A Simple Microphone Array For Source Direction and Distance Estimation

Weisen Jiang, Zebin Cai, Meifeng Luo
College of Automation Science and Engineering,
South China University of Technology,
No. 381, Wushan Road, Guangzhou,
China, 510641
Email: {weson-jhon,773715883,1028798080}@qq.com

Zhu Liang Yu
College of Automation Science and Engineering,
South China University of Technology,
No. 381, Wushan Road, Guangzhou,
China, 510641
Email: zlyu@scut.edu.cn

***Abstract*—In this paper, a simple microphone array is proposed to estimate the direction and distance of a sound source in 2D space. The microphone array consists of three elements and these microphones are arranged in cross form. Based on the geometrical relationship and the estimated time delays between each pair of microphones, a method is proposed to directly get the direction and distance of the sound source in near field. The computational complexity of the proposed method is very low and the implementation of the algorithm is simple. Simulation results show that the algorithm effectively finds the position of the sound source. Therefore, the proposed method is suitable for practical low cost applications, such as steering video camera in video-conference and providing source location in microphone array beamforming, etc.**

I. Introduction

Microphone array has found a lot of applications in practice. It can be used to find the location of a source and extract the target source from the noise and interference corrupted environment [1], [2]. In a very typical application like video-conferencing, microphone array is used to estimate the direction of speaker and this information is further applied to automatically steer a video camera to the speaker. More advanced systems also exploit the localization information to beamform to the target source. With such processing capability, the desired signal can be enhanced and environment noise and interference can be suppressed.

Source localization is a fundamental function of microphone array systems. It exploits the phase differences of the signals arrived at different microphones. Combining the geometrical information of the array system, it is possible to determine where the signal comes. Although in far-field case, only the direction of the target signal can be estimated, in near-field case, the direction as well as the distance of the target signal to the array system can both estimated. The most typical algorithms proposed for source localization includes Maximum Likelihood (ML) [3], Multiple Signal Classification (MUSIC) [4], etc.

In order to achieve better performance, these methods generally require a large microphone array and a huge amount of calculation, which poses a main and major limitation for application. For some portable devices, such as portable computer, this drawback cannot be overcome. A workable method which has low computational load is a must for such practical applications.

In order to overcome the above mentioned problem, a method is proposed in this paper for practical applications, especially for portable systems. The hardware system of the proposed system is quite simple. It only uses three microphones which are arranged in cross form to estimated the distance and direction of target source. An geometry based method is proposed to directly find the direction and distance from the estimated time-delays of each pair of microphones. By exploiting the geometrical relationship of microphones, the proposed method has very low computational cost compared with the other existing methods [1].

This paper is organized as follow. In section II, the estimated Time-Differences Of Arrival (TDOA) based on Generalized Cross-Collection(GCC) [5] algorithm is presented. The method exploiting the geometrical relationship of three microphones is introduced in Section III. In Section IV, some simulation results are presented to show the performance of the proposed method. Finally, a conclusion is given in Section V.

II. Estimation of Time Delay Between Microphones

For a given pair of spatially separated microphones M_1 and M_2, the recorded signals x_1(t) and x_2(t) for a signal $s(t)$, emanated from a remote sound source in a reverberant and noisy environment, can be modelled mathematically as

$$x_i(t) = h_i(t) \star s(t) + n_i(t), i = 1, 2 \tag{1}$$

where $\star$ denotes the convolution operator, $h_i(t)$ is the impulse response function between the ith microphone and the sound source. The additive term $n_i(t)$ summarizes the channel noise in the microphone system as well as the environmental noise for the i^{th} sensor. It is assumed that $n_1(t)$ and $n_2(t)$ are uncorrelated each other. Moreover, they are both uncorrelated with $s(t)$.

In most of the practical applications, we only care about the primary path of the incoming signal. Therefore, we assume that $h_1(t)$ and $h_2(t)$ are both of delta functions and

$$h_1(t) = h_2(t - \tau) \tag{2}$$

978-1-4244-8756-1/11/$26.00

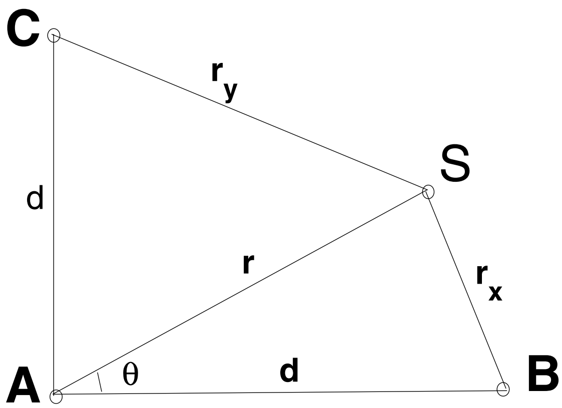

Fig. 1. The geometry of the microphone array and target source.

where τ is the time delay between these two sensors.

In this paper, we adopt the GCC [5] to estimated the TDOA between two sensors. The GCC function is defined as

$$R_{12}(\tau) = \int_{-\infty}^{+\infty} \psi_{12}(\omega) X_1(\omega) X_2^*(\omega) e^{j\omega\tau} d\omega \tag{3}$$

where $X_i(\omega)$ is the Fourier transform of $x_i(t)$, $(\cdot)^*$ denotes the complex conjugate operator, $\psi_{12}(\omega)$ is a weighting function which intends to degrease influences of noise and reverberation and tried to emphasize the GCC value at the true TDOA value τ. In practice, the Phase Transform (PHAT) [5] technique performs well. This PHAT weighting function is define as

$$\psi_{12}^{PHAT}(\omega) = \frac{1}{|X_1(\omega) X_2(\omega)|^\rho} \tag{4}$$

where $0.5 \leq \rho \leq 1$ and the relative time delay τ is estimated as the lag with the global maximum peak in the GCC function R_{12}

$$\hat{\tau} = arg \max_{\tau} R_{12}(\tau). \tag{5}$$

III. The Proposed Source Localization Method

The geometry of the microphone array and target source is shown in Fig. 1, where A, B and C represent the microphones in the array. Point S is the target signal source. According to the Cosine Theorem, we have

$$r_x^2 = r^2 + d^2 - 2rd\cos\theta \tag{6}$$

and

$$r_y^2 = r^2 + d^2 - 2rd\cos(90^o - \theta) \tag{7}$$

where r,r_x and r_y represent the distance between S and A, S and B, S and C, respectively. d is the inter-element distance of the array. θ is the direction of the source relative to the line A-B.

The distance difference can be expressed as

$$r - r_x = ct_x \tag{8}$$

and

$$r - r_y = ct_y \tag{9}$$

where c is the speed of sound, t_x and t_y are the sound propagation times between S and A, S and B, respectively.

Substituting (8) into (6), we have

$$r^2 + (ct_x)^2 - 2ct_x = r^2 + d^2 - 2rdcos\theta \tag{10}$$

then, we can express

$$r = |\frac{(ct_x)^2 - d^2}{2(d\cos\theta - ct_x)}| \tag{11}$$

Using similar derivation, we can express r in terms of t_y as

$$r = |\frac{(ct_y)^2 - d^2}{2(d\sin\theta - ct_y)}| \tag{12}$$

To simplify the expression, here we define two variables α and β as

$$\alpha = (ct_x)^2 - d^2 \tag{13}$$

and

$$\beta = (ct_y)^2 - d^2 \tag{14}$$

Eliminating the variable r in (11) and (12), we have

$$\alpha\sin\theta - \beta\cos\theta = \frac{\beta ct_x - \alpha ct_y}{d} \tag{15}$$

Defining variable k as

$$k = \frac{\beta ct_x - at_y c}{d}, \tag{16}$$

(15) can be expressed as

$$\alpha\sin\theta - \beta\cos\theta = k. \tag{17}$$

From (17), the direction θ of sound source can be estimated as

$$\theta = \arcsin\frac{k}{\sqrt{\alpha^2 + \beta^2}} + \arctan\frac{\beta}{\alpha} \tag{18}$$

Since we only be interested in the $\theta \in [0°, 90°]$, all the estimated θ are converted into this range if not.

Substituting (18) into (11), we can get the estimated distance r_1 as

$$r_1 = |\frac{(ct_x)^2 - d^2}{2(d\cos\theta - ct_x)}| \tag{19}$$

Similarly, we can get the estimated r_2 from (11) as

$$r_2 = |\frac{(ct_y)^2 - d^2}{2(d\sin\theta - ct_y)}| \tag{20}$$

to get a smoothed estimation of r, we take the average of these two estimates

$$\bar{r} = \frac{r_1 + r_2}{2} \tag{21}$$

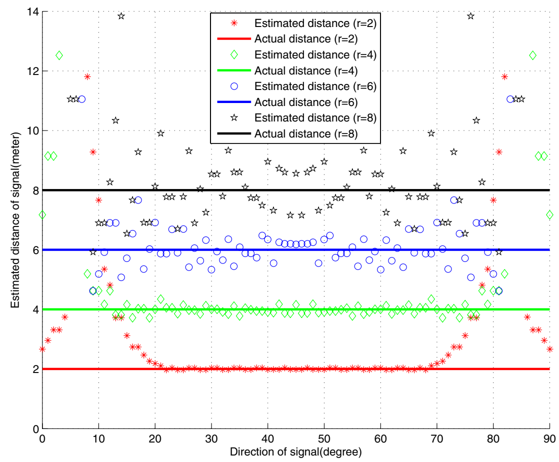

Fig. 2. Simulation results for tracking source at different distances and directions

IV. Simulation Results

In this section, we carried out some simulation experiments to show the performance of the proposed method. In all the simulation experiments, (18) and (21) are used to estimated the direction and distance of the target signal respectively. In the first and second experiments, the inter-element spacing d of microphone array is $d = 1.5m$, and it is $3m$ in the third experiment. The sampling rate is $10k$Hz, and the signal-to-noise ratio (SNR) of the target speech signal is $20dB$. The length of the recorded speech signal used for estimation is $0.265s$. The point A in Fig. 1 is defined as the center of microphone array. Two types of error are considered in the simulations to assess the performance of the proposed method. The first is distance error between estimated distance and actual distance. The second is direction error between estimated direction and actual direction.

In the first simulation, we assume that the source moves in the direction between $0°$ and $90°$, the distance from signal source to the center of microphone array varies from $2m$ to $8m$ in step of $2m$. The estimated location of the target source is shown in Fig. 2, where we can find that the proposed method produces high accuracy on localization when the source is near to the microphone array and its direction is around $45°$. The estimation on direction is more clearly shown in Fig. 3. It is clear that the accuracy of direction estimation is high when the target signal arrives around $45°$. If the target signal arrives in the direction near $0°$ or $90°$, the accuracy is low. This phenomenon can be explained that when the target signal arrives in direction near $45°$, the effective array aperture is maximum so that the accuracy on direction estimation is high. For the distance estimation, from (19) and (20), we find that the accuracy of the estimated distance depends on that of the direction estimation. That explains why the distance estimation has poor performance when the direction of target signal close to $0°$ or $90°$.

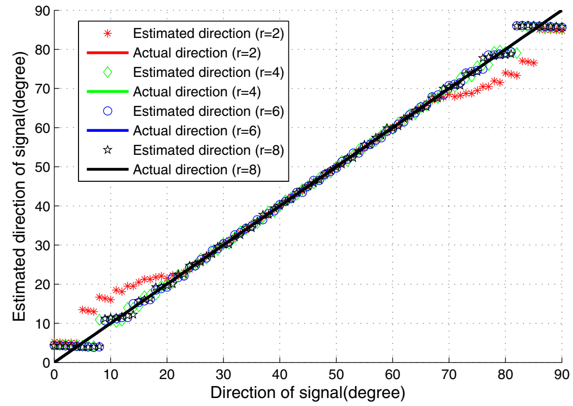

Fig. 3. Direction estimation of a source at different distances

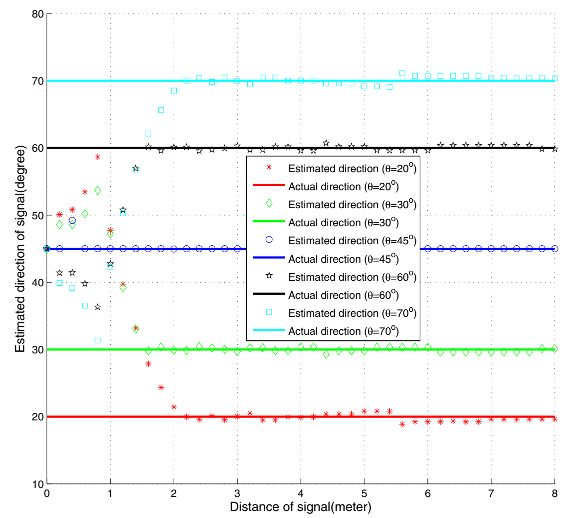

Fig. 4. Direction estimation results of source at fixed directions versus different distances

In the second simulation, in order to reveal the relationship between estimated direction error with the distance of source, we place the source at some specific directions (θ=20°,30°,45°,60°,70°) and then change the distance of these target to the microphone from $0m$ to $8m$. From the simulation results shown in Fig. 4, we can find that the accuracy of direction estimation is much higher when the distance of the source to microphone is large. In other words, in far-field case, the direction estimation is more accurate. In Fig. 5, we also study the distance estimation accuracy with fixed source directions, it is clear that when the target signal is near to the microphone array, the estimation of distance has much higher accuracy. When the source is very close to microphone array, because the difference of time-delay estimation has large error, so that performance of the distance estimation degrades.

In the third simulation, we change the microphone interval

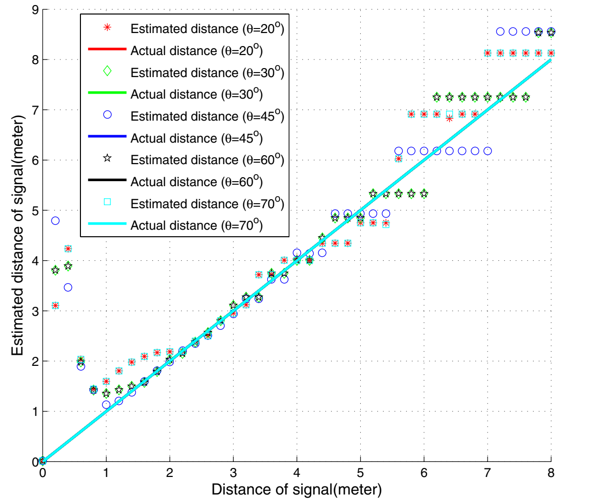

Fig. 5. Distance estimation result of source at fixed directions versus different distances

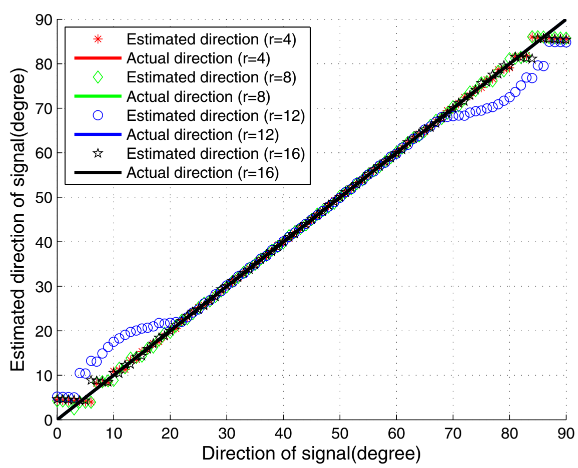

Fig. 6. Simulation results of direction estimation

to $d = 3m$ and repeat the first and second simulations in order to find the relationship between estimation error and the microphone interval spacing d. The simulation results are shown in Fig. 6 and Fig. 7.

These results show that when the d becomes larger, in other words, the aperture of the array is larger, the estimated distance of the source has higher accuracy. Also, the workable estimation range of distance become larger. This simulation result reveals a fact that if the microphone array is design to estimate the distance of a source in large range, the estimation accuracy of the array is guaranteed if the aperture of the array is designed to be larger.

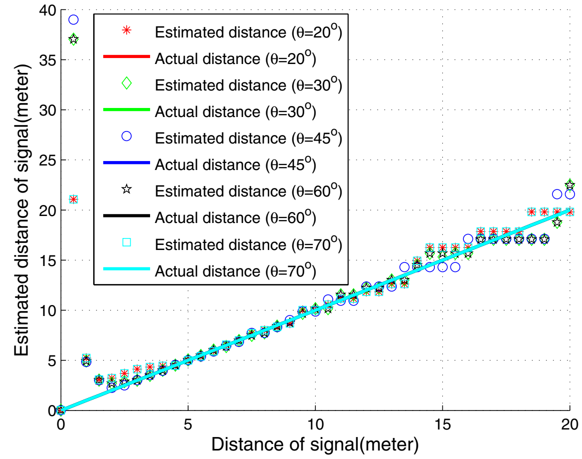

Fig. 7. Simulation results of distance estimation

V. Conclusions

A method for source localization in 2D space with only three microphones has been presented in this paper. Simulation results show that with a designed array structure, the proposed method can produce accurate estimation results in specific range. For example, the direction estimation range is between 20° and 70°. The distance estimation range is more complicated because it depends on the microphone interval and the source direction. For some specific applications, we can design the array size according to the application specifications. The proposed method has very simple hardware structure as well as low computational complexity, these properties support that our method is suitable for many sound source localization applications.

Acknowledgement

This work was supported by the Innovation Foundation 2009 of South China University of Technology under grant CX0705.

References

[1] M. S. Brandstein, "A frame for speech source location," Ph.D. dissertation, Brown University, USA, 1995.
[2] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*. Springer, 2001.
[3] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Speech and Signal Processing*, vol. 36, pp. 1553–1560, Oct 1988.
[4] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation*, vol. 34, 1986.
[5] C. Knapp and Carter.G, "The generalized correlation method for estimation of time delay," *IEEE Trans.Acoustics, Speech and Signal Processing*, pp. 320–327, 1988.