Research and implementation on a real-time microphone array sound source localization system

Kui Peng[1] Xiaopei Wu[2] Yaqin Luo[3] and Xiaoxiao Gong[4]

Abstract. To carry out a real-time sound source localization system with a small amount of microphones, a specific implementation method is presented in this paper. This method comprises two steps: time delay estimation and position calculation. Firstly, the modified cross power spectrum phase algorithm (M-CPSP) is applied to time delay estimation, which has higher precision of time delay estimation in real-time environments. Moreover, the cross array model with four microphones is applied to this system and sound source position is calculated by the geometrical relationship between the spatial coordinates of four microphones and time delay. Experimental results show that the M-CPSP algorithm has a high accuracy of time delay estimation in strong noisy and reverberant environments and the error of distance positioning is less than ± 25cm, the error of angle positioning is less than $\pm 5^{o}$, which meets the needs of practical application.

Keywords: Microphone array. time delay estimation. the cross structure. real time

1.1 Introduction

With the continuous development and research of microphone array technology, sound source localization has broad application prospects in real-time tracking, voice hearing aid devices, non-destructive testing of storage tanks and pressure vessels and electroencephalography[1]. On the basis of a small number of microphones, the paper implements a real-time sound source localization system based on time difference of arrival.

[1] Kui Peng

The Key Lab. Of Intelligent Computing & Signal Processing, Anhui University, Hefei 230039，China

email:15655187593@163.com

[2] Xiaopei Wu (✉)

College of Computer Science and Technology, Anhui university, Hefei, China

email: wxp2001@ahu.edu.cn

[3] Yaqin Luo

The Key Lab. Of Intelligent Computing & Signal Processing, Anhui University, Hefei 230039，China

1047563908@qq.com

[4] Xiaoxiao Gong

The Key Lab. Of Intelligent Computing & Signal Processing, Anhui University, Hefei 230039，China

betttygxx@gmail.com

This technique generally comprises two steps: time delay estimation and position calculation. The accuracy of time delay estimation determines the accuracy of the calculation of the sound source position. Generally speaking, generalized cross correlation (GCC) and least mean square are the most commonly used to estimate time delay. Due to the advantages of lower complexity, GCC is often used in real-time system. In this paper, the M-CPSP algorithm is used to estimate time delay, which has advantages of higher accuracy of time delay estimation in noisy and reverberant environments and low complexity, then calculates the sound source localization based on time delay and the cross array model with four microphones [3].

1.2 System Implementation

In this paper, the sound source localization system comprises three parts that respectively are front-end processing, time delay estimation and position calculation, as Figure.1.1 shown.

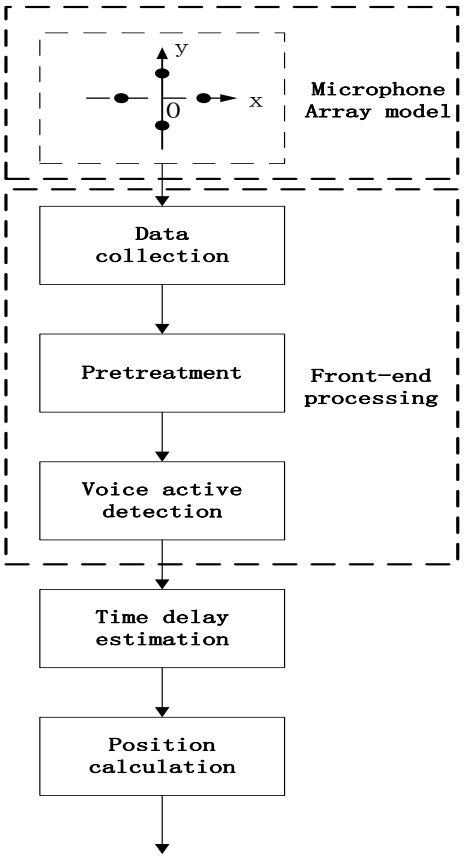

Fig. 1.1 System diagram

1.2.1 Front-end Processing

The system hardware facilities include a personal computer and a microphone array with four electrets microphones. The microphone has the advantages of higher sensitivity wide range of frequency response and small size.

On the other hand, the system software includes signal pretreatment, voice activity detection, time delay estimation and position calculation. The pretreatment includes removal of dc offset and signal filtering. In the system, simply subtract the average of previous block of data to compensate for dc offsets, since dc offsets could influence the validity of time delay. Then a 3-order band-pass filter is designed to restrain random interfering noise. In order to eliminate unnecessary operations, we can analyze the current block of data to determine whether it is a speech signal [2], if so, we estimate the time delay, otherwise we analyze next block of data. When getting time delay, calculate the sound position based on the time delay and the spatial coordinates of microphones.

1.2.2 Time Delay Estimation

Traditional method of cross correlation function for the time delay estimation may produce multiple peaks, even wrong peaks, which are caused by noise and reverberation[4]. Consequently, the basic idea of GCC is weighting the cross power spectrum to sharpen peak and suppress the wrong peaks, as Figure.1.2 shown.

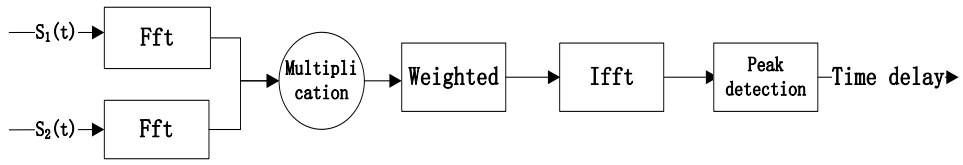

Fig. 1.2 The basic idea of generalized cross correlation

The mathematical expression is shown as following:

$$R_{12}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi_{12}(\omega)X_1(\omega)X_2^*(\omega)e^{jw\tau}dw \qquad (1.1)$$

where $\varphi_{12}(\omega)$ is the weighted function, $X(\omega)$ is Fourier transform of x(t),and * represents the complex conjugate.

In fact, signal to noise ratio (SNR) at different frequency bands is different. Therefore, measures may be used to weaken and block out parts at low SNR and relatively aggrandize the weight of parts at high SNR. It has greater energy and a higher SNR Where there is speech signal。So measures of giving relatively larger weight to the frequency bands that have the larger energy are taking, which is the modified cross power spectrum phase. The weighting function is as following.

$$\varphi^{M}{}_{12}(w) = \frac{1}{|X_1(\omega)X_2^*(\omega)|^{\rho}} \qquad (1.2)$$

where the value ρ is decided by the characteristics of acoustic reflection and noise. In the paper, ρ is the optimal value 0.75.

1.3 Position Calculation

Microphone array is shown as Figure.1.3. M_1、M_2、M_3、M_4 are the four microphones, O is the origin of the coordinate system. S is the target sound source, coordinate is S(x, y, z).The distance between coordinate origin and the target point is r. The angle with the X-axis is θ, with Z-axis is $\emptyset$.The time delays between the microphone i and j are τ_{ij}.

Assuming the distance between microphone and the origin O is d, then the four microphones' coordinates are $M_1(d,0,0)$、$M_2(0,d,0)$、$M_3(-d,0,0)$、$M_4(0,-d,0)$.

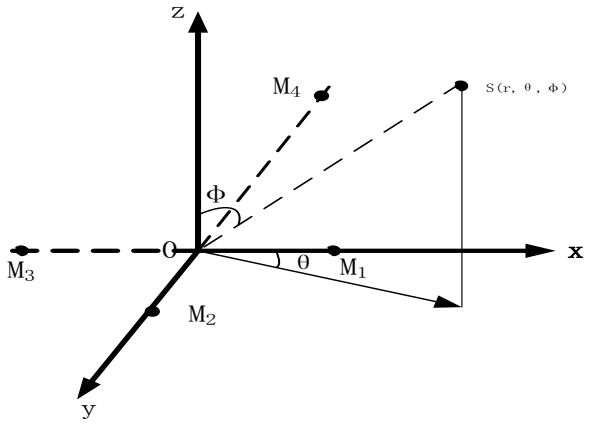

Fig.1.3 Microphone array model

The variable c represents the speed of sound, r_1、r_2、r_3、r_4 represent the distance between the target sound source and the four microphones respectively. By the geometric relationship of the model, we can get the following equations[5]:

$$\begin{cases} x^2+y^2+z^2=r^2 \\ (x-d)^2+y^2+z^2=r_1^2 \\ x^2+(y-d)^2+z^2=r_2^2 \\ (x+d)^2+y^2+z^2=r_3^2 \\ x^2+(y+d)^2+z^2=r_4^2 \end{cases} \tag{1.3}$$

Moreover, based on the relationship between time delay and distance difference, we can also get the following equations.

$$\begin{cases} r_2-r_1=\tau_{12}c \\ r_3-r_1=\tau_{13}c \\ r_4-r_1=\tau_{14}c \end{cases} \tag{1.4}$$

By the above formulas we can obtain the following positioning formula.

$$\begin{cases} r\approx\frac{c}{2}\times\frac{\tau_{12}^2+\tau_{14}^2+\tau_{13}^2}{\tau_{13}-\tau_{12}-\tau_{14}} \\ \theta=\arcsin\left(\frac{c}{2d}\sqrt{(\tau_{12}-\tau_{14})^2+\tau_{13}^2}\right) \\ \emptyset=\arctan\left(\frac{\tau_{14}-\tau_{12}}{\tau_{13}}\right) \end{cases} \tag{1.5}$$

As equation shows, as long as we get time delay between microphone M_1 and M_2、M_3、M_4,we can calculate the spatial coordinates of target. Therefore, it's easily known that the accuracy of

time delay estimation determines the accuracy of target position.

1.4 Experiment Results

Experiments are carried out in a simulated (7m*3.6m*2.75m) rectangular room. The background noise comes from fluorescent lamp, computers and outdoor noise. The sampling rate of signal is 16kHz , the length of a frame is 512, and a 20Hz-20kHz third-order band-pass filter is designed .

1.4.1 Simulation experiments of Time Delay Estimation

As demonstrated above, the accuracy of time delay estimation determines the accuracy of target position. Therefore, it's necessary to verify the robustness of the modified cross-power spectrum phase algorithm. Due to the sharpness of GCC's peaks reflects the accuracy of time delay estimation, the result of simulation is expressed by GCC. To test the performance of the method in the paper, we conduct the simulation in different SNR and reverberation environments.

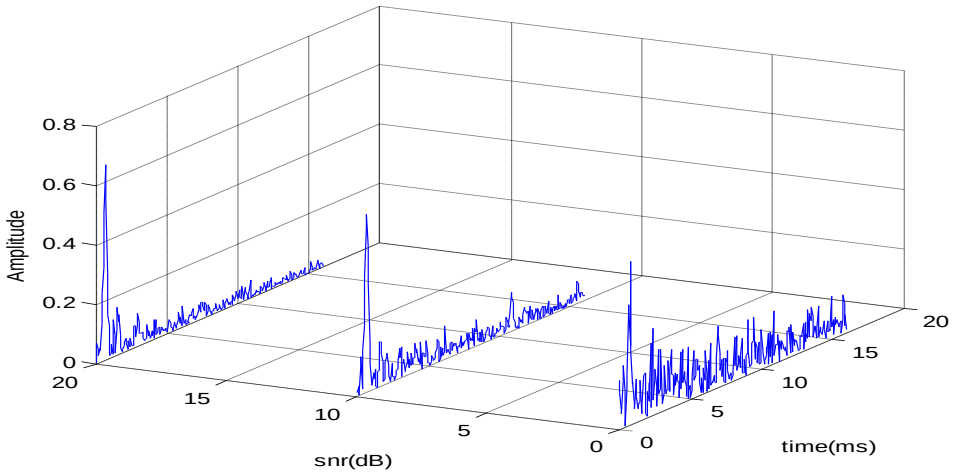

Fig.1.4 Time delay estimation in different signal to noise ratio

As Figure.1.4 shown, x-axis represents SNR, y-axis represents time, and z-axis represents amplitude, no matter SNR is 0dB、10dB or 20 dB, the peaks of the modified cross-power spectrum phase algorithm are apparent and sharp, so it can get a high accuracy of time delay estimation even at a low SNR.

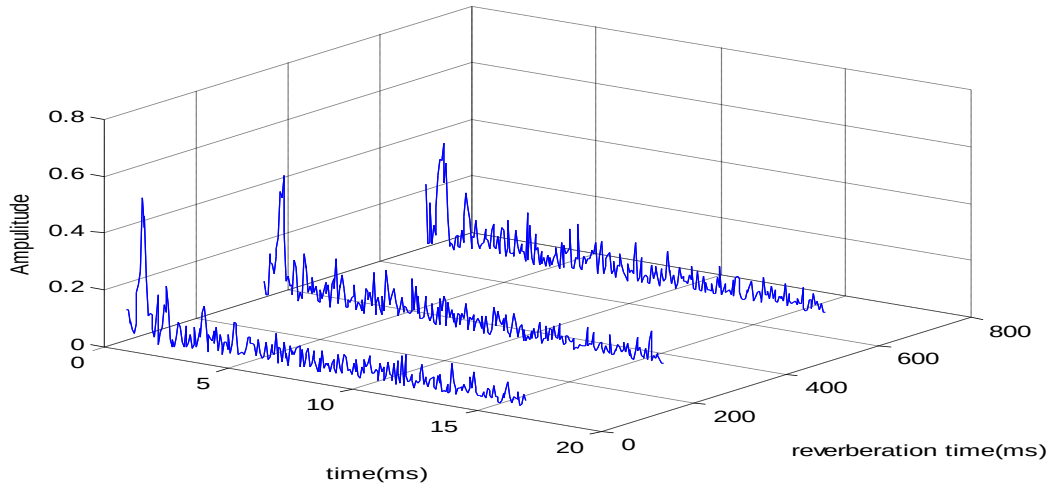

Fig.1.5 Time delay estimation in different reverberation time

As Figure.1.5 shows, x-axis represents time, y-axis represents reverberation, and z-axis represents amplitude. No matter the reverberation time is 50ms or 350ms or 700ms, the peaks are also apparent and sharp.

Simulations show the modified cross-power spectrum phase algorithm can get high accuracy of time delay in strong noise and reverberation environments, and can be apply to practical application.

1.4.2 Positioning Result

The distance between microphone and the Origin is 30cm,so the four microphones coordinate are M_1(30,0,0)、M_2(0,30,0)、M_3(-30,0,0)、M_4(0,-30,0). The height of microphone array is 1m, and the partial results are shown in table 1.

Table.1.1 Experiment results of system.

No.	Real source position	Establish source position
1	(1.0m,30^o,30^o)	(1.14m,27.7^o,32.1^o)
2	(1.0m,40^o,45^o)	(0.91m,42.6^o,47.1^o)
3	(1.5m,50^o,60^o)	(1.32m,53.2^o,64.4^o)
4	(1.5m,60^o,60^o)	(1.38m,63.6^o, 57.1^o)
5	(2.0m,70^o,60^o)	(2.18m,74.6^o,62.2^o)
6	(2.0m,80^o,60^o)	(2.23m,78.8^o,57.6^o)
7	(2.0m,90^o, 60^o)	(1.87m, 90^o, 56.9^o)

From the experiment results, we can summarize that the precise error of distance positioning is less than$\pm$25cm, and the precise error of angle positioning is less than $\pm5^o$ by error analyses. The error is mainly caused by object shelter, the low precision of time delay and angle measurement error. Overall, the positioning error is so small that can meet the needs of practical

application.

1.4 Conclusion

The three-dimensional positioning system is designed to locate sound source by cross model microphone array with four microphones and time delay between them. The positioning system can locate the spatial position of the sound source in real time. From the result of experiments we can get that the precise error of distance positioning is less than $\pm 25\text{cm}$, with the precise error of angle positioning is less than $\pm 5^{\circ}$,which meet the needs of practical application.

1.5 Acknowledgement

The research work described in this paper was supported by national nature science foundation (61271352).

1.6 References

1. Avarvand F S, Ziehe A, Nolte G. Self-Consistent MUSIC algorithm to localize multiple sources in acoustic imaging 4 TH BERLIN BEAMFORMING CONFERENCE[J]. 2012.
2. Fan J, Luo Q, Ma D. Localization estimation of sound source by microphones array[J]. Procedia Engineering, 2010, 7: 312-317.
3. Salvati D, Canazza S. Adaptive Time Delay Estimation Using Filter Length Constraints for Source Localization in Reverberant Acoustic Environments[J]. 2013.
4. Tuma J, Janecka P, Vala M, et al. Sound Source Localization[C]//Carpathian Control Conference (ICCC), 2012 13th International. IEEE, 2012: 740-743.
5. Valin J M, Michaud F, Rouat J, et al. Robust sound source localization using a microphone array on a mobile robot[C]//Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on. IEEE, 2003, 2: 1228-1233.