IAES International Journal of Robotics and Automation (IJRA)
Vol. 15, No. 2, June 2026, pp. 257~266
ISSN: 2722-2586, DOI: 10.11591/ijra.v15i2.pp257-266

A comprehensive review of sound source localization methods for robotics

Muhammad Akmal Aliff[1], Emerson Joseph Raja[2]
[1]Center of Excellence Robotics and Sensing Technology, TM Research and Development, Cyberjaya, Malaysia
[2]Faculty of Engineering Technology, Multimedia University, Melaka, Malaysia

Article Info

Article history:

Received Dec 23, 2024
Revised Mar 10, 2026
Accepted Apr 1, 2026

Keywords:

Artificial intelligence
Classical method
Microphones array
Robotics
Sound source localization

ABSTRACT

Sound source localization (SSL) is a key technology in robotics that allows machines to detect and locate auditory cues in real time. This review provides a thorough examination of SSL techniques classified into classical, artificial intelligence (AI), and hybrid methods. Classical methods, which account for 44% of reviewed studies, excel in computational efficiency and reliability under controlled conditions but have limitations in dynamic environments. AI methods, which account for 16% of studies, use deep learning to adapt to complex scenarios, but they require large datasets and computational resources. Hybrid methods, which combine classical signal processing and AI, are the most robust and accurate, with an average accuracy of 97.45%. The review also looks at the role of microphone arrays in SSL performance, revealing that systems with ten or more microphones achieve the highest accuracy of 99.23%, while single- and dual-microphone systems still perform competitively (97.60% and 97.21%, respectively). These findings suggest that hybrid methods combined with larger microphone arrays are the most effective SSL solution in robotics, balancing precision and adaptability. This paper discusses current SSL trends, challenges, and future research directions, providing insights for the development of advanced auditory systems capable of reliable performance in dynamic, real-world environments.

Corresponding Author:

Muhammad Akmal Aliff
Center of Excellence Robotics and Sensing Technology, TM RND
Cyberjaya, Selangor, Malaysia
Email: akmalaliff@tmrnd.com.my

1. INTRODUCTION

Sound source localization (SSL) has emerged as a pivotal technology in the domain of robotics, enabling machines to perceive and interact with their auditory environment [1]. This capability is particularly crucial for human-robot interaction (HRI), where the ability to detect and locate sound sources enhances a robot's situational awareness, communication abilities, and decision-making [2]. SSL underpins a wide array of applications, including guiding visually impaired individuals, enhancing autonomous navigation, and enabling voice-command-based interfaces even in noisy environments [3].

Over the years, advancements in sensor technology, signal processing, and machine learning have significantly evolved SSL techniques, transitioning from traditional beamforming [4] and time-difference-of-arrival (TDOA) methods [5] to modern deep learning and hybrid approaches [6]. However, integrating SSL into robotic systems presents unique challenges: ensuring real-time processing [7], maintaining robustness in dynamic and noisy environments, and achieving scalability for multi-source localization [8], [9]. The need to

balance computational complexity, hardware constraints, and system adaptability, particularly in cost-sensitive and resource-limited robotic platforms, amplifies these challenges [10].

This review aims to address these challenges by providing a comprehensive examination of contemporary SSL methods. It categorizes the techniques into three primary groups—classical methods, artificial intelligence (AI) methods, and hybrid methods—analyzing their strengths, limitations, and applications in robotics. Additionally, it explores the role of microphone arrays in influencing SSL performance, highlighting trends in the trade-offs between accuracy and hardware configurations. By synthesizing findings from 55 research papers, this work identifies the most effective techniques and configurations for modern robotic systems. The review concludes with insights into emerging research directions, focusing on enhancing SSL accuracy, efficiency, and adaptability for real-world applications.

2. CLASSIFICATION METHOD

This section presents a comprehensive analysis of findings from 55 research papers, examining the distribution and performance of various SSL methods. It categorizes these methods into classical, AI-driven, and hybrid approaches, highlighting their respective strengths, limitations, and real-world applicability. Furthermore, this analysis provides insights into emerging trends, ongoing challenges, and potential future directions in SSL research, offering a valuable reference for both academic and industrial applications.

Figure 1 illustrates the distribution of SSL methods in the reviewed studies. Classical methods account for 44% (24 papers), AI methods for 16% (9 papers), and hybrid methods for 40% (22 papers). Each approach demonstrates unique strengths and limitations:

a. Classical methods: These techniques rely on signal processing approaches such as beamforming, TDOA, and phase-based methods. Known for their simplicity and efficiency, classical methods are ideal for real-time applications but are less effective in noisy or complex environments.
b. AI methods: These utilize deep learning and other data-driven techniques, excelling in dynamic and multi-source scenarios. However, their reliance on large datasets and computational resources limits their scalability for real-time and resource-constrained applications.
c. Hybrid methods: By combining classical feature extraction with AI-driven pattern recognition, hybrid methods achieve the best of both worlds. They are particularly effective in addressing challenges such as noise, reverberation, and complex source dynamics.

The next section provides a detailed overview of the 55 papers analyzed, categorized into three groups based on the methodology employed: classical methods, AI methods, and hybrid methods. Each category is examined in terms of its techniques, applications, and future potential.

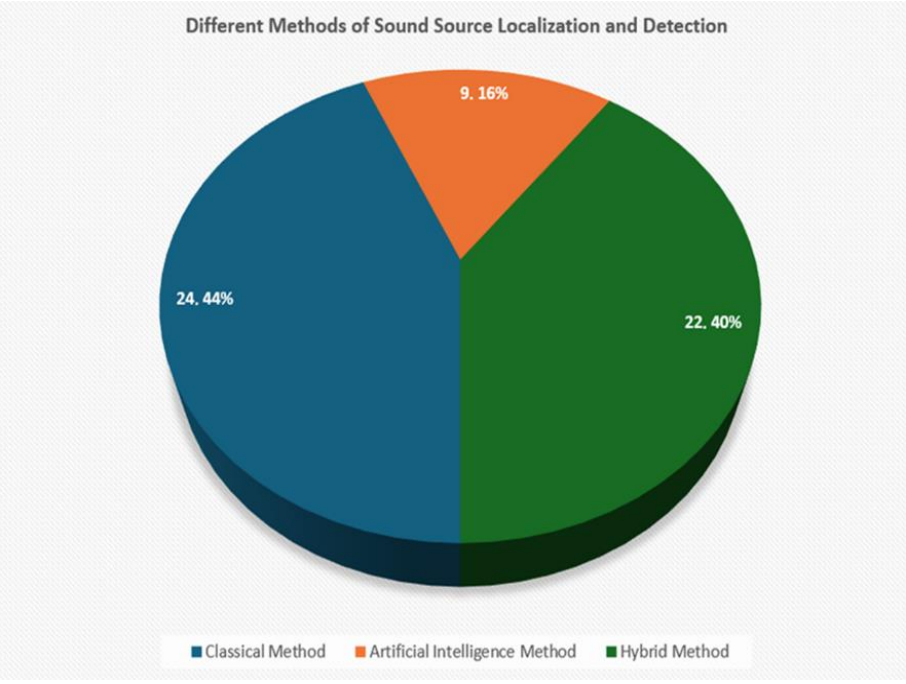

Figure 1. Different methods of sound source localization and detection

2.1. Classical methods

Classical methods, representing 44% (24 papers) of the reviewed studies, rely on established signal processing techniques such as TDOA, phase-based methods, and beamforming. These methods are valued for their mathematical simplicity, low computational demands, and real-time applicability, making them foundational for SSL. While they face limitations in complex acoustic environments, refinements and novel implementations have significantly expanded their utility.

Classical methods excel in resource-constrained and embedded applications. For example, Jamaludin *et al.* [11] implemented SSL using TDOA on field programmable gate array (FPGA) technology, achieving improved accuracy and processing speed. Lin *et al.* [7] optimized SSL on an FPGA SoC platform, balancing power consumption and processing demands for real-time localization. Pamungkas and Rais [12] demonstrated a real-time SSL system using the interaural time difference (ITD) method on the TMS320C6713 board, while Grondin *et al.* [13] proposed the open embedded audition system (ODAS) framework, integrating generalized cross-correlation phase transform (GCC-PHAT) algorithm, delay-and-sum beamforming, and Kalman filtering for robust SSL on platforms like Raspberry Pi.

Beamforming innovations have significantly improved localization accuracy and efficiency. Grondin and Michaud [8] introduced the steered response power phase transform with hierarchical search with directivity (SRP-PHAT-HSDA) algorithm, optimizing hierarchical search and directivity models for real-time robotic applications. Sun *et al.* [4] developed compressed beamforming (CSB-II) with iterative thresholding for enhanced spatial resolution. Salvati *et al.* [14] refined SRP-PHAT with geometrically sampled grids and max-pooling, improving performance in noisy environments. Additionally, Qin *et al.* [15] proposed a compressive sensing-based method that reduced data requirements while maintaining accuracy in reverberant conditions. Gombots *et al.* [16] further extended beamforming by integrating the Helmholtz equation and finite element method (FEM), achieving high-resolution localization in complex environments.

Classical methods have advanced TDOA techniques to improve localization under challenging conditions. Zhao *et al.* [17] enhanced TDOA estimation with PHAT-GCC and a frequency divider, addressing low signal-to-noise ratios. Heydari and Mahabadi [5] demonstrated TDOA localization with high accuracy and low computational complexity, achieving localization in just 360 milliseconds. Still *et al.* [18] introduced real-time TDOA (RTDOA) using Monte Carlo simulations to improve reliability. Lee *et al.* [19] incorporated a diffuseness mask to refine GCC-PHAT in reverberant settings, while Chung *et al.* [20] combined GCC-PHAT with TDOA for precise localization in two-microphone systems, achieving errors as low as 2.3 cm.

Biologically inspired methods have broadened the scope of classical techniques. Yang *et al.* [21] modeled localization on the auditory system of the parasitoid fly Ormia ochracea, achieving high accuracy with compact sensor arrays. An *et al.* [22] proposed a diffraction- and reflection-aware SSL method, leveraging geometric acoustic modeling and ray tracing to estimate the positions of direct and non-line-of-sight (NLOS) sources.

Classical techniques have also been refined for multi-source scenarios. Jia *et al.* [9] introduced diffuseness estimation to isolate single-source time-frequency points, simplifying multi-source localization. Song and Shin [23] combined interchannel phase differences (IPDs) with spectral masks and probabilistic voting to improve direction of arrival (DoA) estimation. Zhou *et al.* [24] utilized phase consistency and outlier removal for robust multi-source localization in reverberant environments.

Classical methods are effective in diverse and specialized scenarios. Hosangadi [3] developed an SSL method for search-and-rescue robots using GCC-PHAT and delay-and-sum beamforming, achieving high angular resolution and efficient triangulation. Sieriebriakov *et al.* [25] proposed a method based on sound intensity and frequency variation for localization in restricted visibility environments. Wang and Zhang [26] combined TDOA with Kalman filtering to achieve stable indoor tracking, demonstrating average localization errors as low as 10 cm across 10 meters.

Adaptive filtering and real-time capabilities further strengthen classical methods. Sewtz *et al.* [27] introduced the motion model enhanced multiple signal classification (MME-MUSIC) algorithm, incorporating motion modeling and noise-aware frequency selection to enhance DoA estimation in reverberant environments. Gala and Sun [28] combined the extended Kalman filter (EKF) with the Hilbert transform for real-time SSL, achieving quick convergence and improved accuracy.

Despite their strengths in mathematical simplicity and real-time performance, classical methods face limitations in dynamic and noisy environments, where reverberation, multi-source interference, and complex acoustic phenomena degrade their accuracy. Incremental advancements, such as adaptive beamforming, biologically inspired designs, and advanced filtering techniques, continue to extend their relevance. However, their reliance on predefined models underscores the growing need for integration with data-driven techniques to address evolving SSL challenges.

2.2. Artificial intelligence methods

AI methods, explored in 16% (9 papers) of the reviewed studies, leverage the capabilities of machine learning and deep learning to address the inherent limitations of classical approaches. By employing data-driven algorithms such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and other advanced models, AI methods excel in localizing sound sources in complex and challenging environments, including multi-source, noisy, and reverberant conditions, where classical techniques often fall short.

AI methods are increasingly being used to study and model complex auditory mechanisms. For example, Ihlefeld *et al.* [29] explored population rate-coding in human sound localization, using psychophysical experiments and computational neural models to demonstrate how sound intensity affects perceived laterality. This research highlighted the potential of AI-driven neural modeling to decode intricate auditory processes and advance our understanding of sound localization dynamics.

Several studies have focused on applying AI techniques to enhance traditional sound source localization tasks. For instance, Tan *et al.* [30] introduced a CNN-regression model (CNN-R) to process interaural phase difference (IPD) features extracted through short-time Fourier transform (STFT). This method achieved high accuracy for angle and distance estimation, demonstrating superior robustness to noise in both simulated and real-world conditions. Similarly, Huang *et al.* [31] employed a backpropagation neural network (BPNN) to process time-delay difference data from acceleration sensors, achieving localization errors as low as 0.01 meters and demonstrating robust performance in structural sound localization.

Zhou *et al.* [32] focused on noise source localization using a deep learning framework based on CNNs. The proposed method effectively localizes sound sources by learning spatial patterns and noise features from acoustic signals. Experimental results showed that the approach is highly robust in noisy environments, outperforming traditional localization methods by improving detection accuracy and generalizing well across varying conditions.

Sakavičius and Serackis [33] focused on 3D localization tasks, employing a CNN to estimate azimuth and elevation using a 2D DoA probability heatmap derived from STFT phase components. This method achieved remarkable precision even in low signal-to-noise ratio (SNR) conditions, underscoring the effectiveness of AI-driven approaches in addressing complex spatial localization challenges.

AI methods have also proven effective in multi-source and pixel-wise localization. Lee *et al.* [34] proposed a fully convolutional neural network (FCN) with an encoder-decoder structure for generating high-resolution source maps without pre-defining the number of sources. This approach outperformed conventional deconvolution methods in both accuracy and efficiency, achieving localization errors as low as 0.020 m.

Innovative AI architectures have been employed to handle specific challenges in sound source localization. For example, Bozkurtlar *et al.* [35] introduced the von-Mises ResNet (vM-B ResNet), which incorporates a novel von-Mises convolutional layer to manage periodic phase information. This model achieved reduced prediction errors in both quiet and noisy environments, outperforming traditional ResNet-based models. Meanwhile, Ko *et al.* [36] focused on real-time applications, demonstrating a multi-stream CNN that processes raw multi-channel acoustic data with high accuracy, tailored for low-power IoT devices such as the Raspberry Pi.

Finally, multimodal approaches are gaining traction in AI-driven SSL. Huang *et al.* [37] introduced audio-visual-language maps (AVLMaps), which combine audio, visual, and language features into a unified 3D spatial map using pre-trained multimodal models like AudioCLIP. This framework enables zero-shot navigation from natural language descriptions, achieving up to 50% higher recall in ambiguous scenarios. Such innovations highlight the transformative potential of AI-based multimodal fusion for robust SSL in real-world robotics applications.

While AI methods offer significant advancements in adaptability and accuracy, they also face challenges such as reliance on large labeled datasets, high computational requirements, and limited generalizability to unseen environments. Despite these hurdles, AI-driven approaches continue to push the boundaries of sound source localization, addressing evolving challenges across diverse and complex scenarios.

2.3. Hybrid methods

Hybrid methods, representing 40% (22 papers) of the reviewed studies, combine classical signal processing techniques with AI models, leveraging the strengths of both domains. Classical methods excel in reliable feature extraction and preprocessing, while AI techniques provide robust pattern recognition and decision-making capabilities. These integrated approaches have shown significant potential in addressing complex and dynamic challenges in SSL.

Hybrid methods have demonstrated exceptional performance in sound event localization and detection (SELD). For instance, Cao *et al.* [38] proposed a two-stage SELD method using GCC-PHAT features with a convolutional recurrent neural network (CRNN), reducing directional angle errors to 9.85°.

Similarly, Krause *et al.* [39] integrated GCC-PHAT and interchannel phase differences with CRNNs, achieving a 4° reduction in localization error and improved SELD scores. Min *et al.* [40] extended this approach by combining GCC-PHAT with principal component analysis (PCA) to enhance feature extraction for SELDnet models, reducing DOA errors to 18.5° and achieving a frame recall rate of 91.7%.

The integration of multiple data modalities is another hallmark of hybrid methods. Chen *et al.* [41] combined visual data and acoustic signals using Fourier-based polar histogram of oriented gradients (HOG) descriptors and hidden Markov models (HMMs), achieving enhanced localization accuracy in reverberant environments. Similarly, Grinstein *et al.* [42] introduced a dual-input neural network (DI-NN) that integrates classical metadata with spectrogram features, reducing localization errors significantly across various acoustic conditions. Liu *et al.* [2] reviewed hybrid approaches in elderly service robots, emphasizing the use of SLAM-based 3D reconstruction and multimodal fusion for seamless human-robot interaction.

Classical preprocessing techniques have been effectively integrated with AI models to enhance spatial resolution and robustness. Boztas [1] employed discrete wavelet transform (DWT) for feature extraction, feeding the results into CNNs and biLSTMs, achieving an R^2 of 0.97. Zhang *et al.* [43] combined conventional beamforming (CBF) with a densely connected convolutional neural network (DCFCN), which enhanced spatial resolution and dynamic range for precise SSL. Zhou *et al.* [44] proposed Acoustic-Net, which uses STFT for feature extraction and integrates RepVGG-B0 with multi-task learning for real-time localization, achieving localization errors of 0.0114 m.

Several studies used TDOA and GCC-PHAT features alongside machine learning to improve localization. Jaddoa *et al.* [45] combined TDOA-based GCC features with Restricted Boltzmann Machine (RBM) and long short-term memory (LSTM) networks, achieving over 99% accuracy in noisy environments. Wang *et al.* [46] enhanced GCC-PHAT with a speech-oriented masking technique, integrating machine learning classifiers like MLP and spiking neural networks (SNNs). Tang *et al.* [47] demonstrated the synergy of GCC and broad learning systems (BLS), achieving robust performance in high-reverberation and low-SNR environments. Li *et al.* [48] proposed GCC-Speaker, which uses speaker-dependent weighting functions derived from SpeakerBeam, improving localization accuracy in multi-speaker scenarios. Liu *et al.* [49] introduced a hybrid approach combining TDOA-based generalized cross-correlation (GCC) with machine learning classifiers such as SVM, KNN, and Naive Bayes, achieving 100% localization accuracy in outdoor field experiments without requiring microphone calibration. Zhang *et al.* [50] combined TDOA with neural networks for nonlinear fitting and Kalman filtering to integrate inertial measurement unit (IMU) data, reducing angular resolution errors from 5.45° to 1.1°, making it highly effective for dynamic applications like smart car navigation.

Hybrid methods have also drawn inspiration from biology to address real-world challenges. Davila-Chacon *et al.* [51] proposed a biomimetic binaural SSL system combining classical interaural time and level differences (ITD/ILD) with spiking neural networks, doubling sentence recognition rates in noisy environments. Similarly, Goto *et al.* [52] combined minimum variance distortionless response (MVDR) beamforming with LiDAR-generated 3D spatial data to improve sound localization and visualization, demonstrating robust performance in both direct and reflected sound scenarios.

Hybrid methods continue to push the boundaries with innovative algorithms. Hu *et al.* [53] combined GCC-PHAT with residual networks and channel attention modules, achieving 86.53% accuracy within a 5° error range. Tang *et al.* [10] used GCC with the incremental broad learning system (Enhance), achieving 97.2% accuracy in simulations under high-reverberation and low-SNR conditions. Feng *et al.* [54] combined improved GCC with CNNs, achieving root mean square errors below 5° in noisy conditions. Bulut *et al.* [55] combined wavelet-transformed acoustic emission signals with CNNs, achieving over 99% validation accuracy in structured environments. Grinstein *et al.* [6] introduced Neural-SRP, integrating steered response power (SRP) with a CRNN, reducing localization errors by 67% in reverberant environments.

Hybrid methods have emerged as a transformative approach to SSL, addressing complex acoustic challenges with unparalleled accuracy, robustness to noise, and computational efficiency. By balancing the simplicity of classical techniques with the adaptability of AI, they represent the most promising direction for advancing SSL, particularly in robotics and real-time applications. However, challenges such as integration complexity and computational overhead remain, highlighting the need for further research to optimize hybrid systems for practical deployment

3. COMPARATIVE ANALYSIS

3.1. Accuracy of SSL method

As shown in Figure 2, the average accuracy of SSL methods varies across categories:

a. Hybrid methods achieve the highest accuracy, averaging 97.45%. Their ability to integrate classical preprocessing techniques (*e.g.*, TDOA, GCC-PHAT) with adaptive AI models allows them to handle diverse acoustic conditions effectively. This makes them well-suited for complex, real-world robotic applications.

b. AI methods follow with an average accuracy of 96.52%. These methods excel in learning complex patterns and adapting to challenging environments. However, their slightly lower accuracy compared to hybrid methods may stem from challenges such as generalization and high computational demands.
c. Classical methods exhibit the lowest average accuracy, at 95.39%. While their mathematical simplicity and computational efficiency make them valuable, they are more susceptible to performance degradation in noisy or reverberant conditions.

The analysis highlights the trend that while classical methods remain competitive, hybrid methods offer the most robust solution by leveraging the strengths of both classical and AI approaches.

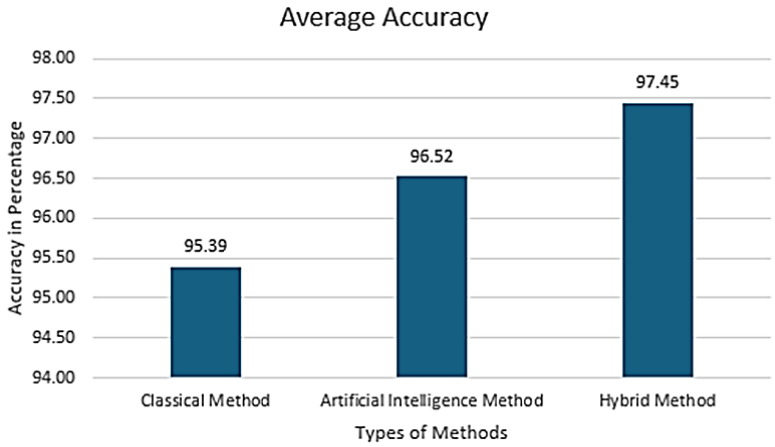

Figure 2. Average accuracy for SSL method according to its classification

3.2. Impact of microphone array size

Figure 3 illustrates the relationship between the number of microphones and SSL accuracy:

a. Single microphone systems achieve a surprisingly high accuracy of 97.60%, demonstrating the potential of optimized algorithms for resource-limited setups. These methods are particularly suitable for portable or low-power devices.
b. Two microphone systems exhibit slightly lower accuracy at 97.21%, indicating potential challenges with phase alignment and noise interference in dual-microphone configurations.
c. Three to ten microphone systems show a further drop in accuracy to 95.90%, likely due to the complexity of processing multi-channel signals and environmental factors such as reverberation.
d. Systems with ten or more microphones achieve the highest accuracy of 99.23%, benefiting from enhanced spatial resolution and robustness to noise. These configurations are ideal for applications requiring high precision in complex environments.

The results indicate that while larger microphone arrays provide significant accuracy improvements, single and dual-microphone systems remain viable for applications with hardware constraints, provided they are paired with well-optimized algorithms.

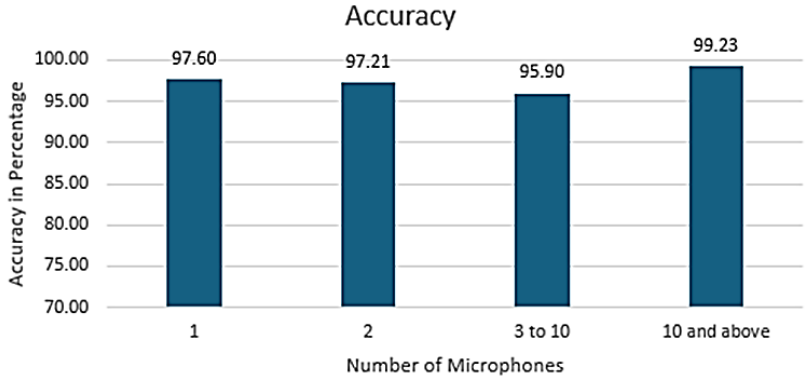

Figure 3. Average accuracy for SSL method according to its number of microphones used in its system

3.3. Practical implications

The findings suggest several practical considerations for SSL system design:

a. Hybrid methods and large arrays for high-performance applications: Hybrid methods, combined with microphone arrays of 10 or more channels, are the most effective for applications requiring high accuracy and robustness, such as autonomous navigation and human-robot interaction in noisy environments.
b. Cost-effective solutions for resource-constrained systems: Single and dual-microphone systems, paired with classical or AI methods, offer competitive accuracy for low-cost and portable devices, such as wearable robots or IoT systems.
c. Trade-offs in multi-microphone systems: Systems with 3 to 10 microphones require further optimization to manage signal complexity and computational loads, particularly for real-time applications in reverberant environments.

3.4. Challenges and future directions

Despite their advancements, each category of SSL methods faces specific challenges:

a. Classical methods:
   - Struggle in dynamic environments with noise and reverberation.
   - Future research should explore adaptive filtering techniques and integration with AI models to enhance robustness.
b. AI methods:
   - Dependence on large datasets and high computational resources limits scalability.
   - Future work should focus on lightweight AI architectures and semi-supervised learning to reduce data and resource requirements.
c. Hybrid methods:
   - Integration complexity and computational overhead remain significant hurdles.
   - Optimization of hybrid systems for real-time processing and energy efficiency will be crucial for broader adoption.

In addition, the exploration of novel sensor designs, such as biologically inspired systems or multimodal approaches, could further enhance SSL capabilities in robotics.

4. CONCLUSION

This review paper presents a comprehensive analysis of SSL methods, categorized into classical methods, AI methods, and hybrid methods. Each category demonstrates unique strengths and limitations. Classical methods are efficient and computationally lightweight but struggle with noise and reverberation. AI-based methods excel in handling complex, dynamic environments but face challenges with scalability and computational intensity. Hybrid methods, integrating classical signal processing with AI-driven models, emerge as the most effective approach, achieving the highest average accuracy (97.45%). Their adaptability and robustness make them particularly suitable for real-world robotic applications.

The review also evaluates the impact of microphone configurations on SSL performance. Systems with 10 or more microphones achieve the highest accuracy (99.23%), benefiting from spatial redundancy and robust signal processing. However, single and dual-microphone systems, when paired with optimized algorithms, deliver competitive accuracy, offering cost-effective solutions for resource-constrained applications.

In conclusion, hybrid methods combined with larger microphone arrays offer the most reliable and precise solutions for SSL in robotics and other advanced applications. Future research should focus on optimizing hybrid approaches for real-time and energy-efficient performance, enabling their deployment in dynamic and noisy environments. Additionally, exploring multimodal hybrid systems or lightweight AI techniques with minimal microphones could yield cost-effective solutions for a wider range of practical SSL applications. By addressing these challenges, SSL technologies will continue to evolve, driving advancements in robotics and auditory systems.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the Center of Excellence (CoE) Robotics and Sensing Technologies team of TMRND for their invaluable guidance and support throughout the development of this research paper.

FUNDING INFORMATION

This research was funded by Project SOLARIS (RDTC241129), a collaborative research program between Telekom Malaysia Research & Development (TMRND) and Multimedia University (MMU), Malaysia. The funding supported the design, implementation, and analysis of the study.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Muhammad Akmal Aliff	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Emerson Joseph Raja	✓	✓		✓	✓	✓		✓		✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest regarding the publication of this paper.

DATA AVAILABILITY

All data used in this study are obtained from publicly available research articles cited in this paper. The processed data, including the compiled dataset of reviewed studies and analysis results, are available from the corresponding author upon reasonable request.

REFERENCES

[1] G. Boztas, "Sound source localization for auditory perception of a humanoid robot using deep neural networks," *Neural Computing and Applications*, vol. 35, no. 9, pp. 6801–6811, Mar. 2023, doi: 10.1007/s00521-022-08047-x.

[2] X. Liu, C. Huang, H. Zhu, Z. Wang, J. Li, and A. Cangelosi, "State-of-the-art elderly service robot: environmental perception, compliance control, intention recognition, and research challenges," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 10, no. 1, pp. 2–16, Jan. 2024, doi: 10.1109/MSMC.2023.3238855.

[3] R. Hosangadi, "A proposed method for acoustic source localization in search and rescue robot," in *Proceedings of the 5th International Conference on Mechatronics and Robotics Engineering*, Feb. 2019, pp. 134–140. doi: 10.1145/3314493.3314510.

[4] J. Sun, P. Li, Y. Chen, H. Lu, D. Shao, and G. Chen, "Sound source identification algorithm for compressed beamforming," *Journal of Mechanical Science and Technology*, vol. 38, no. 4, pp. 1627–1634, Apr. 2024, doi: 10.1007/s12206-024-0301-z.

[5] Z. Heydari and A. Mahabadi, "Scalable real-time sound source localization method based on TDOA," *Multimedia Tools and Applications*, vol. 82, no. 15, pp. 23333–23372, Jun. 2023, doi: 10.1007/s11042-022-14256-2.

[6] E. Grinstein, T. van Waterschoot, M. Brookes, and P. A. Naylor, "The Neural-SRP method for positional sound source localization," *arXiv:2403.09455*, 2024, doi: 10.48550/arXiv.2403.09455.

[7] Z. Lin, K. Itoyama, K. Nakadai, and H. Amano, "FPGA-based low power acceleration of HARK sound source localization," in *2024 IEEE Symposium in Low-Power and High-Speed Chips (COOL CHIPS)*, Apr. 2024, pp. 1–6. doi: 10.1109/COOLCHIPS61292.2024.10531180.

[8] F. Grondin and F. Michaud, "Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations," *Robotics and Autonomous Systems*, vol. 113, pp. 63–80, Mar. 2019, doi: 10.1016/j.robot.2019.01.002.

[9] X. Jia, M. Jia, L. Li, and Y. Jia, "Multiple sound source localization by using diffuseness estimation," in *2020 7th International Conference on Information Science and Control Engineering (ICISCE)*, Dec. 2020, pp. 874–877. doi: 10.1109/ICISCE50968.2020.00181.

[10] R. Tang, Y. Zhang, Y. Zuo, B. Lin, and M. Liang, "Sound source localization algorithm of microphone array based on incremental broad learning system," *Circuits, Systems, and Signal Processing*, vol. 43, no. 3, pp. 1549–1571, Mar. 2024, doi: 10.1007/s00034-023-02521-0.

[11] N. C. Jamaludin, I. S. A. Halim, S. L. M. Hassan, and W. F. H. Abdullah, "Real-time FPGA-based sound source localization," in *2024 IEEE 14th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, May 2024, pp. 260–264. doi: 10.1109/ISCAIE61308.2024.10576608.

[12] Y. Pamungkas and Y. Rais, "Implementation of real-time sound source localization using TMS320C6713 board with interaural time difference method," in *2022 2nd International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*, Dec. 2022, pp. 269–274. doi: 10.1109/ISMODE56940.2022.10180999.

[13] F. Grondin *et al.*, "ODAS: Open embeddeD audition system," *Frontiers in Robotics and AI*, vol. 9, May 2022, doi:

10.3389/frobt.2022.854444.
[14] D. Salvati, C. Drioli, and G. L. Foresti, "Acoustic source localization using a geometrically sampled grid SRP-PHAT algorithm with max-pooling operation," *IEEE Signal Processing Letters*, vol. 29, pp. 1828–1832, 2022, doi: 10.1109/LSP.2022.3199662.
[15] M. Qin, D. Hu, Z. Chen, and F. Yin, "Compressive sensing-based sound source localization for microphone arrays," *Circuits, Systems, and Signal Processing*, vol. 40, no. 9, pp. 4696–4719, Sep. 2021, doi: 10.1007/s00034-021-01692-y.
[16] S. Gombots, J. Nowak, and M. Kaltenbacher, "Sound source localization – state of the art and new inverse scheme," *e & i Elektrotechnik und Informationstechnik*, vol. 138, no. 3, pp. 229–243, Jun. 2021, doi: 10.1007/s00502-021-00881-6.
[17] J. Zhao *et al.*, "A sound source localization method based on frequency divider and time difference of arrival," *Applied Sciences*, vol. 13, no. 10, p. 6183, May 2023, doi: 10.3390/app13106183.
[18] L. Still, M. Oispuu, and W. Koch, "Simultaneous sensor and sound source localization in urban environments," in *2023 26th International Conference on Information Fusion (FUSION)*, Jun. 2023, pp. 1–8. doi: 10.23919/FUSION52260.2023.10224230.
[19] R. Lee, M.-S. Kang, B.-H. Kim, K.-H. Park, S. Q. Lee, and H.-M. Park, "Sound source localization based on GCC-PHAT with diffuseness mask in noisy and reverberant environments," *IEEE Access*, vol. 8, pp. 7373–7382, 2020, doi: 10.1109/ACCESS.2019.2963768.
[20] M.-A. Chung, H.-C. Chou, and C.-W. Lin, "Sound localization based on acoustic source using multiple microphone array in an indoor environment," *Electronics*, vol. 11, no. 6, p. 890, Mar. 2022, doi: 10.3390/electronics11060890.
[21] M. Yang, X. Zhu, Y. Zhang, N. Ta, and Z. Rao, "Estimation of sound source directions using a biological coupled sensor array with a multistage iteration method," *Applied Acoustics*, vol. 177, p. 107960, Jun. 2021, doi: 10.1016/j.apacoust.2021.107960.
[22] I. An, Y. Kwon, and S. Yoon, "Diffraction- and reflection-aware multiple sound source localization," *IEEE Transactions on Robotics*, vol. 38, no. 3, pp. 1925–1944, Jun. 2022, doi: 10.1109/TRO.2021.3118966.
[23] H. Song and J. W. Shin, "Multiple sound source localization based on interchannel phase differences in all frequencies with spectral masks," in *Interspeech 2021*, Aug. 2021, pp. 671–675. doi: 10.21437/Interspeech.2021-1178.
[24] L. Zhou, M. Jia, L. Li, and L. Tao, "Sound source localization by combining phase consistency and angle deviation," in *Proceedings of the 2023 9th International Conference on Computing and Artificial Intelligence*, Mar. 2023, pp. 619–623. doi: 10.1145/3594315.3594381.
[25] A. Sieriebriakov, O. Hospodarchuk, D. Lakhtyr, V. Simakhin, and S. Bondar, "Technology of sound source localization based on incoming sound intensity," in *2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT)*, Oct. 2023, pp. 1–5. doi: 10.1109/CSIT61576.2023.10324229.
[26] K. Wang and M. Zhang, "Sound source localization system based on TDOA algorithm," in *Advances in Transdisciplinary Engineering, IOS Press BV*, 2024. doi: 10.3233/ATDE231207.
[27] M. Sewtz, T. Bodenmuller, and R. Triebel, "Robust MUSIC-based sound source localization in reverberant and echoic environments," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2020, pp. 2474–2480. doi: 10.1109/IROS45743.2020.9340826.
[28] D. Gala and L. Sun, "Moving sound source localization and tracking using a self rotating bi-microphone array," in *ASME 2019 Dynamic Systems and Control Conference, DSCC 2019, American Society of Mechanical Engineers (ASME)*, 2019. doi: 10.1115/DSCC2019-9024.
[29] A. Ihlefeld, N. Alamatsaz, and R. M. Shapley, "Population rate-coding predicts correctly that human sound localization depends on sound intensity," *eLife*, vol. 8, Oct. 2019, doi: 10.7554/eLife.47027.
[30] T.-H. Tan, Y.-T. Lin, Y.-L. Chang, and M. Alkhaleefah, "Sound source localization using a convolutional neural network and regression model," *Sensors*, vol. 21, no. 23, p. 8031, Dec. 2021, doi: 10.3390/s21238031.
[31] X. Huang, R. Xu, W. Yu, and T. Peng, "Research on structural sound source localization method by neural network," *EURASIP Journal on Advances in Signal Processing*, vol. 2023, no. 1, p. 54, May 2023, doi: 10.1186/s13634-023-01017-y.
[32] J. Zhou *et al.*, "Noise source localization using deep learning," *Geophysical Journal International*, vol. 238, no. 1, pp. 513–536, May 2024, doi: 10.1093/gji/ggae171.
[33] S. Sakavičius and A. Serackis, "Estimation of Azimuth and Elevation for multiple acoustic sources using tetrahedral microphone arrays and convolutional neural networks," *Electronics*, vol. 10, no. 21, p. 2585, Oct. 2021, doi: 10.3390/electronics10212585.
[34] S. Y. Lee, J. Chang, and S. Lee, "Deep learning-based method for multiple sound source localization with high resolution and accuracy," *Mechanical Systems and Signal Processing*, vol. 161, p. 107959, Dec. 2021, doi: 10.1016/j.ymssp.2021.107959.
[35] M. Bozkurtlar, B. Yen, K. Itoyama, and K. Nakadai, "Real time sound source localization using von-Mises ResNet," in *2024 IEEE/SICE International Symposium on System Integration (SII)*, Jan. 2024, pp. 466–471. doi: 10.1109/SII58957.2024.10417224.
[36] J. Ko, H. Kim, and J. Kim, "Real-time sound source localization for low-power IoT devices based on Multi-Stream CNN," *Sensors*, vol. 22, no. 12, p. 4650, Jun. 2022, doi: 10.3390/s22124650.
[37] C. Huang, O. Mees, A. Zeng, and W. Burgard, "Audio visual language maps for robot navigation," in *Experimental Robotics*, 2024, pp. 105–117.
[38] Y. Cao, Q. Kong, T. Iqbal, F. An, W. Wang, and M. Plumbley, "Polyphonic sound event detection and localization using a two-stage strategy," in *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019)*, 2019, pp. 30–34. doi: 10.33682/4jhy-bj81.
[39] D. Krause, A. Politis, and K. Kowalczyk, "Feature overview for joint modeling of sound event detection and localization using a microphone array," in *2020 28th European Signal Processing Conference (EUSIPCO)*, Jan. 2021, pp. 31–35. doi: 10.23919/Eusipco47968.2020.9287374.
[40] Y. Min, P. Xin, C. Xu, and H. Liu, "Detection and localization of sound events based on principal components analysis," in *2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, Jan. 2022, pp. 507–511. doi: 10.1109/ICCECE54139.2022.9712717.
[41] J. Chen *et al.*, "Multimodal fusion for indoor sound source localization," *Pattern Recognition*, vol. 115, p. 107906, Jul. 2021, doi: 10.1016/j.patcog.2021.107906.
[42] E. Grinstein, V. W. Neo, and P. A. Naylor, "Dual input neural networks for positional sound source localization," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2023, no. 1, p. 32, Aug. 2023, doi: 10.1186/s13636-023-00301-x.
[43] G. Zhang, L. Geng, and X. Chen, "Sound source localization method based on Densely connected convolutional neural network," in *2022 5th International Conference on Information Communication and Signal Processing (ICICSP)*, Nov. 2022, pp. 743–747. doi: 10.1109/ICICSP55539.2022.10050682.
[44] G. Zhou, H. Liang, X. Ding, Y. Huang, X. Tu, and S. Abbas, "Acoustic-Net: A novel neural network for sound localization and quantification," *ArXiv:2203.16988*, 2021.
[45] S. Jaddoa, R. Ali, M. Najm Abdullah, and B. F. Abed, "Localization of speaker using fusion techniques and neural network algorithms," *Wasit Journal for Pure sciences*, vol. 3, no. 2, pp. 172–184, Jun. 2024, doi: 10.31185/wjps.399.
[46] J. Wang, X. Qian, Z. Pan, M. Zhang, and H. Li, "GCC-PHAT with speech-oriented attention for robotic sound source

localization," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, May 2021, pp. 5876–5883. doi: 10.1109/ICRA48506.2021.9561885.
[47] R. Tang, Y. Zhang, T. Lu, and M. He, "Sound source location prediction method based on broad learning," in *2023 11th International Conference on Information Technology: IoT and Smart City (ITIoTSC)*, Aug. 2023, pp. 135–138. doi: 10.1109/ITIoTSC60379.2023.00030.
[48] G. Li, W. Xue, W. Liu, J. Yi, and J. Tao, "GCC-Speaker: target speaker localization with optimal speaker-dependent weighting in multi-speaker scenarios," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2023, pp. 1–5. doi: 10.1109/ICASSP49357.2023.10095551.
[49] S. Liu, H. Li, Q. Yu, Y. Zhao, and W. Zhang, "Research on sound source localization of multiple fixed targets based on machine learning and distributed arrays," *Journal of Physics: Conference Series*, vol. 2718, no. 1, p. 012069, Mar. 2024, doi: 10.1088/1742-6596/2718/1/012069.
[50] Y. Zhang, Z. Zeng, and D. Tang, "Application of neural network and Kalman filtering in sound source localization," in *2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, Aug. 2021, pp. 5–8. doi: 10.1109/AEECA52519.2021.9574193.
[51] J. Davila-Chacon, J. Liu, and S. Wermter, "Enhanced robot speech recognition using biomimetic binaural sound source localization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 138–150, Jan. 2019, doi: 10.1109/TNNLS.2018.2830119.
[52] M. Goto *et al.*, "Utilizing LiDAR data for 3D sound source localization," in *ACM SIGGRAPH 2023 Posters*, Jul. 2023, pp. 1–2. doi: 10.1145/3588028.3603682.
[53] F. Hu, X. Song, R. He, and Y. Yu, "Sound source localization based on residual network and channel attention module," *Scientific Reports*, vol. 13, no. 1, p. 5443, Apr. 2023, doi: 10.1038/s41598-023-32657-7.
[54] H. Feng, H. Zhang, Z. Shen, Z. Qiu, X. Zhang, and Z. Tao, "A study of sound source localization based on improved time delay features and the convolutional neural network," in *2024 20th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Jul. 2024, pp. 1–9. doi: 10.1109/ICNC-FSKD64080.2024.10702302.
[55] E.-B. Bulut, G. Manthei, I. Sirbu, and M. Guckert, "Research of acoustic emission characteristics and applicability of artificial intelligence for source localization," *e-Journal of Nondestructive Testing*, vol. 29, no. 10, Oct. 2024, doi: 10.58286/30243.

BIOGRAPHIES OF AUTHORS

Muhammad Akmal Aliff is a researcher currently working with TM Research and Development, Malaysia. He holds a Master of Engineering (MEng) in electronics and communication engineering from the University of Nottingham, United Kingdom. With a strong foundation in cutting-edge technologies, Akmal's professional and research interests focus on Internet of Things (IoT), robotics, and sensing technologies. He can be contacted at akmalaliff@tmrnd.com.my.

Emerson Joseph Raja is currently an assistant professor in the Faculty of Engineering and Technology at Multimedia University (MMU), Malaysia. His technical expertise centers around applying artificial intelligence techniques to monitor the health of machines and enhancing accuracy in robotics. He has published several conference and journal papers in his area of research. He is a senior member of IEEE. He was honored with the notable commendation award by his alma mater, SRM University, India. He was also honored with the best executive award and group CEO merit award for the year 2014 from TM, the leading integrated telecommunication company in Malaysia. His profile can be found at https://mmuexpert.mmu.edu.my/emersonraja. He can be contacted at emerson.raja@mmu.edu.my.