GRD Journal for Engineering | Volume 2 | Issue 10 | September 2017
ISSN: 2455-5703

# Analysis of Beamforming Methods for Speech Enhancement

**C Anisha**
*M. Tech*
*Department of Electronic and Communication Engineering*
*B V Raju Institute of Technology, Telangana*

**A S N Murthy**
*Professor*
*Department of Electronic and Communication Engineering*
*B V Raju Institute of Technology, Telangana*

## Abstract

Nowadays, hands-free technology has been rapidly increasing in communication systems. It allows a natural form of communication, as if the communication partner was right next to you, the major problem in these systems is the addition of background noise. In this paper implemented beamforming methods for noise reduction in hands-free communication. A beamformer is an array of microphones, which can do spatial filtering. Speech data has been collected using linear microphone array and processed using delay-and-sum beamformer (DSB) and generalized side-lobe canceller (GSC) algorithms. The performance of these techniques analyzed using objective and subjective methods. It has been observed that the beamforming techniques are useful for hands-free communication.

**Keywords- Speech Enhancement, Beamforming, Delay and Sum Beamformer, Generlized Sidelobe Canceller, Time Difference of Arrial**

## I. INTRODUCTION

Microphone array technologies [1], [2] provide new opportunities in hands-free voice communication. A microphone array consists of a set of microphones positioned in a way that the spatial information is well capture.

A Beamformer [3] is a signal processing technique used together with a microphone array to provide the capability of spatial filtering. The microphone array produces spatial samples of the propagating wave, which are then manipulated by the signal processing methods [4] to produce the output signal. The noise reduction systems play an important role in communications.

In this paper the beamforming techniques have been implemented using Matlab programming and the simulations are carried out at various noise environments. The results have been analyzed using subjective and objective methods.

## II. MICROPHONE ARRAY DESIGN

Uniform Linear microphone array is one of the most commonly used microphone array setup. It is shown in the Figure 2.1; the microphones are placed in a straight line with equal distance between them.

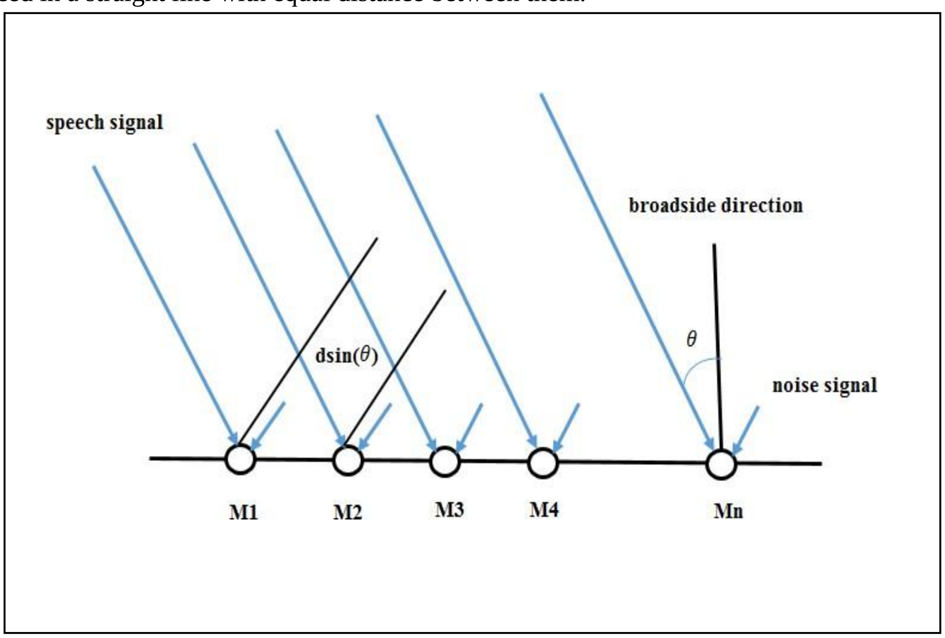


Fig. 2.1: Linear-Microphone Array Setup

To avoid aliasing effect, the distance between the pair of microphones, d should be less than half of the smallest wavelength.

$$d \leq \frac{\lambda_{min}}{2} \qquad (2.1)$$

$\lambda_{min}$ , is the minimum wavelength corresponding to the maximum frequency $f_{max}$

$$\lambda_{min} = \frac{c}{f_{max}} \qquad (2.2)$$

c is the velocity of sound
Array processing consists of two main procedures: source localization and the other is beamforming.
In this paper we have used TDOA estimation for source localization process [5]-[7]. Time-Delay Estimation aims to measure the relative time difference of arrival (TDOA) among microphones, which are spatially arranged.
TDOA between the two microphones (i,j) is shown in the Figure 2.2. The time $t_i$, is the amount of time required for a signal *s(t)* to propagate from the source to the $i^{th}$ microphone in an array.

$$t_i = \frac{\|m_i - s\|}{c} \qquad (2.3)$$

The TDOA for a given pair of microphones and the source is defined as the time difference between the signals received by the two microphones. It is computed using the spatial positions of the source and microphone.
From the Figure $t_i$ is the time taken by the signal to reach microphone $i$ and $t_j$ is the time taken by the signal to reach microphone $j$.

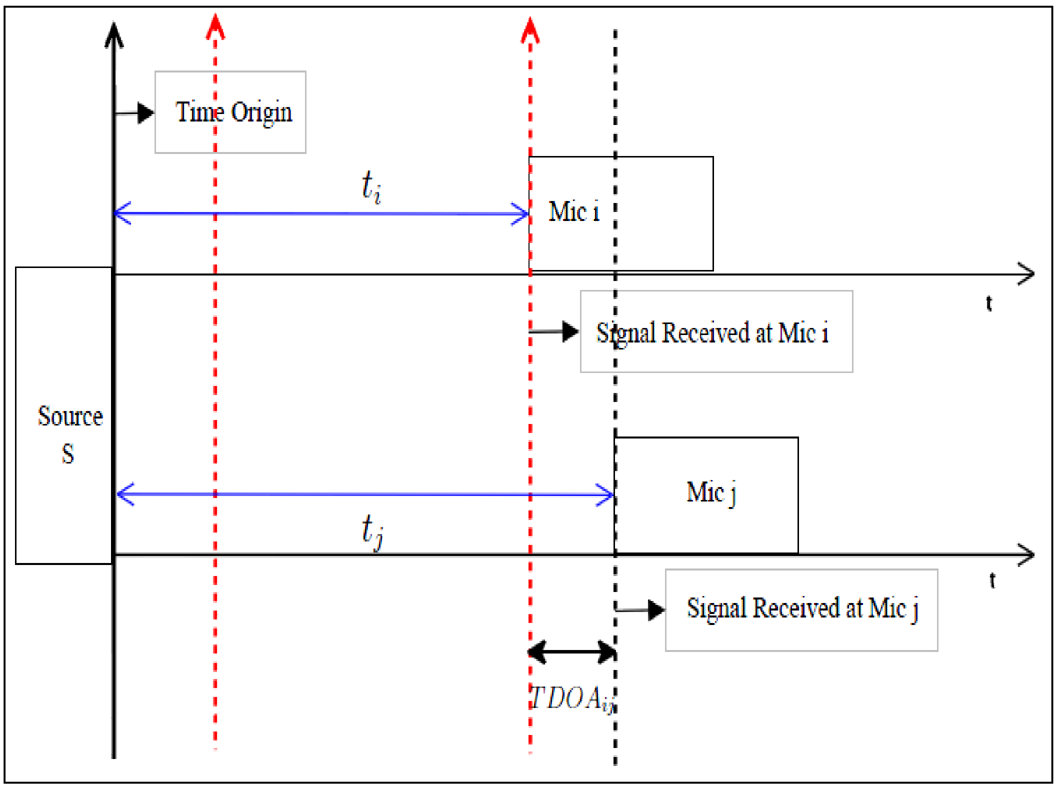


Fig. 2.2: Schematic Representing the TDOA

The time difference of arrival $TDOA_{ij}$ between $i^{th}$ and $j^{th}$ microphones is given by

$$TDOA_{ij} = \frac{(\|m_i - s\| - \|m_j - s\|)}{c} \qquad (2.4)$$

## III. BEAMFORING ALOGORITHMS

The beamforming is the process to enhance the signal [8] coming from the desired direction and rejecting from the other directions. In this paper two algorithms are implemented for speech signal processing application.

### *A. DSB*

The basic idea behind this algorithm is to delay the output of each microphone by a proper amount of time so that the signal components from the desired source are synchronized across all the sensors. The delayed signals are then added together. As they add up together coherently, the desired signal components are reinforced and noise was eliminated as they are added together destructively. The structure of the delay and sum beamformer is shown in the Figure 3.1.

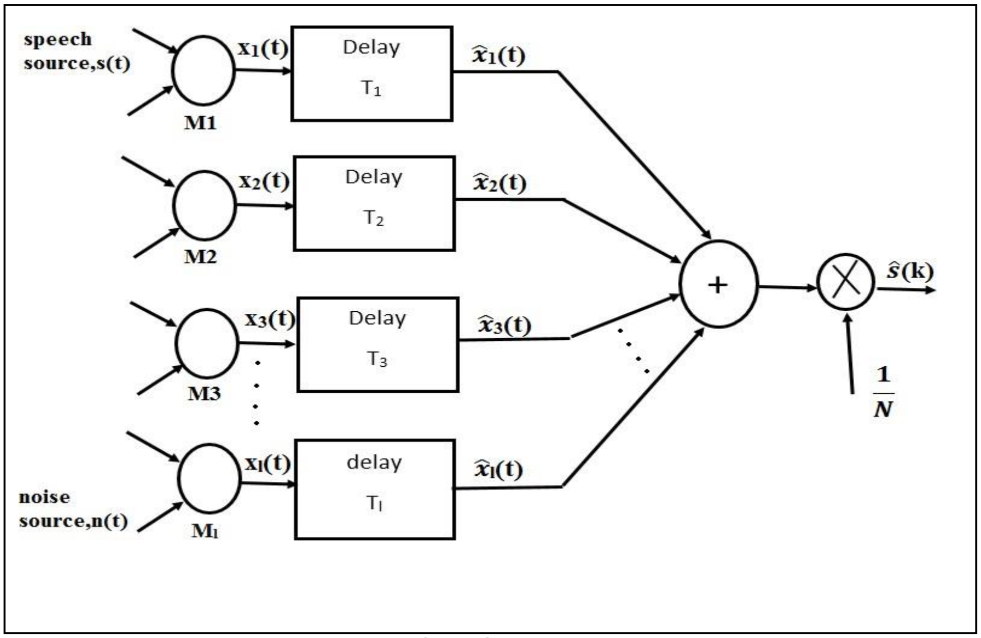


Fig. 3.1: DSB structure

The signal arriving at the microphone is the combination of both the desired signal and the noise signal, which is expressed as

$$x_l(t)=s_l(t)+n_l(t) \quad (3.1)$$

The delayed signal can be expressed as

$$\hat{x}_l(t) = s_l(t - T_l) + n_l(t - T_l) \quad (3.2)$$

Where, $T_l$ is the delay to the $l^{th}$ signal.

The aligned signals are added together to form the beamed output $\hat{s}(kT)$. The output signal is expressed as

$$\hat{s}(kT) = \frac{1}{N}\sum_{l=1}^{N} s_l(kT) + \frac{1}{N}\sum_{l=1}^{N} \hat{n}_l(kT) \quad (3.3)$$

***B. GSC***

Adaptive beamformer [9] alters the direction pattern in according to the changes in the acoustic environment, thus provides a better performance than fixed beamforming.

The structure of GSC with N microphones as shown in the Figure 3.2, which consists of fixed beamformer (DSB), blocking matrix (BM) and adaptive noise canceller (ANC).

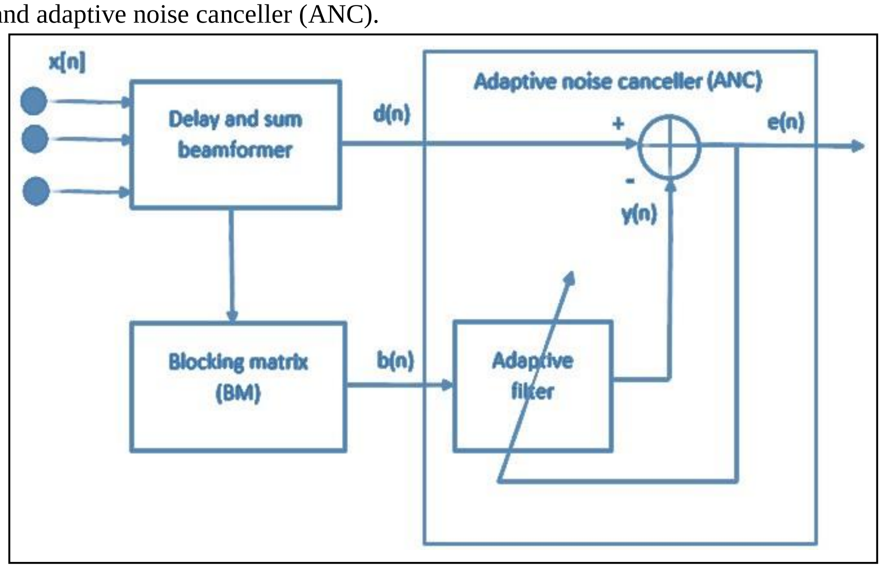


Fig. 3.2: GSC structure

The fixed beamformer used in this algorithm is DSB. The DSB reinforce the signal in steered direction and reduces the noise. The output is

$$d(n)=\frac{1}{M}\sum_{k=1}^{M} x_k[n] \quad (3.4)$$

Ideally the BM produces noise reference signals by blocking the desired signal. This is possible by subtracting the pairs of time-aligned signals. The BM consists of pair wise differences of tracks which is shown below

$$W_s = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \qquad (3.5)$$

The BM output is the product of time aligned signals and elements of blocking matrix.

$$b(n) = W_s * x(n) \qquad (3.6)$$

The Adaptive noise canceller is provided with the primary and reference signals. The enhanced signal is obtained by subtracting the approximate noise $y_b(n)$ from the DSB output.

$$y(n) = y_c(n) - y_b(n) \qquad (3.7)$$

The adaptive filter coefficients are updated using least mean square algorithm, which is given by

$$w[n+1]=w(n)+\mu b(n)y(n) \qquad (3.8)$$

Where $\mu$ is step size and its range is from 0 to 1. The main drawback of the generalized side lobe canceller is that the desired signal may leak into the blocking matrix output if the speech signals are not aligned perfectly. This leads to suppression of the speech signal.

## IV. Speech Quality Measures

The quality of the processed signals can be done by either using subjective listening tests or objective quality measures. [10],[11]

### A. Objective Analysis

The quality testing is done using mathematical formulae. It does not require human listeners and is less expensive and less time-consuming. Out of the available measures SNR is the popular method. The SNR can be calculated as follows:

$$SNR=10\log_{10}\frac{\sum_{n=1}^{N} x^2(n)}{\sum_{n=1}^{N}\{x(n)-\hat{x}(n)\}^2}\ [dB] \qquad (4.1)$$

Where, $x(n)$ is the clean speech, $\hat{x}(n)$ the distorted speech, and $N$ *is* the number of samples.
As speech is non-stationary signal, the SNR does not correlate well with speech quality. Hence the SNR was calculated in short frames, and then averaged. This measure is called the segmental SNR, and can be defined as:

$$SNR_{seg}=\frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{n=Lm}^{Lm+L-1} x^2(n)}{\sum_{n=Lm}^{Lm+L-1}\{x(n)-\hat{x}(n)\}^2} \qquad (4.2)$$

Where $L$ is the frame length (number of samples), and $M$ the number of frames in the signal ($N = ML$).

### B. Subjective Analysis

Subjective Analysis is the judgment of human listeners. It is based on the opinion of a panel of listeners. The measure of speech quality on a one-dimensional scale, i.e., a numerical value that rate the quality of speech.
Mean Opinion score (MOS) is one of the most widely used subjective evaluation method.

*Table 4.1: Five-Point scale of MOS*

| *Rating* | *Speech quality* | *Degradation* |
|---|---|---|
| *5* | *Excellent* | *Imperceptible* |
| *4* | *Good* | *Just perceptible* |
| *3* | *Fair* | *Perceptible* |
| *2* | *Poor* | *Annoying* |
| *1* | *Unsatisfactory* | *Very annoying* |

The listeners rate the speech sample into one of the five quality categories. Each category is assigned a numerical value, shown in the Table 4.1. The resulting MOS value is the average value of all listeners for each of the speech under test. Subjective evaluation is an efficient method but time consuming.

## V. Results And Discussion

The DSB and GSC algorithms are implemented by using Matlab programming. A Uniform linear array of 5 sensors with inter-element spacing, d is used to evaluate the performance of the systems. These are tested under various noise sources such as white noise, pink noise, babble noise, train noise, car noise. The recorded male voice utterance "The shells will bear of both jam or crackers. A joy to every child is the swan boat." is the clean speech source having duration of 8ms is used for the experiments. The speech signal and noisy signals are sampled at sampling frequency 16 kHz. The performance has been evaluated by subjective and objective analysis.

### A. White Noise

The clean speech signal has been degraded by white Gaussian noise and processed by DSB and GSC algorithms. The input and processed output waveforms are shown in Figures 5.1, 5.2 and 5.3.

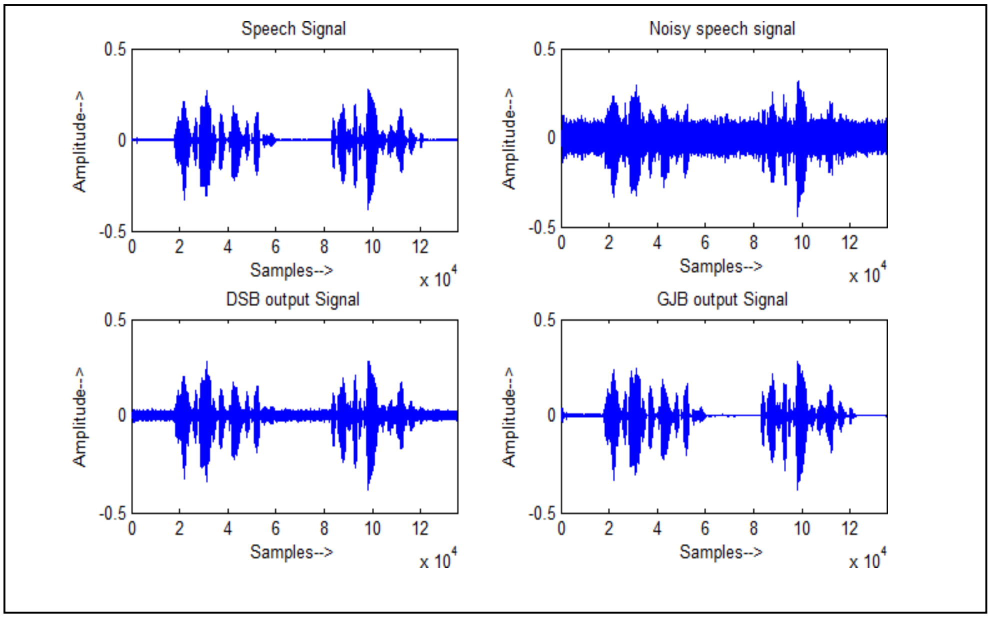


Fig. 5.1: clean Speech signal (top left); Noisy speech (top right); DSB filter (bottom left); GSC filter (bottom right)

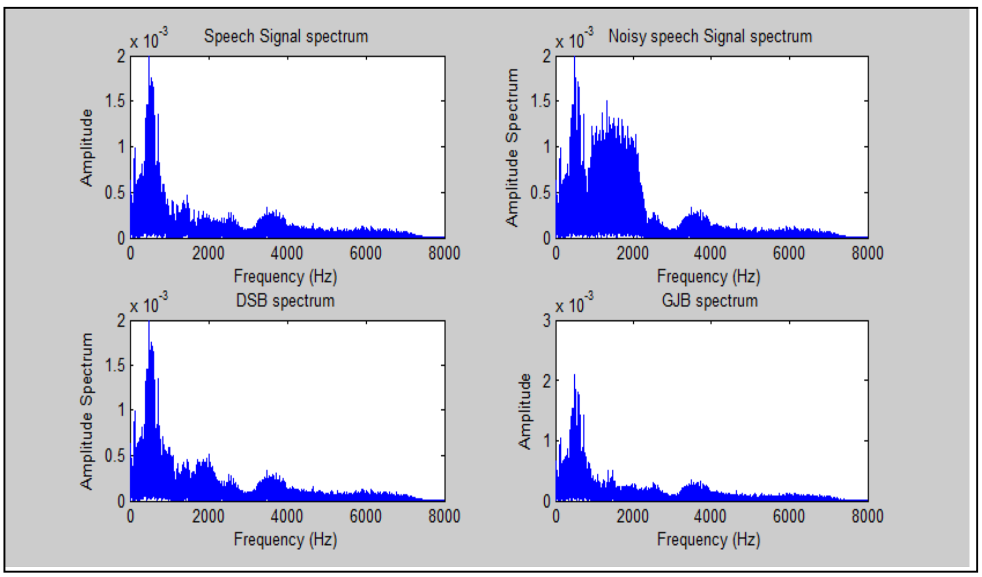


Fig. 5.2: Spectrum s; clean signal (top left); noisy speech (top right); DSB filter (bottom left); GSC filter (bottom right)

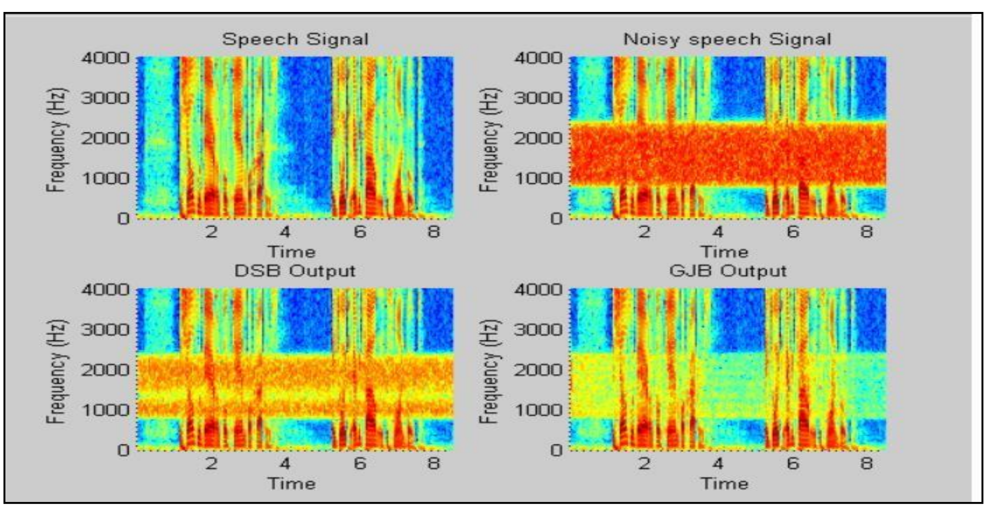


Fig. 5.3: Spectograms; speech signal (top left); noisy speech (top right); DSB filter (bottom left); GSC filter (bottom right).

It was observed that the GSC processed signal reduce more noise than DSB. As white noise is not highly correlated to speech, so we can see a significant improvement in quality of both the systems.

### *B. Babble Noise at 10db SNR*

Next the noisy speech with babble noise is processed. The results are shown in Figures 5.4 and 5.5.

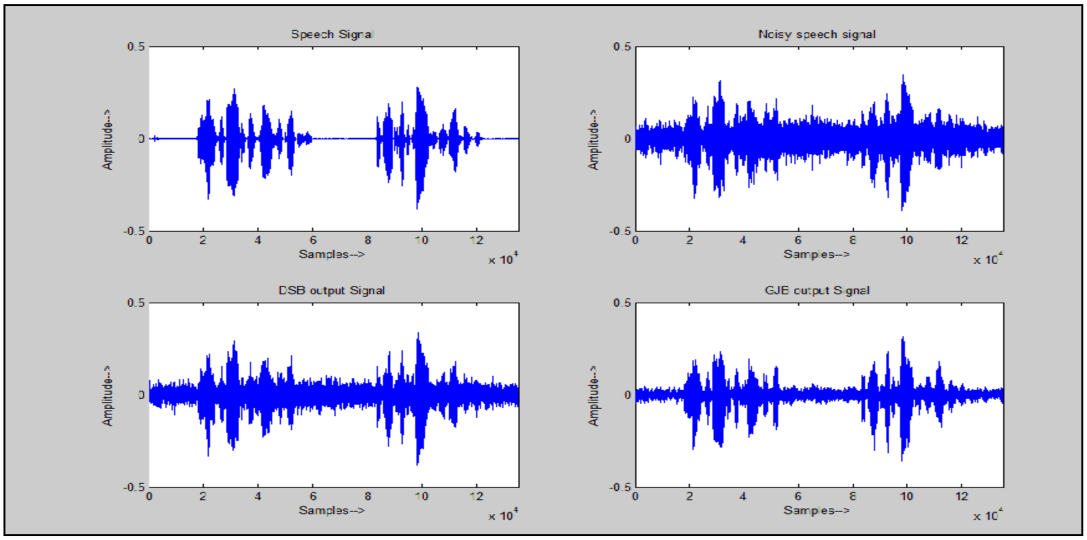


Fig. 5.4: clean speech signal (top left); noisy speech (top right); DSB filter (bottom left); GSC filter (bottom right)

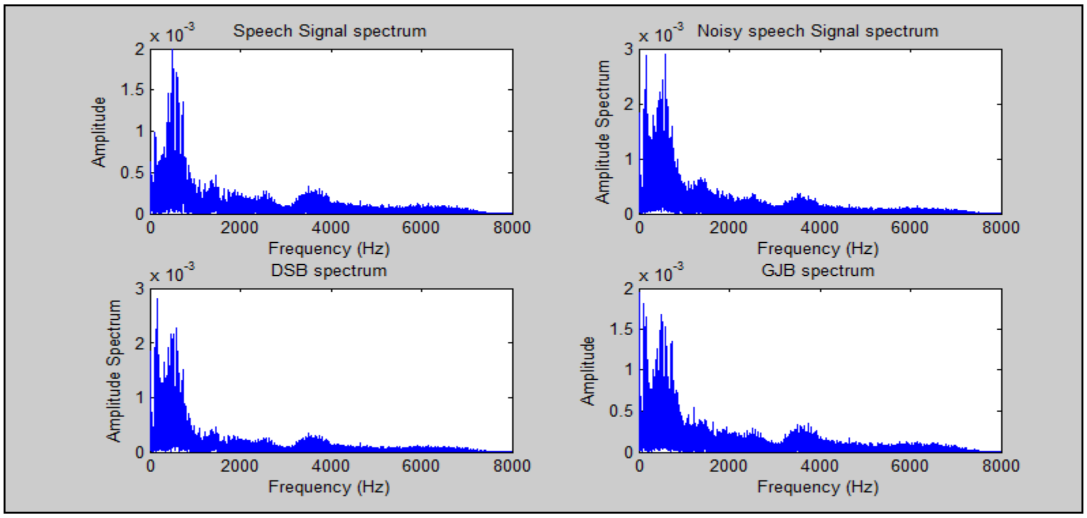


Fig. 5.5: Spectrum of speech signal, noisy speech, DSB filter, GSC filter

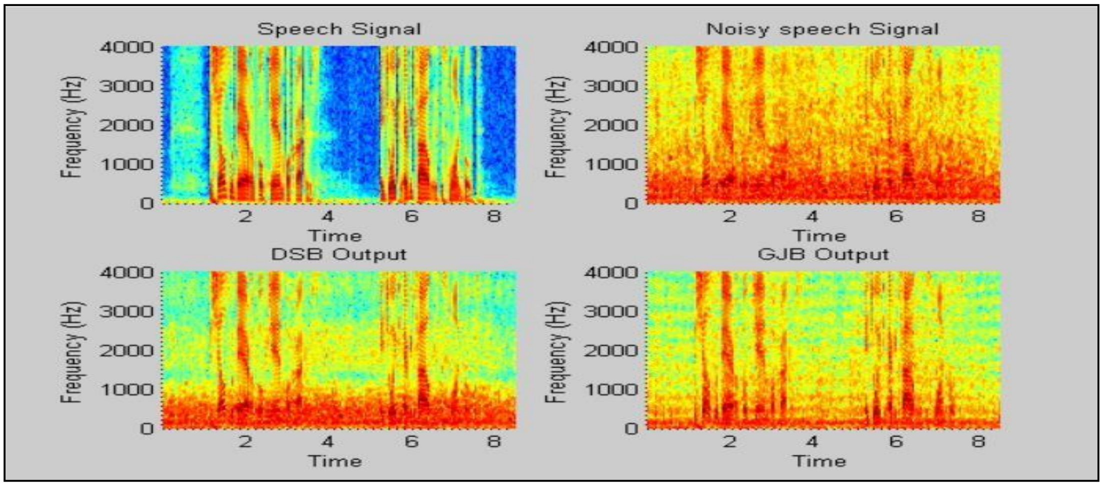


Fig. 5.6: Spectograms; speech signal (top left); noisy speech (top right); DSB filter (bottom left); GSC filter (bottom right)

The babble noise is the noise, which is generated when a group of people talking simultaneously. The SNR factor 0.96 dB and 7.63 dB obtained for DSB and GSC respectively. There is no much improvement in the performance of both the systems because babble noise is highly correlated to speech.

### *C. Pink Noise*

In the third experiment speech signal corrupted by pink noise is processed and the resulted signals are shown in Figures 5.7 and 5.8. The results are not much encouraging in this case because it is much correlated to speech signal. Pink noise has a frequency range from 0 Hz - 8000 Hz.

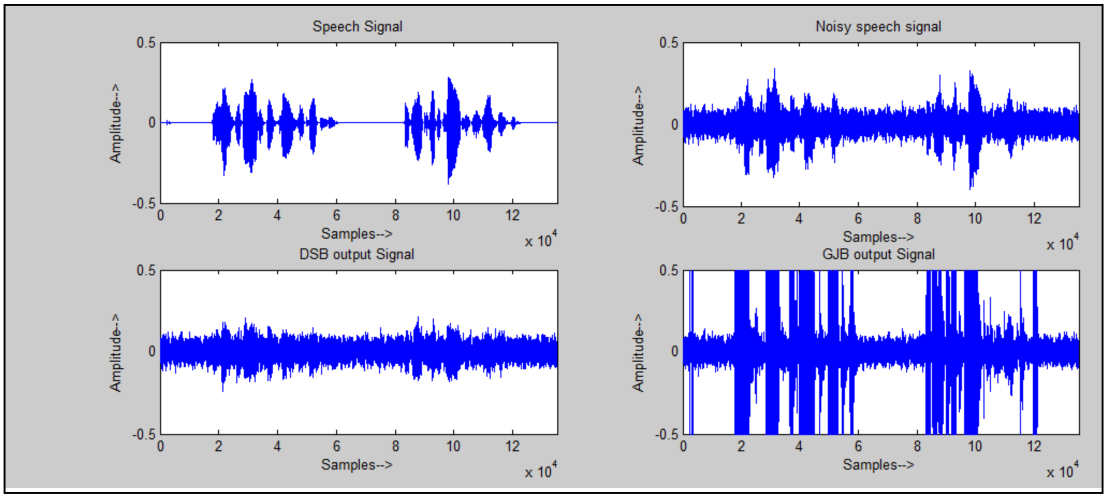


Fig. 5.7: clean speech signal (top left); noisy speech (top right); DSB filter (bottom left); GSC filter (bottom right)

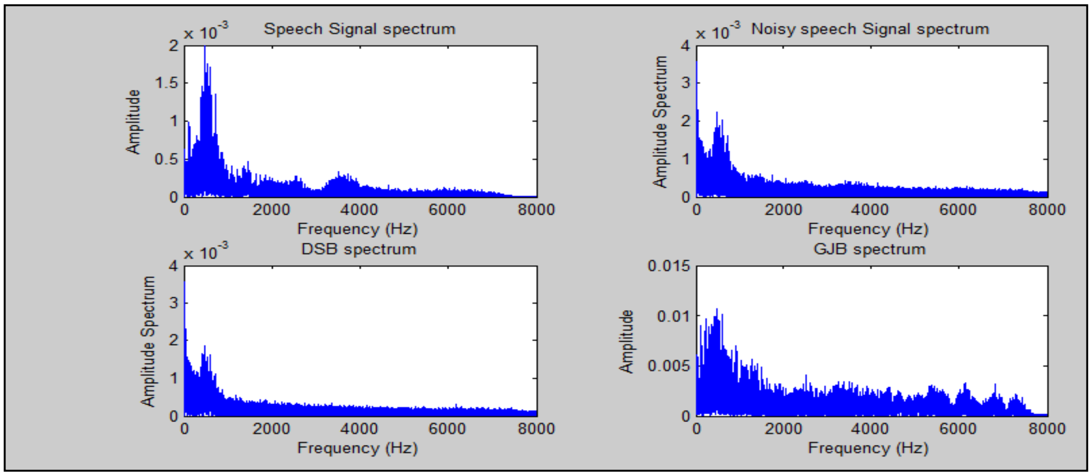


Fig. 5.8: Spectrums; speech signal (top left); noisy speech (top right); DSB filter (bottom left); GSC filter (bottom right)

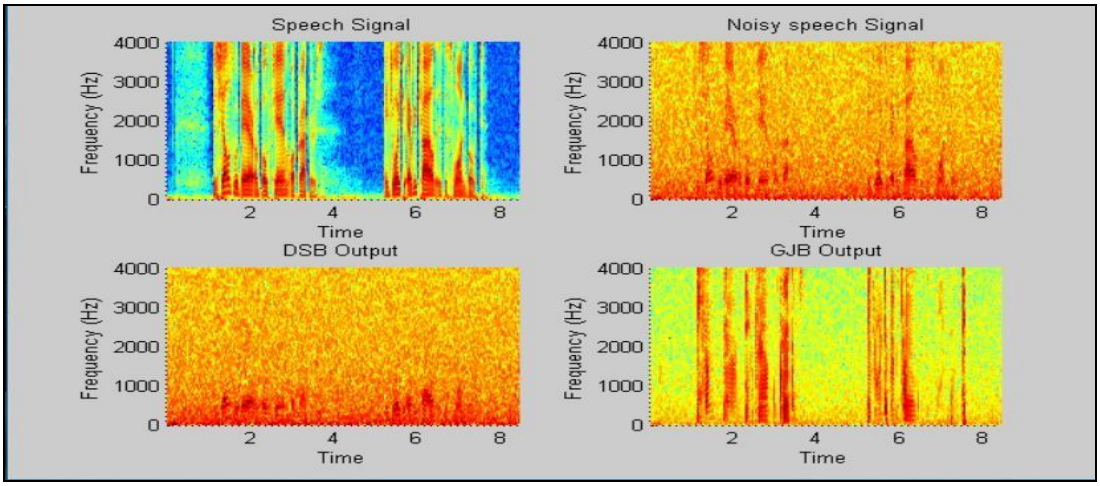


Fig. 5.9: Spectograms; speech signal (top left); noisy speech (top right); DSB filter (bottom left); GSC filter (bottom right)

### D. *Effect on Varying Aperture*

In this experiment the performance is observed while the number of microphones has been increased. A white noise is considered which is at an angle of 45 degrees to the array. The performance data has been plotted in the Figure 5.10

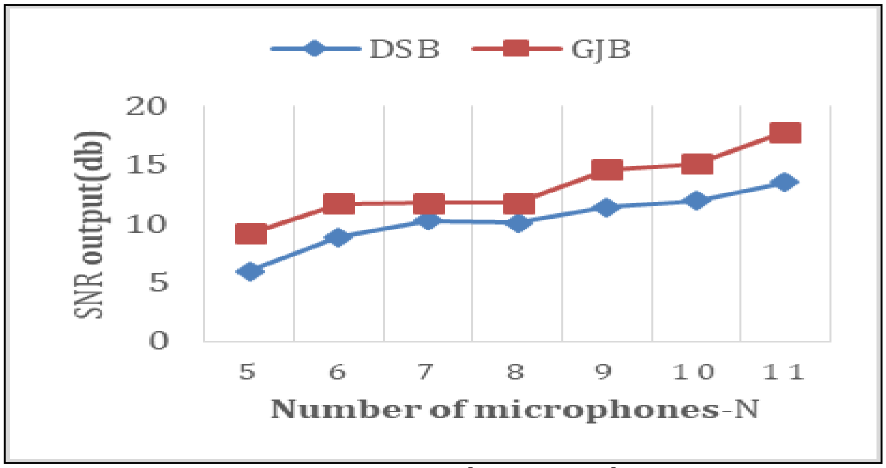


Fig. 5.10: Comparison plots: DSB and GSC

From the figure it is clear there is an increase in SNR of both the systems as we increase the number of microphones.

### E. *Effect on Varying the Source Angle*

In this experiment the systems performance has been observed by changing the direction of the speech signal and kept the noise signal at a constant angle. The experiments are conducted at angles 0° and 40° and the noise is kept at -45°. The data has been plotted and is shown in Figure 5.11

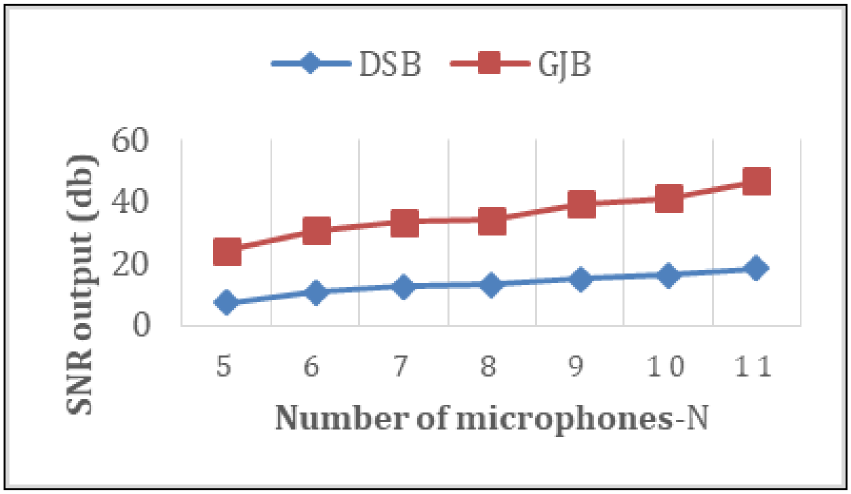


Fig. 5.11: The source at an angle of 0°

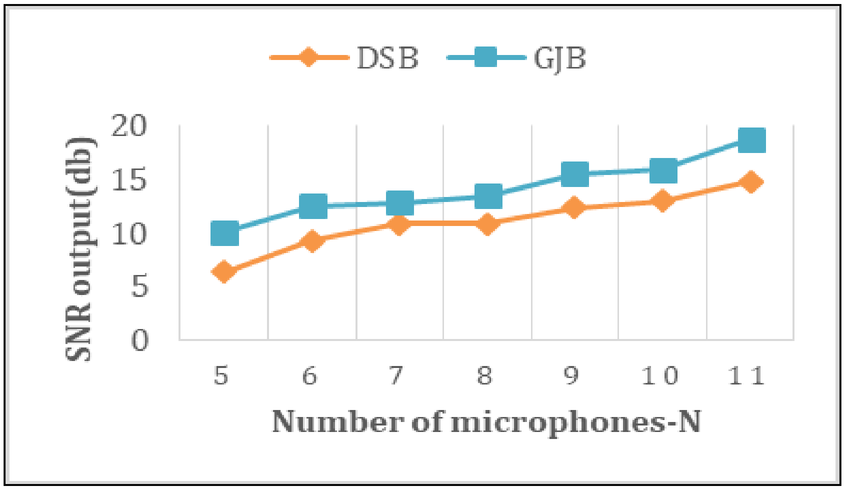


Fig. 5.12: The source at an angle of 40°

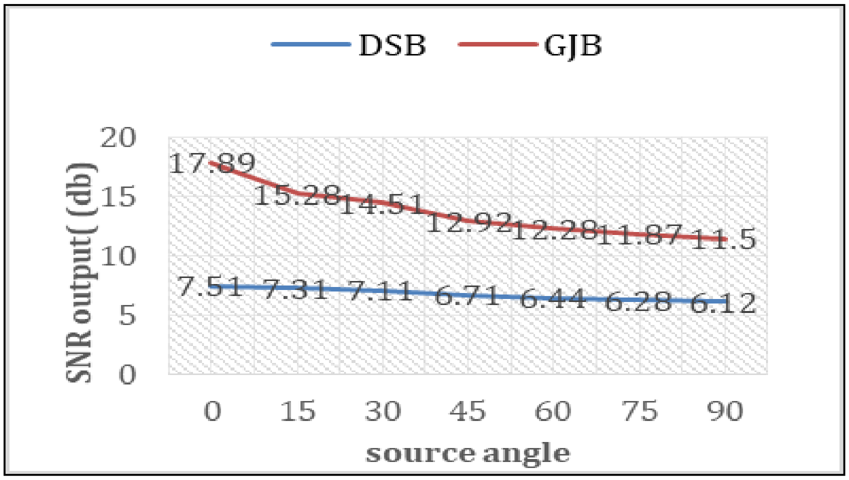

Fig. 5.13: Plot between SNR and Angles

The GSC system response is better than the DSB system and the graph is shown in Figure 5.13.

## VI. Conclusion

In this paper DSB and DSC systems performance has been analyzed for speech enhancement application. Our results have demonstrated that these methods with a microphone array are a reliable approach. The GSC provides better SNR improvement when comparing to DSB. We also found that as increasing the number of microphones more noise reduction in both the systems. It is shown that the use of multiple microphones and incorporating beamforming systems can provide speaker tracking capability and improvement in SNR.

## References

[1] Jacob Benesty, Jingdong Chen, Yiteng Huang, “Microphone array signal processing”
[2] Stephen Oh, Vishu Viswanathan, Panos Papamichalis, “Hands free voice communication in an automobile with a microphone array” IEEE Trans., 281-284 vol 1, March 1992
[3] Hidri Adel, Meddeb Souad, Abdulqadir Alaqeeli, Amiri Hamid,”Beamforming Techniques for Multichannel audio Signal Separation “
[4] Ahmed Abdalla, Peiyu He, “A Study of a various Acoustic Beamforming Techniques Using a Microphone Array”
[5] Ashok Kumar Tellakula, IISC Bangalore, “Acoustic Source Localization Using Time Delay Estimation”
[6] Muhammad Asim, Akbar Ali, “Source Localization and Speech Enhancement for Speech Recognition for Real time Environment “
[7] Smitha Paulose, Elizabeth Sebastian, Dr. Babu Paul, “Acoustic Source Localization “, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 2, February 2013
[8] Sharon Gannot, David Burshtein, Ehud Weinstein, “Signal Enhancement Using Beamforming and Non-stationary with Applications to Speech “ IEEE Trans., 1614-1626, August 2001
[9] Osamu Hoshuyama, Akihiko Sugiyama, Akihiro Hirano, “A Robust Adaptive Beamformer for Microphone Arrays with a Blocking Matrix Using Constrained Adaptive Filters “ IEEE Trans., 2677-2684, October 1999
[10] Sukhvinder kaur, Anil garg, “An effective evaluation study of objective measures using spectral subtractive enhanced signal “International Journal of Research in Engineering and Technology
[11] KazuhiroKondo,”Subjective Quality Measurement of Speech “