ICIC Express Letters
Volume **11**, Number **2**, February **2017**

ICIC International ©2017 ISSN 1881-803X
pp. 475–481

INTELLIGENT DIRECTIONAL SYSTEM BASED ON MICROPHONE ARRAY

Xinyu Liu[1], Lianming Sun[1], Jingyang Meng[2] and Han Lin[2]

[1]Faculty of Environmental Engineering
The University of Kitakyushu
1-1 Hibikino, Wakamatsu-ku, Kitakyushu 808-0135, Japan
sun@kitakyu-u.ac.jp

[2]College of Information Science and Electronic Engineering
Zhejiang University
No. 866, Yuhangtang Road, Hangzhou 310058, P. R. China

Received July 2016; accepted October 2016

Abstract. *The design of an intelligent directional system using microphone array is investigated in this paper for high-accuracy intelligent positioning. The system is composed of a six-element microphone array and a PTZ control block, where the signals obtained by microphone array are utilized in the estimation algorithm to estimate the difference of time delays between the microphone elements. Combined with digital signal processing (DSP) unit and the corresponding interface, the system can illustrate the estimated sound source's position, and the positioning can be executed in real time by using PTZ. Moreover, the system is validated in a positioning experiment and the experimental results show that the positioning system has high accuracy when the distance between the target sound source and the center of microphone array is within 1.2m in the horizontal direction and within 0.7m in the vertical direction.*

Keywords: Intelligent directional system, Delay estimation, Microphone array, PTZ, DSP

1. **Introduction.** Sound is an essential way to obtain the circumference information outside and transmit the message to the destination, and it has become a significant part in human's life; naturally, both the larger quantity and the higher quality of the acoustic information are desired to be captured from the sound with the aid of information technology. Therefore, the ways to accurately capture the information through the technologies of acoustic information have attracted many researchers' interests, and these technologies are expected in various fields such as household appliances, transportation, aerospace, and engineering machinery.

Identifying and correctly locating the sound source are a premise of capturing information in many application problems. Consequently, the localization technology for sound source becomes a more and more important issue to capture the acoustic information. In the practical acoustic field, it is the technique that locates the sound source by using acoustic and electronic devices to improve the performance of signal receiver and signal processing in the acoustic field.

Many sound localization technologies commonly adopt the microphone array to capture, analyze and process the signals with respect to their time, space and frequency properties. The existing methods for sound source localization based on microphone array can basically be divided into the following categories: spatial information based methods [1], adaptive signal processing techniques [2], neural network for complex nonlinear optimization [3], correlational analysis to detect the time difference or the semblance [4, 5]. Since the correlational analysis has considerable accuracy with low computational complexity, it becomes a more and more widely utilized method. On the other hand, the placement of the microphone arrays is typically divided into the linear array, the planar array and three-dimensional array where more than four elements are used. Up to now, the four-element and five-element microphone arrays that are represented by three-dimensional array have mainly been studied [6, 7].

For the current sound source localization systems based on microphone array, there are three main issues that need to be improved. Firstly, under a complicated environment, the existence of sound reflection and noise makes the sound signal processing become a difficult and challenging problem, where the reflection and noise will greatly affect the accuracy of sound source localization. Secondly, the placement of microphone array has an effect on the positioning accuracy; hence, it should be investigated further for the practical applications. Finally, developing a system to treat with the multiple sound sources positioning is expected in many practical applications.

In this paper, the design of an intelligent directional system using microphone array is investigated for high-accuracy intelligent positioning. The sound source localization through the time delay detected from cross-correlation [2, 5] is investigated. Furthermore, in order to acquire the target source's omnibearing information just with simple computation and improve the accuracy of positioning, a six-element microphone array is considered in the proposed system. Moreover, the system is composed of a sound source positioning block and a PTZ control block, where the two blocks are connected through the serial communication [8]. In order to meet the requirements of real-time location, the designed system introduces DSP unit such as DSP2812 [9] and C52, whose powerful processing capability makes the sound source localization be more flexible.

The paper is organized as follows. In Section 1, the general issues of the associated study are addressed. In the second section, the localization algorithm and the design of microphone array are investigated. In the third section, the design of the hardware of the system is illustrated. In Section 4, the analysis of experimental results is given. The last section is the conclusion that summarizes the major points of this paper.

2. Key Elements of the Intelligent System.

The estimation for the difference of time delay and the design of microphone array, which are the key elements in the intelligent system, are illustrated in this section.

2.1. Estimation of time delay.

First let the signals received by any two microphones, e.g., M_1 and M_2 in the array, be denoted as $X_1(t)$ and $X_2(t)$, respectively. They can be approximated by

$$X_1(t) = a_1 S(t) + n_1(t) \tag{1}$$

$$X_2(t) = a_2 S(t - t_{12}) + n_2(t) \tag{2}$$

where $S(t)$ is the sound source, a_1 and a_2 are the attenuation coefficients, while $n_1(t)$ and $n_2(t)$ are the corresponding noise terms that are assumed as the white Gaussian ones independent of $S(t)$. t_{12} is the time difference between two microphones M_1 and M_2. Define the correlation function of $X_1(t)$ and $X_2(t)$ as follows:

$$R_{X_1X_2}(\tau) = E\left[X_1(t)X_2(t+\tau)\right] \tag{3}$$

Then, substituting Equations (1) and (2) into Equation (3) yields that

$$\begin{aligned} R_{X_1X_2}(\tau) &= E[a_1 S(t) + n_1(t)][a_2 S(t - t_{12} + \tau) + n_2(t+\tau)] \\ &= a_1 a_2 E[S(t)S(t - t_{12} + \tau)] + a_1 E[S(t) n_2(t+\tau)] \\ &\quad + a_2 E[S(t - t_{12} + \tau) + n_1(t)] + E[n_1(t) n_2(t+\tau)] \end{aligned} \tag{4}$$

Without loss of generality, assume that attenuation coefficients are 1. Moreover, following the whiteness and independence assumption of the noise terms, Equation (4) can be rewritten as

$$R_{X_1X_2}(\tau) = E\left[S(t)S(t - t_{12} + \tau)\right] = R_{SS}(\tau - t_{12}) \tag{5}$$

Following Equation (5), it can be seen that the difference of time delay can be detected when the correlation function reaches to its maximum value [4, 5].

In the practical noisy environment, the signals need to be pre-processed to reduce the affection of the noise term. Then the processed signals are input into a cross-correlation block to calculate the correlation function. Finally, through the peak detection, the time delay can be obtained by detecting the peak in the correlation function. The principle of the estimation for time delay is shown in Figure 1.

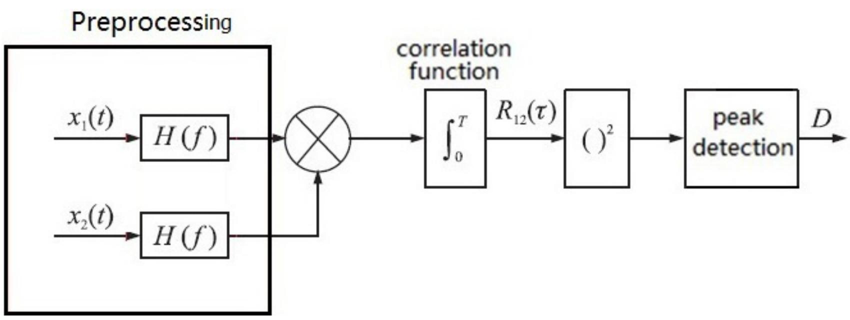

Figure 1. Principle of time delay estimation

On the other hand, from the relationship between the cross-correlation function and the cross power spectrum, the cross-correlation function can be rewritten as the inverse Fourier transform

$$R_{X_1X_2}(\tau) = \int_{-\infty}^{\infty} W(f)X_1^*(f)X_2(f)e^{j2f\pi\tau} \tag{6}$$

where $W(f) = H(f)H^*(f)$ is the weighting function, and $H(f)$ is the filter for pre-processing. If $H(f)$ is a whitening filter for $X_1(t)$ or $X_2(t)$, the peak in the correlation function becomes sharp since the sidelobes are reduced largely, so it is easily detected even using the data record with a finite data length. Correspondingly, the weighting function can be chosen as

$$W(f) = \frac{1}{\left|G_{X_1X_2}(f)\right|} \tag{7}$$

where $G_{X_1X_2}(f)$ is the cross power spectrum of $X_1(t)$ and $X_2(t)$.

2.2. **Design of microphone array.** The placement of the six-element microphone array is shown in Figure 2. Here the coordinates of M_0, M_1, M_2, M_3, M_4, M_5 are $(0,0,0)$, $(d,0,0)$, $(0,d,0)$, $(-d,0,0)$, $(0,-d,0)$, $(0,0,d)$, respectively. Assume that S is the sound source with the coordinate (x,y,z).

Let the distances between S and M be indicated as r_0, r_1, r_2, r_3, r_4, r_5, while the travel time of sound wave from S to M_0 be denoted as t_0. When the delay time between M_i and M_j is denoted as τ_{ij}, then the delay time from M_0 to M_1, M_2, M_3, M_4, M_5 are τ_{10}, τ_{20}, τ_{30}, τ_{40}, τ_{50}, respectively. Denote the sound travel velocity as C. The geometric relationships

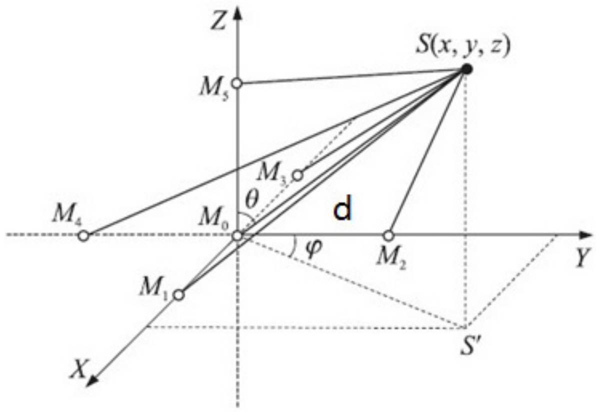

FIGURE 2. Six-element microphone array

yield the following equations:

$$\begin{cases} x^2 + y^2 + z^2 = r_0^2 \\ (x-d)^2 + y^2 + z^2 = r_1^2 \\ x^2 + (y-d)^2 + z^2 = r_2^2 \\ (x+d)^2 + y^2 + z^2 = r_3^2 \\ x^2 + (y+d)^2 + z^2 = r_4^2 \\ x^2 + y^2 + (z-d)^2 = r_5^2 \end{cases} \tag{8}$$

where

$$\begin{cases} r_0 = Ct_0 \\ r_1 = C(t_0 + \tau_{10}) \\ r_2 = C(t_0 + \tau_{20}) \\ r_3 = C(t_0 + \tau_{30}) \\ r_4 = C(t_0 + \tau_{40}) \\ r_5 = C(t_0 + \tau_{50}) \end{cases} \tag{9}$$

Following the symmetries of the array placement, the following relation

$$r_1^2 + r_3^2 = r_2^2 + r_4^2 \tag{10}$$

holds true. Substituting Equation (9) into (10) leads to

$$(t_0 + \tau_{10})^2 + (t_0 + \tau_{30})^2 = (t_0 + \tau_{20})^2 + (t_0 + \tau_{40})^2 \tag{11}$$

Then t_0 can be obtained

$$t_0 = \frac{r_2^2 + r_4^2 - r_1^2 - r_3^2}{2(r_1 + r_3 - r_2 - r_4)} \tag{12}$$

Furthermore, substituting Equation (12) into (8) yields the estimation of source location.

$$\begin{cases} x = \dfrac{d^2 + r_0^2 - r_1^2}{2d} \\ y = \dfrac{d^2 + r_0^2 - r_2^2}{2d} \\ z = \dfrac{d^2 + r_0^2 - r_5^2}{2d} \end{cases} \tag{13}$$

From Equation (13), it is seen that the source location can be estimated easily by using the placement of the designed six-element array.

3. **Design of System Hardware.** In order to implement the sound source localization and PTZ control, the system is designed as a system framework shown in Figure 3, where the sound signals acquired by the microphone array are transformed into the digital ones through the A/D conversion. The digital signals in the DSP will be pre-processed first, and then the filtered digital signals are input into the time delay estimation block. Using the time delay estimates, the sound source's coordinates are calculated from the spatial positioning equations, and they are also shown by LCD. The PTZ control adopts C52 to perform the signal decoding and drive motor's turning through the serial communication with the DSP.

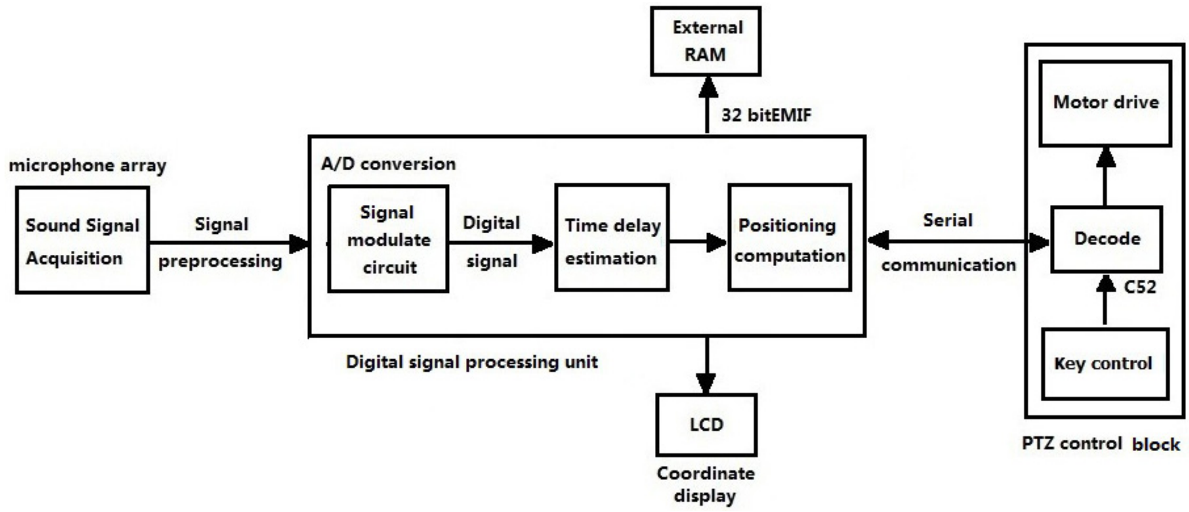

FIGURE 3. Framework of intelligent directional system

In order to fulfill the practical real-time requirements and multi-channel performance, the system uses the Tl DSP2812 unit as the core processor, while the PTZ control uses C52 as its core processor that can perform the signal decoding and drive motor's turning to make the system have a smart search function. The circuit of PTZ control block is shown in Figure 4.

4. **Experiment Example.** The location experiment is performed to illustrate the practical application of the intelligent system. It is carried out in the laboratory to reduce the affection of the sound reflection, where the sound source is given as a sinusoidal signal whose frequency is 500Hz. This intelligent system is tested in both the horizontal and vertical directions.

The experimental results are shown in Figure 5, where the distance errors between the actual target and the estimated ones are plotted with respect to r_0. It shows that the positioning performance varies with the distance. In the horizontal direction, let z be fixed as a constant, the experiment is performed by moving the target sound source along the horizontal xy plane, and the positioning system can obtain higher accuracy when x and y are less than 1.2m, respectively. Correspondingly, in the vertical direction, let x and y be fixed, the experiment is performed by moving the target sound source along the z direction, and the positioning system has higher accuracy when z is less than 0.7m.

Figure 6 illustrates the standard deviation of the experimental results. It is seen that the positioning system only guarantees high accuracy within the effective range around the microphone array. The estimation error may mainly be caused by the following reasons. (1) The sensitivity of microphones may affect the experimental results. (2) In the process of modulation, the sound wave of the circuit output has a small clipping distortion, which will influence the peak detection of correlation function. (3) Although the experiment is

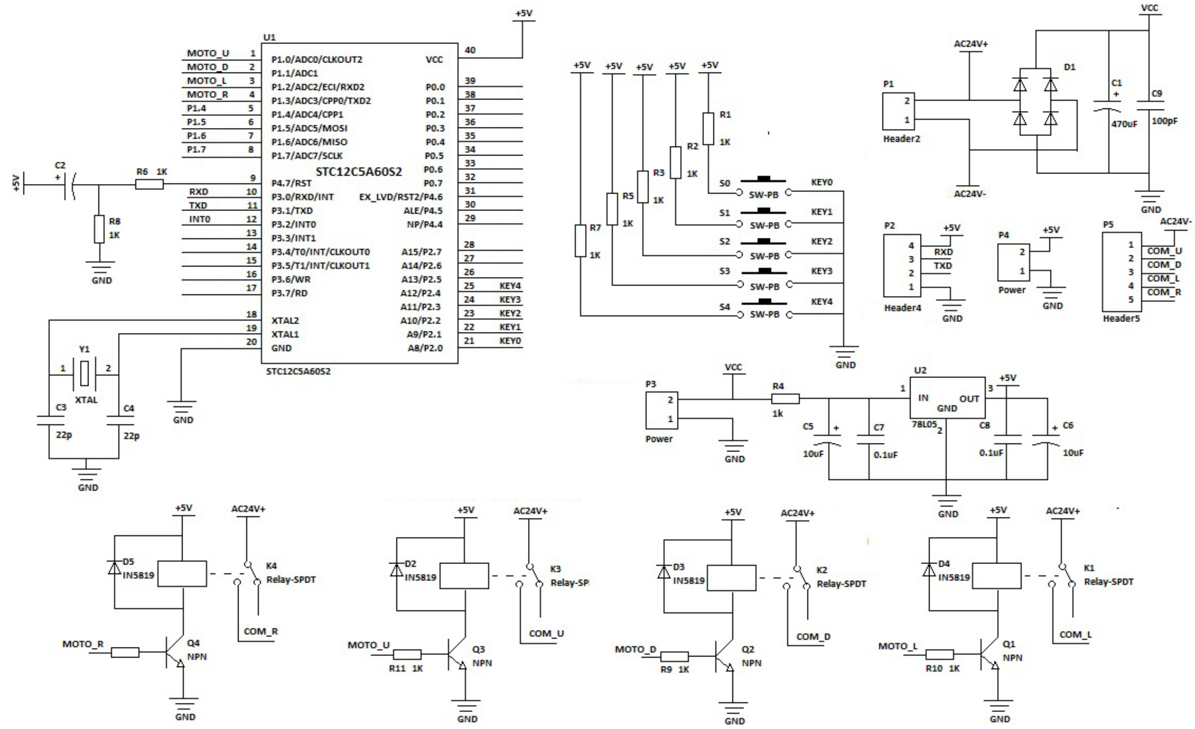

FIGURE 4. PTZ control block

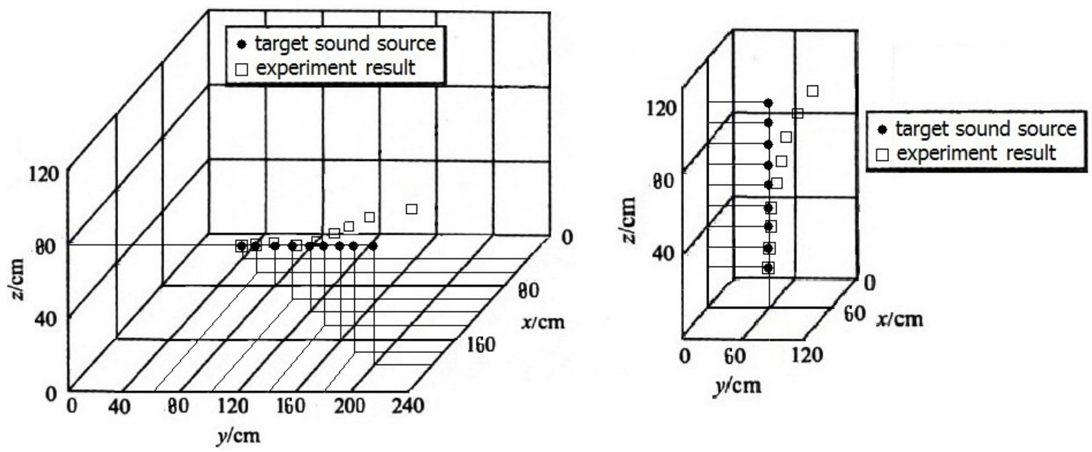

FIGURE 5. Experimental results

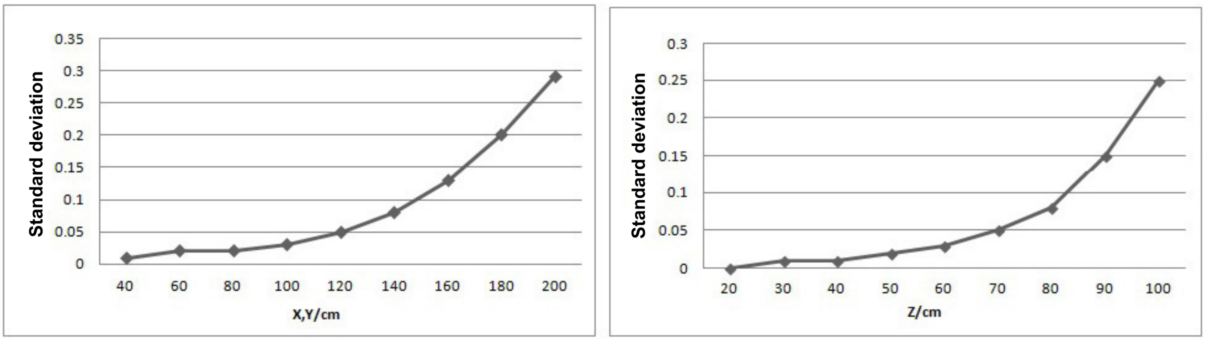

FIGURE 6. Standard deviation of experimental results

carried out in the laboratory, the sound reflection still exists, and it may greatly affect the accuracy of time delay estimation.

5. **Conclusions.** The intelligent directional system based on microphone array is designed by using a six-element microphone array and PTZ control block. Combined with DSP unit and the corresponding interface, the system can illustrate the estimated sound source's position, and the positioning can be executed in real time by using PTZ. In the practical application, the system has high accuracy within the effective range. The remained issue is the problem caused by the sound reflection, which arises if the distance between the target sound source and array becomes too large. It is the main factor that influences the accuracy of the system, and should be solved by using an appropriate preprocessing block. The implementation for the sound reflection will be investigated in the further research work.

REFERENCES

[1] Z. Yin, *Research on Music Sound Source Localization Algorithms Based on Microphone Arrays*, Master Thesis, Harbin Institute of Technology, 2008.
[2] K. Wang, Z. Zhong, X. Sun and W. Shi, Study on improved adaptive eigenvalue decomposition algorithm for acoustic source localization, *Chinese Journal of Scientific Instrument*, vol.34, no.6, pp.1241-1246, 2013.
[3] R. Guo and Z. He, Research on passive acoustic localization algorithm based on multistage neural network, *Application Research of Computers*, vol.28, no.6, 2011.
[4] X. Wan and Z. Wu, Sound source localization based on discrimination of cross-correlation functions, *Applied Acoustics*, vol.74, no.1, pp.28-37, 2013.
[5] Y. Sun, K. Xu and H. Yang, The analysis of time delay estimation algorithm based on generalized cross correlation, *Computer and Digital Engineering*, vol.12, no.1, pp.33-34, 2013.
[6] J. Lu, *Sound Source Localization Based on Microphone Array*, Tianjin University, 2008.
[7] F. Wang, *Three-Dimensional Sound Source Localization System Based on Microphone Array*, Hebei University of Engineering, 2010.
[8] Z. Cheng, Realize PTZ control system with the serial communication, *Chinese Scientific Papers*, 2008.
[9] X. Liu, D. Wang and S. Zhao, The data sampling transmission system based on TMS320F2812, *International Conference on Machine Learning and Cybernetics*, pp.1036-1040, 2010.