Sound Source Localization Using 3D Microphone Array

Dr. D. Bhavana, Dr. K. Kishore Kumar, Y. Vijay Kumar, N. Naga Bhanu Maitreyee, R. Naga Jyothirmayi, D. Ravi Tej

Abstract: Source localization and tracking with the microphone arrays have become a major interest in room acoustics, teleconference systems and tracking of sound producing objects. The current methods to estimate the source localization depend on conventional time-delay estimation techniques between microphone pairs, however, ignoring the ambient noise, reflections from surrounding and reverberation in the closed space. There are three basic and important methods for finding the direction of arrival (DOA) in a far field environment for sound sources. The first two approaches are based on Beamforming techniques: Delay and Sum Beamformer and Minimum Variance Distortion-less Response Beam former (MVDR). The third approach is a subspace method that uses the well-known algorithm, Multiple Signal Classification (MUSIC). The main goal is to prices-lylocate the direction of a sound source (azimuth, elevation) using recordings from an microphone array. This task is quite-demanding because of a high volume of acoustic noise produced by the UAV, causing negative signal-to-noise ratios (SNR). The resulted or obtained noise consists of harmonics components which are related to the speed of propellers and structural noise and also sometimes atmospheric noise due to the UAV’s movements and propellers rotations. Another problem comes from the reality that a UAV is moving constantly, sometimes with quickshiftsindirections, resulting in very complex and comparable source trajectories in the microphone array’s which is used as a frame for reference.

1. INTRODUCTION
RECENT advancements in acoustics source localization and tracking, using microphone arrays, have numerous applications in room acoustics measurements, teleconference systems, automotive industry and tracking of sound producing objects for surveillance systems. The main aim of any localizer is to precisely locate or estimate the direction of arrival (DOA) of a single or more active sound sources simultaneously, in order to point out the listening stream of sound field. A sound localizer is a core part of any capture system that uses microphone arrays of different configurations and beam-steering. The accuracy of sound source localization critically depends on the ambience and reverberation of indoor environments, such as small rooms, offices and auditorium. Traditionally, the localization methods are based on two ways to estimate the sound arrival direction.

- Dr. D. Bhavana, Associate professor, Department of Electronics and communication, koneru Lakshmaiah Education foundation, vaddeswaram, Guntur, A.P, India
- Dr. K. Kishore Kumar, Associate professor, Department of Mechanical Engineering, koneru Lakshmaiah Education foundation, vaddeswaram, Guntur, A.P, India
- Y. Vijay Kumar, N. Naga Bhanu Maitreyee, R. Naga Jyothirmayi, Graduating students, Department of Electronics and communication, koneru Lakshmaiah Education foundation, vaddeswaram, Guntur, A.P, India
- D. Ravi Tej, Assistant Professor, SRK institute of technology, Enikepadu, Vijayawada, 521108.
- Dr. D. Bhavana, Associate professor, Department of Electronics and communication, koneru Lakshmaiah Education foundation, vaddeswaram, Guntur, A.P, India
- Dr. K. Kishore Kumar, Associate professor, Department of Mechanical Engineering, koneru Lakshmaiah Education foundation, vaddeswaram, Guntur, A.P, India
- Y. Vijay Kumar, N. Naga Bhanu Maitreyee, R. Naga Jyothirmayi, Graduating students, Department of Electronics and communication, koneru Lakshmaiah Education foundation, vaddeswaram, Guntur, A.P, India
- D. Ravi Tej, Assistant Professor, SRK institute of technology, Enikepadu, Vijayawada, 521108.

The first approach is based on time difference of arrival estimation (TDOA) using paired combinations of microphone pairs, whereas, the second approach maximizes the source steered power response (SPR) at the delay output and sum beamformer. Both approaches use single frame of captured sound for estimation, producing poor results in precision and also require additional post-processing algorithms to track multiple sources in real time applications. Recent developments in simultaneous multiple source localization, such as, MUSIC and ESPRIT resolve this problem up to satisfactory extent, however, these techniques are only applicable for narrowband sources. In past few decades, microphone array based processing has been investigated for sound localization and tracking in order to emulate the existing techniques for multi-channel processing. Different microphone array configurations have been studied and proposed for estimation process, such as, direction of arrival (DOA) estimation and localizing the sources, using correlation among the two pairs of microphones. In our study, we investigated both methods, i.e. TDOA and SPR. In time delay estimation techniques, we have employed generalized cross correlation (GCC) method in frequency domain using several combinations of microphone pairs of the array with weighting functions. The array used for this purpose consists of six microphones with spherical geometric configurations in evenly distributed manners at its imaginary surface. Some of the weighting functions like, 'phase transform (PHAT)', 'smoother coherence transform (SCOT)' and the 'maximum-likelihood (ML)' for GCC algorithm are evaluated in reverberant and noisy environment. A comparative study and evaluation of these weighting functions is performed and presented; therefore, 'PHAT' weighting function is proposed for optimum detection of source in the presence of reverberant environment, especially for multiple sources. For SPR, minimum variance distortion less response weighting is evaluated and proposed for accurate source tracking application at getting high SNR of the steered signal in reverberation conditions. In addition, a practical ASLT...
based on six channel open sphere spherical microphone array is proposed with voice activity detector. The voice activity detector is used for ignoring the ambience noise and capturing only the active speaker or sound source. Several useful algorithms essential for real-time implementation are developed in order to derive and evaluate an appropriate novel localization estimator. An experimental evaluation was performed to verify the algorithmic chain of the processing and to validate the results under different environmental conditions. Therefore, a measurement setup was installed in a reverberant chamber. The results obtained from measurement process and those known beforehand showed a good agreement with each other, achieving high accuracy in results as compared to the traditional methods.

2. MICROPHONE ARRAY AND CAPTURING SYSTEM:
A microphone array is a set of spatially distributed transducers aligned in specific configurations on real or imaginary surfaces. Different microphone array configurations have been studied in recent years, such as, linear arrays, planner arrays, circular arrays and spherical arrays with their own pros and cons. The choice of configurations depends on the use of microphone arrays and the accuracy requirements.

3. Cube Shaped Array:
The orientation of the eight different microphones in an array’s coordinate frame is kept centred at the barycentre of an array. The structure holding the mics is cubical, while the microphone pattern forms two horizontal squares, each square is rotated in opposite angles in the azimuthal plane.

4. THEORETICAL ANALYSIS
TDOA (Time Difference of Arrival)
Time Difference of Arrival is the nextwellfavoured and accepted technique, and it is more adaptable than the ToA (Time of Arrival). This technique does not need the information about the time at which source sent the signal. It requires only the time of the received signal and also it requires the speed at which the signal is transmitted. Whenever the signal is captured at two respective locations which are taken as reference points, the difference in between the time of arrival can be used to find the difference in interval or separation between the source and the two reference points. This difference can be estimated using the below equation:

\[ \Delta d = c \times (\Delta t) \]

Where \( c \) is the speed of light and \( \Delta t \) is the difference in time of arrivals at each reference locations. In 2D, this equation can be written as:

\[ \theta = \arcsin \frac{\Delta d}{d_{12}} \]

In the above equation \( \theta \) is the arrival angle, \( \Delta d \) is the difference between the wave plane and \( d_{12} \) distance between different mics or microphones.

As mentioned above delay in time is obtained for all the pairs of microphones. For estimating delay time in microphones pair. We use cross-correlation between the

\[ R_{12}[k] = \frac{1}{N} \sum_{n=0}^{N-1} x_1[n] \cdot x_2[n + k] \]
previously obtained signals on both the microphones. It can be given as:

Where $x_1[n]$ is sound signal from first mic, $x_2[n]$ sound signal from second mic and $N$ is number of samples.

By applying cross-correlation on two signals we can get $k$. The obtained $k$ value which is termed as index describes the point which is inside graph at the global maximum. When you already know sample frequencies of received sound signal you can obtain a delaytime of microphone pairs follow:

$$\tau_{12} = \frac{1}{f_{oz}} \arg \max(R_{12}[k])$$

The acronym MUSIC stands for Multiple Signal Classification. The main theme of DOA estimation by MUSIC algorithm is that the narrowband signal captured by microphone array gives a covariance matrix of a rank equal to number of signal sources and can be transformed into two orthogonal subspaces namely signal subspace and noise subspace. The signal subspace is represented by

$$P_{MUSIC}(\theta) = \frac{1}{\sum_{i=k+1}^{M}} |\beta_i^d A(\theta)|^2$$

Eigen Vectors that are associated to high power Eigen Values and noise subspace is represented by Eigen Vectors corresponding low power Eigen Values. The signal subspace corresponds to array manifolds and thus the dot product of array manifold matrix $A(\theta)$ and noise subspace will be minimum (zero) in the direction of true DOA. The covariance matrix of the observed signal by the microphone array is given by. This covariance matrix can be expressed in terms of Eigen values and corresponding Eigen vectors. The Eigen Vectors corresponding to maximum Eigen values represent signal subspace and the Eigen Vectors corresponding to minimum or equal Eigen values represent noise subspace. Thus if it is assumed that there are K sources from which speech signals are arriving at the array, the largest Eigen values of $R$ represents a power function of every different K sources, while the eigenvectors are used to stretch the K-dimensional subspace related to signal of $R$. The least (M-K) Eigen value is the representation of noise power, and theoretically they are said to be equal, under the assumption of white noise. The Eigenvectors which are correlated with these Eigen values are used to stretch the M-K dimensional subspace related to noise of $R$. It has been proven that the eigenvectors correlated with the least M-K Eigen values are orthogonal to the direction of vectors which are equivalent to the arrival angles of the sources. This algorithm computes function $PMUSIC$ as the indicator of DOA given by Where the $\beta_i$ represents the Eigenvectors corresponding to noise subspace, and $A(\theta)$ is a array vector manifold for every element in the array which corresponds to signals subspace. The function $PMUSIC$ is computed for the different values of $k$. When value of $k$ becomes equal to that of DOA the denominator becomes zero and $PMUSIC$ becomes maximum. Obviously, graph of $k$ versus $PMUSIC$ will show peaks for the DOAs. This is the general method of computing DOA using MUSIC algorithm. Its application in DOA estimation for the broadband signals such as speech can be done in the frequency domain. The spectrogram of the speech signal reveals that the signal energy is distributed in different bands of frequencies over the wide range of frequencies.

5 MUSIC ALGORITHM FOR DOA ESTIMATION

6 EXPERIMENTAL RESULTS

Static:

Flight:
6.1 Static:
Processing audio sample 01/03:
-- frame: 01/01
Input signal duration: 2.20 seconds
Input signal truncated at 96256 samples with respect to window length used
Block 1 / 1
Src: 1
Az.: 90
El.: 0
Elaps. time : 6.967571 sec.
90.00 90.00
0.00 0.00

Processing audio sample 02/03:
-- frame: 01/01
Input signal duration: 2.20 seconds
Input signal truncated at 96256 samples with respect to window length used
Block 1 / 1
Src: 1
Az.: 75
El.: -1
Elaps. time : 6.122288 sec.
75.00 75.00
0.00 -1.00

Processing audio sample 03/03:
-- frame: 01/01
Input signal duration: 2.20 seconds
Input signal truncated at 96256 samples with respect to window length used
Block 1 / 1
Src: 1
Az.: 45
El.: -27
Elaps. time : 5.998592 sec.
45.00 45.00
-30.00 -27.00

6.2 Flight:
Processing audio sample 01/05:
-- frame: 01/15
Input signal duration: 0.50 seconds
Input signal truncated at 20480 samples with respect to window length used
Block 1 / 1
Src: 1
Az.: -88
El.: -10
Elaps. time : 1.970961 sec.
-86.52 -88.00
-12.06 -10.00

-- frame: 02/15
Input signal duration: 0.50 seconds
Input signal truncated at 20480 samples with respect to window length used
Block 1 / 1
Src: 1
Az.: -76
El.: -5
Elaps. time : 1.962465 sec.
-77.31 -76.00
-4.84 -5.00

-- frame: 05/15
Input signal duration: 0.50 seconds
Input signal truncated at 20480 samples with respect to window length used
Block 1 / 1
Src: 1
Az.: -71
El.: -5
Elaps. time : 1.944919 sec.
-73.66 -71.00
-3.61 -5.00

-- frame: 06/15
Input signal duration: 0.50 seconds
Input signal truncated at 20480 samples with respect to window length used
Block 1 / 1
Src: 1
Az.: -70
El.: -3
Elaps. time : 2.014733 sec.
-70.92 -70.00
-4.05 -3.00

-- frame: 07/15
Input signal duration: 0.50 seconds
Input signal truncated at 20480 samples with respect to window length used
Block 1 / 1
Src: 1
Az.: -70
El.: -3
Elaps. time : 1.980753 sec.
-69.06 -70.00
7 DISCUSSION OF RESULTS
In the static case, the position of the drone carrying the microphone array is kept static at a position. Taking three audio samples the Azimuthal and Elevation angles are estimated for each of the audio sample. In the flight case, the drone is moving, for each audio signal the Azimuthal and Elevation angles are estimated for different time frames. The estimated and the original trajectories are almost similar. The function Mbss_pre-process values needed to compute and aggregate the angular spectrum over all microphone pairs. The output parameters of this function are pairId, dMic, alpha, alphaSampled, tauGrid.

- nMicPair-28 (Number of pairs of mics)
- pairId - nMicPair x 2, All microphone pair indexes
- alpha - array of angles(angle of arrival ) for each microphone pair corresponding to all (θ, φ) to be tested.
- taugrid - time difference of arrival.

8 CONCLUSIONS
We presented a framework for difference of arrival in time estimation and sound source localization using eight channel cube microphone array based on PHAT-GCC weighted algorithms. By the TDOA method the estimated and actual position of the source differ by within 10 degrees. There is slight change or a difference between a frame workand the direction previously estimated. It is done because of the problem with the sampling frequency and also rounding. It was also observed that the increase in size of the array microphone which improves accuracy with respect to the DOA estimate but also the computational cost increases.

9 REFERENCES