sasaki, Y., Takahashi, R., & Tsubouchi, H. (2018). Online spatial sound perception using microphone array on mobile robot. In Proceedings of IROS 2018.

Sattar, O. A., Fath Elrahman, M., & Taha, S. (2013). Comparison between MUSIC and ESPRIT direction of arrival estimation algorithms for wireless communication systems. In Proceedings First International Conference on FGCT (pp. 79–83).

Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. IEEE Transactions on Antennas and Propagation, 34(3), 276–280. https://doi.org/10.1109/TAP.1986.1143830

Sheng, X., & Huang, Y.-H. (2003). Energy based acoustic source localization. LNCS, 285–300.

Shin, S.-C., Seo, B.-R., Kim, W.-G., & Kim, J.-G. (2008). ERFS: Enhanced RSSI value filtering schema for localization in WSN. IFIP International Federation for Information Processing, 281, 266–272). https://doi.org/10.1007/978-0-387-85563-4_35

Shin, Y.-H., & Lee, J.-H. (2021). Prefiltering approach of adaptive eigenvalue decomposition method for an improved time delay estimation in room environment. IEEE Access.

Song, T., Chen, J., Zhang, D., Qu, T., & Wu, X. (2016). A sound source localization algorithm using microphone array with rigid body. In Proceedings of the 22nd International Congress on Acoustics. Buenos Aires, Argentina.

Szwajcowski, A., Makuch, T., & Celniak, W. (2023). An iterative approach to sound source localization based on spherical beamforming. Vibrations in Physical Systems, 34(2), Article 2023216. https://doi.org/10.21008/j.0860-6897.2023.2.16

Tan, T.-H., Lin, Y.-T., Chang, Y.-L., & Alkhaleefah, M. (2021). Sound source localization using a convolutional neural network and regression model. Sensors, 21(8031). https://doi.org/10.3390/s21238031

Taraktaş, K. F., Çetin, O., & Yağcı, B. (2010). Received signal strength based localization in sensor network. In SIU2010 - IEEE 18. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, Diyarbakır, Turkey.

Tu-Ly, N., Tran, T. L., & Mai, L. (2016). Power adaptive-resampling based on particle filters for wireless patient tracking in health care system. In Studies in Health Technology and Informatics (Vol. 282, pp. 148–151). IOS Press.

Tuma, J., Janecka, P., Vala, M., & Richter, L. (2012). Sound source localization. In Carpathian Control Conference (ICCC), 2012 13th International.

Van Veen, B. D., & Buckley, K. M. (1988). Beamforming: A versatile approach to spatial filtering. IEEE ASSP Magazine, 5(2), 4–24. https://doi.org/10.1109/53.665

Wang, H., Li, J., Xu, F., & Liu, Y. (2022). 3-D sound source localization with a ternary microphone array based on TDOA-ILD algorithm. IEEE Sensors Journal, 22(24), 23991–23999. https://doi.org/10.1109/JSEN.2022.3211429

Weinstein, E., Sabnis, K., Agarwal, A., & Glass, J. (2007). LOUD: A 1020-node microphone array and acoustic beamformer. In Proceedings of the 14th ICSV, Cairns, Australia.

Williams, F., Jose Tom, A., Martin, T., & Fookes, C. (2023). Enhancing emitter localization accuracy through integration of received signal strength in direct position determination. In 2023 IEEE Statistical Signal Processing Workshop (SSP). IEEE.

Xiao, J., Zhou, Z., Yi, Y., & Ni, L. M. (2016). A survey on wireless indoor localization from the device perspective. ACM Computing Surveys, 49(2), 1–31. https://doi.org/10.1145/2933232

Xu, E., Ding, Z., & Dasgupta, S. (2011). Source localization in wireless sensor networks from signal time-of-arrival measurements. IEEE Transactions on Signal Processing, 59(6), 2887–2897. https://doi.org/10.1109/TSP.2011.2116012

Yang, F., & Sun, R. (2023). A review of sound source localization research in three-dimensional space. In Proceedings of the DDCLS 2023, Xiangtan, China. IEEE.

Yang, T., Chafouk, A. (2021). A survey of recent indoor localization scenarios and methodologies. Sensors, 21. https://doi.org/10.3390/s21238086

Yılmaz, M., & Günel, B. (2016). Investigation of acoustic source localization algorithm performances for gunshot localization on helicopters. In 2016 24th SIU.

Zhang, Y., Huang, W., & Allen, A. (2005). A comparative study of time-delay estimation techniques using microphone arrays (School of Engineering Report No. 619).

Zhang, Y., Zhang, H., & Liu, H. (2021). An improved multiple sound source localization method using a uniform concentric circular microphone array. In 11th International Conference on Intelligent Control and Information Processing, Dali, China.

Zhou, H., Liu, Z., Luo, L., Wang, M., & Song, X. (2023). An improved two-stage spherical harmonic ESPRIT-type algorithm. Symmetry, 15(2), 368. https://doi.org/10.3390/sym15020368

Zou, Y., & Li, H. (2020). A simple and efficient iterative method for TOA localization. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Zu, X., Huang, F. G., Huang, J., Zhao, Q., Liu, H., Li, B., & Yuan, X. (2017). Design of an acoustic target intrusion detection system based on small-aperture microphone array. Sensors, 17(514). https://doi.org/10.3390/s17030514