DOA Estimation Techniques using Microphone array –A comparative study

Dr.Navin Kumar, Dr. Alka Singh

Center Coordinator, IGNOU center for Engineering, Muzaffarpur, Bihar, India

Assistant Professor, Dept. of Pure &Applied Physics, GGD Central University, Bilaspur, Chhattisgarh, India

ABSTRACT: Acoustic-source localization based on microphone-arrays has been a mainstream research topic for over two decades. The solution available in the literature can be broadly categorized into three categories mainly those based on maximizing the steered response power (SRP) of a beamformer , those based on High-resoluiton spectral estimation (HRSE) methods; and those based on time difference of arrival (TDOA) algorithm source localization methods of the second categories are all based on the analysis of the spatial convergence matrix (SCM) of the array sensors 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 algorithm based on HRSE are minimum variance beamformer and multiple signals classification (MUSIC) algorithm. These algorithms can be extended to wide band signals ex. Speech by transforming the signal into narrow band signal. Each narrow band signal can be processed individually (incoherent method) or a universal focusing SCM can be generated to perform a coherent localization.

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

I. INTRODUCTION

Microphone array is a series of microphones in a spatial pattern that process acoustic signal. An array of sensors is often used in many fields of science and engineering, particularly when the goal is to study propagating wave-fields. It captures spatio-temporal characteristics of signal or quantity of interest. The main goal of array signal processing is to do source localization (In Radar, Sonar, Seismology, etc), Source waveform estimation(In communication etc),source characterization(in seismology) or imaging of scattering medium(In medical diagnosis or seismic exploration).The estimation of direction of arrival (DOA) in array signal processing is a mean of determining the direction of each source signal using the information contained in the signal. In acoustic signal processing DOA estimation plays an essential part in microphone array technology as a pre-process in speech enhancement or noise robust recognition.

Time delay estimation (TDE)-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. Though other algorithms, like the steered response power with phase transform (SRP-PHAT), are available that perform better than TDE based algorithms, the huge computational load required for this algorithm makes it unsuitable for applications that require fast refresh rates using short frames. In addition, the estimation errors that do occur with SRP-PHAT tend to be large. This kind of performance is unsuitable for an application such as video camera steering, which is much less tolerant to large errors than it is to small errors. Various factors affect the accuracy of the DOA estimates obtained using the microphone array. Accuracy of the hardware used to capture the array signals, sampling frequency, number of microphones used, reverberation and noise present in the signals, are some of these factors. This is a requirement no matter what method is used for DOA estimation. Also, the more microphones we use in the array the better the estimates are that we get effect of reverberation on different DOA estimation techniques are analysed and compared statistically in this paper.
The basic theory behind the estimation of direction of arrival using Microphone Array is to make use of the phase information present in the signals picked by sensors which are spatially separated. When the microphones are spatially separated the sound source signal arrive at them with time differences. For known array geometry, these time- delayed signals are dependent on the direction of arrival of the signal.

In so far as the estimation of direction of arrival for narrowband sources are concerned, the theory is well established and lots of literature is available. Within many direction of arrival algorithms, MUSIC (Multiple Signal Classification) [1,2] has been most widely studied. The MUSIC algorithm is based on Eigen-value Decomposition (EVD) method. The EVD method divides the cross-correlation matrix of the array signals into signal and noise subspaces. The popularity of MUSIC algorithm is due to its generality i.e. it is applicable to arrays of arbitrary but known configurations and response, and can be used to estimate multiple parameters per source. The condition is that array response must be known for all the possible combinations of source parameters.

The narrowband MUSIC produces a sharp beam patterns, but requires the frequency bin to have high SNR. In general, any narrowband direction of arrival technique will not exploit the wideband nature of the acoustic sources. To exploit as much of the multispectral content from the acoustic source as possible, improve accuracy and stability of the direction of arrival estimates, a wideband direction of arrival algorithm is required.

One approach is to implement wideband MUSIC algorithm using Incoherent Signal- Subspace method [3] in frequency domain. This approach is useful if there is sufficient or high SNR in multiple frequency bins, so that narrowband MUSIC algorithm yields good results independently for each bin. Over each processing interval it is assumed that a single frequency bin is occupied by a single source only. This takes advantage of the non stationary nature of the source and simplifies the algorithmic complexity of the algorithm. This assumption is justified because different wideband sources are not likely to occupy all of the same bins in any given processing interval and keeps on changing bins as function of time.

The effects of multipath are encountered in a received signal when the source signal reflects off of surrounding objects and gets added to the direct path signal with a delay. The larger the number of surrounding objects, the more reflected signals are added to the direct path signal. For acoustic sources and microphone arrays placed inside a room, this effect can be quite large. The sound reflects off the walls, floor and ceiling of the room, multiple times, and these reflected signals get added to the direct signal. This effect is called room reverberation. Reverberation causes drastic changes to the time delay estimates derived from signals at the different microphones of an array. These changes are of a local nature with respect to time, which means that at certain instants of time there could be strong reflections and at certain other instants the reflections could be weak. Because of this, if we estimate the time delays using a short frame of signal data, the estimates keep changing over time. This introduces a significant challenge to algorithms performing DOA estimation.

Sound may be considered as a traveling wave that is generated by the vibrations of a plane surface that is in contact with a medium. The vibrations of the plane surface cause the layer of molecules of the medium close to the surface to compress and expand alternately. These compressions and expansions are then transferred to the next layer of molecules and so on. This way the sound generated by a vibrating body is transferred through a medium. At any point in time, the space surrounding the vibrating plane will consist of waves of compressed or expanded molecules of the medium. Such a space, which has moving sound in it, is called a sound field. The compressions and expansions of the medium at any point cause the pressure at that point to keep changing instantaneously. This variation in pressure at any point in the medium is what is heard as the sound signal. When considering the sound field in an enclosed room, the use of the wave model can become quite challenging. Apart from considering the effect of superposition of numerous reflected waves, one also needs to take into account the particle velocity normal to the wall of reflection. This effect, which is characterized by the specific impedance of the wall, has not been considered in the discussion in Section 3.2. A simpler approach to take is to take the limiting case of very small wavelengths (high frequencies) and thus replace the sound wave with a sound ray and then use geometrical acoustics. This simplification is justified for wavelengths that are arbitrarily small when compared to the dimensions of the room and distances traveled by the sound wave. For frequencies around the medium range (1000 *Hz*, 34 *cm* wavelength) this approximation is valid for typical rooms. Several other assumptions are made when using this approach. The sound ray originates from a certain point and has a well-defined direction of propagation. It has a finite velocity of propagation and follows the law of reflection when it encounters a rigid wall. The medium in the room is assumed to be homogeneous, i.e. there are no sudden changes in density in the medium, and thus refraction is assumed to be non-existent and the sound rays travel in straight lines until they encounter reflecting walls. Also, since sound rays do not change directions while traveling in the medium, diffraction is also assumed to be non-existent.

Under these circumstances there are three effects that determine the acoustics of a room, viz. finite velocity of sound, absorption of sound energy by the walls during each reflection and absorption of sound energy by the medium. The finite velocity of sound causes reflected signals to arrive at the listener with finite delays and these signals get added to the original source signal. The reflections at walls can be simplified by making the reflection coefficients real valued and independent of either frequency or angle of incidence. A mean value can be used to represent reflection coefficients for all frequencies and all angles. The propagation of sound in a medium is not ideal. Sound is transmitted from one layer of the medium to the next by mechanical collisions between adjoining molecules. These collisions are not ideal and some energy is lost as heat, which goes towards increasing the temperature of the medium.

The other consideration while simulating reverberation for a room is the duration of reverberation or the reverberation time. Formally, the reverberation time is defined as the time required for the intensities of reflected sound rays to be down 60 dB from the direct path sound ray.An empirical formula, known as Eyring's formula [18] can be used to relate thereverberation time, T_r, to the reflection coefficient, R, of the walls.
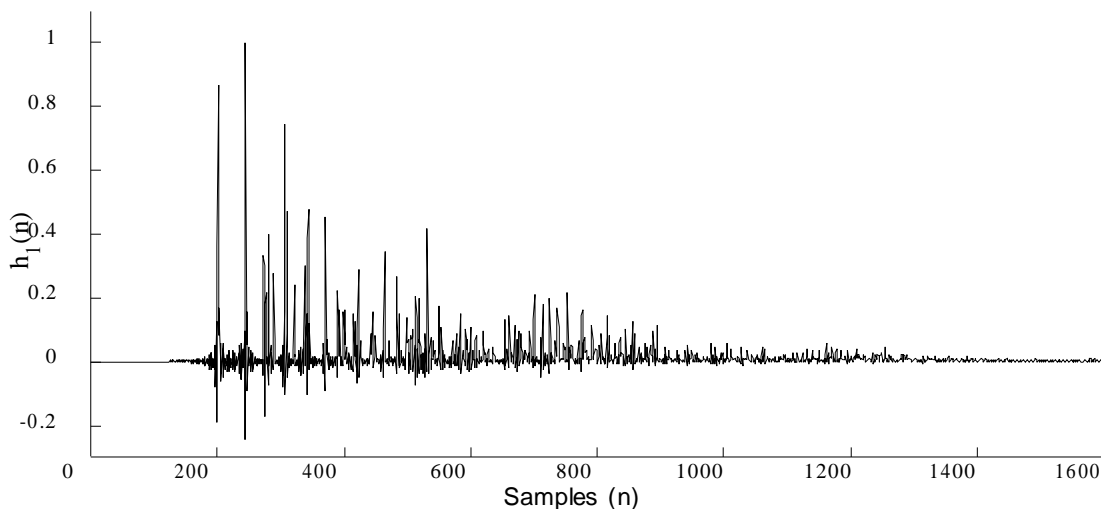

Figure 2 Simulated impulse response for Mic-1.

In this method, narrow pulses are sounded through a loudspeaker. The microphone in the array records the actual response between the source and itself to that narrow pulse. Under the assumption that the narrow pulse that was sounded was a good approximation of an impulse, the recorded response can be assumed to the impulse response of the system. The narrow pulse that was sounded was generated on a computer and was of a width of one sample. The signal was played at 4 *kHz* sampling frequency.

We can now look at the effect of reverberation on the DOA estimation techniques that were discussed in. We will again use the same room that was used in Section with dimensions $5m \times 3m \times 3m$. The source is again placed at $[0.5\,0.05 \quad 1.5]^T$ with respect to one of the corners of the room. We will be using a 4-element ULA with a spacing of 10 *cm* to perform the DOA estimation. The microphones are located at the following locations:

$$Mic \square 1: [4.5 \quad 1.35 \quad 1.5]^T$$

$$Mic \square 2: [4.5 \quad 1.35 \quad 1.5]^T$$

$$Mic \square 3: [4.5 \quad 1.35 \quad 1.5]^T$$

$$Mic \square 4: [4.5 \quad 1.35 \quad 1.5]^T$$

This setup sets the true DOA to $\square 19.93^0$.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

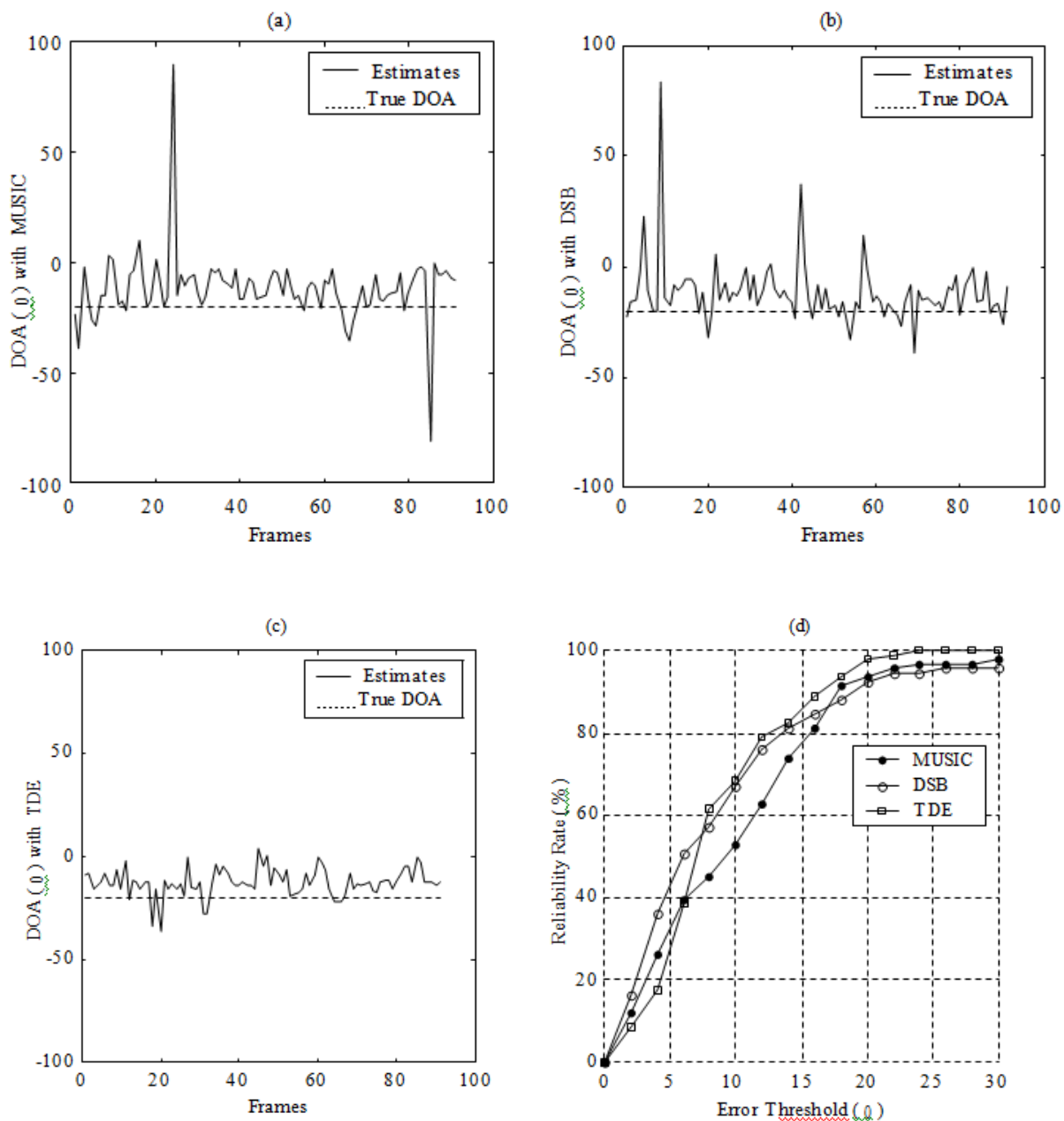*Website: www.ijircce.com*

Vol. 4, Issue 12, December 2016

A speech signal sampled at 16 *kHz* was used as a source and the signals at each microphone were simulated assuming a reverberation of 100 *ms* for the room. The SNR of the simulated signal was set to 30 *dB*. The signals from the microphones were divided into frames of 512 samples each (32 *ms*) and DOA estimates were performed using all three methods for all the frames. For the case of MUSIC, multiple metrics were obtained for several different frequencies. These metrics were added and the angle at which the sum maximized was used as an estimate of the DOA. The same procedure was followed for the DSB based method where the PSD obtained for several different frequencies were added and the angle at which this cumulative PSD maximized was used as an estimate of the DOA.

Figure 3 shows the results obtained from this simulation. For comparison purposes Figure 3.16 shows the results of the simulation for the same setup except that this time the room was assumed to have no reverberation. Only frames with energy greater than or equal to $1 \square 10^{\square 6}$ were used for the simulations. This is the reason why we get 91 frames for the case with reverberation and 58 frames for the case without reverberation. Table1 lists the standard deviations and means of the estimated DOA for all three methods for both scenarios. The presence of multiple reflections increases the power of the reverberated signals. Clearly the presence of reverberation degrades the performance of all three methods.

The table shows two things. First there is far more variation around the mean in the presence of reverberation. Second, there is a bias towards 0 in the estimates in the presence of reverberation.

Table 1 *Standard deviations and means of DOA estimates over all frames.*

Method	Standard Deviation		Mean	
	$T_r = 100$ *ms*	$T_r = 0$ *ms*	$T_r = 100$ *ms*	$T_r = 0$ *ms*
MUSIC	15.51	5.03	-11.88	-18.41
DSB	14.67	5.18	-11.79	-19.23
TDE	6.75	0.59	-12.69	-20.17

Table.1   Nature and Effects of Room Reverberation

Figure 3 *Framewise DOA estimates using*

(a)MUSIC, (b)DSB and

(c) DE for 100 ms reverberation time and

(d) reliability-rates.

REFERENCES

[1] Y. Huang, J. Benesty, and G. W. Elko, "Microphone Arrays for Video Camera Steering," Acoustic Signal Processing for Telecommunications, ed. S. L. Gay and J. Benesty, Kluwer Academic Publishers, 2000.

[2] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *Proc. on Automatic Face and Gesture Recognition*, 1996, pp. 51-56.

[3] M. S. Brandstein and S. M. Griebel, "Nonlinear, model-based microphone array speech enhancement," Acoustic Signal Processing for Telecommunications, ed. S. L. Gay and J. Benesty, Kluwer Academic Publishers, 2000.

[4] J. H. DiBiase, H. F. Silverman, and M. Brandstein, "Robust Localization in Reverberant

Rooms," Microphone Arrays, Springer-Verlag, 2001.

[5] B. V. Veen and K. M. Buckley, "Beamforming Techniques for Spatial Filtering," CRC Digital Signal Processing Handbook, 1999.

[6] H. Kamiyanagida, H. Saruwatari, K. Takeda, and F. Itakura, "Direction of arrival estimation based on non-linear microphone array," *IEEE Conf. On Acoustics, Speech and Signa1 Processing*, Vol. 5, pp. 3033-3036, 2001.

[7] C. H. Knapp and G. C. Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Trans. Acoustics, Speech and Signa1 Proc.*, vol. ASSP-24, No. 4, August 1976.

[8] K. Varma, T. Ikuma, and A. A (Louis) Beex, "Robust TDE-based DOA estimation for compact audio arrays," *IEEE Sensor Array and Mu1tichanne1 Signa1 Proc. Workshop (SAM)*, August, 2002.

[9] J. P. Ianiello, "Time delay estimation via cross-correlation in the presence of large estimation errors," *IEEE Trans. Acoust. Speech, Signa1 Processing,* vol. 30, no. 6, pp.

998-1003, December 1982.

[10] L. B. Jackson, Digital Filters and Signal Processing, pp. 462-464, Kluwer Academic

Publishers, 1996.

[11] DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus, NIST Speech Disc 1-1.1, Oct. 1990.