DUAL-MICROPHONE SOURCE LOCATION METHOD IN 2-D SPACE

Weiwei Cui, Zhigang Cao

Department of Electronic Engineering
Tsinghua University, Beijing, China 100084
cuiww03@mails.tsinghua.edu.cn
czg-dee@tsinghua.edu.cn

Jianqiang Wei

Panasonic R&D Center (China) Co., Ltd.
Beijing,China 100080
weijianqiang@cn.panasonic.com

ABSTRACT

Array based source location attracts a growing interest nowadays, which is frequently used in videoconference, hearing aids and hands free telephone systems to detect a speaker's position. Time delay estimation (TDE) based dual-step source location is assumed to be classical method in this field. Recently, a new technique is developed based on interaural level difference (ILD) method, which determines the source position by the energy ratio from microphone pairs. However, all these localization techniques need at least three sensors to obtain a 2D source locus. In this work, we combine the former two techniques, i.e., ILD and TDE based techniques, to present a novel localization approach by an array of only two microphones, and further provide its closed form solution. Our final simulation confirms that such method, which is thought to be more suitable for the equipments with small size, can achieve a good result under normal conditions.

1. INTRODUCTION

In many speech communication applications, much more attention is paid on array based source location techniques, such as videoconference, hearing aids, and hands free telephone devices. The spatial position of a speaker is used not only for microphone array beamforming, but also for automatic video camera steering and tracking. The digital pen, another application of source location technique, has provided us a newly developed computer input device. Its combination with MS Office or other word processing tools enables us to take notes offline and send colorful handwritings to our friends.

Specially, the related algorithms can be divided into two groups: single and dual step approaches. In single step approaches, the source position is determined directly from the received signals at microphone array, while the latter obtains location estimates after applying two algorithmic stages.

Single step location approaches can be further divided into two groups. The first is steered-beamformer localizing method, which is based on maximizing the output power of a steered-beamformer. The second group refers to modern beamforming methods adapted from the field of high resolution spectral analysis: autoregressive (AR) modelling, minimum variance (MV) spectral estimation, and the variety of eigenanalysis based techniques (e.g., MUSIC). Since these methods either suffer from additional model assumption (plane wave from narrow band signal) or cannot resist reverberations or directional interference, coupled with high computational complexity, they are rarely used in today's location systems.

For the second group, time difference of arrival (TDOA) based locator is the popular dual step localization technique. It first estimates TDOA from spatially separated microphone pairs, and then these measurements are used for producing source locus at the second stage. Recently, [1] gives a distinct dual step location method, which is based on ILD instead of TDOA cues. We know that, during the signal propagation, its energy attenuates according to inverse-square-law, so the energy at each microphone is different if their ranges from source is unequal. ILD method explores this information for the location processing. However, all these developed methods mentioned above, when used for actual source location, require at least three sensors.

In this work, a simple source location technique is proposed. It evaluates the source position by two microphones, rather than three or more. On the other hand, unlike the triangular measurement approach, it needn't explore additional signal (e.g., infrared) as the reference, to determine the absolute propagation time [2]. Furthermore, when the application is extended to 3D scenario, spatial points can be determined by 3 microphones of nonlinear array. This simplifies the previous location system where at least 4 microphones are needed. The paper is organized as follows. Section 2 gives the signal model and the proposed location method. In section 3, we simulate this approach under different reverberation environments to demonstrate our algorithm's feasibility. At last, the conclusion is given in section 4.

2. MODEL AND ALGORITHM DESCRIPTION

Suppose we have two microphones and a source signal . According to the so-called inverse-square-law, the signal received by the *i*-th microphone can be modeled as

$$x_i(t) = s(t - \tau_i)/d_i + n_i(t) \tag{1}$$

where $n_i(t)$ is additive noise. d_i and τ_i are the distance and time delay from source to *i*-th microphone, respectively.

As ILD method [1] concentrates on the energy difference rather than the relative time shift between signals, here we ignores the time delay part in (1). A further assumption is made that the sound source was audible and fixed during the time interval [0,*W*]. Then energy received by the *i*-th microphone can be obtained by integrating the square of the signal over this time interval,

$$\begin{aligned} E_i &= \int_0^W x_i^2(t)\mathrm{d}t \\ &= \frac{1}{d_i^2}\int_0^W s^2(t)\mathrm{d}t + \int_0^W n_i^2(t)\mathrm{d}t \end{aligned} \qquad i = 1,2 \tag{2}$$

From above equations, a simple relationship between the energies and distances can be obtained,

$$E_1 d_1^2 = E_2 d_2^2 + \eta \tag{3}$$

where $\eta = \int_0^W [d_1^2 n_1^2(t) - d_2^2 n_2^2(t)]\mathrm{d}t$ is error term. Let (x_i, y_i) be the coordinates of *i*-th microphone and (x_s, y_s) be the coordinates of the sound source, with respect to the origin located at array center. Then,

$$d_i = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2} \tag{4}$$

As far as time delay is considered, the signals acquired by each microphone can be assumed as delayed replicas of the source signal. Thus, to localize a source is to estimate TDE between signals received by two microphones. Here, the time delay τ_i is more important than d_i, which will be ignored in TDE based model. Once time delay is measured, source Cartesian coordinate will satisfy the hyperbolic equation [3] as,

$$\sqrt{(x_1 - x_s)^2 + (y_1 - y_s)^2} - \sqrt{(x_2 - x_s)^2 + (y_2 - y_s)^2} = c\tau_{12} \tag{5}$$

where c is sound speed, approximately 340m/s under normal temperature, and τ_{12} is the TDOA of mic1 and mic2. From equation (3) and (4), we can get

$$(x_1 - x_s)^2 + (y_1 - y_s)^2 = \frac{1}{\gamma^2}\left[(x_2 - x_s)^2 + (y_2 - y_s)^2\right] + \frac{\eta}{E_1} \tag{6}$$

where $\gamma = \sqrt{E_1/E_2}$. In a high SNR environment, the noise term η/E_1 can be neglected. Then, insert (6) into (5), and after some algebraic manipulations, we derive

$$(x_s - x_1)^2 + (y_s - y_1)^2 = (\frac{c\tau_{12}}{1-\gamma})^2 \tag{7}$$

Similarly, another equation will be obtained as

$$(x_s - x_2)^2 + (y_s - y_2)^2 = (\frac{c\tau_{12}\gamma}{1-\gamma})^2 \tag{8}$$

The solution to equation set composed by (7) and (8) gives the exact source position. We can see that the source lies on the intersection of two circles determined by (7) and (8), with center (x_1, y_1), (x_2, y_2) , and radius $c\tau_{12}/(1-\gamma)$, $c\tau_{12}\gamma/(1-\gamma)$, respectively. In terms of the geometric relationship, the existence of solution relies on the following expression,

$$(\frac{c\tau_{12}}{1-\gamma})|1-\gamma| \le d \le (\frac{c\tau_{12}}{1-\gamma}(1+\gamma)) \tag{9}$$

where d is the distance between two circle centers. If τ_{12} is greater than zero, that is, source reaches mic1 later than mic2. On the other side, γ herein will be less than one. Therefore, τ_{12} and $(1-\gamma)$ are simultaneously positive or negative. If we further assume $E_1 \ne E_2$, (9) can be rewritten as

$$c|\tau_{12}| \le d \le c|\tau_{12}|\frac{1+\gamma}{|1-\gamma|} \tag{10}$$

Considering $0 < c|\tau_{12}| \le d$ and $d \le d_1 + d_2 = c|\tau_{12}|\frac{1+\gamma}{|1-\gamma|}$, so equation (10) will satisfy automatically. However, if $E_1 = E_2$, both the hyperbola and the circle determined by (5) and (6) degenerate to a line, i.e., perpendicular bisector of microphone pair. Consequently, there will be no intersection to determine source position.

In the following, we try to obtain a closed form solution to this problem. Again, note the equation set of (7) and (8),

$$\begin{aligned} (x_s - x_1)^2 + (y_s - y_1)^2 &= (\frac{c\tau_{12}}{1-\gamma})^2 = r_1^2 \\ (x_s - x_2)^2 + (y_s - y_2)^2 &= (\frac{c\tau_{12}\gamma}{1-\gamma})^2 = r_2^2 \end{aligned} \tag{11}$$

Transform the expression by

$$x_i x_s + y_i y_s = \frac{1}{2}(K_i - r_i^2 + R_s^2); \;\; i = 1,2 \tag{12}$$

where

$$\begin{aligned} K_i &= x_i^2 + y_i^2; \;\; (i = 1,2) \\ R_s &= \sqrt{x_s^2 + y_s^2} \end{aligned} \tag{13}$$

Rewrite (12) into matrix form,

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \frac{1}{2}\left\{R_s^2\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} K_1 - r_1^2 \\ K_2 - r_2^2 \end{bmatrix}\right\} \tag{14}$$

Then,

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}^{-1} \times \frac{1}{2}\left\{R_s^2\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} K_1 - r_1^2 \\ K_2 - r_2^2 \end{bmatrix}\right\} \tag{15}$$

If we define,

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}^{-1}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{16}$$

and

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}^{-1} \begin{bmatrix} K_1 - r_1^2 \\ K_2 - r_2^2 \end{bmatrix} \quad (17)$$

Then source coordinate can be expressed with respect to R_s,

$$\mathbf{x} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} b_1 + a_1 R_s^2 \\ b_2 + a_2 R_s^2 \end{bmatrix} \quad (18)$$

Insert (18) into (13), the equation associated with R_s will be

$$(a_1^2 + a_2^2) R_s^4 + 2(a_1 b_1 + a_2 b_2 - 1) R_s^2 + b_1^2 + b_2^2 = 0 \quad (19)$$

Then the solution to R_s^2 is obtained as,

$$R_s^2 = \frac{c_b \pm c_c}{c_a} \quad (20)$$

where

$$\begin{aligned} c_a &= a_1^2 + a_2^2 \\ c_b &= 1 - a_1 b_1 + a_2 b_2 \\ c_c &= \sqrt{(1 - a_1 b_1 + a_2 b_2)^2 - (a_1^2 + a_2^2)(b_1^2 + b_2^2)} \end{aligned}$$

The positive root gives the square of distance from source to origin. Substituting R_s into (18), the final source coordinate will be achieved. However, a rational solution requires prior information of evaluation regions.

3. SIMULATION AND EXPERIMENT RESULTS

As for our newly proposed dual-microphone localization algorithm, we first test it in non-reverberant environment. The applied source signal is 8KHz speech recordings, which last for nearly 1 second. While in the simulation, we treat this entire signal as one frame.

Localization experiment is performed in a simulated (4m$\times$ 6m$\times$3m) rectangular room with plane reflective surface and uniform, frequency-independent reflection coefficients. A coordinate system with the origin in one corner and axes parallel to the walls is set up to be the reference. The basic unit of length is defined as meter. Then, speech source s0 is fixed at (1,3,1). Two microphones mic1 and mic2 are located at (2,3.75,1) and (2,4,1), respectively.

Under this environment, the calculated source position by our method is shown in **Fig. 1**. Where hyperbola is obtained according to equation (5), and TDOA is estimated by phase transform (PHAT) weighed generalized cross correlation (GCC) method [4]. While the circle is drawn based on equatoin (3) [1]. We also denote the position of actual source and two microphones by positive triangular and circle in this figure. It is known to us that, by using a linear array, two mirror points will be produced simultaneously. The same defect turns up in our method, e.g., two intersections (noted by inverted triangular) in **Fig. 1**. To eliminate this directional ambiguity, we can restrict the testing range to the area of interest. For example, when the array is mounted on the wall in one enclosed room, we only need to determine the location in the front-end region of microphone array, while the back-end is not what we care. However, we can see as well that one of these two intersections matches actual source position well.

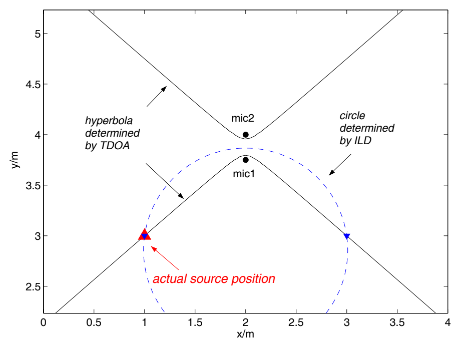


Fig. 1. Localizing result by ILD & TDOA technique

In the sequel, experiments are carried out under various reverberant scenarios. Room impulse responses are generated with image method proposed by Allen and Berkley [5]. Reflection coefficient is considered to be uniform at each plane. Specially, reverberation time RT_{60} in seconds is estimated by Sabine's formula,

$$\begin{gathered} RT_{60} = 55 \frac{V}{c \times Se} \\ Se = \alpha_1 S_1 + \alpha_2 S_2 + \cdots \end{gathered}$$

Where *V* is the room volume in cubic meters, and *Se* is the total absorption area in square meters. For a given reflection coefficient β_i associated with each wall, Sabine energy absorption coefficient is of the form,

$$\alpha_i = 1 - \beta_i^2$$

Microphone pair here is placed at (2,2.8,1) and (2,3.2,1). A Gaussian white noise sequence is used as source signal, which is located one meter away from array center and along 10, 45 and 80 degrees with respect to array normal direction. In these cases, reflection coefficient is changing from 0.1 ($RT_{60} = 108ms$) to 0.9 ($RT_{60} = 568ms$), and the achieved results are shown in **Fig. 2**.

We can see that under low reflected situations, our proposed method could achieve the source position correctly in the region of interest. When reverberation increases to a certain degree, the performance degrades. Compared to TDOA based hyperbolic location method, ILD method is more affected by reverberation. When the bearing angle is close to zero, with reverberation's increasing, information from energy is unreliable to determine source position. A slight error introduced by reverberation will result in the variation in energy ratio, i.e., γ in equation (7) and (8), resultantly make an erroneous result. This is also reflected in the reverse of circle relative to array normal (**Fig.2**-c). However, if we can estimate the signal energy of its direct path, accurate results will be obtained as well, where some de-reverberant techniques are required to be applied.

A special case is that, once sound source lies in mid-sagittal line, then $E_1 = E_2$ and $\tau_{12} = 0$, both the circle and hyperbola degenerate to array's perpendicular bisector, this implies that we are unable to fix the accuracy position of the source. In this case, we can slightly change the array's direction so as to avoid source appearing in the mid line, and then perform our proposed algorithm to localize.

4. SUMMARY AND DISCUSSION

In this paper, we propose a special dual-microphone source location method. Since the preceding localization technique, when applied to actual location processing, either employ multiple microphone pairs, or resort to special signals (infrared) as the reference, the principal innovation of our method is that we combine the TDE and ILD techniques, and then implement source location based on the array of only two microphones, which didn't appear in the former localization techniques. Besides, the closed form solution enables us to perform real time localization so as to track a moving source. Finally, our experiments testify its feasibility in low reverberant environments. Although it has some performance degradation when reverberation grows to a certain degree, its reduced array size will be an attractive feature to speech communication devices of small scale.

5. REFERENCES

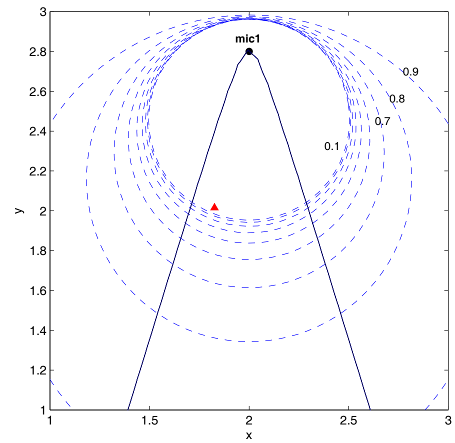
[1] S. T. Birchfield and R. Gangishetty, "Acoustic localization by interaural level difference," *proceeding of ICASSP*, pp. 1109–1112, 2005.

[2] G. F. Rusell, B. A. Smith, and T. G. Zimmerman, "Digital pen using ultrasonic tracking," United States Patent, March 2004, US6703570 B1.

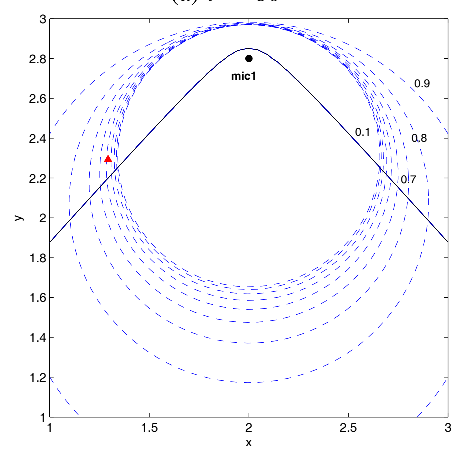
[3] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Transactions on signal processing*, vol. 42, pp. 1905–1915, 1994.

[4] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, pp. 320–327, 1976.

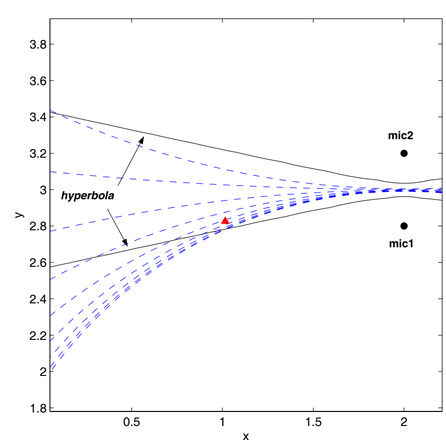
[5] J. B. Allen and D. A. Berkely, "Image method for efficiently simulating small room acoustics," *Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, 1979.



(a) $\theta = 80°$



(b) $\theta = 45°$



(c) $\theta = 10°$

Fig. 2. Simulation results for proposed location method. The source is placed at three different directions with respect to array normal direction, and depart from array center one meter away. The reflection coefficient varies from 0.1 to 0.9