International Journal of Electronics Engineering Research.
ISSN 0975-6450 Volume 9, Number 4 (2017) pp. 545-556
© Research India Publications
http://www.ripublication.com

Study Of Sound Source Localization *Using Music Method In Real Acoustic Environment*

Dr. Navin Kumar[1]

[1]*Centre Coordinator, IGNOU Centre for Engineering, Muzaffarpur, Bihar, India.*

Abstract

Sound source localization is very important and challenging problem in speech signal processing because it is very difficult in real acoustic environment and is of great practical importance [1] [5] . Microphone array is imitation and extension of human listening process with two ears. Basic idea involved in estimation of Direction of Arrival (DOA) of sound source is to find out some distinctions in signals observed at different sensor points. Microphone array does the same by capturing spatio-temporal characteristics of the speech signal which is used to estimate DOA. Acoustic source localization is very important technique required in many areas of practical applications such as teleconferencing, human machine spoken communication, active audition and development of spoken interfaces etc. Human beings are capable in extracting much information such as about speaker, distance and direction of active speaker in addition to language dependent information from the sound waves reaching at earsIn this research paper study on one of the very popular subspace based techniques, namely MUSIC algorithm, for DOA estimation in real acoustic environment has been described. The acronym MUSIC stands for Multiple Signal Classification and is a subspace based technique for DOA estimation. There have been developments of different techniques for DOA estimation The performance of MUSIC algorithm for DOA estimation is presented under reverberation. It has been found that performance of algorithm heavily deteriorates with increasing reverberation.

Keywords: Acoustic-source localization; microphone-arrays; Beamforming; Music;

INTRODUCTION

Acoustic source localization based on microphone arrays has been one of the mainstream research topics for the last two-three decades. The basic theory behind the DOA estimation is that the signal captured by different array elements are delayed in time and thus suffers phase shift. For the known geometry of the microphone array, the phase information at different sensors depends on direction of arrival of the signal and thus using the same information, DOA can be estimated. The solutions available in the literature for DOA estimation can be classified into three broad categories namely (a) methods based on maximizing the Steered Response Power (SRP) of a beamformer, (b) method based on High-Resolution Spectral Estimation (HRSE) methods and (c) methods based on Time-Difference of Arrival (TDOA) estimation algorithms. SRP-based localization methods rely on a focused beamformer, which steers the array to various locations in space, and look for peaks in the detected output power [38] . In its simplest implementation, the steered response can be obtained through a Delay-and-Sum process performed on the signals acquired by a microphone array. Source localization methods of the second category are all based on the analysis of the Spatial Covariance Matrix (SCM) of the array sensor signals. The SCM is usually unknown and needs to be estimated from the acquired data. Such solutions rely on high resolution spectral estimation techniques. Popular algorithms based on HRSE are Minimum Variance beamformer and Multiple Signal Classification (MUSIC) algorithms etc. For the same short time Fourier transform is used to estimate SCM on narrowband parts of the captured signal. Time delay estimation based algorithms for estimation of direction of arrival (DOA) have been most popular for use with speech signals. This is due to their simplicity and low computational requirements. TDOA methods extract information on the source location through the analysis of a set of delay estimates. Methods based on TDOA algorithms have two steps [21] . First, they estimate the TDOAs. The most popular method for TDOA estimation is the cross correlation approach [39] [40] .

MUSIC ALGORITHM FOR DOA ESTIMATION

The acronym MUSIC stands for Multiple Signal Classification and was developed by R.O. Schmidt in the late 70s that laid foundation for subspace based array signal processing and subspace based frequency estimation [42] . It has been used to estimate DOA of multiple sources. The basic idea 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 decomposed into two orthogonal subspaces namely signal subspace and noise subspace. The signal subspace is represented by Eigen Vectors corresponding 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 $A(\ )$ 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

$$R \underline{\underline{def}} E[X(t)\overset{*}{X}(t)] = \lim_{N\to\infty} \frac{1}{N}\sum_{K=1}^{N} xk(t)\,x\overset{*}{k}(t). \qquad (1)$$

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 there *K* sources from which speech signals are arriving at the array, the largest *K*

Eigenvalues of **R** represent a function of the power of each of the *K* sources, while their eigenvectors are said to span the *K*-dimensional signal subspace of **R.** The smallest (*M-K)* Eigenvalues represent the noise power, and theoretically they are equal, under the white noise assumption. The Eigenvectors that are associated with these Eigenvalues are said to span the *M-K* dimensional noise subspace of **R.** It has been shown that the eigenvectors associated with the smallest *M-K* Eigenvalues are orthogonal to the direction vectors corresponding to the arrival angles of the sources. The MUSIC algorithm computes function $P_{MUSIC}(\theta)$ as the indicator of DOA given by

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

where the β_i represent the Eigenvectors corresponding to noise subspace , and

$\mathbf{A}(\theta)$ represents a vector of array manifold for each array element which corresponds to signal subspace. The function P_{MUSIC} is computed for the different values of (θ). When value of θ becomes equal to that of DOA the denominator becomes zero and P_{MUSIC} becomes maximum. Obviously, graph of θ versus P_{MUSIC} 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 an arbitrary speech signal is shown in the Figure 3.2. The spectrogram of the speech signal revels that the energy of the signal is distributed in different frequency bands over the wide range of frequencies. This clue hints that these frequency bins are most suitable for the DOA estimation. The application of MUSIC algorithm requires stationarily in the speech signal but it is not so.

EXPERIMENTS AND RESULTS

The two element linear microphone array with inter-element spacing of 4 cm was used to collect speech signal. The captured signal was sampled at sampling frequency of 8 kHz. The signal was also simulated for different reverberation time. The true DOA of a speech source is at $_40°$. The speech signals observed at both microphone are shown in Figure 3.6. The TFSS of speech signal were obtained as described above in Figure 3.3 using 1024 point DFT and Hanning window for frame length of 20 ms with 50% overlap. The cross-sensor covariance matrix was estimated in each frequency bin. The MUSIC algorithm was used to estimate DOA in each frequency bin. The signal and noise spaces were estimated using Eigenvalue decomposition of R. The Eigen Values for the frequency bin f=1500 Hz is shown in

the Figure 7. The Eigenvalues were arranged in decreasing order and Eigenvectors corresponding to Eigen Values were selected for the signal and noise subspaces. The eigenvector corresponding to zero Eigenvalues were taken as noise subspace. Then value of P as per Eq was estimated. The plot of P_{MUSIC} and θ is shown in Figure 3.8 for different values of RT. In this figure peak of the curve shows estimated of DOA under different reverberated conditions. It can be observed how the performance of algorithm degrades with increasing reverberation. The estimates of DOA in all frequency bins are not same. The estimate of DOAs in some selected frequency bins are shown in Figure 3.9. Next the speech signal was simulated for two speakers using two microphones. The speech signals, shown in Figure 3.10 are for two element microphone array, simulated for the two speakers situated at $_30°and27°$ at the distance of 1m. In such a situation when the number of sources is equal to number of sensor, all the eigenvalues for such a case, as shown in Figure 3.14 for two sensor two speaker case, represent signal subspace and noise subspace cannot be estimated. Thus the number of used sensor was increased. The captured speech signals for 3, 4 and 5 element ULAs are shown in Figure 6, Figure and Figure 7 and estimated corresponding eigen structure of cross sensor covariance matrices are also shown in Figure 3.15, Figure 8 and Figure 9 respectively. The estimates of DOA using MUSIC algorithm as per Eq.(3.11) are plotted in Figure 3.18 for all the cases of number of microphones considered here. Obviously, with increasing number of sensors peaks in the curve shifts towards true DOA and estimated DOA are more accurate. In Figure 3.19 , DOA estimates, for source positions at -40 degree, in different frequency bins for the speech data captured by two element linear microphone array are shown for different values of RTs. It can be observed in that figure that with increasing RT the estimation of DOA by MUSIC method becomes less accurate. Next the speech signals were simulated for five different speakers including male and female subjects using two elements ULA. The DOA in each frequency bins were estimated as per Eq. (3.12) for each speaker for different RTs. The average value of DOA for different values of RT estimated as per Eq.(3.12) are shown in Figure.

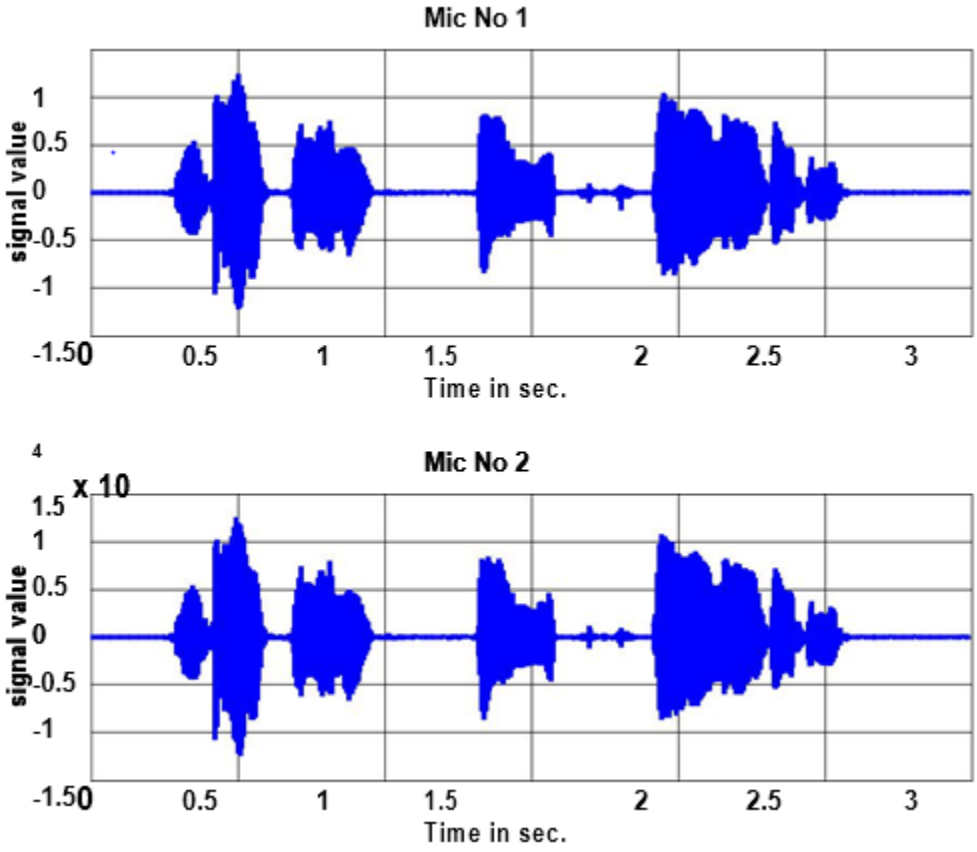


Figure 1 Speech signal captured by two element microphone array from single speaker

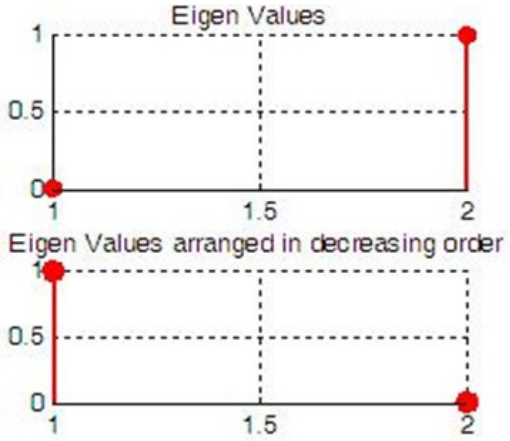


Figure 2 Eigenvalues of covariance matrix.

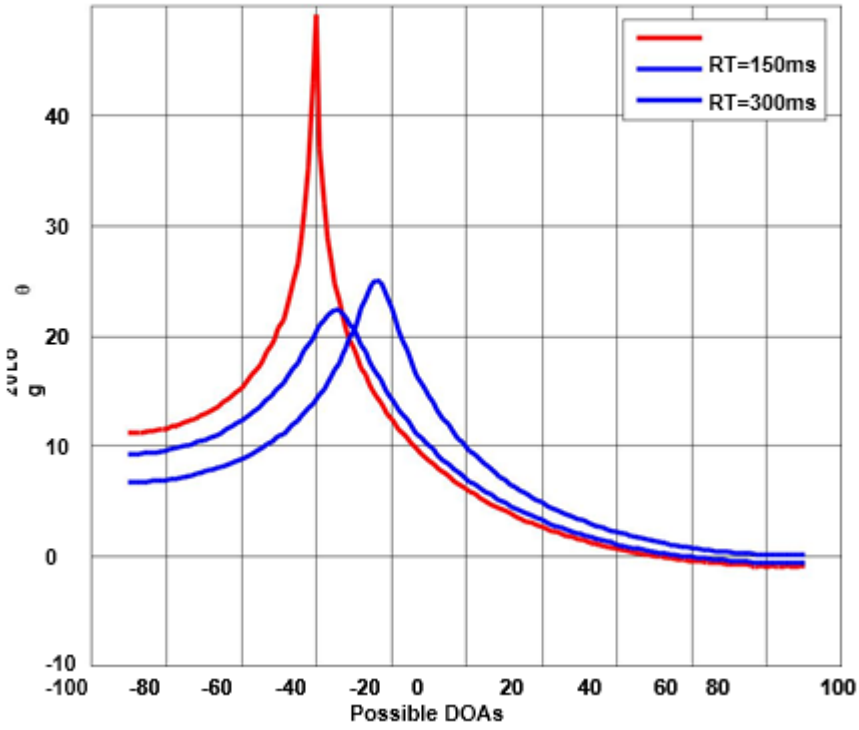


Figure 3: DOA estimates for different values of RT in frequency bin f=1.5 kHz. (True DOA=-40°)

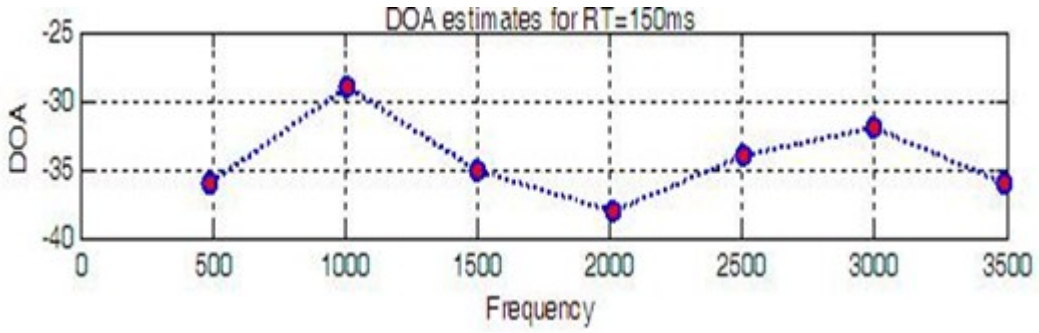


Figure 4 DOA estimation in some selected frequency binsfor RT=150 ms. (True DOA=-40°)

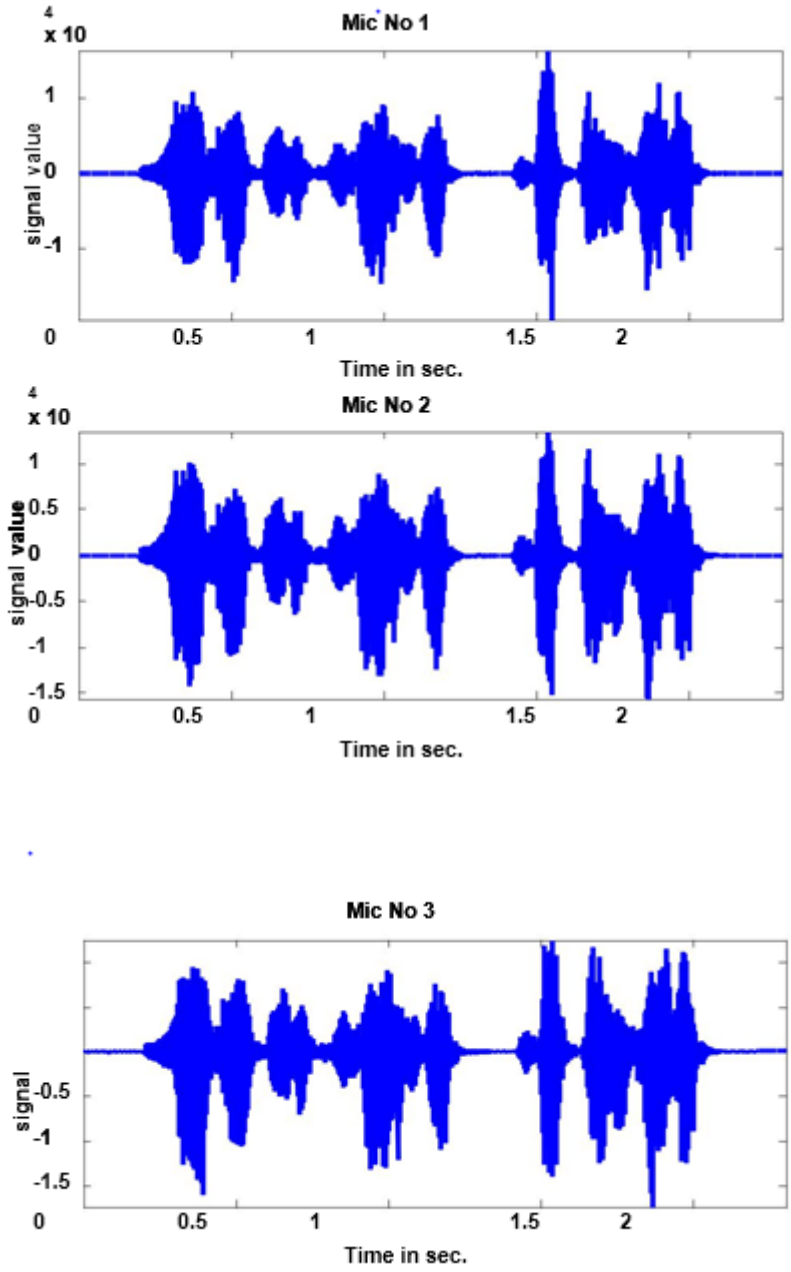
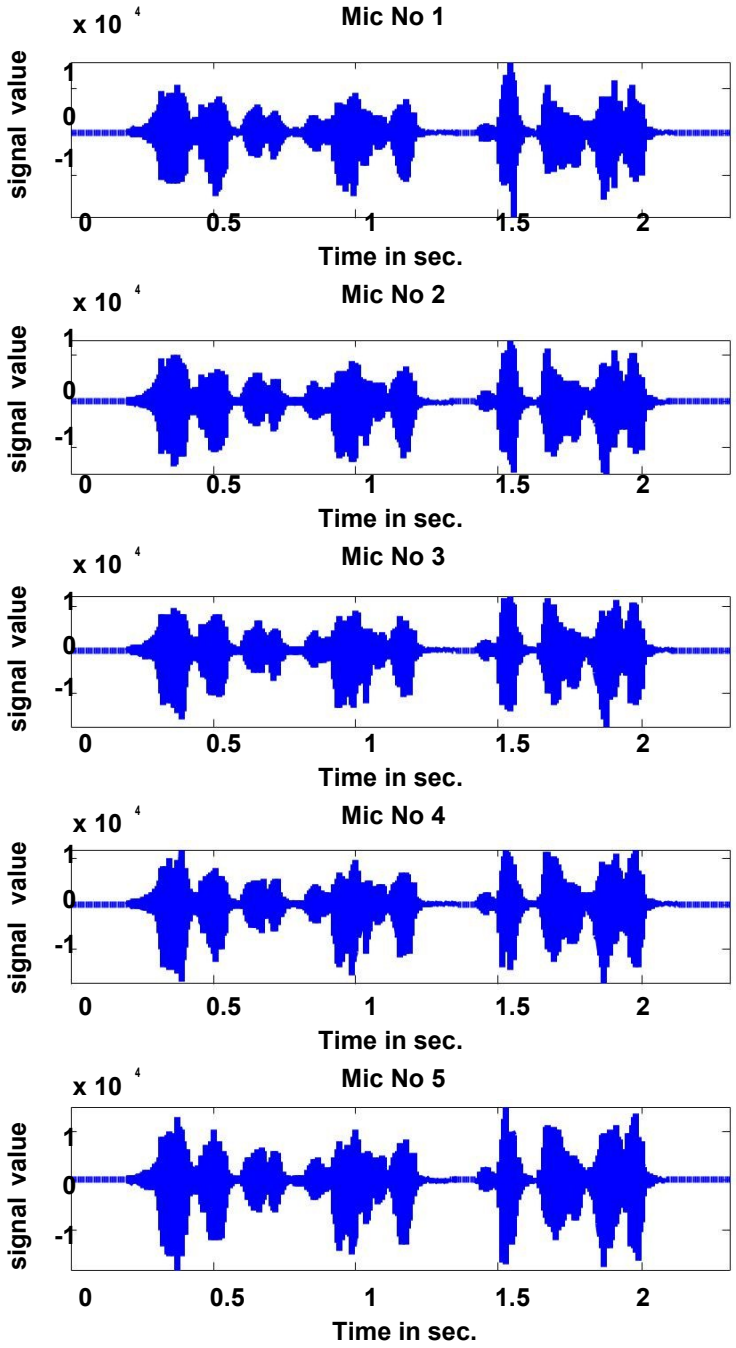


Figure 5. Speech signals for two speakers captured by 3-element ULA.

Figure 6. Speech signals for two speakers captured by 5-element ULA.

Figure 7. Speech signals for two speakers captured by 4-element ULA.



4

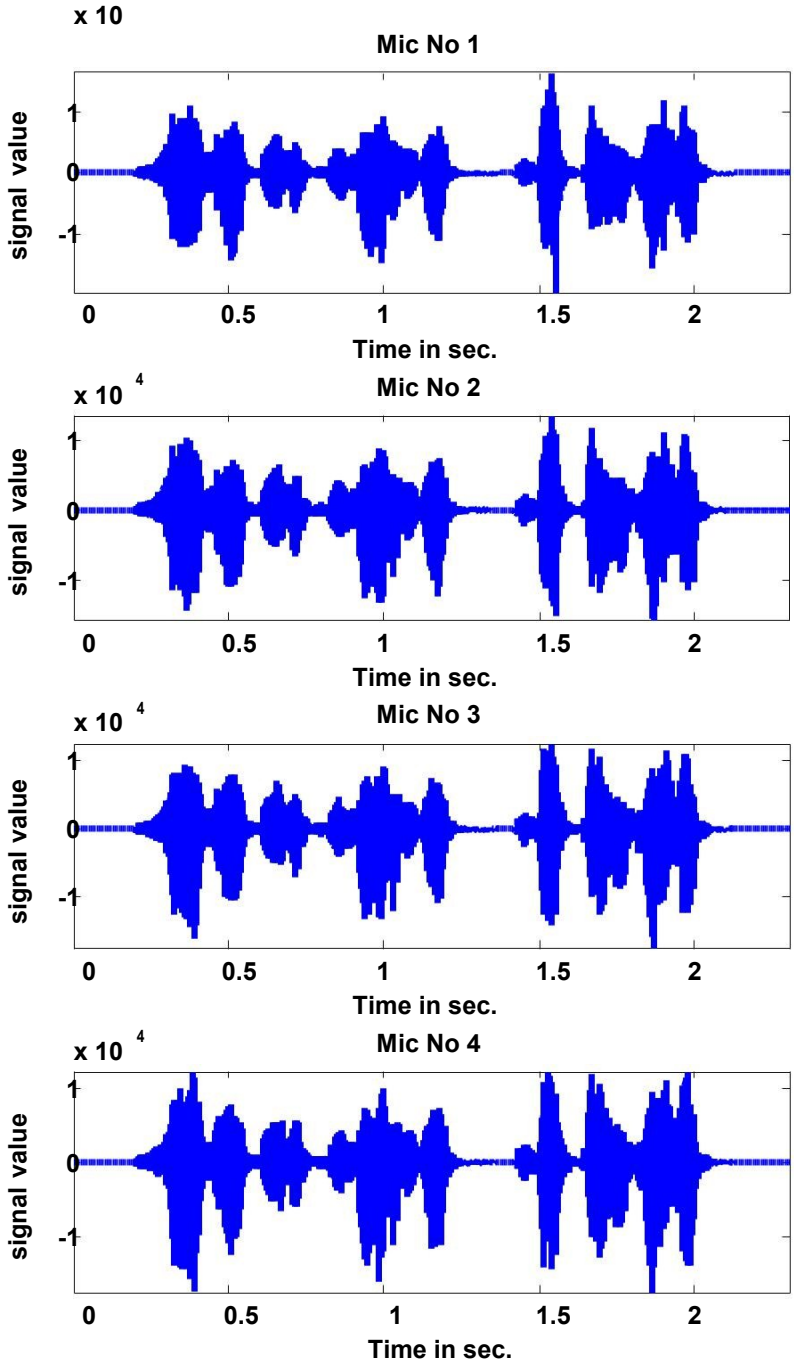

x 10
Mic No 1
signal value
1
0
-1
0
0.5
1
1.5
2
Time in sec.
x 10 4
Mic No 2
signal value
1
0
-1
0
0.5
1
1.5
2
Time in sec.
x 10 4
Mic No 3
signal value
1
0
-1
0
0.5
1
1.5
2
Time in sec.
x 10 4
Mic No 4
signal value
1
0
-1
0
0.5
1
1.5
2
Time in sec.

Figure 8. P-θ curve for the case of 2 speakers and 3-element,4-element and 5-element microphone arrays (for f=2000 Hz).

For RT=0 ms

Averaged value of DOA for five different speakers at -40 degree

Figure 9. Averaged DOA estimated by MUSIC algorithm for five different speakers for location -40 degree (In the bar graph absolute values of true and estimated DOAs are shown to invert the bar graph) for different values of RTs.

CONCLUSION

In this Paper the basic concept of microphone array and it application in DOA estimation of active acoustic source have been presented. How the performance of the MUSIC algorithm deteriorates with increasing reverberation has also been shown. The real acoustic environment is very dynamic in the sense that position of speaker and sensor may change with time and presence of noise, reverberation, coherency of sources etc. may exist . One needs to develop algorithm that can cope up with such variations. It was also observed that the increase in size of the microphone array improves accuracy of the DOA estimate but the computational cost also increases. The evaluation results in the present work are based on the raw DOA estimation results, so that a post-processing for example by grouping and interpolating the detection results will probably increase the accuracy numbers. Post-processing of the detected DOA is scope of our next work.

REFERENCE

[1] J.Benesty (Eds) .et.al.: Microphone array signal processing, Springer Topics in Signal Processing, Springer, 2009.

[2] J.E. Greenberg et.al.: Microphone array in hearing aids, M. Brandstein (Ed.) Microphone array, Springer Verlag,2001.

[3] T. J. Shepherd S. Haykin, and J. Litva, editors. : *Radar array processing.* Springer-Verlag, New York, 1992.

[4] B.D. Steinberg.: Principles of aperture and array system design. Wiley, New York, 1976

[5] S. Haykin. : *Array signal processing.* Prentice-Hall, Englewood Cliffs, New Jersey, 1985.

[6] R. J. Mailloux. : *Phased array antenna Handbook.* Artech House, Boston, Massachusetts, 1994.

[7] D. H. Johnson and D. E. Dudgeon.: Array signal processing. Prentice-Hall, Englewood Cliffs, New Jersey, 1993.

[8] J. C. Hassab: Underwater signal and data processing. CRC Press, Boca Raton, Florida, 1989.

[9] L. J. Ziomek: Fundamentals of acoustic field theory and space-time signal processing. CRC Press, Boca Raton, Florida, 1995.

[10] Okada et.al. :3-dimensional sound source localization and voice-separation by three microphones. J of institute of systems, control and information engineers, vol.6, vol3, pp-149-155, 1993

[11] D.C More: Speech enhancement using microphone array, university of technology, 2000.

[12] M. Brandstein: Time delay estimation of reverberated signal exploiting harmonic structure. J. of Acoustic society of America, Vol 105, no.5 Pp. 2914 – 2919, 1999.

[13] J. chan et. al.:, Time-delay estimation using spatial correlation techniques. More DWAENC 2003, Japan.

[14] T. Nishitra et. al.: Localization of multiple sound sources based on a CSP analysis with a microphone array.

[15] M. Omologo et. al, Acoustic event localization using CSP based technique, froc ICASP94 99 273-76, 1994

[16] M. Omologo et. al.: Speech recognition with microphone array, M.Brandstein (Ed.) Microphone array, Springer Verlag,2001.

[17] W.L. Kellerman: Acoustic echo cancellation for beamforming microphone array, M.Brandstein (Ed.) Microphone array, Springer Verlag,2001.

[18] S.C. Douglas: Blind separation of acoustic signals, M. Brandstein (Ed.) Microphone array, Springer Verlag,2001.

[19] R.K.Prasad et.al.: Enhancement of speech signal separated from their convolutive mixture by FDICA algorithm, International Journal of DSP, Vol 19 (1) Elsevier,2009.