Proceedings of APCC2008 copyright © 2008 IEICE 08 SB 0083

Sound Source Localization in 3-D Space by a Triple-Microphone Algorithm

Nima Yousefian, Mohsen Rahmani, Ahmad Akbari
ni_yousefiyan@comp.iust.ac.ir, {m_rahmani, akbari}@iust.ac.ir
Audio and Speech Processing Lab., Department of Computer Engineering
Iran University of Science and Technology, Tehran, Iran

***Abstract*- In this paper we present a novel technique which localizes sound source with acceptable accuracy and low time complexity. In this technique, contrary to former localization techniques in 3-D space, only three microphones are sufficient. Instead of searching the entire space domain for the sound source, we first estimate an approximate point of the sound source according to the time delays and the ratio of energy received by each pair of microphones. Then, we enhance our estimation by applying an error criterion to points in vicinity of the estimated point. Simulation results show performance of our work, especially when considering the processing time of algorithms as a metric.**

Keywords: time delay- accuracy of estimation- intersection method

I. INTRODUCTION

Precise source localization, a technology commonplace in many smart meeting rooms, is useful in a variety of domains, including experience recording, meeting analysis, and camera steering[1]. In general, the precision of the location estimation is dependent upon various factors. The quantity and quality of present microphones, microphone placement relative to each other, the ambient noise and reverberation levels, and the number of active sources and their spectral content are the most important of those factors. The performance of localization techniques generally improves with the number of microphones in the array, particularly when adverse acoustic effects are present [2].

In many speech array applications, such as automatic cam-era steering video-conferencing system, processing time of the localization algorithm is not a critical issue and also long data segments for processing guarantee accurate localization of the sound source. However, in other applications such as tracking multiple talkers, where higher estimate rates are required, the length of speech data segments become shorter and the processing time of algorithms become a major factor for the evaluation. In such systems, usually the precision of estimation declines, especially when dealing with even moderately adverse acoustic conditions. Many of the methodologies for localization, such as ML-TDOA, SRP, and SRP-PHAT require searching the whole problem space to find the position of the sound source. An overall view for these techniques can be found in [2]. It is obvious that these approaches are extremely time consuming and when the volume of room exceeds a threshold they become completely inefficient.

There are only a few approaches to sound source localization and almost all of the algorithms proposed so far in the literatures are based on them. Of those approaches, Time Delay of Arrival (TDOA) based locator is a popular one. So many authors believe that the problem of localizing the sound source is equivalent to estimating the time delays between the signals received [8]. These approaches first estimates TDOA from spatially separated microphone pairs, and then uses these measurements for obtaining a source locus. Recently, in[3], authors introduced a different localization approach which is based on Interaural Level Differences (ILD) instead of TDOA cues. During signal propagation, its energy attenuates according to inverse-square-law; therefore, the energy at each microphone is different given that their distances from the source are unequal. In[4], a combination of TDOA and ILD has been proposed for a 2-D space with dual-microphone system. This method estimates the position of the source as the intersection of two hyperbolas and a circle which are calculated from TDOA and ILD separately. The downside of this approach is that according to simulation results, accuracy of this method is not always in the region of interest, especially in high reverberation environments. Some methods such as that of[5], only use TDOA for calculating several circles in the space and consider their intersections as sound source. Such Methods require at least four microphones in a 3-D space to obtain accurate source localization.

In this paper, we propose a new 3-D source localization technique with three microphones. The first step of the proposed algorithm is based on the combination of ILD and TDOA which is an extension of the above mentioned 2-D method. The important difference of our approach is that instead of using the intersections, we change the equations of source localization to linear equations and use Linear least square method to solve the sound source localization problem. Subsequently, we aim to increase the accuracy of the estimations by an extra processing on the adjacent points of the previously estimated position. In this step, an error criterion is defined based on TDOA. This error measure is calculated for all the adjacent points. A log likelihood approach calculates the final point of estimation using these errors.

II. 3-D INTERSECTION METHOD

Suppose we have three 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

$$x_i(t) = s(t-\tau_i)/d_i + n_i(t) \qquad (i = 1,2,3) \tag{1}$$

where $n_i(t)$ is the additive noise, d_i and τ_i are the distance and time delay from the source to *i-th* microphone, respectively. Let us show the position of the *i-th* microphone by (x_i, y_i ,z_i) and the sound source by (x_s , y_s , z_s). The distance between *i-th* microphone and the source is calculated by the following formula:

$$d_i = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2 + (z_i - z_s)^2} \qquad (2)$$

Assuming a fixed location for the sound source during the time interval [0,W], where W is the window size, the energy received by the i-th microphone can be obtained by integrating the square of the signal over this time interval.

$$E_i = \int_0^W x_i^2(t)dt = \frac{1}{d_i^2}\int_0^W s^2(t)dt + \int_0^W n_i^2(t)dt \qquad (3)$$

Considering the above equation, a simple relationship between energies and distances can be given as:

$$E_i d_i^2 = E_j d_j^2 + \eta \qquad (4)$$

where $\eta = \int_0^W [d_i^2 n_i^2(t) - d_j^2 n_j^2(t)]dt$. We can ignore the term η/E_i in high SNR environments. That gives us:

$$\frac{d_j}{d_i} = \gamma_{ij} \qquad (5)$$

where $\gamma_{ij} = \sqrt{\frac{E_i}{E_j}}$. Then, by considering the time delay between each pair of microphones, we have:

$$d_i - d_j = c\tau_{ij} \qquad (6)$$

where c is the sound speed and τ_{ij} is the time delay between i-th and j-th microphones.

As we are considering three microphones, it derives from (5) and (6) that totally six equations exist with only 3 variables of d_1, d_2, and d_3. However, as we have neglected the noise parameter in our equations based on ILD in (4), one of these three equations is repeated and ineffective. As the equations in (5) and (6) are linear in respect to d_i, it is reasonable to solve them by linear least square method[6]. Linear least squares method, also known as linear regression analysis, is a form of least squares analysis, used to find an optimal solution for an over determined system of linear equations. The idea behind linear least squares is that in order to obtain good estimates of the parameters of a given linear model, one should use more observations than there are parameters to be determined. This situation is true for our problem as we have five observations and three parameters to be determined. This gives us a 5*3 matrix for coefficients that is calculated from combining (5) and (6) as follow

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ \gamma_{12} & -1 & 0 \\ \gamma_{13} & 0 & -1 \end{bmatrix} \qquad (7)$$

And a 5*1 matrix of measurements:

$$M = \begin{bmatrix} c\tau_{12} \\ c\tau_{13} \\ c\tau_{23} \\ 0 \\ 0 \end{bmatrix} \qquad (8)$$

If we denote the three elements vector of desirable distances by D= $[d_1\, d_2\, d_3]^T$, then it is desirable to solve D from A*D=M. From linear least square method it derives:

$$D = (A^T A)^{-1} A^T M \qquad (9)$$

By obtaining D from the above formula, now we can calculate the position of the source from(2).

III. Improving the Intersection Method

In this section, we first introduce an error criterion based on TDOA. Then, we calculate this error for all of the points adjacent to the source point estimated from intersection method. And finally enhance the accuracy of localization using these calculated errors.

A. Error Criteria

In this work, error is defined as the difference of the real and estimated TDOA, related to each microphone pair. Consider a pair of microphones with spatial coordinates denoted by the 3-element vectors $p_1(x_1,y_1,z_1)$ and p_2 (x_2,y_2,z_2). For a signal source with known spatial location, q_s, the true TDOA to this sensor pair will be denoted by T({p_1 , p_2} , q_s) , and is calculated from the expression:

$$T(\{p_1, p_2\}, q_s) = \frac{(|q_s - p_1| - |q_s - p_2|)}{c} \qquad (10)$$

The estimate of this true TDOA, is the result of the time delay estimation procedure involving the signals received at the two microphones, will be given by τ. In practice, the TDOA estimate is a corrupted version of the true TDOA and in general $\tau \neq T(\{p_1, p_2\}, q_s)$ [2]. Various time delay estimation procedures have been proposed and implemented over the years, including enveloped cross-correlation functions, unit impulse response calculations, smoothed coherence transforms, maximum likelihood estimates and others [7]. All of procedures mentioned above generally can be viewed as inverse Fourier transforms of appropriately normalized weighted cross-spectral density function measurements which yield delay estimates directly in time domain terms. Sometimes it can be advantageous to base time delay estimates on properly interpreted phase data directly in the frequency domain. Here TDOA is estimated by phase transform (PHAT) weighted generalized cross correlation (GCC) method.

Given a set of M estimates of TDOA, derived from the signals received at multiple pairs of microphones, the problem remains as how to best estimate the true source location, q_s .With no ideal solution available, the source location must be estimated as the point in space which best fits the sensor TDOA data or more specifically, minimizes an error criterion. Therefore, the maximum likelihood location estimate can be shown to be the position which minimizes the least squares error criterion, defined as follows.

$$E_{ML}(q)=\sum_{i-1}^{M}(\tau - T(\{p_{i1},p_{i2}\},q))^2 \qquad (11)$$

It is obvious that there are M= $\begin{pmatrix} N \\ 2 \end{pmatrix}$ selections for microphone pairs, where N is the number of microphones. This method is known as ML-TDOA and is one of the widely-used algorithms for sound localization. We will compare our proposed method with ML-TDOA in the evaluation result section.

We calculate this error for all neighborhoods of the estimated point, resulted from the intersection step. For determining vicinity, a sphere is defined, with the estimated point from the intersection method as center and a predefined radius length.

B. *Final Estimation Procedure*

After calculating the errors for all the abovementioned points, as opposed to picking the point with minimum error as in ML-TDOA, we apply a weighted error approach for defining the final source estimation. Consider the K best points with minimum error from all the points in the sphere. As the error values in our simulation results are insignificant and near each other, it is better to use the logarithm of errors as weights. We can formulate this procedure by

$$q=\frac{\sum_{i=1}^{K} -\log(E_{ML}(p_i)) * p_i}{\sum_{i=1}^{K} -\log(E_{ML}(p_i))} \qquad (12)$$

Where q(x,y,z) is the final point of estimation and p_i (x,y,z) is an arbitrary point between the K best points.

To justify that our approach of using weighed errors of K points is better than selecting only the point with minimum error, we compare the performance of these two techniques in the next section.

IV. Evaluation Resulst

The proposed algorithm has been verified in a room with reverberation simulated by Allen's image method [13]. The simulation parameters of the simulated room are listed in Table 1. A random impulse-like signal is generated by a sound source with 16 KHz sampling rate for less than 200 milliseconds. Two white noise sources are simulated at the corners of the room, simulating effects of a fan or a window. These sources represent coherent noises.

Table 1
Simulated room properties

Room property	Parameters
Length and width	6 m
Height	3m
Surface reflection coefficients	0.7 (uniform)
Sound speed	348 m/s
Air absorption factor	-32 µdB/(mHz)
Spacing factor	1 cm

Two types of error should be measured during the simulation to assess the performance of different algorithms. The first is distance error, the distance between estimated sound source position and the real source. The second is angular error, the angle between estimated and real sound source in respect to the center of the three microphones. The latter is same as the error in estimation of Direction of Arrival (DOA). There are two independent variables for comparing above-mentioned errors: the distance between microphone array center and the real sound source as well as the angle between microphone array center and the real source. The general expectation is that by increasing the distance between microphone array center and sound source, the distance error increases and the angular error decreases [9][10]. This is true in ideal cases, but in situations where reverberations are asymmetric, the accuracy of the estimation decreases significantly to the point where the above expectation does not hold anymore.

In this simulation, microphones array center is fixed at the center of the room. Three microphones are set on a circle with radius of 0.3m, along x-y plane with a spacing factor of 120^o. The sound source position changes during simulation according to scenarios that will be described later. Because our simulations are in 3-D space, it is better to consider spherical coordinates. Figure 1 illustrates this coordinates for the sound source localization problem. Some references name θ as Azimuth and φ as elevation angles [2][9][11].

The radius of sphere that defines the candidate points for the second step of the algorithm is set to 25cm. This value is chosen because during simulations we found that the distance error of the intersection method in each dimension is usually less than 25cm.We set K, the number of selected points for the second step, to 150 in our simulation. Other values for k can potentially result in a better accuracy in other situations.

Figure2 illustrates the angular error of estimated point with three different methods: using only intersection method, ML-TDOA, and our proposed algorithm. The errors are calculated for different distances of the source to the microphone array center (ρ) varying from 0.5m to 1.5m, when $\theta = 45^o$ and $\varphi = 65^o$ in spherical coordinates. As expected, by increasing ρ, the mean angular error decreases. Figure 3 illustrates results of another simulation, where we fix ρ to 0.85m and φ to 65^o and change θ (the angle between microphone array center and sound source in the x-y plane) from 0^o to 90^o. These simulations are done in a room with SNR=15 dB. Figure 3 demonstrates that there is no justifiable relation between the mean angular and θ. It is evident from figures 2 and 3 that the proposed method outperforms alone 3-D intersection method and ML-TDOA.

In the next simulation, we compare the three algorithms by considering the distance error as a metric. Here, ρ changes from 0.5m to 1.5m, 10 cm in each step, θ is fixed to 45^o and φ varies between 30, 45 and 60 degrees each time. Simulation results show that as φ declines, the error rises. This is due to the fact that when φ is close to 0, the symmetrical positions of microphones with respect to the source causes the time delay between each pair of microphones to decrease (it reaches zero at $\varphi=0^o$); in such a situation there is no logical estimation for the source position. The averages of distance errors for the above simulation for different SNR values in the room are listed in Table 2. It is clear from the table that with low SNR, accuracy of ML-TDOA is superior to both the 3-D intersection and the proposed method. This superiority comes from ignoring the noise parameter in (4). However, as SNR value increases, the localization accuracy achieved by the proposed method is more reliable than ML-TDOA. It can be concluded from the table, results of the proposed method is about 5.7% better than ML-TDOA in average.

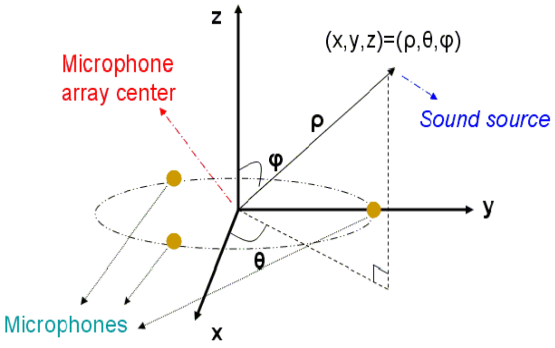

Figure 1. Spherical coordinates for microphone array

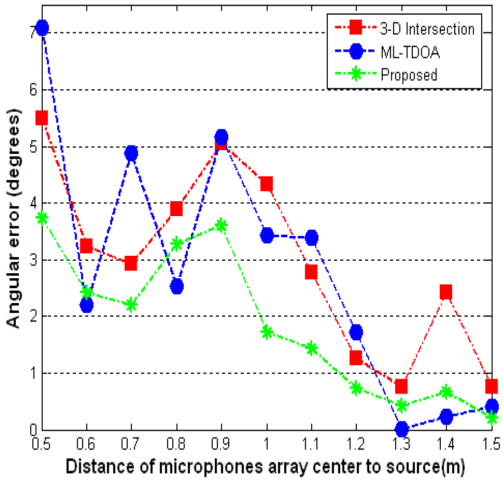

Figure 2. Comparison of angular error for three algorithms by assuming θ=45° and φ=65° and ρ as a variable

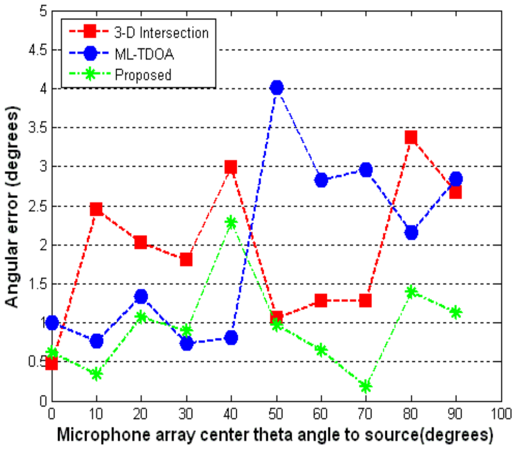

Figure 3. Comparison of angular error for three algorithms by assuming ρ=0.85m and φ=65° and θ as a variable

Table 2
Averaged distance error (cm) resulted from different algorithms

SNR (dB)	ML-TDOA	3-D Intersection	Proposed
-5	24.05	25.87	24.93
0	20.91	22.32	20.94
5	20.03	21.28	18.32
15	19.72	20.90	17.54
20	19.53	20.75	17.41
30	19.33	20.54	17.33

Finally, it is interesting to examine the processing time of the proposed algorithm. Room dimension, clearly, has no effect on the time complexity of the algorithm. This is contrary to algorithms such as ML-TDOA or SRP-PHAT that require searching the whole space of the room. The processing time of our algorithm, depends on K, number of candidate points, and the radius of the sphere used for defining adjacent points in the second step. This implies that our algorithm presents a trade-off between the precision of localization and the processing time. In our simulations, using MATLAB, on a desktop computer (CPU: Athlon 1.8 GHz, RAM: 1GB) the 3-D intersection method processing time is about 210ms (considering above-mentioned values for K and the radius of the sphere of vicinity). Enhancing estimations of the 3-D intersection method by the proposed method adds about 250ms to processing time of the algorithm. These processing times are insignificant in comparison with ML-TODA, which requires about 2.9 seconds to estimates the sound source position. In all of the above simulations, a 25ms data segment of signal is adequate for the localization.

V. CONCLUSION

In this paper we introduced a novel dual-step method for source localization in 3-D space with only three microphones. Simulation results show that the accuracy of proposed method in average is about 6% better than ML-TDOA, one of the most prominent algorithms for sound source localization. Furthermore, the processing time of the proposed technique is insignificant in comparison to many other methods, which require searching the all areas of the room. Although some algorithms have been suggested for the localization with low processing time, they require at least four microphones for localization in 3-D space to achieve desirable accuracy for the localization. Moreover, our method is capable of obtaining the source position from short data segments of input signals. These properties make our method suitable for the semi real-time source localizers.

REFERENCES

[1] X. Chen, W. Jiang, and Y. Shi, "Speaker Tracking and Identifying Based on Indoor Localization System and Microphone Array," in Proc. of the 21st International Conference on Advanced Information Networking and Applications Workshops, vol. 2, 2007, pp. 347-352.

[2] J. DiBiase, H. Silverman and M. Brandstein, Robust Localization in Reverberant Rooms, Microphone Arrays: Signal Processing Techniques and Applications: SPRINGER VERLAG, 2001.

[3] S.T. Birchfield and R. Gangishetty, "Acoustic localization by interaural level difference," in Proc. ICASSP 2005, Pennsylvania, U.S.A, pp. 1109-1112.

[4] W. Cui, Z. Cao and J. Wei, "Dual-Microphone Source Location Method in 2-D Space," in Proc. ICASSP 2006, Toulouse, France, pp. 845-848.

[5] C. Y. Tong, C. H. YAU and P.C. CHING," Linear and approximate maximum likelihood localization from TOA measurements," in Proc. Seventh International Symposium on Signal Processing and Its Applications, Paris, France, 2003, pp. 295-298.

[6] C.L. Lawson and R.J. Hanson, Solving Least Squares Problems: PRENTICE-HALL, 1974.

[7] C. Knapp, G. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 24, pp. 320-327, 1976.

[8] M. Omologo and P. Svaizer, "Talker localization and speech enhancement in a noisy environment using a microphone array based acquisition system," in Proc. EUROSPEECH 93, Berlin, Germany, pp. 605-608.

[9] J.M. Valin, F. Michaud, J. Rouat and D. Letourneau, "Robust Sound Source Localization Using a Microphone Array on a Mobile Robot," in Proc. International Conference on Intelligent Robots and Systems, Nevada, U.S.A, 2003, pp. 1228-1233.

[10] Y. S. Lim, J. S. Choi and M.S. Kim, "Probabilistic Sound Source Localization," International Conference on Control, Automation and Systems, Seoul, Korea, 2007 pp. 1925-1928.

[11] L.S Smith and S. Collins, "Determining ITDs Using Two Microphones on a Flat Panel During Onset Intervals With a Biologically Inspired Spike-Based Technique," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 24, pp. 320-327, 1976.

[12] K.D. Donohue, J. Hannemann and H.G. Dietz, "Performance of phase transform for detecting sound sources with microphone arrays in reverberant and noisy environments," Signal Processing 87, pp. 1677-1691, 2007.

[13] J.B Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics," J.Acoust. Soc. Am Vol 65, no. 4, pp 943-950, 1979.