Real Time Sound Source Localization using Beagle Bone Black

Akshay A. J.[1] G. Pavan Kumar[2] Navaneeth P.[3] Paul V. Sebastian[4] Ranjith R. Nair[5]
[1,2,3,4,5]UG Student
[1,2,3,4,5]Department of Electronics & Communication Engineering
[1,2,3,4,5]Govt. Model Engineering College

Abstract— This paper presents the implementation of sound source localization algorithms on a low cost embedded platform. The estimation of sound source location is achieved using multilateration that uses Time Difference Of Arrival (TDOA) and Quadrant estimation. A simple algorithm and system design is proposed that overcomes the drawbacks of current estimation systems including the use of complex techniques and high end hardware. The paper focuses on estimating sound source location in real-time using BeagleBone black.

Key words: BeagleBone Black, Cross-correlation, Multilateration, TDOA, SSL

I. INTRODUCTION

Sound Source Localization (SSL) is still an upcoming technology and has not yet made it’s way into mainstream applications. Some of the reasons for this are cost of implementation, accuracy and complexity which limits this technology to mainly research areas and high end products. This technology if implemented properly, is a value addition to existing products.

Sound localization is the process of determining the location of a sound source. The brain utilizes subtle differences in intensity, spectral, and timing cues to allow us to localize sound sources. Localization can be described in terms of three-dimensional position: the azimuth or horizontal angle, the elevation or vertical angle, and the distance. Among the cue available for localization, Time difference of arrival(TDOA) cues are the most commonly used. Most algorithms are therefore developed based on the estimation of TDOA. Some of the TDOA Estimation algorithms are Cross correlation, Generalized Cross correlation Phase Transform (GCC-PHAT), Maximum Likelihood (ML) method, Average Square Difference Function(ASDF) method and Least Mean Square (LMS) Adaptive filter method .

Multilateration is a technique that uses multiple omnidirectional sensors in order to isolate the unknown position of a signal emitter in two- or three-dimensional Euclidian space. Two examples of sensors and signals could be microphones listening for sharp noises, or radio receivers listening for radio signals. In any case, these sensors are located at unique known positions where they listen for what is called a signal event. Such events are timestamped based on a synchronized or centralized clock common to all sensors. The signal from an emitter is registered by all sensors only once as the signal wave expands spherically in all directions with constant propagation speed. The time difference between when two sensors register the signal event is called the time difference of arrival (TDOA). Based on the TDOA and the location of each registration, i.e., sensor positions, we can deduce the location of the signal emitter.

The challenge of high implementation cost of localisation systems can be solved by using low cost high performance embedded boards like the BeagleBone Black. The BeagleBone Black is a credit-card-sized development platform with good supportfrom a fast growing community. The BeagleBone Black differs slightly from the regular version by providing an onboard micro HDMI port, 512MB of DDR3L DRAM, 4GB onboard ash memory, an AM3358 processor at 1GHz, and making JTAG optional with a user supplied header.

This paper focuses on implementing the SSL in BeagleBone Black. The different algorithms used are outlined and finally the obtained results are discussed.

II. ALGORITHM

The first approach to estimate the time delay is to compute the cross correlation between the received signals at two microphones

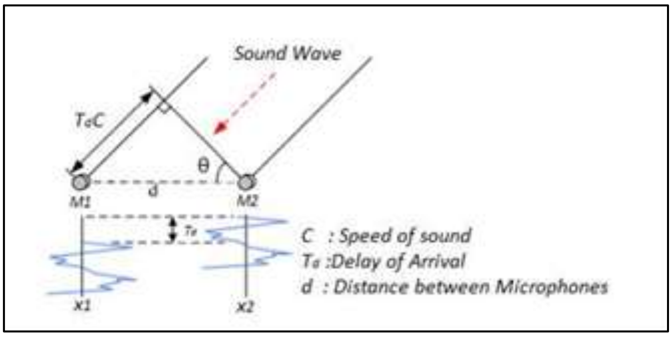

Fig. 1: TDOA estimation

Peaks of the cross correlation plot give the Time difference of arrival (Td). The location of the maximum peak cross correlation result represents the estimated time delay(Td). The sound source direction is then given by in reference to the Fig. 1. A basic frequency domain cross-correlator does an FFT on each signal, conjugates one of them, multiplies one versus the other, then inverse transforms. A peak in the result indicates a time delay.

The obtained TDOA can be used to compute the location of the sound source using the multilateration algorithm. The algorithm solves for the unknown coordinates of source (x,y) by solving the distance equations. The distance equations are written using the known coordinates and the path difference obtained from the TDOA values. The unknown coordinates are solved by converting the distance equations to an approximated linear form as described in [3].

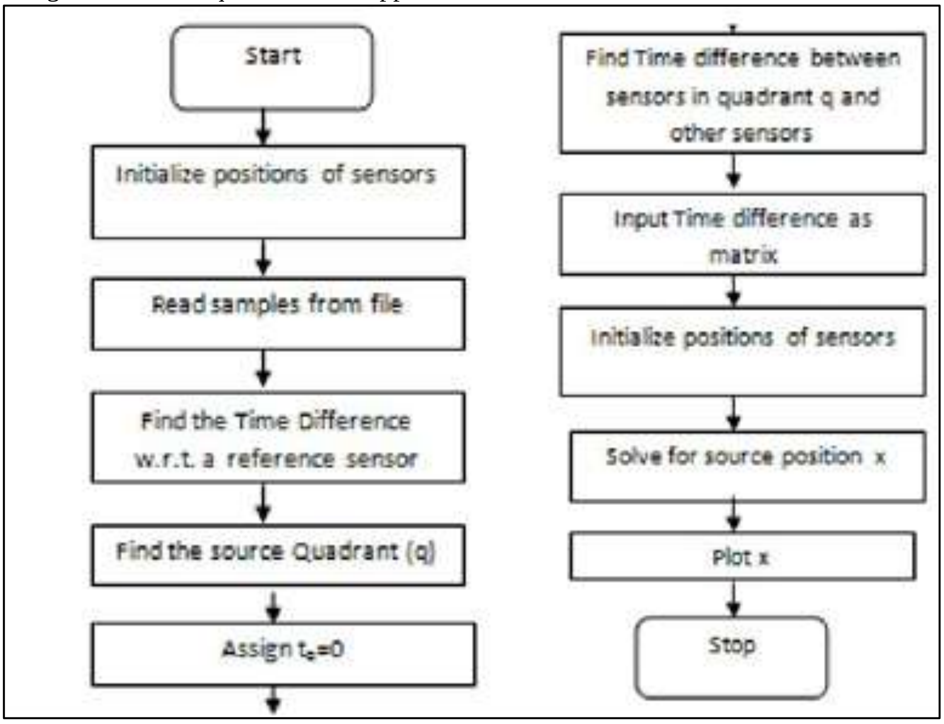

Fig. 2: Multilateration Implementation

Source position is estimated using multilateration technique that uses multiple sensors to locate a source based on the time difference of arrival of signal (from the source to each of the sensors) and the sensor positions. The input to the algorithm are the time differences to sensors based on a fixed sensor, which is the one closest to the source. This requires that the sensor closest to the source be identified initially and time difference using cross-correlation be calculated with respect to this sensor. So a Quadrant Estimation technique was developed to aid in identification of closest sensor from the source. Sensors assumed to be placed in a four quadrant system are assigned fixed representations r, 1, 2, and 3 respectively where r refers to the mic which is nearest to the source. The time differences between each sensor with respect to sensor r are calculated. Based on the values of these time differences the quadrant of source and in turn the sensor closest to source can be found. Time differences for multilateration code are then calculated with respect to this sensor.

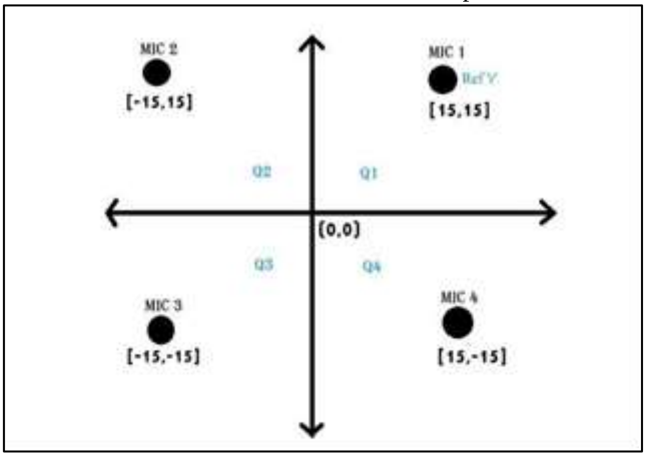

Fig. 3: Mic positions in Quadrant Estimation

III. SYSTEM OVERVIEW

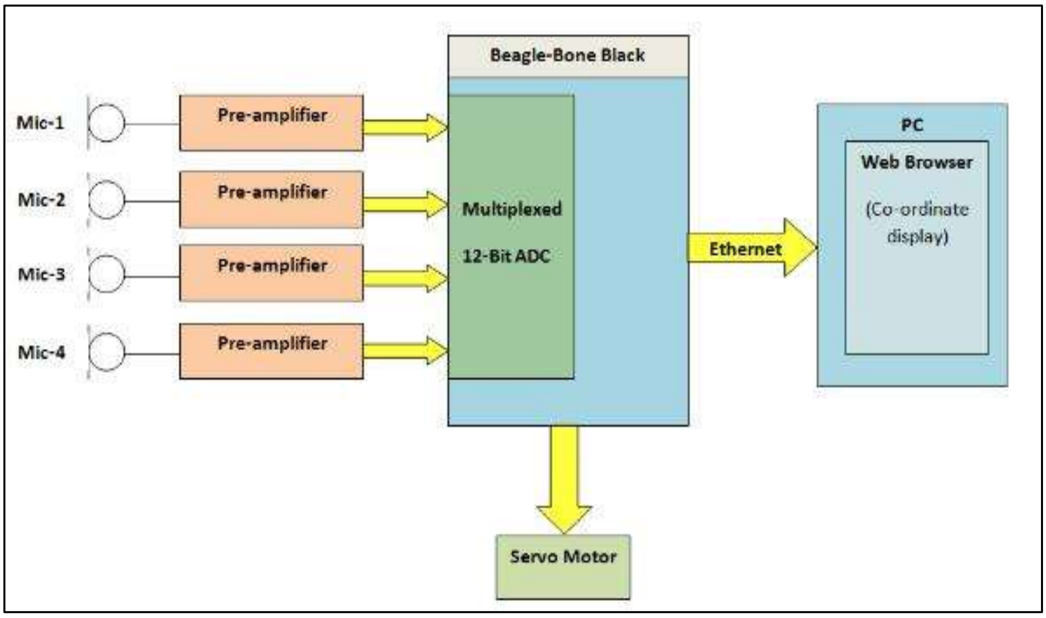

Fig. 4: Block diagram of system

The block diagram of the system is shown in fig.4. It consists of a microphone array. Signal captured by each of the microphones is pre-amplified and fed to the Multiplexed ADC pins of BeagleBone Black. The microphone array is a uniform circular array with four condenser microphones. After preamplifiers, simultaneous sampling ADCs are used to capture signals from the microphones. The synchronized filters and amplifiers mean that a strict demand on the consistency of the four channels is requested. With the availability of an 8 input channel multiplexed 12 bit ADC in the BeagleBone, each amplified signal is processed with minimum delay. For application purposes of this system like real time video camera surveillance, a servo motor is interfaced with BeagleBone which has inbuilt provisions for the same. Also, a graphical user interface can be setup which can be accessed by the user remotely via web.

IV. IMPLEMENTATION AND RESULTS

Fig 5. Shows the arrangement of the SSL system. Four microphones are placed at the edges of a square of size 20cm and the centre of the square is assumed to be the origin of the coordinate system based on which measurement is made and mic 1, 2, 3 and 4 are placed in quadrants 1, 2, 3 and 4 respectively. The microphone's four output signals are connected to the BeagleBone's ADC pins

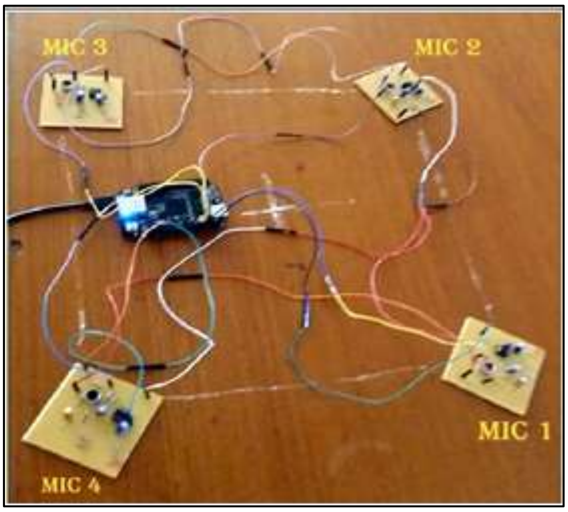

Fig. 5: Real time estimation using 4 microphones

AIN0, AIN1, AIN2,AIN3 respectively. The input ADC values are stored by Programmable Real Time Unit (PRU) into the shared RAM and fed into a buffer which is accessed by Octave. The PRU is a coprocessor system to the main ARM processor in the AM3358 SoC of the BeagleBone Black. It is programmed independent of the ARM processor and runs parallel to the ARM processor. The PRU has been used to fetch values from ADC to the system's RAM from where the Octave program accesses it's input samples.

Initially a frame size of 0.5 seconds was used to access the buffer values. Therefore after every 0.5 second at 20 Khz sampling frequency and 4 channels 40000 samples are stored in memory. The computation with 40000 samples takes approximately 0.55-0.6 seconds which is slightly in excess of the 0.5 second interval in which buffer is being filled. Therefore after around 2.5 to 3 seconds a frame buffer may be lost. A magnified view of the signals captured by mic 3 and mic 4 is shown in Fig. 6. The recorded sound is a clap sound made in quadrant 4, i.e. near Mic 4. The time difference of arrival between the different mic's and the intensity difference is clearly visible.

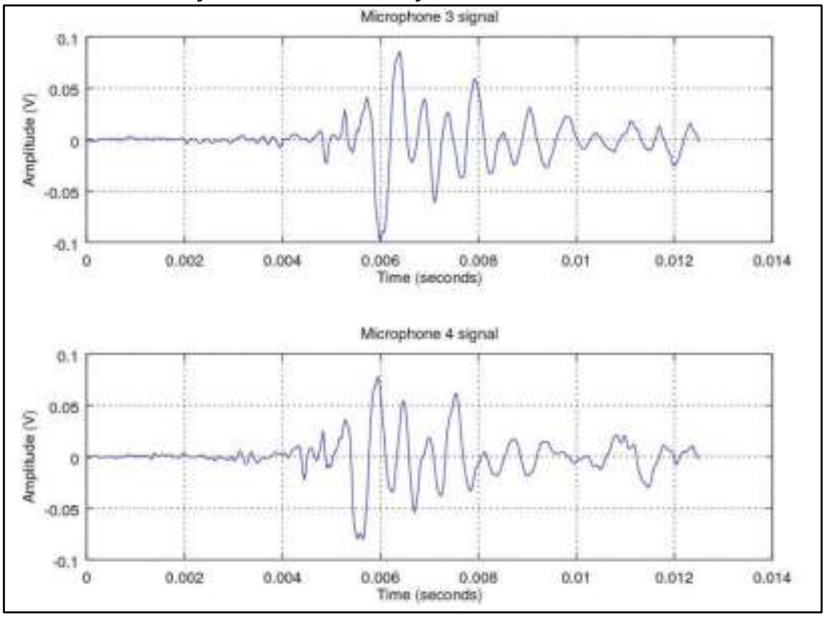

Fig. 6: Mic 3 and 4 signals showing TDOA

The cross correlation plot from which the TDOA's are calculated is shown in Fig. 7.

The results of the implementation using clap sound recorded at 20Khz are shown below in Table 1 and Table 2. The 2D location is specified in the Cartesian coordinate system (x,y) as well as the cylindrical coordinate system (r,Θ) where r is the square root of the sum of squares of x and y and Θ is the tan inverse of y/x. Table 1 consists of readings taken by performing a single recording of a clap.

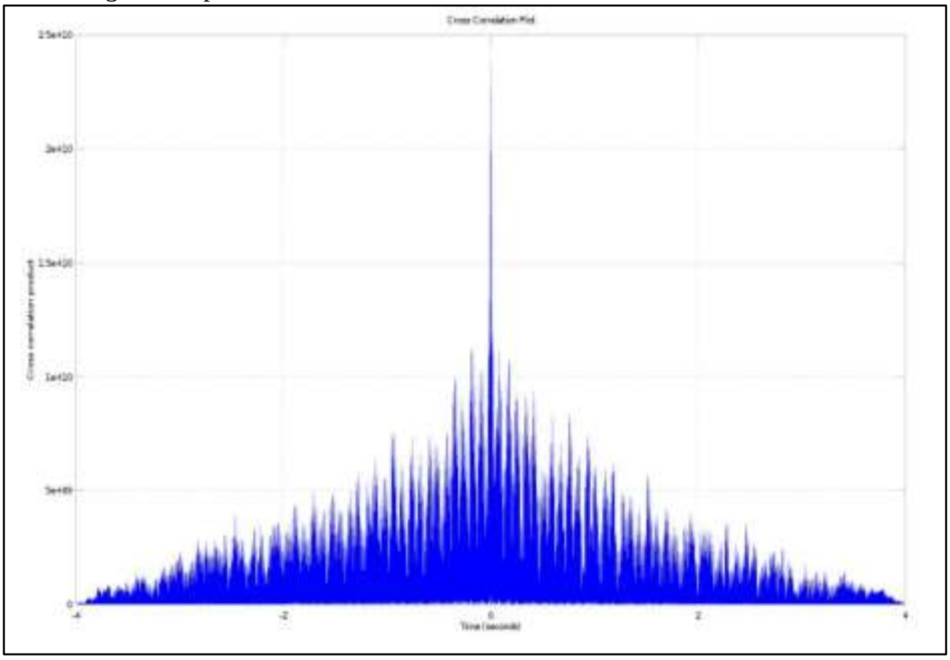

Fig. 7: Cross correlation of mic 3 and 4 signals

In Table 2 a location was fixed and multiple recordings of clap sound at that location were performed. The different estimations computed are within a consistent error margin. The location plot is shown in Fig. 8

Coordinate	Actual Reading	Observation
x (cm)	75	81.679
y (cm)	75	63.399
r (cm)	106.066	103.396
θ(degree)	45	37.819

Table 1: Readings for Fig. 8(a).

Coordinate	Actual Reading	Observation 1	Observation 2	Observation 3
x (cm)	70	55.271	78.03	72.069
y (cm)	50	40.099	49.076	35.764
r (cm)	86.023	68.285	92.18	80.455
$\theta(degree)$	35.538	35.961	32.167	26.393

Table 2: Readings for Fig. 8(b).

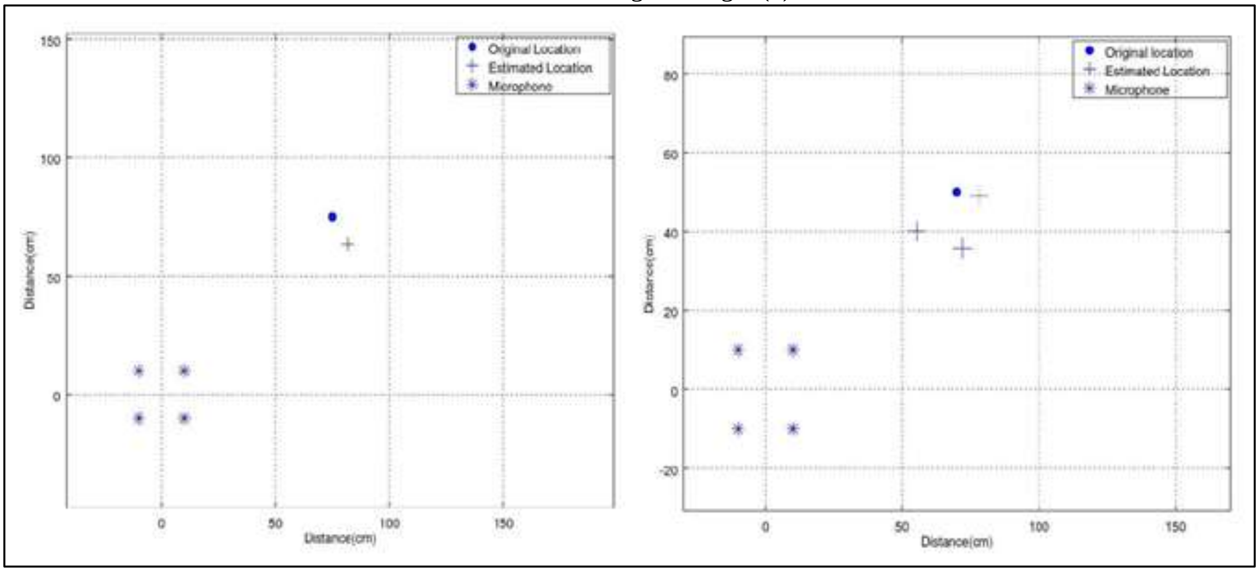

Fig. 8: Estimated locations :(a) Source at(75,75), (b) Source at (70,50)

V. CONCLUSION

This paper points out that sound source localisation can be implemented in a low cost embedded like BeagleBone and obtain results with considerable accuracy is obtained. These results can be further improved and implemented in a wide variety of applications like Robotic Auditory systems, Multi Speaker Conferencing and Surveillance in the future.

REFERENCES

[1] D. Wang and G. Brown, Computational auditory scene analysis, 1st ed. Hoboken, N. J.: Wiley-Interscience, 2006, pp. 147-185.
[2] J. Stachurski, L. Netsch, R. Cole, Sound Source Localization for Video Surveillance Camera", 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, Krakow, 2013, pp. 93-98.
[3] C. Andersen, \Comparative analysis of multilateration methods for signal emitter positioning," Tech. Rep., Jun. 2011
[4] F. Keyrouz, "Advanced Binaural Sound Localization in 3-D for Humanoid Robots", IEEE Transactions on Instrumentation and Measurement, vol. 63, no. 9, pp.2098-2107, 2014.