Postprint: Research on Spatial Microphone Localization Method Based on Source Arrays

Authors: Chen Xiaohui, Sun Hao, Zhang Heng, Zhai Baoshuo

Date: 2019-04-01T00:00:00+00:00

Abstract

To address the issue that existing microphone position estimation methods are limited in number and only perform position estimation for planar microphone arrays, this paper proposes a spatial microphone localization method based on a sound source array. This method arranges four sound sources in a regular tetrahedron configuration, leveraging its symmetry to ensure omnidirectional localization capability while calculating the distances between the microphone and each sound source using a time-delay-estimation-based ranging approach. Subsequently, the microphone position coordinates are determined based on geometric relationships, thereby achieving online spatial microphone position estimation based on a sound source array. The feasibility and effectiveness of the proposed method are verified through MATLAB simulation experiments.

Full Text

Preamble

Vol. 37 No. 5
Application Research of Computers
ChinaXiv Cooperative Journal

Microphone Spatial Positioning Method Based on Sound Source Array

Chen Xiaohui, Sun Hao†, Zhang Heng, Zhai Baoshuo
(School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin 300130, China)

Abstract: Current microphone position estimation methods are scarce and primarily limited to planar microphone arrays. To address this limitation, this paper proposes a spatial microphone positioning method based on a sound source array. The method arranges four sound sources in a regular tetrahedron configuration, leveraging its symmetry to ensure omnidirectional positioning capability. Based on time delay estimation ranging techniques, the distances between microphones and each sound source are calculated, after which microphone position coordinates are determined through geometric relationships to achieve online spatial microphone position estimation using a sound source array. MATLAB simulation experiments validate the feasibility and effectiveness of the proposed method.

Keywords: sound source array; time delay estimation; time difference of arrival; microphone positioning

0 Introduction

With the rapid advancement of science and technology, sensor applications have expanded into increasingly complex environments. Accurate microphone positioning facilitates distributed microphone array signal processing, which plays a crucial role in speech enhancement, sound source localization, and source separation applications. Although research on microphone array signal processing has spanned nearly a decade, most work has been conducted abroad and remains in the exploratory stage, while domestic research in this area has been limited, yielding few substantial results. Precise microphone position coordinates can also address numerous challenges in robotics. For instance, when a robot operates in an unknown environment without knowledge of its own pose, simultaneous self-localization and mapping (SLAM) has long been a hot yet difficult topic in intelligent robotics research. From the perspective of robotic auditory systems, determining the position coordinates of microphones to be localized and integrating robotic technology with distributed microphone positions in the environment can help solve SLAM problems, thereby enabling fully autonomous robot navigation. Consequently, determining microphone positions within microphone arrays represents a critical technical challenge that must be resolved.

Reference [11] employs three calibration sound sources at different positions to correct microphone array positions. This approach places a single sound source at three locations sequentially, performing signal acquisition and calculation three times before fusing the results to obtain microphone positions. However, this method involves cumbersome signal acquisition operations, suffers from low efficiency, and only supports offline calculation, making online microphone position estimation impossible. Reference [12] achieves online microphone position estimation by selecting three sound signals in different frequency bands as source signals, enabling simultaneous acquisition and separation of three sound sources during the estimation process. This method combines energy-based and time delay-based ranging for coarse and fine distance estimation, respectively, and uses a triangular centroid algorithm to obtain microphone positions in the array, with corresponding handling for various unsolvable cases. This approach ensures low computational complexity and high accuracy while maintaining continuity and stability of positioning results. However, it only addresses planar microphone arrays and cannot estimate spatially distributed microphone positions.

To overcome these limitations, this paper proposes a spatial microphone positioning method based on a sound source array that achieves high accuracy, low computational complexity, and online positioning capability. Within a defined spatial region, a certain number of microphones are randomly distributed. The sound source distribution unit consists of four sound sources arranged in a tetrahedral configuration—specifically, a regular tetrahedron. The sound sources emit signals sequentially, and microphones that receive all four signals are designated as microphones to be localized. Using time delay estimation, the time difference of arrival (TDOA) between the reference microphone and each microphone to be localized is obtained. Finally, the position of each microphone is determined through geometric relationships. This method leverages the symmetry of the regular tetrahedron to ensure omnidirectional positioning capability while guaranteeing uniform signal reception by microphones, thereby satisfying positioning accuracy requirements and ensuring continuous and stable performance.

1 Sound Source Array Model

This paper arranges the sound source distribution unit in a tetrahedral configuration by establishing a three-dimensional coordinate system and spatially distributing four sound sources in a regular tetrahedron formation. This arrangement utilizes the symmetry property to ensure uniform signal reception by microphones, thereby satisfying positioning accuracy requirements. All subsequent spatial microphone position representations use this coordinate system as a reference. The sound source array model is illustrated in [Figure 1: see original paper].

In the three-dimensional coordinate system (x, y, z) with origin o, S , S , S , and S represent the four sound sources. C is the reference microphone, located within 10 meters of each sound source. M denotes the first microphone to be localized—one of the target microphones for positioning. L represents the known and fixed distance from reference microphone C to sound source S . Similarly, the distances from C to the other sound sources (S , S , S), designated as L , L , and L , are also known and fixed. d is the distance from the first microphone to be localized (M) to the first sound source (S), calculated from the time difference between S ' s arrival at C and M multiplied by the speed of sound, then combined with the known distance L . By extension, the distances from M to the other sound sources can be computed—a necessary condition for determining M ' s spatial position.

2 Time Delay Estimation Algorithm

Time delay estimation primarily employs signal processing theory and methods to estimate the time difference between signals received by different sensors, thereby determining related parameters such as source distance, direction, velocity, and movement [13,14]. Based on measurement environment, conditions, and signal characteristics, time delay estimation methods can be categorized into phase-based methods, bispectral methods, correlation methods, and adaptive filter parametric model methods. Among these, the correlation method is the most classical and commonly used.

2.1 Basic Cross-Correlation Method

The sound source distribution unit generates sound signals sequentially, which are received by both reference and target microphones. We assume an ideal signal model where only environmental noise is considered, approximated by Gaussian white noise that is mutually uncorrelated between noise components and between noise and source signals. The mathematical model of the received signal is given by Eq. (1).

Let the source signal be $s(t)$, and the signal received by the i-th microphone be $x\ (t)$, where , , and $n\ (t)$ represent the amplitude attenuation coefficient, time delay, and environmental noise for the source signal reaching the i-th microphone, respectively. The cross-correlation function between the i-th and j-th microphones is obtained by substituting Eq. (1) into Eq. (2). $R\ (\ )$ is the autocorrelation function of $s(t)$, $R\ (t)$ is the cross-correlation function between noise components, and R and R are cross-correlation functions between noise and source. represents the relative time delay between signals received by the i-th and j-th microphones. Since noise components and noise-source pairs are uncorrelated, Eq. (3) simplifies to Eq. (4). Based on the properties of autocorrelation functions, $R\ (\ -\ )$ reaches its maximum when $=$ [15]. Therefore, the time delay estimate can be obtained from the peak location of the cross-correlation function between the two microphone signals.

2.2 Generalized Cross-Correlation Method

The generalized cross-correlation (GCC) algorithm is currently the most widely used time delay estimation method [16]. In practical environments with significant noise and reverberation, the basic cross-correlation function may exhibit indistinct or multiple peaks. GCC offers better performance in peak detection by suppressing noise and reverberation. Applying the Fourier transform to Eq. (4) yields Eq. (5), where $(\ )$ is the power spectrum of $R\ (\ )$, and $(\ )$ is the cross-power spectrum between $x\ (t)$ and $x\ (t)$.

Subsequently, applying inverse transformation after weighting Eq. (5) produces the generalized cross-correlation function shown in Eq. (6), where $(\ )$ is the generalized weighting function [17]. Common weighting functions include Roth, SCOT (Smoothed Coherence Transform), and PATH (Phase Transform), each with distinct characteristics suitable for different applications. Due to its superior noise and reverberation suppression capabilities compared to basic cross-correlation, this paper adopts the GCC time delay estimation algorithm. The PATH weighting function exhibits low fluctuation under varying SNR conditions and provides good noise resistance [18], making it suitable for our positioning simulation analysis.

3 Microphone Position Estimation

In practical environments, accurately placing multiple microphones in the same plane is challenging, introducing positioning errors for planar arrays. In contrast, randomly distributing microphones in space is relatively easier, reducing assembly difficulties. Furthermore, spatial microphone position estimation can help solve numerous problems in sensor technology and robotics. The spatial microphone position estimation system described in this paper comprises a sound source distribution unit, a reference microphone, microphones to be localized, and a microphone positioning unit. Within a defined spatial region, multiple microphones are randomly distributed. The sound source distribution unit consists of four sound sources arranged as described in Section 1, emitting signals sequentially to produce four sound signals. Microphones in the space that receive all four signals are designated as microphones to be localized. The time differences (TDOAs) between the reference microphone and each microphone to be localized are transmitted to the microphone positioning unit. After obtaining TDOAs through the GCC method described above, the positioning unit determines the positions of the microphones to be localized based on these time differences and their geometric relationships with the reference microphone.

The flowchart of the spatial microphone positioning method based on the sound source array is shown in [Figure 2: see original paper]. As illustrated, S , S , S , and S represent the four sound sources generating signals received by the reference microphone and microphones to be localized (M , M , ···, M). The classic GCC method estimates the time differences of arrival (TDOAs) between the reference microphone and each microphone to be localized for the four sequentially emitted sound sources, designated as TDOA , TDOA , TDOA , and TDOA , where t represents the emission time of each source and i = 1,2,3,4 indexes the four sources. The microphone positioning unit takes these TDOAs as input and outputs the three-dimensional coordinates of each microphone to be localized (M , M , ···, M), thereby completing the spatial microphone positioning based on the sound source array.

The positioning unit operates as follows: Let ΔT denote the time difference between the reference microphone and the i-th microphone to be localized when receiving the j-th sound signal. With C representing the speed of sound in air, multiplying the time difference by C yields the distance difference ΔL between the j-th sound source and the reference microphone versus the i-th microphone to be localized, as expressed in Eq. (7).

Based on the known distance L from the reference microphone to the j-th sound source and the distance difference ΔL , the distance d from the i-th microphone to be localized to the j-th sound source is obtained by $d = L + \Delta L = L + \Delta T \times C$.

According to geometric relationships, the position of the i-th microphone in space is determined by solving for the intersection point of four spheres centered at the four sound source positions with radii equal to the corresponding distances from the microphone to each source. This yields the system of equations:

$$\begin{cases} (x_i - x_{S1})^2 + (y_i - y_{S1})^2 + (z_i - z_{S1})^2 = d_{i1}^2 \\ (x_i - x_{S2})^2 + (y_i - y_{S2})^2 + (z_i - z_{S2})^2 = d_{i2}^2 \\ (x_i - x_{S3})^2 + (y_i - y_{S3})^2 + (z_i - z_{S3})^2 = d_{i3}^2 \\ (x_i - x_{S4})^2 + (y_i - y_{S4})^2 + (z_i - z_{S4})^2 = d_{i4}^2 \end{cases}$$

Since the four sound sources are distributed in a regular tetrahedron, the vectors formed between any pair of sources are linearly independent, ensuring that this system has a unique solution. This solution represents the intersection point of the four spheres, which corresponds to the spatial coordinates of the i-th microphone to be localized. [Figure 3: see original paper] illustrates this geometric model: S , S , S , and S are the four sound sources, and d , d , d , and d are the distances from the first microphone to be localized (M) to each source. Constructing spheres centered at the source positions with radii equal to the corresponding distances, the linear independence of source vectors guarantees a single intersection point in space, which is the coordinate of microphone M (shown as the black triangle). [Figure 4: see original paper] provides a more intuitive three-dimensional visualization of the four spheres intersecting at a single point, with the red dot representing a microphone to be localized.

4 Experiments and Analysis

To verify the feasibility and effectiveness of the proposed method, MATLAB simulation experiments were conducted based on TDOA data obtained through the GCC method. The simulation configured a spatial region of 60 m × 60 m × 20 m, as shown by the coordinate axes in [Figure 5: see original paper]. Within this region, four hollow circles ( ) mark the positions of the sound sources at coordinates (6, 0, 0), (-3, √3, 0), (-3, -√3, 0), and (0, 0, 6√2) m. The coordinate origin (0, 0, 0) serves as the reference microphone position ( ). The space contains 200 randomly distributed microphones (+), among which two received all four sound signals and were designated as microphones to be localized ( ). Since these two microphones are part of the 200 randomly distributed microphones, their symbols ( ) overlap with two of the "+" symbols in [Figure 5: see original paper]. Using the proposed method, the positions of these two microphones were estimated, with results indicated by " " symbols in [Figure 5: see original paper]. The " "positions represent the estimated locations, while the " "positions show the actual ground-truth locations of the microphones to be localized.

The spatial microphone positioning results are presented in . The table demonstrates that under certain SNR conditions, the three-dimensional coordinates of microphones can be accurately estimated with small positioning errors.

5 Conclusion

This paper addresses the microphone position estimation problem by proposing a spatial microphone positioning method based on a sound source array. The method employs four sound sources arranged in a regular tetrahedron as the distribution unit. Using the generalized cross-correlation time delay estimation method, it obtains the time differences of arrival between the reference microphone and microphones to be localized. These TDOAs are used to calculate distances between microphones and each sound source, after which geometric relationships determine microphone positions by solving for the intersection point of four spheres centered at the sound sources. The symmetrical distribution of the four sources ensures omnidirectional positioning capability and uniform signal reception, thereby satisfying accuracy requirements and ensuring continuous, stable performance. This method achieves high precision, low computational complexity, and online positioning capability for spatial microphone position estimation based on a sound source array. Spatial microphone positioning represents a novel and highly promising technology, and its application to fields such as acoustics and robotics warrants further investigation.

References

[1] Sun Hao. Study on sound source localization for mobile robot based on time delay of arrival [D]. Tianjin: Hebei University of Technology, 2012.

[2] Chen Guoliang, Xu Yang, Huang Xiaoqin. Sound source localization system for robot based on rectangular pyramid structure [J]. Application Research of Computer, 2019, 36(5).

[3] Yook D, Lee T, Cho Y. Fast sound source localization using two-level search space clustering [J]. IEEE Trans on Cybern, 2017, 46(1): 20-26.

[4] Li X, Liu H. Sound source localization for hri using FOC-based time difference feature and spatial grid matching [J]. IEEE Trans on Cybernetics, 2013, 43(4): 1199-1212.

[5] Benesty J, Chen J, Huang Y, et al. On microphone-array beamforming from a MIMO acoustic signal processing perspective [J]. IEEE Trans on Audio Speech & Language Processing, 2007, 15(3): 1053-1065.

[6] Silva B D, An B, Steenhaut K, et al. Design considerations when accelerating an FPGA-based digital microphone array for sound-source localization [J]. Journal of Sensors, 2017 (4): 1-20.

[7] Zhang Leiyue, Zhang Xinggan, Liu Chao. The improvement of time delay estimation in the microphone array sound localization system [J]. Journal of Nanjing University: Natural Sciences, 2015, 51(1): 25-30.

[8] Ke Wende, Cai Zesu, Li Jialan. Research of simulation localization and mapping in robot [J]. Application Research of Computers, 2010, 27(4): 1216-1219.

[9] Luo Yuan, Su Qin, Zhang Yi, et al. Simultaneous localization and mapping implementation based on optimized RBPF [J]. Journal of Huazhong University of Science and Technology: Natural Science Edition, 2016, 44(5): 30-34.

[10] Evers C, Naylor P A. Acoustic SLAM [J]. IEEE/ACM Trans on Audio Speech & Language Processing, 2018, PP(99): 1-1.

[11] Xiao Hua. Research on correction method of microphone array [D]. Chengdu: University of Electronic Science and Technology of China, 2008.

[12] Wang Shuwen. Study on distributed microphone array localization method [D]. Dalian: Dalian University of Technology, 2013.

[13] Hang Hongyan, Tang Juan. Analysis and survey of algorithms for time-delay estimation [J]. Technical Acoustics, 2008, 27(1): 110-114.

[14] Chen Z, Li Z, Wang S, et al. A microphone position calibration method based on combination of acoustic energy decay model and TDOA for distributed microphone array [J]. Applied Acoustics, 2015, 95: 13-19.

[15] Hang Hongyan, Zhao Shouguo, Di Jizheng, et al. Adaptive realizing form of generalized correlation algorithm for time delay estimation [J]. Journal of Xi' an Petroleum Institute: Natural Science Edition, 2001, 16(6): 47-50.

[16] Mao Huida, Zhang Linghua. Research on generalized cross correlation algorithm for time delay estimation in sound source localization [J]. Computer Engineering and Applications, 2016, 52(22): 138-142.

[17] Zhang Li. Research on sound source localization method based on microphone array [D]. Chengdu: University of Electronic Science and Technology of China, 2007.

[18] Sun Yang, Xu Kai, Yang Hailiang. Performance analysis of time-delay estimation based on generalized cross-correlation algorithm [J]. Computer and Digital Engineering, 2013, 41(1): 33-34.

Note: Figure translations are in progress. See original paper for figures.