Multiple Sound Sources 3D Localization System Using TDOA and Tetrahedral Microphone Array

Cong Xu, Ravi Sankar

iCONS Research Lab, Department of Electrical engineering, University of South Florida,Tampa 33647, US.

Abstract: In this paper, we propose an acoustic positioning and tracking system using time difference of arrival (TDOA) for multiple sound source positions in a three-dimensional (3D) space. The microphone array uses a regular tetrahedron structure and only requires four sensors to realize the position analysis of the sound source. We develop a fast cross-correlation algorithm to obtain characteristic peaks quickly from the TDOA data, which can effectively distinguish the phases of different sound sources and locate them in a multi-source environment. The optimized TDOA algorithm uses a minimum number of 3D sound source capture sensor arrays, which can capture the 3D coordinates of every sound source while performing low-latency multiple signal classification (MUSIC) calculations. The development of this technology will have a wide range of applications in real-time sound source coordinate acquisition in the field of multi-source acoustic visualization.

Keywords: Multi-Source Localization; TDOA; Fast Feature Peak Discrimination; Multiple Signal Classification; 3D Sound Source Visualization

1. Introduction

The main research directions of sound source location identification are divided into two areas: two-dimensional (2D) source capture techniques based on azimuth and source depth [2][4][5][6], and three-dimensional (3D) source location identification considering height space coordinates [1][3][7][8][9]. Since 3D acoustic localization can locate sound sources more accurately in various situations, it is more general in practicality than 2D sound source identification. However, since 3D localization usually uses a large number of sensors to build a mesh array, it becomes extremely challenging to process multiple sets of microphone information in real time, making it difficult to distinguish and localize multiple source signals simultaneously while ensuring low latency.

2. Multi-source acoustic recognition system

2.1 System overview

The sound source recognition system is mainly composed of a hardware system based on a stereo array of microphones and a digital signal processing module, and a software system for analyzing sound source information and coordinate calculation.

Generally, a characteristic peak refers to a high-energy signal that is significantly different from other sound signals in a segment of audio information. The system can estimate the phase difference between the two signals in the same channel by comparing the differences in the positions of the characteristic peaks between different signals. The differentiated signals are combined with the sound source characteristic peaks obtained in the first step to further determine the initial phases of the sound sources in different channels and use these phase differences to obtain the TDOA value according to the sampling frequency setting.

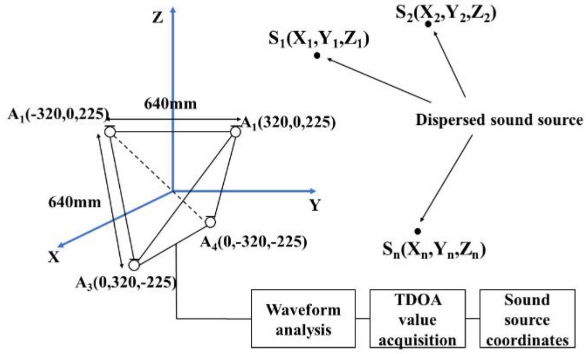

Fig 1 Overview of 3D sound source localization system

2.2 Sound Source 3D coordinate fitting algorithm

In this study, we focus on using the time difference generated during the sound transmission process to calculate the sound source position and fit the specific 3D coordinates of the sound source in the system. Without considering the effects of thermodynamic air convection on sound propagation, air can be approximately regarded as a uniform medium in the process of sound source propagation in the air. In a homogeneous medium, the sound wave spreads along the sphere along the radial direction, and the air propagation velocity equation is shown in equation (1).

$$v_{air} = 331 \times \sqrt{1 + \frac{T}{273}}(m/sec)/$$

Where v_{air} is the propagation speed of sound waves in the air, T is the specific temperature of the air medium in the environment in degrees Celsius. From the sound velocity information in Equation (1), we can derive the trajectory equation of sound wave conduction to various places in3D space at any specific moment by Equation (2) for each individual sound source S_i (x_i, y_i, z_i).

$$\sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} = v_{air} \times t_{trans}/$$

The time difference of arrival (TDOA) refers to the corresponding delay when the sound source receives the information of the sound source due to the difference of the linear distance between the sensors in the microphone array and the sound source during the transmission of sound. Combined with the relative Cartesian orthogonal coordinates of each sensor for the origin of the data, the relationship between the time difference of arrival between the sensors and the position information of the sound source can be deduced, as shown in Equation (3).

$$\begin{cases} S_{Si-A1} - TDOA_{0-A1} * V_{air} - S_{Si-A0} \\ S_{Si-A2} - TDOA_{0-A2} * V_{air} - S_{Si-A0} = 0/ \\ S_{Si-A3} - TDOA_{0-A3} * V_{air} - S_{Si-A0} \end{cases}$$

Where S represents the distance between the sound source and each sensor. It is worth noting that Ai does not represent the number of the sensor, but the order in which the sound source signal arrives between the sensors.

3. Experiment

3.1 Multi-source acquisition and separation

In the coordinates of the sound source recognition system, the center of the sensor array is taken as the origin of the coordinates. When the unit length is 1mm, the positions of the four microphones are: A1(0, -320, 225), A2 (0, 320, 225), A3(320, 0, -225) and A4(-320, 0, -225). The distance between two adjacent microphone sensors is 680mm.

The two sound sources are respectively located at S1 (-2000, 0, 0) and S2 (0, 1000, -1000) in the three-dimensional rectangular coordinate system in the system. The specific site layout is shown in Figure 4, the two sound sources It is composed of Bluetooth speakers remotely connected to the host computer to achieve simultaneous emission of different sound waves. The sound source signal is two different percussion instruments (big cymbals and cloud gongs) playing different audio information [10] in two Bluetooth speakers at the same time under computer control.

Active signal extraction uses Equation (4) to detect power anomalies in the microphone channel, and intercept audio information when the instantaneous power is higher than the average power. Since the distance between each microphone pair is 640mm, the signal window width is set to 20 based on the preset sampling frequency of the system.

The comparisons of the waveforms between the four channels and the short-time Fourier transform spectrograms shown in Figure 5 could roughly estimate the amount of audio present in the respective channel. Two different spectral density distribution curves can be identified from Figure 5, in which the vocal frequency of big cymbal (S1) is slightly higher than that of cloud gong (S2). However, because the TDOA interval between the channels is ephemeral, it is difficult to directly observe the phase difference between the channels from the spectrogram. After the system triggers the collection of sound source information, it records audio information with a length of one second in four channels at the same time, which is used to separate the sound source information. Since the sound source information in the channels overlaps and cannot be intuitively obtained from a single sound source information, it is necessary to adopt Multi-source Signal Differentiation Algorithm (MSDA) to strip the single sound source signal in each channel.

The principle of MSDA is to estimate the amplitude of the signal sequence based on the sequence and time difference of the active signal detected in the sensor. Here N represents how many channels receive the sound source signal at the same time when the sensor detects the active signal for the first time. The logic analysis of the arrival sequence using MSDA can iteratively calculate the signal difference with the arrival time difference in the sensor array and separate independent signals. Due to the presence of passive environmental noise and electromagnetic interference in the channel, the accuracy of the source information estimated by MSDA is highly positively correlated with the signal-to-noise ratio in the channel. In this experiment, the MSDA iterative calculation was used to compare the results of the original sound source signal waveform. The specific results are shown in Table 1.

Table 1 Comparison of original signal and MSDA results

Source signal Verses MSDA results		
Evaluation Parameters	***S1 estimation***	***S2 estimation***
MSE	0.001762	0.001741
RMSE	0.041983	0.04172
R SQURE	0.665201	0.629829

The reliability of the algorithm in terms of audio source separation in outdoor environments can be confirmed by comparing the two sets of signals estimated by MSDA with the original source signals. The part that differs from the MUSIC algorithm for signal separation is that MSDA performs logical iterative reasoning by cross-ratio correlation of sources between channels, which can significantly increase the computational speed in cases where multi-channel audio data needs to be processed in real time.

3.2 TDOA-based 3D multi-source localization

Once acoustic source estimation information is obtained, TDOA values are calculated for each channel according to the corresponding positions of the different sources between the channels. Due to the structural characteristics of the sensor array, signal aliasing is generated in each channel according to the time difference in the arrival of different sound sources at each of the sensor. Therefore, the correlation coefficient and delay difference between the sound source information and the channel signal can be obtained by comparing the correlation coefficient and delay difference between the sound source information and the channel signal. The different arrival timings between the two sound sources in the channel are processed. Therefore, TDOA value could be calculated for each microphone arrival coordinate successively.

During this experiment, the TDOA values of the two sound sources in four different signal channels were determined by frame-by-frame calculation between each channel. The estimation of the position of different sound sources is based on the calculation of the three-dimensional coordinates represented by the different TDOA information of the respective signals in these four channels. Because the data of each channel represents the sound pressure information collected at a specific coordinate in the 3D space at the same time, the $S1_{arrival}$ phase value and $S2_{arrival}$ phase value of each of the four channels are used to know the difference between each channel using Equation 3. Since $S1_{arrival}$ and $S2_{arrival}$ respectively represent the starting point of the time index when the sound source in the sound channel first arrives at the sensor, noise interference will cause errors in the location of the sound source starting point when calculating the sound source coordinates. The range of error generated by TDOA estimated coordinates is negatively correlated with sampling frequency, and the error value can be reduced by increasing the system sampling frequency.

Table 2 Multi-source detection reliability table

Multi-source detection reliability table				
	1 Source	2 Sources	3 Sources	4 Sources
Dispersion interval	10.2cm	13.6cm	40.8cm	81.6cm
Estimated Accuracy	94.9%	93.2%	79.6%	59.2%
Experimental samples	30 samples per test			
System recognition accuracy	3.4 cm			
Speed of sound	340 m / sec			
Sound source level	80 dB			
Environmental noise	35 dB (by average)			

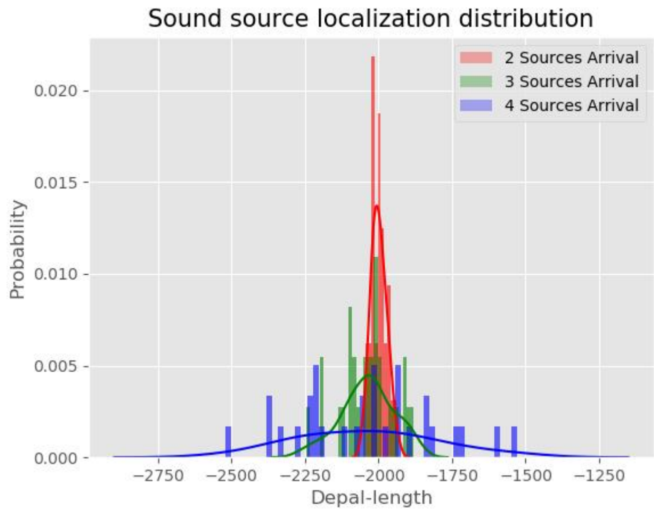

Fig 2 Multi-source MSDA coordinate fitting results distribution statistics

4. Conclusion

The multi-sound source separation and localization system is established using TDOA and MSDA algorithms. The detection of the three-dimensional position information of multiple sound sources and the separation from simultaneous arrival of multiple sound sources were conducted. From the results, a three-dimensional acoustic sensor with a minimum number of detectors based on a regular tetrahedron can effectively detect the coordinates of multiple sound sources in three-dimensional space. The experiment also proves that multi-source sound sources can be captured and located in space with a minimum of four sensors. Although the accuracy of the system to detect the three-dimensional coordinates of multiple sound sources under outdoor conditions needs to be improved, the application of this technology will be able to provide effective assistance in the field of miniaturized three-dimensional sound source positioning.

References

[1] Michaud S, Faucher S, Grondin F, et al., “3D Localization of a Sound Source Using Mobile Microphone Arrays Referenced by SLAM” 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020 , pp. 10402–10407.

[2] Iyama1 T, Sugiyama1 O, Otsuka1 T, et al., “Visualization of auditory awareness based on sound source positions estimated by depth sensor and microphone array” 2020 I 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014), 2014, pp. 1908–1913.

[3] Sasaki Y, Tanabe R and Takemura H, “Probabilistic 3D Sound Source Mapping using Moving Microphone Array” 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016, pp. 1293–1298.

[4] Alex C, Jose KM, Joseph A, “Sound Localization and Visualization Device” IEEE 2013 Global Humanitarian Technology Conference, 2013, pp. 362–364.

[5] Wakatsuki N, Ebihara T, Mizutani K and Aoki T, “Sound Source Visualization System in Reflective Environment Using Time-Reversal Wave” 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), 2019, pp. 307–310.

[6] Moravec M, Badida M, Pinosová M, Dzuro T and Badidová A, “Innovative Application Options of Sound Visualization Tools” 2019 International Council on Technologies of Environmental Protection (ICTEP) , 2019, pp. 191–194.

[7] Pourmohammad A, and Ahadi SM, “Real Time High Accuracy 3-D PHAT-Based Sound Source Localization Using a Simple 4-Microphone Arrangement.” IEEE Systems Journal, 2012, pp. 455–468.

[8] Kawanishi M, Maruta R, Ikoma N, Kawano H and Maeda H, “Sound Target Tracking in 3D using Particle Filter with 4 Microphones” SICE Annual Conference 2007, 2007, pp. 1427–1430.

[9] Kunin V, Jia W, Turqueti M, Saniie J and Oruklu E, “3D Direction of Arrival Estimation and Localization Using Ultrasonic Sensors in an Anechoic Chamber” 2011 IEEE International Ultrasonics Symposium Proceedings, 2011, pp. 756–759.