Journal of Basic and Applied Engineering Research
Print ISSN: 2350-0077; Online ISSN: 2350-0255; Volume 1, Number 11; October-December 2014 pp. 54-58

http://www.krishisanskriti.org/jbaer.html

# Sound Source Localization using GCC-PHAT with TDOA Estimation

**Nitesh Kumar Chaudhary[1], Subir Verma[2] and Anshu Aditya[3]**

*Department of Electronics & Communication Engineering LNM Institute Of Information Technology , Jaipur (India)*
*E-mail: nitesh.chaudhary@lnmiit.ac.in[1], subir.verma48@gmail.com[2], annshu0641@gmail.com[3]*

**Abstract:** *This paper mainly focuses on the Localization of the sound source in 2-D plane using the concept of Time Delay of Arrival of Signals to the respective microphones. TDE between replicas of signals is intrinsic to many signal processing applications and DOA estimation of acoustic signals using a set of spatially separated microphones has many practical applications in everyday life. DOA estimated from the set of microphones can be used to automatically steer cameras to the speaker in a conference room. From the known array geometry, the DOA of the signal can be obtained from the measured time-delays. The time-delays are estimated for each pair of microphones. To read the signals and transform them into frequency domain Cross correlation (CC),GCC and GCC-PHAT algorithms are available. The idea is to use GCC-PHAT based TDE algorithm. In order to robustify the estimation process and to overcome the the spatial aliasing ambiguity occurring at higher frequencies the GCC-PHAT function is summed over all frequencies and then it is reduced to a single dimension to obtain the angular spectrum from which the TDOA's are estimated. This has been performed through Pooling function. The obtained peaks after maximizing the function which is our desired TDOA is used in Clustering method. This method re-estimates the contribution of source to each time-frequency bin(clusters), given current estimates of TDOAs. Hyperbolic estimation for localization of source position is adapted. It takes the TDOA and the position co-ordinates of the microphones as an input argument to create the solution region for the source.*

**Keywords:** *Acoustic source, Direction of arrival, Time Delay Estimation, Short time Fourier Transform, Time delay of Arrival, cross-correlation(CC), Generalized Cross-correlation(GCC),GCC-PHAT, Pooling and Clustering.*

## 1. INTRODUCTION

In order to localize the position of the sound source in a 2-D plane we have considered a set of 3 microphones which are placed arbitrarily and the position/co-ordinates are known to us. Figure1 depicts the stepwise process .The notion is to detect the DOA of the signals from the source which is to be detected. It has been assumed that there is no other noise source and the solely source is point sized. The source is omni-directional in nature. The noise source has been considered stationary and the reflections from the bottom of the plane and from the surrounding objects are negligible. As we have considered our region of interest in 2-D plane and number of microphone to be 3, the first pair of microphones gives us the DOA of the source along the axis of the pair. The third microphone when inserted in 'L' fashion with the previously present microphones the DOA with respect to the axis containing the third and the center microphone is obtained.

The DOA further allows us for TDOA analysis. TDOA is actually the time delay which the signal takes to get perceived by the adjacent microphones. There are algorithms such as GCC, GCC-PHAT and DFSE for TDOA estimation. We have considered GCC-PHAT for our TDOA estimation as PHAT method has better results in simulated Gaussian distribution noise and actual noisy environments. The GCC-PHAT is also less computationally intensive and vital in real time analysis. Now to make our algorithm to cope up with the errors during plotting final peaks and reduce the spatial aliasing at higher frequencies the PHAT function has been summed over all frequencies and reduced to a single dimension to obtain the angular spectrum from which TDOA is estimated. Clustering method has been adopted to re-estimate the contribution of each source to each frequency bin(clusters), given current estimates of the TDOAs. This gives the true TDOA. Further this TDOA is passed as an argument to our localizer function which localizes our source by hyperbolic estimation method.

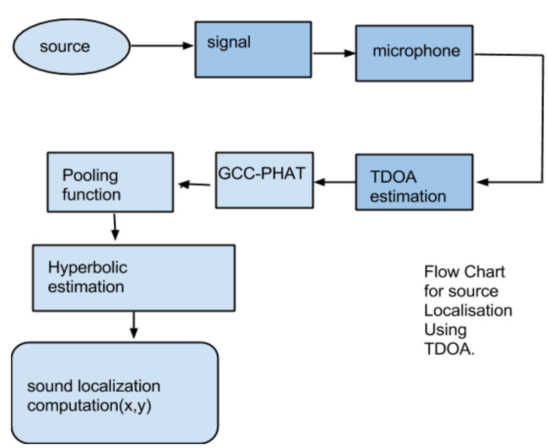


**Fig. 1**

## 2. BASIC CONCEPTS OF DOA AND SPATIALLY ARRANGED MICROPHONES

As the geometry of array of microphones is known to us we can use it to determine time delay for each pair. Accuracy of the DOA[1] estimates obtained using the TDE[2] based algorithms depends on various factors. The hardware used for data acquisition, sampling frequency, number of microphones used for data acquisition and noise present in the signals captured, determine the accuracy of the estimates.

The process of determining the location of an acoustic source relative to some reference frame is known as acoustic source localization. Acoustic source present in the nearfield of the time difference of arrivals (TDOAs) measured with pairs of microphones. The speed of sound in the medium in which the acoustic source is present is known to us.

Following figure-2 illustrate the specification of TDOA estimation.

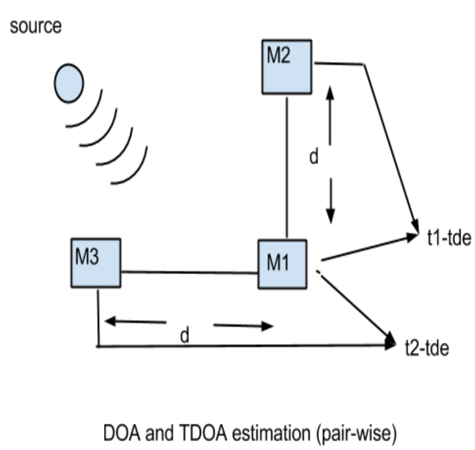


**Fig. 2**

## 3. TDOA ALGORITHM

Acoustic source localization is based on time delay of arrival estimation (TDOA) for the following reasons-

- Such systems are conceptually simple.
- They are reasonably effective in moderately reverberant environments.
- Moreover, their low computational complexity makes them well-suited to real-time implementation with several sensors.

We have implemented this algorithm which constitutes of mainly two step procedure:

- The first stage involves estimation of TDOA[3].
- solving hyperbolic equations to get their solution region.

The TDOA between all pairs of microphones is estimated, typically by finding the peak in a cross-correlation[4] or generalized cross-correlation[5] function. But we have used GCC-PHAT[6] implementation for TDOA. The peak of this function gives more precise TDOA.

## 4. FREQUENCY DOMAIN BASED GCC-PHAT

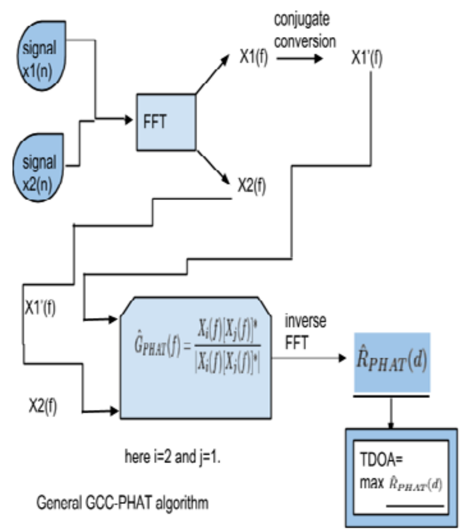


**Fig. 3**

In order to compute the TDOA between the reference channel and any other channel it is usual to estimate it as the delay that causes the cross-correlation between the two signals segment to be maximum. In order to increase the robustness we use GCC with PHAT. The GCC-PHAT( Fig. 3) of two signals is given by:

$$GCC_{PHAT}(f)= [\ Y_0(f)\times Y_1(f)^*\ ] \div [\ ||\ Y_0(f)\ ||\ ||\ Y_1(f)||]$$

Where, $GCC_{PHAT}(f)$ is frequency domain generalized cross-correlation phase transformation $Y_n(f)$ is the STFT of $y_n(n)$ and is equal to $|\ Y_n(f)|exp(j\ arg[Y_n(f)])|$.

Two input signals were considered as the sound source where one was more of a noise kind. Shown in figure-4..This signal is received by the microphones arranged in the 2-D plane arbitrarily with a delay. The time domain based signals were changed into frequency domain using the STFT[10] concept. The frequency domain based signals are used in TDOA estimation. GCC-PHAT based TDOA estimation is used .now to omit the spatial aliasing occurring at higher frequencies GCC-PHAT function is summed over all frequencies. Then it is reduced to single dimension to obtain the angular spectrum from which the TDOAs are estimated.

$$\phi^{\text{sum}}(\tau) = \sum_{t=1}^{T} \sum_{f=1}^{F} \phi(t, f, \tau).$$

where $\phi(t, f, \tau)$. is a GCC-PHAT function

A limitation of this approach is that it makes it difficult to localize a source that is active only within few time frames. This can be addressed by taking the maximum over all the time frames. basically done by summing it all over the time frames. It is basically done with pooling [7].

$$\phi^{\max}(\tau) = \max_{t} \sum_{f=1}^{F} \phi(t, f, \tau).$$

The peak of the above mentioned GCC-PHAT function gives the TDOA as shown in Fig. -5.

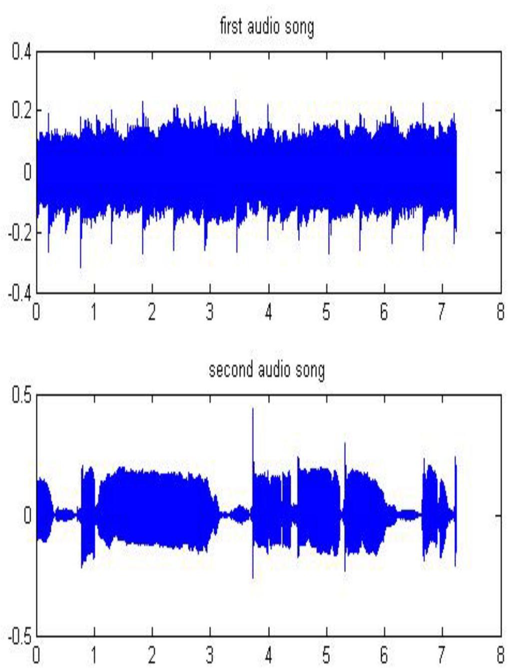


**Fig. 4**

The result is more redefined as compared to earlier CC implementation and simple GCC.

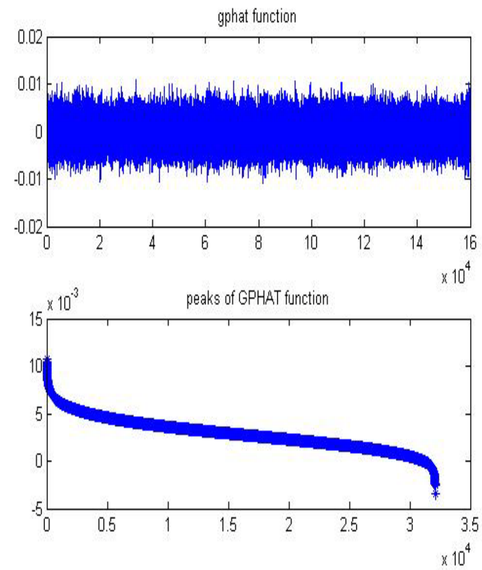


**Fig. 5**

## 5. CLUSTERING METHOD USED FOR IMPROVED TDOA

Cluster analysis or clustering[8] is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).

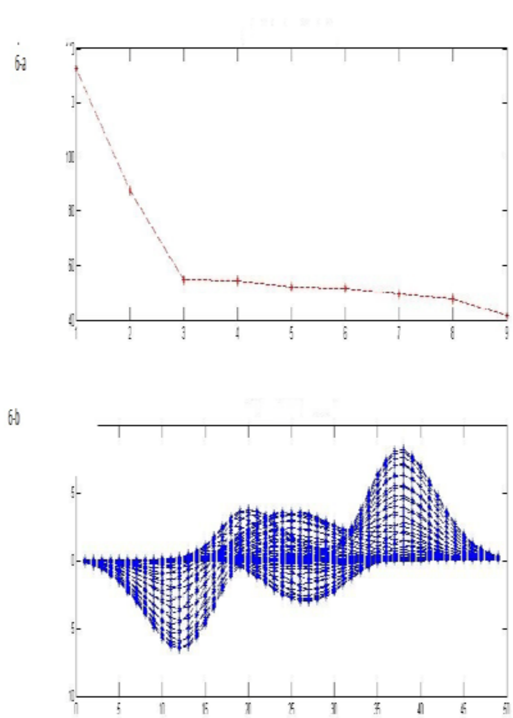

**Fig. 6**

The principle of clustering based method, proposed by Swada[11], consists of estimating the measure of source activity given some estimates of the TDOAs, then re-estimating the TDOA relying on this measure.

Figure-6 clearly shows the effect of clustering method which gives well defined peaks within a particular range of frequency bins reducing the effect of disturbance in signal and the peak value corresponding to the maximized function is TDOA between any pair of microphones considering one as a reference.

## 6. HYPERBOLIC ESTIMATION FOR LOCALISATION OF SOURCE POSITION.

Our implementation is based on the consideration that microphones are located at any arbitrary position.

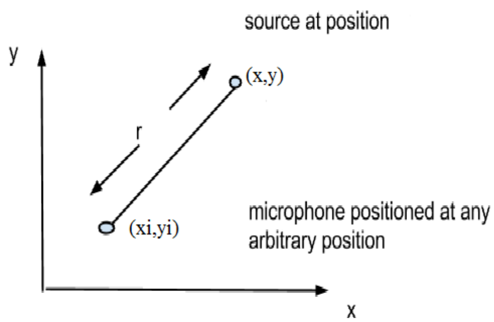


**Fig. 7**

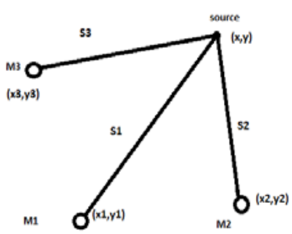


**Fig. 8**

For more realistic view ,considering the position of source to be (x,y)[say]. Placed at an distance of 'r' from any microphone which is positioned at any arbitrary point ($x_i, y_i$) which is known to us while estimating the position of source.

Let $t_1$, $t_2$ and $t_3$ be time taken by signal to reach the respective microphones from the source. Let $T_3$, $T_2$ be the estimated TDOAs with respect to the first microphone i.e $T_3=t_3-t_1$ and $T_2=t_2-t_1$.

$S_3-S_1= r_{(3,1)}= cT_3$ and $S_2-S_1= cT_2 = r_{(2,1)}$

$$\sqrt{(x-x_3)^2+(y-y_3)^2} - \sqrt{(x-x_1)^2+(y-y_1)^2} = cT_3$$

$$\sqrt{(x-x_2)^2+(y-y_2)^2} - \sqrt{(x-x_1)^2+(y-y_1)^2} = cT_2$$

The above mentioned equations can be solved for (x,y) by the formula given by:

$$\begin{pmatrix} x \\ y \end{pmatrix} = - \begin{pmatrix} x(2,1) & y(2,1) \\ x(3,1) & y(3,1) \end{pmatrix}^{-1} \times \left\{ \begin{pmatrix} r(2,1) \\ r(3,1) \end{pmatrix} r1 + 1/2 \begin{pmatrix} r(2,1)^2-K1+K2 \\ r(3,1)^2-K3+K1 \end{pmatrix} \right\}$$

$$K(i)= x(i)^2 + y(i)^2$$

**where $S_1=r_1$.**

The above equation has been achieved by solving the Hyperbolic equations [9] and non-linear equations using Taylor series [10] expansion.

As this above formula for estimating (x,y) give rise to a quadratic equation we get two values of (x,y). To get rid of this ambiguity we can restrict the transmitter to be in the same region of interest.

## 7. CONCLUSION AND FUTURE WORK

In this paper we discussed the solution to localize the position of sound source in 2-D space using 3 microphones using TDE. We have used the proposed method of improved GCC-PHAT for TDOA estimation. GCC-PHAT has been chosen because it works very well in high SNR environment. Pooling function is added to robustify the estimation process and to localize a source that is active only within few time frames. Clustering was implemented for better data analysis as proposed by Sawada et al. Estimated TDOAs has been used in Hyperbolic location estimator as given by Y.T. chan and K.C.Ho. The results are encouraging, but there are still improvements to be done to facilitate tracking and identification of sound sources.

## REFERENCES


[1] G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] Acoustic sound source localisation using time delay estimation by Ashok kumar tellakula.

[3] Robust Sound Source Localization Using a Microphone Array on a Mobile Robot Jean-Marc Valin, Franc¸ois Michaud, Jean Rouat, Dominic Letourneau ´ LABORIUS - Research Laboratory on Mobile Robotics and Intelligent Systems Department of Electrical Engineering and Computer Engineering Universite´ de Sherbrooke Quebec, ´ Canada.

[4] Real-Time Sound Source localisation, Project Report for EE 586 by Jin Fu Liou,Kunal Joshi,Gaurang Vador. B. C. Dalton, “Audio-based localisation for ubiquitous sensor networks,” Master’s thesis, Massachusetts Institute of Technology, September 2005.

[5] Functions for engineering applications, part-1, basic theory by M. R. Belmont School of Engineering, University of Exeter, Harrison Engineering Building, North Park Road, Exeter EX4 4QF, UK.

[6] GCC-PHAT based head orientation estimation by Carlos Segura,Javier Hernando Universitat Politecnica de Catalunya, Barcelona, Spain ` Herta Security, S.L., Barcelona, Spain

[7] Multi-source TDOA estimation in reverberant audio using angular spectra and clustering by Charles Blandin, Alexey Ozerov and Emmanuel Vincent.

[8] [8] Multi source TDOA estimation in reverbrant audio using angular spectra and clustering by charles baldin, Alexy Ozerov and Emmanuel vincent

[9] [9] A simple and efficient estimator for Hyperbolic location by Y.T. chan and K.C.Ho.

[10] [10] A dynamic synchrophasor estimation algorithm for online application by R. K. Mai, Z. Y. He, L. Fu, B. Kirby, and Z. Q. Bo, Senior Member, IEEE.

[11] [11] H. Sawada, S. Araki, R. Mukai, S. Makino, Grouping separated frequency components by estimating propagation model parameters in frequency-domain blind source separation, IEEE Transactions on Audio,

[12] Speech, and Language Processing 15 (5) (2007) 1592–1604.

[13] [12] J.-M. Valin, F. Michaud, and J. Rouat, “Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering,” Robotics and Autonomous Systems Journal, vol. 55, no. 3, pp. 216–228, 2007.