J. Inf. Technol. Appl. Manag. 23(3): 87~97, September 2016
http://dx.doi.org/10.21219/jitam.2016.23.3.087

ISSN 1598-6284 (Print)
ISSN 2508-1209 (Online)

A TDOA Sign-Based Algorithm for Fast Sound Source Localization using an L-Shaped Microphone Array

Mariam Yiwere* · Eun Joo Rhee**

Abstract

This paper proposes a fast sound source localization method using a TDOA sign-based algorithm. We present an L-shaped microphone set-up which creates four major regions in the range of 0°～360° by the intersection of the positive and negative regions of the individual microphone pairs. Then, we make an initial source region prediction based on the signs of two TDOA estimates before computing the azimuth value. Also, we apply a threshold and angle comparison to tackle the existing front-back confusion problem.

Our experimental results show that the proposed method is comparable in accuracy to previous three microphone array methods; however, it takes a shorter computation time because we compute only two TDOA values.

Keywords : Time Difference of Arrival, L-shaped microphone set-up, Sign of TDOA

Received : 2016. 05. 25. Revised : 2016. 06. 29. Final Acceptance : 2016. 09. 19.

※ This research was supported by the research fund of Hanbat National University in 2015.
* Department of Computer Engineering, Hanbat National University, e-mail : myiwere@yahoo.co.uk
** Corresponding Author, Department of Computer Engineering, Hanbat National University, e-mail : ejrhee@hanbat.ac.kr

1. Introduction

1.1 Background and Objective

Sound source localization is the process of estimating the direction from which a sound originates. A basic sound source localization system takes in audio signals as input and produces an angle as the estimated direction of the sound. With the increasing advancement in technology it has become necessary for systems to be able to determine the directions of sound sources for the automation of activities that will otherwise require the audition of humans, and that is when sound source localization comes in handy.

Over the past decades, this method has proven successful and it has been used in various fields including human-robot interaction allowing the robot to determine the position of the person speaking, video surveillance where the surveillance camera automatically rotates to change its field of view in the event of an incident that generates a loud sound, hearing aid systems improving hearing for people with impaired audition, lecture archiving, video conferencing, etc. In spite of its success in performance, there has always been room for improvement of the method in terms of accuracy, robustness and speed among others.

1.2 Related Work

Different algorithms have been developed for sound source localization and they can be categorized into three main groups [Tashev, 2009]; that is, i) the Eigen value based algorithms, ii) the Steered Power Response based algorithms, and iii) the Time Difference of Arrival based algorithms.

Due to Eigen value decomposition, the Eigen value based algorithms are more complex to use in practice. Likewise, the steered power response based algorithms use complex rotations together with large number of microphones to estimate the sound source location, which makes them more expensive. The most common methods in sound source localization are the TDOA based methods because they mimic the Interaural Time Difference cues used by mammals in their auditory cortex [McAlpine and Grothe, 2003] and they use fewer microphones.

The TDOA-based methods are mostly implemented using an array of microphone, the simplest of which consists of two microphones. Using two microphones, the range of localization is $-90°\sim+90°$, which is equivalent to $0°\sim180°$ and it brings about a front-back confusion, i.e. difficulty to distinguish between angle in front of the microphone array and angles behind it. Many methods [Pourmohammad et al., 2012; Canclini et al., 2013; Pavlidi et al., 2013; Rafaely et al., 2005; Alameda-Pineda et al., 2014; Hwang et al., 2005; Usagawa et al., 2011; Lee et al., 2005; Sreejith et al., 2015; Kwon et al., 2007] have been proposed for $0°\sim360°$ range sound source localization. These include non-linear microphone array geometry methods [Pavlidi et al., 2013; Rafaely et al., 2005; Alameda-Pineda et al., 2014], Head Related Transfer Function methods [Hwang et al., 2005; Usagawa et al., 2011], and energy-based methods [Lee et al., 2005] among others.

For instance, Pavlidi et al. [2013] proposed a circular microphone array geometry, and Rafaely

[2005] designed a spherical microphone array, however, one thing common to both approaches is that they use more microphones than the basic three microphone approaches thereby increasing the hardware costs.

Also, Hwan et al. [2005] used the phase and magnitude information present in their empirically obtained HRTF database to create a criteria to be satisfied by the signals received at the dummy head's microphones; but the presence of noise in any real environment can easily affect the magnitudes of real-time signals which may mislead the algorithm to make a wrong estimation of the sound source location. Usagawa et al. [2011] used an artificial neural network to estimate sound source localization on a sagittal coordinate, but the use of artificial neural networks consumes a large initial computation time.

Ji-Yeoun Lee et al. [2005] used signal energy and zero crossing rates to select one microphone pair in the correct region for sound source localization. With this approach, they compute only one TDOA value which saves computation time; however, in noisy environments, it is very easy to compute wrong energy estimates due to the microphones picking up unwanted signals in which case the microphone pair selection will not reflect the true region of the sound source.

The use of TDOA signs for sound source localization does not require signal energy estimation which eliminates the risk of wrong energy estimation; however, in their work, Sreejith et al. [2015] used four microphones which can be reduced to minimize the hardware cost. Also, Byoungho Kwon et al. [2007] combine the azimuth values computed from three pairs of microphones to obtain the actual direction of the sound source. This is among the commonly used methods however, performing TDOA computation for three different microphone pairs adds up to the overall computational time.

1.3 Suggestion

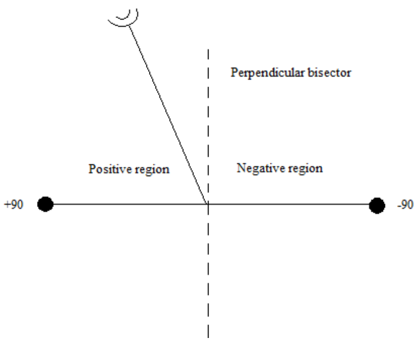
By conducting a survey of different approaches, we found that using more microphones makes the task of localization easier and improves the accuracy at the expense of hardware and computational cost. On the other hand, using fewer microphones is likely to reduce the accuracy of localization.

In order to reduce the computation time in sound source localization, we propose a simple TDOA Sign-based algorithm using three microphones set up in an L-shape which creates four major regions. We make an initial prediction of the sound source region based on the TDOA signs before we compute the actual azimuth value. Although there are three microphone pairs, we focus on only two of them so we compute only two TDOA values. Our proposed method showed results comparable in accuracy to previous three microphone methods but it spends a shorter computation time due to the fewer TDOA computations.

The organization of this paper is as follows; section 2 describes the Microphone Array's Region formation based on sign of TDOA, section 3 describes the L-shape microphone array design, TDOA estimation is described in section 4, the azimuth computation is described in section 5 experiment is presented in section 6 and section 7 presents conclusion.

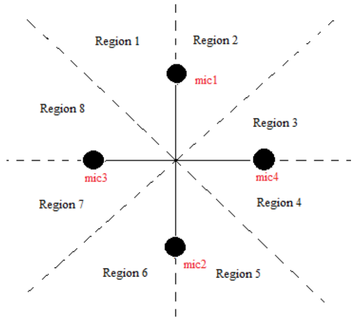
2. Microphone Array's Region Formation based on Sign of TDOA

Time difference of arrival is simply the difference between the times a signal from a single source reaches two different sensors-microphones in the case of sound source localization. The distance between a pair of microphones is divided equally by a perpendicular bisector at the point where the TDOA (delay) value is zero, that is, when the signal received by microphone 1 is identical to that received by microphone 2.



〈Figure 1〉 Two Regions based on TDOA Sign

All TDOA values at one side of the bisector are positive values while all TDOA values at the other side of the bisector are negative values [Yiwere and Rhee, 2015]. We label these sides positive and negative regions respectively as shown in <Figure 1>. With additional microphones [Sreejith et al., 2015], these regions can be used to estimate the location of a sound source in the range of 0°~360°. <Figure 2> shows a symmetric array of four microphones creating different regions based on the signs of the TDOA values [Sreejith et al., 2015]. The perpendicular bisectors of the microphone pairs are represented by the dotted lines.

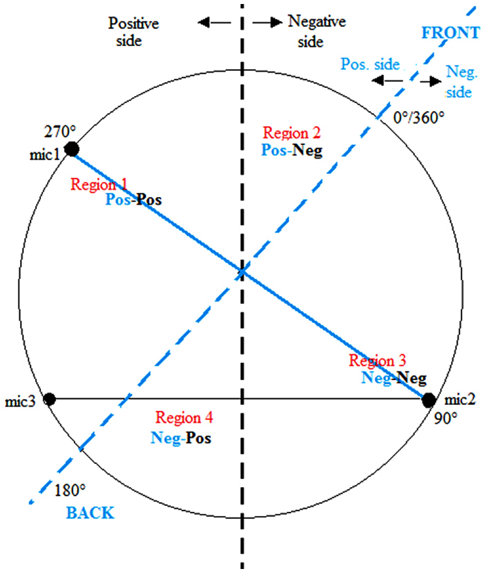


〈Figure 2〉 Region Formation with a Symmetric Four Microphone Array

The positive and negative regions (as illustrated in <Figure 1>) of each of the four microphone pairs intersect one another to create eight different regions. By checking the TDOA signs at all the microphone pairs, the source region of the sound can be predicted. This symmetric array uses more microphones than the minimum number required for a 0°~360° sound source localization which increases the hardware cost as well as the computation time. In order to reduce the computation time and the number of microphones, we present a TDOA Sign-based algorithm for fast sound source localization using an L-shaped microphone array.

3. L-Shape Microphone Array Design

<Figure 3> shows our suggested microphone array design in which microphones 1, 2 and 3 represent the L-shape set-up. Using three microphones, we set up two microphone pairs-major and minor-such that the three microphones are in an 'L' shape. The reason for creating an L-shape is to intersect the perpendicular bisectors of the two microphone pairs at the center of an imaginary circle using only three



〈Figure 3〉 L-shape Microphone Array Design

microphones.

The two microphone pairs are spaced differently so that their perpendicular bisectors intersect each other at the center of the major microphone pair, forming four regions (sectors) around the major microphone pair. The microphones 1 and 2 form the major microphone pair and the microphones 2 and 3 form the minor pair. The major microphone pair is a distance of 42cm apart while the second microphone pair is 30cm apart.

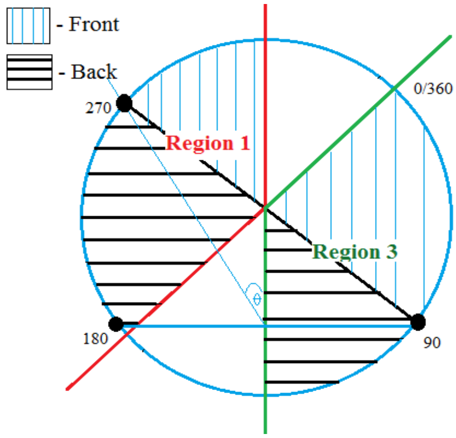
Firstly, we compute TDOA values for the two microphone pairs. The two TDOA values are expected to have certain characteristic signs depending on which region the sound source is located in. The signs of the TDOA values with respect to the sound source locations (regions) are shown in <Table 1>. Using <Table 1>, the region of the sound source is initially predicted after which the major microphone pair's TDOA value is converted to an azimuth value and then that angle is in turn converted to a 360° equivalent. The second conversion to 360° equivalent is based on which region is predicted to have the sound source.

〈Table 1〉 Microphone Pairs TDOA Signs in the Four Regions

	TDOA Sign For The Two Microphone Pairs	
Region	Major Microphone Pair	Minor Microphone Pair
1	positive(+)	positive(+)
2	positive(+)	negative(−)
3	negative(−)	negative(−)
4	negative(−)	positive(+)

In regions 1 and 3, there is some ambiguity resulting from the front-back confusion of the major microphone pair. In view of this, we compute the minor microphone pair's azimuth in order to distinguish between the front and back angles in the two regions. To tackle this ambiguity, we use the following steps; In region 1, a simple threshold angle Θ, as shown in <Figure 4> is used to distinguish between the front and back angles.

For a sound source in Region 1, if the minor microphone pair's angle is greater than Θ, then the sound source is located in the back side of the major microphone pair, else it is located in the front side. In region 3, the major and minor microphone pairs' angles are compared to each other to distinguish between the front and back angles. If the major angle is greater than the minor angle, then the sound source location is in the front side but if the minor angle is greater than the major angle, then the sound source location is at the back side.



〈Figure 4〉 Front-Back Confusion in Regions 1 and 3

4. TDOA Computation

TDOA, as defined above, is the difference between the times a signal from a single source reaches two different sensors. It is estimated using cross correlation which can be implemented in either time or frequency domain. The time domain implementation of cross correlation is more time consuming than the frequency domain implementation; however, in terms of computational complexity, the time domain implementation is sometimes more efficient and preferable. This is because the algorithmic or computational complexity is directly related to the hardware cost and power consumption.

For instance, to develop a sound source localization algorithm in a small system, which implies smaller microphone arrays, only a few cross correlation coefficients are relevant. Therefore, it is not useful to spend resources to compute all possible coefficients as in the Generalized Cross Correlation method [Knapp and Carter, 1976], which uses Fast Fourier Transform (FFT) computations, especially when using high sampling rates.

We compute our TDOA values using the time domain cross correlation function defined in equation 1 to reveal the point at which the two signals received at two different sensors are at their maximum correlation.

$$\mathrm{xCorr}(\mathrm{x},\mathrm{y})(\mathrm{j}) = \sum_{k=0}^{N-1} x_{(j+k)} y_{(k)} \tag{1}$$

where x and y represent the two signal and N is the length of one signal. Equation 2 shows the length of the new array z, produced by the cross correlation function.

$$\mathrm{length}(\mathrm{z}) = length(x) + length(y) - 1 \tag{2}$$

$$\mathrm{Range} = [\min\tau, \max\tau] = \left[\frac{-\mathrm{df}}{\mathrm{v}}, \frac{\mathrm{df}}{\mathrm{v}}\right] \tag{3}$$

By computing the relevant range of cross correlation coefficients, that is, from the minimum possible delay value (min τ) to the maximum possible delay value (max τ), the length of z can be significantly reduced to save time. The sampling frequency of the signals (f), the velocity of sound (υ) and the distance between the two microphones (d) are used to compute the relevant number of correlation coefficients as shown in equation 3.

5. Azimuth Computation

Each azimuth value, θ, is first computed using the delta Δ, i.e. the time increment for sampling sound, velocity of sound, υ, and the distance between the two microphones in question, d; equations 4-6. We used a sampling rate of 44.1 kHz, therefore our delta value is 2.2676×10^{-5}, shown in equation 3. The velocity of sound is assumed to be 343m/s at room temperature and atmospheric pressure

$$\text{Delta} = \frac{1}{44100} = 2.2676\times10^{-5} \tag{4}$$

$$t = \Delta\times\tau \tag{5}$$

$$\theta = \text{arcsine}\frac{vt}{d} \tag{6}$$

Converting the major microphone pair's angle to an angle in the 360 degree range depends on the selected region.

6. Experiment and Discussion

The set-up for our system includes three dynamic cardioid microphones set up in an L-shape in a laboratory environment so that the perpendicular bisector of the minor microphone pair intersects the major microphone pair at its center point. The major microphone pair is a distance of 42cm apart while the second microphone pair is 30cm apart. All three microphones are connected to a TASCAM US 4×4 multichannel audio interface through its XLR ports. The experimental data were audio signals captured in real-time using our L-shape microphone array with a sampling rate of 44.1 kHz. The sound is generated at a distance of at least one meter from the microphone set-up by clapping.

<Figure 5> shows a flow diagram of the overall proposed method. First, a three-channel audio signal is captured using the three microphones. The first and second channels are used to compute the first TDOA value for the major microphone pair. Then the second and third channels are used to compute the second TDOA value for the minor microphone pair. If the signs of the two TDOA values are not the same, use the first TDOA value to compute the azimuth value and convert it to the 360 range, else, compute the second TDOA to compute the minor azimuth.

In the positive-positive region, compare the minor angle to a threshold angle to differentiate between front and back angles. Then, in the negative-negative region, compare the major angle to the minor angle to differentiate between the front and back angles.

<Table 2> shows a comparison of our proposed method to previous methods in terms of the number of microphone pairs and the number of TDOA computations. Previous methods which used three microphones mostly compute three TDOA values, one for each pair of microphones, however in our method, though there are three microphone pairs in the set-up, we make use of only two of them. There we compute just two TDOA values. This reduces the computation time as compared to the previous methods.

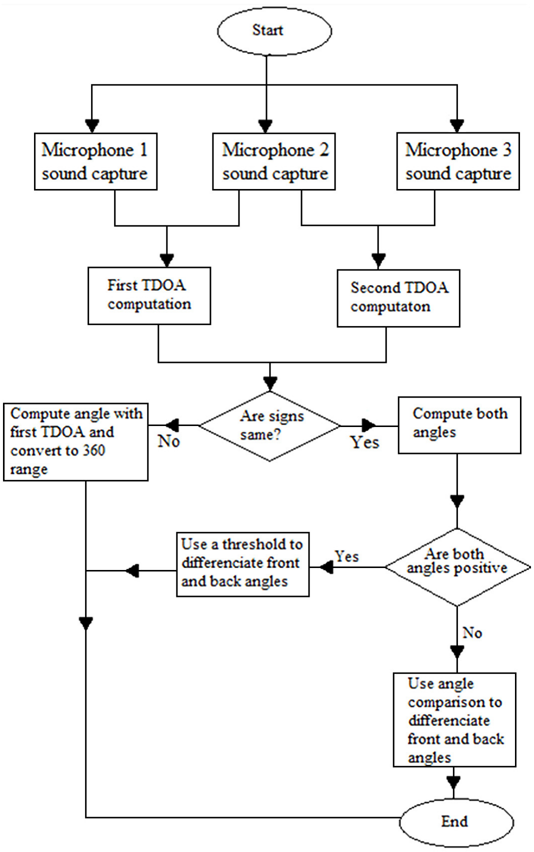
〈Table 2〉 Comparison of the Mic Pairs and TDOA Computation

Method	No. of Microphone Pairs	No. of TDOA Computations
Previous	3	3
Proposed	3	2

<Table 3> shows the accuracy of our proposed method in angle estimation experiments.

We tested the method for nine different angles shown in the first column as the actual angles. For each of the nine angles, we repeated the tests at least ten times and recorded the average of all the ten angles as the estimated angles in the second column. The last column of the table records the errors associated with the estimated angles in relation to the actual angles.

This shows that our method records a maximum error of 2.965° and a minimum error of 0.63666 which is comparable to the accuracy of previous three microphone method. The results of our experiments show that our method gives satisfactory results in terms of accuracy.



〈Figure 5〉 Flow Diagram for Sound Source Localization

〈Table 3〉 Results of Angle Estimation

Actual Angle (°)	Estimated Angle (°)	Error (°)
0	0.636666	0.636666
30	28.91993	1.08007
60	57.035	2.965
120	117.1151	2.8849
150	148.1321	1.8679
210	211.1259	1.1259
240	237.4295	2.5705
300	301.8919	1.8919
330	331.819	1.819

7. Conclusion

This paper describes a TDOA Sign-based algorithm for fast sound source localization. We have presented a simple three microphone array set up in an L-shape to reduce the computation time spent by previous methods. Using two microphone pairs, we created four major regions in the range of 0°~360°. The regions are formed by the intersection of the positive and/or negative regions of one microphone pair with that of the other's.

To tackle the front-back confusion of the major microphone pair, we used i) a threshold angle in region 1 and ii) angle comparison in the region 3 to distinguish between the front and back angles. With this method, there is no need for signal energy estimation. Also, by using three microphones the hardware cost is reduced, compared to other TDOA Sign-based methods which use four microphones.

Our experiments show satisfactory results, comparable in accuracy to previous methods which compute three TDOA values. It spends a shorter computation time because of the fewer

TDOA computations; therefore this method can be used in real time to improve the speed of sound source localization systems.

Generally, the presence of noise in the environment may reduce accuracy of TDOA-based sound source localization. In future, we will include a noise suppression method as a preprocessing step in order to improve the accuracy of the method.

References

[1] Alameda-Pineda, X. and Horaud, R., "A geometric approach to sound source localization from time-delay estimates", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, No. 6, 2014, pp. 1082-1095.

[2] Canclini, A., Antonacci, E., Sarti, A., and Turbo, S., "Acoustic source localization with distributed asynchronous microphone networks", *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 21, No. 2, 2013, pp. 439-443, 2013.

[3] Hwang, S., Park, Y., and Park, Y., "Sound Source Localization using HRTF database", Proc. Int. Conf. on Control, Automation, and Systems (ICCAS2005), 2005, pp. 751-755.

[4] Knapp, C. H. and Carter, G., "The Generalized correlation method for estimation of time delay", *IEEE Transactions on Acoustic Speech And Signal Processing*, Vol. 24, No. 4, 1976, pp. 320-327.

[5] Kwon, B., Kim, G., and Park, Y., "Sound Source Localization Methods with Considering Microphone Placement in Robot Platform", 16th IEEE International Conference on Robots and Human Interactive Communication, 2007, pp. 127-130.

[6] Lee, J.-Y., Chi, S. Y., Lee, J. Y., Hahn, M., and Cho, Y. J., "Real-time Sound Source Localization using Time Difference for Human Robot Interaction", IFAC Proceedings Volumes (IFAC-Papers Online), Vol. 16, 2005, pp. 54-57.

[7] McAlpine, D. and Grothe, B., "Sound Localization and delay lines-do mammals fit the model?", *TRENDS in Neuroscience*, Vol. 26, No. 7, 2003, pp. 347-350.

[8] Pavlidi, D., Griffin, A., Puigt, M., and Mouchtaris, A., "Real-time multiple sound source localization and counting using a circular microphone array", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, No. 10, 2013, pp. 2193-2206.

[9] Pourmohammad, A. and Ahadi, S. M., "Real time high accuracy 3-D PHAT-based sound source localization using a Simple 4-Microphone Arrangement", *IEEE Systems Journal*, Vol. 6, No. 3, 2012, pp. 455-468.

[10] Rafaely, B., "Analysis and design of spherical microphone arrays", *IEEE Transactions on Speech and Audio Processing*, Vol. 13, No. 1, 2005, pp. 135-143.

[11] Sreejith, T. M., Joshin, P. K., Harshavardhan, S., and Sreenivas, T.V., "TDE Sign Based Homing Algorithm for Sound Source Tracking Using a Y-shaped Microphone Array", 23rd European Signal Processing Conference (EUSIPCO), 2015, pp. 1207-1211.

[12] Tashev, I., Sound Capture and Processing, John Wiley and Sons, 2009.

[13] Usagawa, T., Saho, A., Imamura, K., and Chisaki, Y., "A Solution of Front-back Confusion within Binaural Processing by an Estimation Method of Sound Source Direction on Sagittal Coordinate", TENCON 2011-2011 IEEE Region 10 Conference, 2011, pp. 1-4.

[14] Yiwere, M. and Rhee, E.-J., "Fast Time Difference of Arrival Estimation using Partial Cross Correlation", *Journal of Information and Communication Management*, Vol. 22, No. 3, 2015, pp. 106-114.

Author Profile

Mariam Yiwere

She received her B.S. degree in Computer Science from the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana in 2012. In 2015, she received her M.S. degree in Computer Engineering from Hanbat National University, Daejeon, Korea. She is currently conducting research in the area of sound source localization in the Artificial Intelligence and Computer Vision Lab in the Graduate School of Information and Communications, Hanbat National University. She is interested in computer vision, digital signal processing and artificial intelligence.

Eun Joo Rhee

He is a Professor of Department of Computer Engineering at College of Information Technology, Hanbat National University, Daejeon, Korea. He has the degree of Ph.D in Electronics Engineering from Chungnam National University. His research interests include in image processing, pattern recognition, computer vision and artificial intelligence. His papers have appeared in IEICE Trans. on Information and Systems, Journal of KIISE, Journal of the Institute of Electronics and Information Engineers, Journal of Information Technology Applications and Management, Journal of Information Technology Application, Journal of the Modern Linguistic Society of Korea, Journal of Korea Multimedia Society, Journal of the Korea Academia-Industrial cooperation Society.