Multi-Sound Source Localization in Time Domain Using Voting Mechanism

Shih-Tsung Yang, Kai-Wen Liang, and Pao-Chi Chang
Department of Communication Engineering
National Central University
Taoyuan City, Taiwan
{styang.vaplab@gmail.com, kwistron@gmail.com, pcchang@ce.ncu.edu.tw}

***Abstract*— This work proposes a source localization method that can detect multiple source locations without knowing the number of sources in advance. For a fixed microphone array arrangement, we can estimate the relative delay relationship between each microphone pair. The arrival angle can be calculated by the correlation operations based on the relative delay relationship. Then, the signals received by the microphone array are calculated and used to search for the most relevant angle patterns. Finally, the multi-angle sound source is estimated based on the voting system.**

***Keywords*— Microphone array, Correlation calculation mode, Voting mechanism**

I. INTRODUCTION

Technologies for locating sound sources using microphone arrays have been widely used in many applications, such as speaker tracking, teleconferencing systems, and autonomous driving technologies. The direction of arrival (DOA) technique is mainly used to estimate the angle and direction information of the sound source through the relative time delay between the microphone elements for the sound source signal.

Many DOA estimation methods exist, such as SRP-PHAT, GCC-PHAT, TDOA, *etc*. The time delay is estimated by calculating the correlation between the sound source signals, and most of them are processed in the frequency domain[1][2]. However, traditional DOA estimation methods[3][4] may cause confusion or even errors when multiple sound sources exist, especially when the number of sources is unknown. Unlike common DOA estimation methods, this study only performs the product between signals in the time domain, and combines this product with the time delay estimated by TDOA to establish an angle model. Then, the received signal is used to search for the most relevant mode from each angle. Multiple sound sources can easily be detected without increasing the computational complexity. Finally, the proposed voting mechanism is used to accumulate the relative number of votes and then estimate the angles of the sound source signals.

II. METHODOLOGY

A. Proposed system block diagram

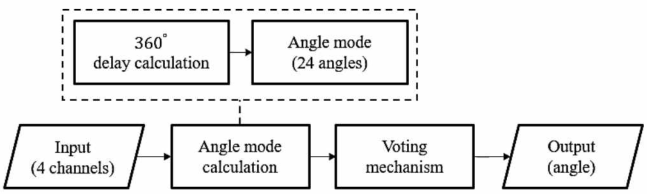


Fig. 1. Block diagram of the proposed method

We use a microphone array with four microphone elements for illustration as shown in figure 1. The input is a four-channel sound source signal. Then, the received signal is sent to the preset angle model to search for the angle patterns with the highest correlations. Finally, the angles of the sound source signals are generated through the voting mechanism.

B. Delay estimation of angle mode

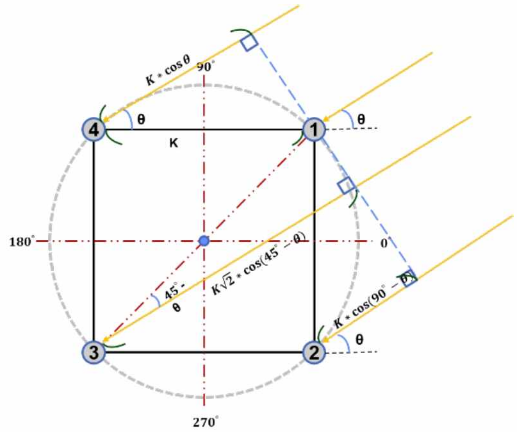


Fig. 2. Delay estimation for a microphone array

The microphone array is arranged in equidistant rectangles, and assuming that the signal is a far-field sound source, as shown in Fig. 2. Taking the sound source in the first quadrant as an example, the microphone in the first quadrant will receive the sound signal first, and the other three microphones will receive signals with a relative delay from the first microphone as in (1), that is

$$L_{12} = \mathrm{K}\cos(90^{\circ} - \theta)$$
$$L_{13} = \mathrm{K}\sqrt{2}\cos(45^{\circ} - \theta) \quad (1)$$
$$L_{14} = \mathrm{K}\cos\theta$$

where K is the distance between two adjacent microphones, $\boldsymbol{\theta}$ is the angle at which the sound source is located, and L_{ij} is the relative delay distance of the i-th and j-th microphones. For example, when the sound source is at 0 degree, channel 1 and channel 2 receive the signal first, and channel 3 and channel 4 receive the signal after a delay of 3 sampling points. The schematic diagram is shown in figure 3.

Then the delay sample is obtained from the delay distance as in (2), where $\boldsymbol{v}$ is the sound speed.

$$delay\ sample = \frac{L}{v} \times sample\ rate \quad (2)$$

According to the delay sample of all angles, the relative delay relationship between the microphones for the sound sources of all angles can be established as in (3),

$$mic1[i+n_1] \times mic2[i+n_2] \times mic3[i+n_3] \times mic4[i+n_4] \quad (3)$$

where *mic[k]* represents the value of *k* sampling points of

978-1-7281-9802-6/20/$31.00 ©2020 IEEE

each microphone, i is the number of sampling points contained in each frame, $i=1,\ldots,256$, n is the number of sampling points delayed at this angle.

Feeding the received signal into all angle modes for synchronization, alignment, and multiplication, the output is the correlation of the angle mode as shown in figure 4. The angle mode with the highest correlation is close to the angular position of the sound source.

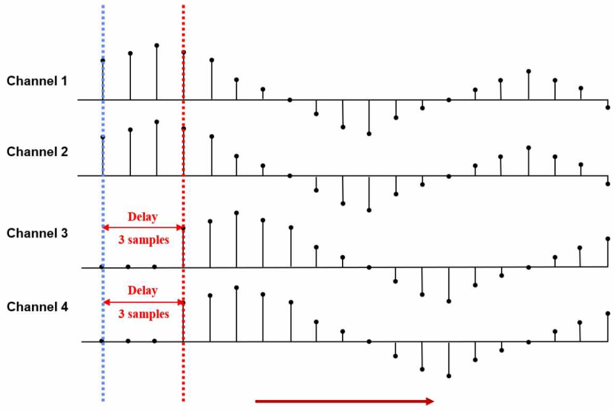


Fig. 3. Schematic diagram of the delay relationship of the sound source at 0 degree.

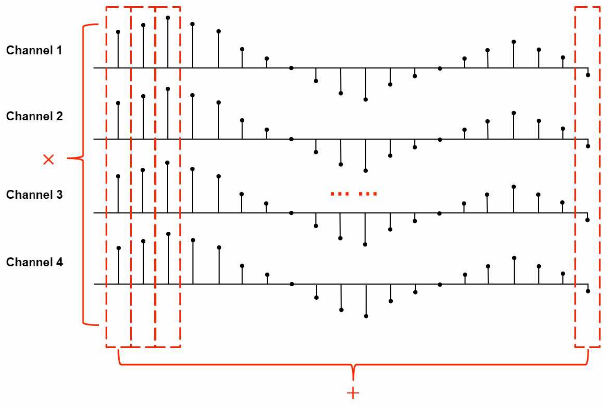


Fig. 4. The method of obtaining the correlation of the angle mode, using synchronization, alignment, and multiplication.

C. *Voting mechanism*

The voting mechanism is used to accumulate the angle patterns with the highest correlation from each frame, that is, when the frame gets the kth angle pattern, the kth angle pattern gets a vote. When the number of votes accumulates to a threshold, the final judgment is made. The judgment mechanism with a 50-frame window as an example can be described as follows and shown in Table 1:

i. ① The angle of the highest number of votes is much larger than that of other angles (n>25).
   – Output: The angle of the highest vote.

ii. The top three votes are similar (10≤n<25).
   ② Three angles adjacent.
   – Output: The angle in the middle position.
   ③ Two angles are adjacent, one is not adjacent.
   – Output: Two direction. (a) The middle angle of two adjacent angles. (b) The non-adjacent angle.
   ④ All three angles are not adjacent.
   – Output: Three direction.

iii. The top two votes are close, and much larger than the third.
   ⑤ The first two angles are adjacent.
   – Output: The middle angle of two adjacent angles.
   ⑥ The first two angles are not adjacent.
   – Output: Two direction.

Table 1. Examples of voting mechanisms. Every 30 degrees is an interval to generate 12 angle patterns.

	The angle of the top three votes			Output angle
	No.1 Angle(vote)	No.2 Angle(vote)	No.3 Angle(vote)	
①	30° **(26)**	60° (7)	0° (5)	30°
②	30° **(15)**	60° **(12)**	90° **(11)**	60°
③	30° **(15)**	60° **(12)**	270° **(11)**	45°/270°
④	30° **(15)**	120° **(12)**	270° **(11)**	30°/120°/270°
⑤	30° **(15)**	60° **(12)**	90° **(3)**	45°
⑥	30° **(15)**	120° **(12)**	90° **(3)**	30°/120°

It can be observed from the example of Table1 that the middle angle can be generated from two adjacent angles through the voting mechanism, so that the angle mode can be extended from 12 to 24 angle modes.

The voting mechanism can also be regarded as a way to judge by sliding windows. Taking 50 frames as an interval, the 51st frame will replace the votes and information of the 1st frame to update it. Therefore, it can judge the sound source signals continuously.

III. EXPERIMENT

A. *Experiment setup*

We use *ReSpeaker 4-Mics Pi HAT* as the radio receiver, as shown in figure 5, with four microphones, the distance between adjacent microphones is 6.2 cm.

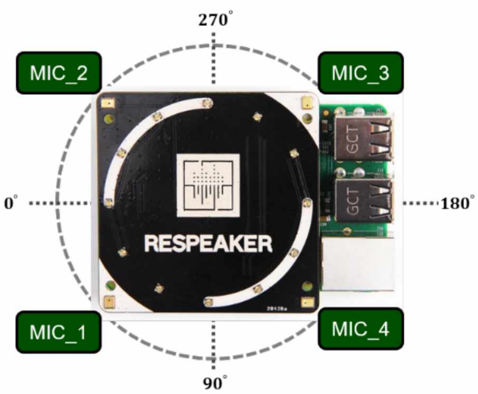


Fig. 5. ReSpeaker 4-Mics Pi HAT.

Every 30 degrees is used as an interval to establish 12 angle patterns, as shown in Table 2, and the voting mechanism is used to extend 24 angles (every 15 degrees as an interval).

978-1-7281-9802-6/20/$31.00 ©2020 IEEE

The sample rate is 16kHz, each frame contains 256 sampling points and overlaps by half. The audio file length is 5 seconds and containing two directions is used as the test, and it is divided into high and low sound source overlap rates for analysis.

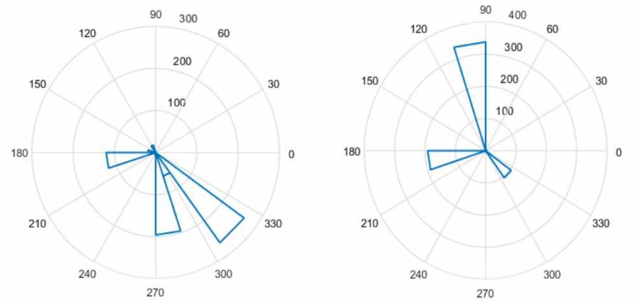
Table 2. Relative delay relationship between microphones for sound sources at various angles.

Mic / Degree	1	2	3	4
0°	0	0	3	3
30°	0	1	4	3
60°	0	3	4	1
90°	0	3	3	0
120°	1	4	3	0
150°	3	4	1	0
180°	3	3	0	0
210°	4	3	0	1
240°	4	1	0	3
270°	3	0	0	3
300°	3	0	1	4
330°	1	0	3	4

B. Experimental results

Case 1: Low overlap

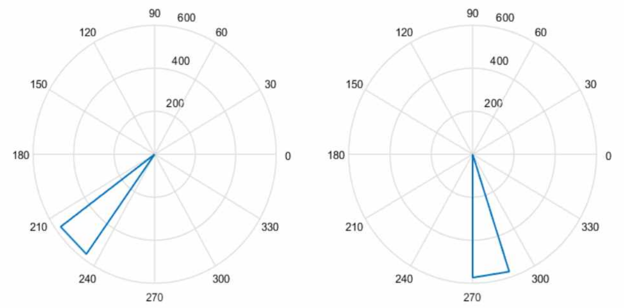
When the sound sources in two directions are interlaced with each other without too much overlap, the proposed method can effectively estimate the angles of multiple sound sources, as shown in Fig.6.



(a) (b)

Fig. 6. Low overlap sound sources (a) 180° and 315°sound sources. (b) 90° and 180° sound sources.

Case 2: High overlap



(a) (b)

Fig. 7. High overlap sound sources (a) 45° and 225° sound sources. (b) 0° and 270° sound sources.

When the sound sources in the two directions appear at the same time and almost overlap, only the angle of the sound source signal with larger energy or closer distance can be estimated.

According to the results of case 1 and case 2, the method proposed in this study can indeed locate multiple sound sources, but there are still some limitations. Since the overall structure is processed in the time domain, when in a noisy environment, or when sound sources overlap with each other, it may cause the voting mechanism to misjudge.

IV. CONCLUSION

This study proposes a relative majority voting mechanism that uses the correlation product of signals received by each microphone to estimate the location of multiple sound sources based on the relative number of votes. According to the experimental results, the voting decision mechanism can indeed estimate the multi-directional sound source. But in the case of noise interference and excessive overlap of sound sources that lead to misjudgment, it is also a goal that we must improve in the future.

V. REFERENCES

[1] M. R. Bai, S. Lan, and J. Huang, "Time difference of arrival (TDOA)- based acoustic source localization and signal extraction for intelligent audio classification," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, Jul. 2018, pp. 632–636.

[2] M. Gaber, T. Elnady, A. Elsabbagh, "Sound Source Localization in 360 Degrees Using a Circular Microphone Array", *Euronoise2018* conference, Crete, Greece, May 27-31, 2018, pp. 2613-2620.

[3] M. Imran, A. Hussain, N. M. Qazi, and M. Sadiq, "A Methodology for Sound Source Localization and Tracking: Development of 3D Microphone Array for Near-Field and Far-Field Applications," *13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, January 2015.

[4] G. Liu, S. Yuan, J. Wu, R. Zhang, "A Sound Source Localization Method Based on Microphone Array for Mobile Robot", *2018 Chinese Automation Congress (CAC)*, January 2019.