MARQUETTE UNIVERSITY

Speech Signal Enhancement
Using A Microphone Array

A THESIS

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree of

MASTER OF SCIENCE

Field of Electrical and Computer Engineering

by

Heather Elaine Ewalt, B.S.

Speech and Signal Processing Lab

Milwaukee, Wisconsin

December 2002

Preface

This thesis describes the design and implementation of a speech enhancement system that uses microphone array beamforming and speech enhancement algorithms applied to a speech signal in a multiple source environment. The goal of the system is to improve the quality of the primary speech signal.

Beamformers work by means of steering an array of microphones towards a desired look direction through utilizing signal information rather than physically moving the array. They accomplish this through minimizing the energy of interference sources and noise in non-look directions while increasing the energy of the signal in the look direction. In this research, two beamforming methods are examined: the delay and sum (DS) beamformer and the minimum variance distortionless response (MVDR) beamformer. The input signals are first split into frequency bands so that narrowband beamforming techniques can be used.

Multiple source Wiener filtering and multiple source spectral subtraction enhancement algorithms are incorporated into the two methods of beamforming. The algorithms utilize signal estimates of each source obtained from the initial beamforming algorithms as inputs. These multiple source enhancement algorithms result in iterative techniques to improve those estimates while improving the signal to noise ratio of the primary source.

The experimental setup presented here consists of both two and three speech sources using a linear microphone input system. The algorithms are performed on both simulated experimental setups and on data obtained from a data acquisition system in an acoustically treated sound room.

To measure the improvement in quality of the enhanced signal, overall SNR and segmental SNR improvement is determined for the original, beamformed, and enhanced signal. In addition to these quality improvement metrics, listener opinion testing is performed.

Acknowledgments

I would like to thank my husband, Jerry Ewalt, and my advisor, Dr. Mike Johnson, for their support of me performing this research. I would also like to especially thank the GAANN fellowship program and Frank Rogers Bacon fellowship program for sponsoring my studies and funding this research. The past few years have been challenging and rewarding, and I have always considered it a blessing to be given the opportunity and the ability to carry out this research. It is my hope that I can be an example for other young women who are contemplating engineering and research possibilities.

This thesis is dedicated to:

My son, Andrew, who was my constant companion for nine months of this research.

My mother, Pamela Lee Kusnierz, who made me who I am and instilled in me my ability to learn and love.

Table of Contents

List of Figures

List of Tables

List of Symbols and Acronyms

DFT	Discrete Fourier Transform
DS	Delay and Sum (Beamformer)
MVDR	Minimum Variance Distortionless Response (Beamformer)
PDS	Power Density Spectrum
sSNR	Segmental Signal to Noise Ratio
SNR	Signal to Noise Ratio
TDOA	Time Delay Of Arrival
TIMIT	Texas Instruments & Massachusetts Institute of Technology speech corpus

d	Distance between microphone pairs (m)
d_i,g_i,h_i	Delay vector
f	Frequency of interest (Hz)
n	Noise signal
r_{nf}	Nearfield radius
v	Velocity of sound (m/s)
w	Filter weights
y	Microphone signal
z	Beamformer signal estimate

L	Overall length of a microphone array; the aperture size (m)
M	Number of microphones in an array
$\mathbf{R}$	Autocorrelation matrix

ϕ	Angle of arrival of signal
λ	Wavelength of interest (m)
φ_s	Phase information of signal s

Chapter 1 Introduction

The ability to separate or enhance a primary speech signal in an environment with many speakers, the so called "cocktail party" effect, is an important issue, especially in recent years with the number of people with hearing damage dramatically on the rise and with the expansion of global businesses requiring the use of more sophisticated video and teleconferencing equipment. Healthy human hearing is capable of identifying a single conversation among the noise of other conversations due to the binaural characteristic of human hearing in which the brain's cognitive processing abilities utilize time differentials between signal inputs from each ear. However, people who have hearing damage often compromise their binaural abilities (Plomp, 1986).

Most common hearing aids work through amplifying all sounds and do not attempt to isolate the primary signal of interest, and recently, a hearing aid designed with a microphone array has shown tremendous results in increasing the ability of hearing impaired persons to understand speech in noisy environments using a fixed beamformer (Widrow, 2001). Similarly, teleconferencing and hands free telephony equipment have traditionally amplified all sounds in a room. Thus, in a room with multiple speakers, these systems output a fusion of sounds where the primary speech source signal is difficult to recognize and understand, and binaural information is lost.

Beamforming algorithms have shown great promise in noise reduction, through utilizing the spatial information of the noise and primary source signals. As the number of microphones in an array increases, increasing the aperture size, the ability of beamforming algorithms to extract the primary source using spatial information improves (Brandstein, 2001; Dundgeon, 1993).

The research presented here focuses on microphone arrays with a small number of microphones, up to eight, and a small aperture size, up to 0.4 meters, as would be required for hearing aid applications where users could comfortably wear the array (Widrow, 2001). Smaller arrays are also more portable and affordable for applications with teleconferencing and hands free telephony. These smaller arrays have less ability to extract the primary signal using beamforming algorithms and thus are amenable to improvement through the use of further speech enhancement algorithms (Brandstein, 2001).

While methods exist for a variety of beamforming techniques (Brandstein, 2001; Dundgeon, 1993) as well as for multi-source filtering in stationary noise (Saruwatari, 2000), theory has yet to be developed for integrating spatial filtering with additional enhancement methods to deal with non-stationary interference, such as in multiple speaker interference environments. This research addressed this need by creating methodologies to enhance speech signals with simultaneous, nonstationary noise sources. The primary contribution of this research work is to extend traditional speech enhancement algorithms such as spectral subtraction and Wiener filtering into the multiple-source domain. By incorporating multiple parallel beamformers with algorithms that iteratively improve the spectral magnitude estimates of each source, substantial improvement in overall signal separation can be obtained.

For the non-stationary noise sources present in the multiple speaker scenario being investigated here, the method must be implemented on a frame-by-frame basis over the primary speech signal, allowing the noise source spectra to be continuously re-estimated. Specifically, the problem of enhancing a primary speech signal with one and two interfering speech sources and known source locations is addressed here. In addition to nulling the directions of the interfering sources to extract a primary signal using one beamformer, as in (Widrow, 2001; Brandstein, 2001; Omologo, 1997), this research develops a new method of utilizing multiple beamformers, with coupled post-processing enhancement algorithms, to extract each speech source signal. The fixed beamformers used initially have narrowband, far-field assumptions. The spacing of the microphones as related to the distance to the sources is chosen appropriately for the far-field assumption as given in (Ryan, 1997) and discussed in Section 2.1.4. Despite the fact that speech signals are broadband signals, narrowband assumptions can be approximated with the use of filter banks applied to each microphone input (McCowan, 2001).

1.1 Thesis Statement

This research addresses the problem of primary source enhancement in a multiple source environment. It is important to note that enhancement techniques addressing speech signals contaminated with nonstationary speech as noise are not yet fully developed. To improve the quality and recognition of the speech signal of interest, a microphone array along with beamforming and speech enhancement algorithms can be used to separate the primary speech signal from the interfering speech signals. The novel approach of using multiple beamformers to estimate each source signal and using those estimates in traditional speech enhancement algorithms adapted to a multiple source problem is implemented in this research. Thus, it is the goal of this research to enhance the quality of the primary speech signal of interest through the development and implementation of multiple source beamforming and enhancement algorithms.

1.2 Thesis Overview

In Chapter 2, microphone array background information along with issues in the array geometry setup are discussed. The algorithms used with microphone arrays to create and implement beamformers are explained in detail, particularly the delay and sum beamformer and the minimum variance delayed response beamformer. In addition, traditional speech enhancement techniques are described. Finally, the quality measures used to determine the amount of improvement performed by the algorithms developed in this research are examined.

Chapter 3 presents the newly developed iterative multiple source enhancement algorithms that are at the heart of this research. The issues with multiple speech sources as noise are discussed in association with the methods of implementing traditional enhancement methods into a multiple source environment. To test the developed algorithms, the creation of simulated multiple speaker environments is detailed in Chapter 4 along with information on the geometries of the sound room, microphones, and speakers.

Chapter 5 presents the experimental hardware setup with the data acquisition system and details of the sound booth environment. The hardware and software used are described. Chapter 6 outlines the experimental results of both the simulated experiments and the sound booth experiments. The results are divided into two source and three source experimental setups. Chapter 7 discusses the results of the research, and Chapter 8 gives recommendations to future direction related to this research.

Chapter 2 Background

2.1 Microphone Array Fundamentals

Although research has shown that single channel, as compared to multiple channel, signal separation algorithms have limitations in their ability to improve signal quality, implementation of multiple channel algorithms was difficult to perform in the early multiple input systems research. This was due to the expense of purchasing many microphones and multiple input data acquisition hardware, in addition to the tremendous computing power required by the multiple dimension complexities. However, with the advent of faster and greater computing power along with more affordable multiple input systems, microphone array signal processing is becoming a more feasible option. This is beneficial to the speech processing field for the reason that multiple input systems are able to utilize beamforming algorithms. These algorithms use spatial and temporal differences in the input signals to better improve the signal quality as compared to the improvements shown using single channel systems.

2.1.1 Geometry

Before developing and implementing beamforming algorithms with microphone arrays, the geometry of the microphone array must be addressed. The number, the spacing, and the arrangement of the microphones all need to be determined. An array shape can be designed as linear, square, circle, logarithmic, or many other microphone arrangements. Optimal microphone placement depends upon the specific enhancement and quality assessment algorithms that will be used with the array in addition to the type of speech signal being analyzed (Rabinkin, 1997; Wang, 1996).

In microphone array research, the most common and practical geometries examined are linear and square arrays. These arrangements allow signal processing algorithms to be more easily implemented. Square arrays have the advantage over linear arrays because they can operate in a three dimensional space. Although linear arrays allow for only a two dimensional domain problem, operating in two dimensions requires less computation time and power as compared to three dimensions. In a multiple speaker environment, it is commonly the case that the speaker locations are located in a roughly two dimensional plane since speakers are speaking at similar elevations.

This research utilizes a linear array of microphones, and thus the signal sources are located at the same elevations. For a linear, equally spaced array, the time it takes a speech signal to arrive at a given microphone is given by:

$$\Delta t = \frac{(m-1)\mathrm{d}\sin\phi}{v} \qquad [1]$$

where m is the microphone number (one through M) so that microphone one has a Δt of zero, ϕ is the angle of arrival of the speech signal, d is the distance between microphones, and v is the speed of sound. Figure 1 shows the graphical presentation of this equation.

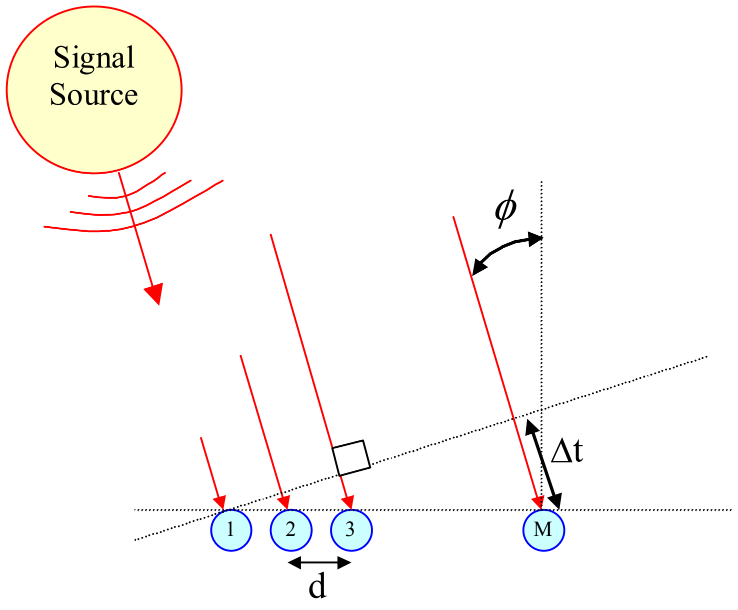


Figure 1: Propagating far-field sound wave with the microphone array

In addition to the shape of the array, the number of microphones needs to be determined, which also determines the aperture, or end to end length, of the array. With linear arrays, the overall length of the microphone array defines the aperture. Increasing the number of microphones, which in turn increases the aperture size, will increase the resolution of the respective spatial filter that can be created by that array. With an increasing resolution, the spatial filter becomes more able to extract a signal from a more precise location. Hence, an infinitely long aperture is able to discriminate or separate signals that are infinitesimally close together (Dundgeon, 1993). Infinitesimally large aperture arrays, however, can obviously not be implemented and it is necessary to decide upon a practical aperture size that will allow for the signal of interest to be effectively filtered from other interfering direction signals.

Microphone spacing is the final, crucial design parameter in microphone array geometry and much research has been performed in this arena. Spatial filtering, which is the basis of beamforming, is utilized when the function of a microphone array is to extract a signal from a specific location. Similar to the Nyquist theory for frequency filtering of signals, spatial filtering must conform to a spatial aliasing criterion related to the highest frequency found in the signal of interest. The spatial equivalent can be given by:

$$\mathrm{d} = \frac{v}{2f} \qquad [2]$$

where d is the maximum spacing between the microphones, *f* is the highest frequency of the signal being detected by the microphones, and *v* is the velocity of sound waves. The velocity of sound waves used for this research is 345 meters per second, which is approximately the velocity at standard atmospheric conditions at sea level and 22 degrees Celsius. To prevent spatial aliasing, the above equation requires the microphone spacing to be small enough to prevent aliasing of the highest frequency being analyzed.

For example, speech signals' highest frequency is approximately 20,000 Hz, resulting in a microphone spacing of 8.625 millimeters. Microphones physical characteristics will not allow for spacing this small. In this research, the highest frequency content to be analyzed is 7000 Hz, which results in a microphone spacing design of about 2.5 centimeters to have no spatial aliasing.

2.1.2 Source Localization

A major field of array speech signal processing is dedicated to source localization and detection. In localization research, the microphone array can be used to determine the location of a speaker, angular direction of a signal, and number of speakers and additionally used to track speaker positions (Svaizer, 1997; Brandstein, 1995; Rabinkin, 1996). The ability to locate a speaker in an environment is crucial to many teleconferencing and videoconferencing applications and can be used as a front end process for the beamforming source separation algorithms discussed in section 2.2.

When utilizing microphone arrays for source location applications, the spatial aliasing criteria need not be followed in most setups. This is because source location algorithms are usually only interested in time differentials between microphone pairs to determine the location of a source. Therefore, they do not use spatial filtering to extract the source signal. As long as the velocity of the signal and the spacing of the microphones are known, the time differentials between microphone pairs yield the information necessary to locate the direction of the signal and the source in space. With greater time differentials, less resolution and less computational load is required to determine the location of the signal source (Svaizer, 1997). To achieve greater time differentials between microphone pairs, the design of source location arrays requires a larger microphone spacing design.

As described above, the basis of the microphone spacing design parameter differs in source location arrays versus source extraction arrays. Source location microphone arrays benefit from larger spacing through increased resolution whereas source extraction arrays benefit from smaller spacing to prevent aliasing at the highest frequency of interest. Thus, the ability to design a combined source location and source extraction array is problematic. The arrays in this research have *a priori* knowledge of the source locations and are designed to perform source extraction algorithms, focusing instead on beamforming enhancement. In the future, source localization and tracking technology will be able to be coupled to the enhancement algorithms being investigated here.

2.1.3 Speech Signal Broadband Issues

The first array signal processing algorithms were developed for sonar equipment using narrowband signals. Today, in a majority of array signal applications, such as sonar and telecommunication applications, the signal of interest is a narrowband frequency signal. Consequently, much narrowband research has been conducted and narrowband sensor array algorithms have been well developed. Much of the array signal processing theory is thus based upon a narrowband frequency assumption. One of the challenges of microphone array signal processing applications is the fact that speech signals are broadband signals spanning the frequency band of human acoustic perception, approximately 20 to 20,000 Hz.

To be able to use narrowband frequency theories, the broadband speech signal can be broken into frequency bands. The smaller each frequency band range is defined, the more accurately a narrowband signal is approximated. However, more analysis is required when using small frequency bands. With more bands used, there is a proportional increase in the size of the overall speech array model and consequently an increase in computational complexity.

The model used in speech array signal processing must balance the number of frequency bins and the span of each bin when using narrowband model assumptions. As the bandwidth of a frequency band increases, aliasing occurs and introduces artifacts in the signal when resynthesized. Therefore, a compromise between the bandwidth of the frequency bins and the ability of the narrowband model to remain valid must be created. (Kellerman, 1988; Weiss, 1998).

To minimize the computational complexity associated with a large amount of frequency bands over the 20 to 20,000 Hz range, researchers can take advantage of the fact that the perception of human hearing is primarily focused on a frequency span of approximately 200 to 8000 Hz, meaning that this frequency range has more importance when listening to and understanding a speech signal. For example, telephone signals have a frequency range of 300 to 3000 Hz. Although this is a much smaller range of frequencies compared to the 20 to 20,000 range, it is still generally acceptable and intelligible. This research will focus upon a frequency range of 300 to 7000 Hz with ten frequency bins.

Although breaking the broadband speech signal into narrowband frequency bins creates more computational load, this setup also easily allows for the use of sub-band enhancement and sub-band recognition algorithms, which have recently shown great interest in speech processing research (McCowan, 2001; Kajita, 1996; Wu, 1998). The basic methodology of this research is shown in Figure 2. The enhancement algorithms are processed on each frequency sub-band signal, thus allowing for more or less emphasis to be placed on particular frequency bands in the signal. The ability to process frequency bands separately better emulates human hearing where lower frequencies are given more perceptual emphasis and also allows systems to focus on frequencies with lower noise, leading to more robust systems.

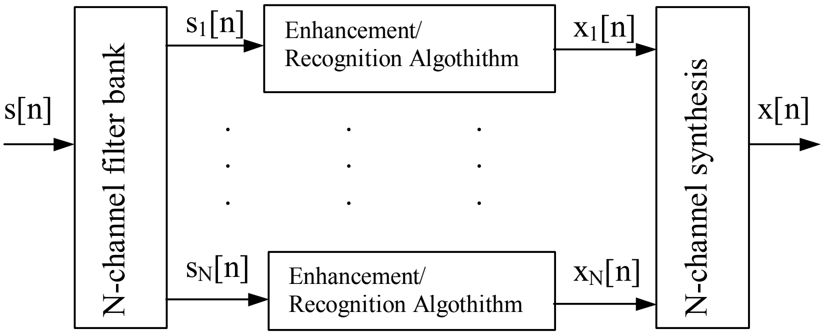

s[n]
N-channel filter bank
s1[n]
Enhancement/ Recognition Algothithm
x1[n]
sN[n]
Enhancement/ Recognition Algothithm
xN[n]
N-channel synthesis
x[n]

Figure 2: Sub-band speech recognition and enhancement

In (McCowan, 2001), microphone array technology is integrated with sub-band speech recognition models through the creation of a speech recognition system for each frequency sub-band. The resulting beamformed sub-band recognition systems outperformed both single channel sub-band systems and full band beamformed recognition systems.

As previously stated, the algorithms developed here filter the microphone signals into frequency bins to be able to use the established narrowband array signal processing methods. This allows the enhancement algorithms developed here to be incorporated in sub-band methods.

2.1.4 Nearfield/Far-field Approximations

Most traditional array research has been performed using far-field approximations of signal waves where the signal of interest's wave is planar upon reaching the array as shown in Figure 3.

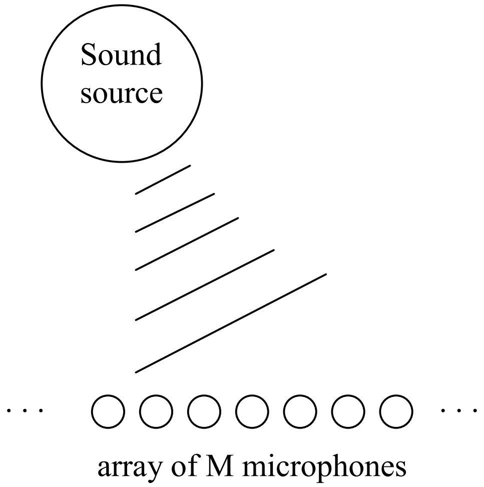


Figure 3: Far-field planar sound wave propagation

Although the planar wave assumption is used extensively in array signal processing, the true wavefront of a speech signal is spherical, as shown in Figure 4. The curvature, however, becomes less pronounced as the wave travels, which leads to a more planar wavefront.

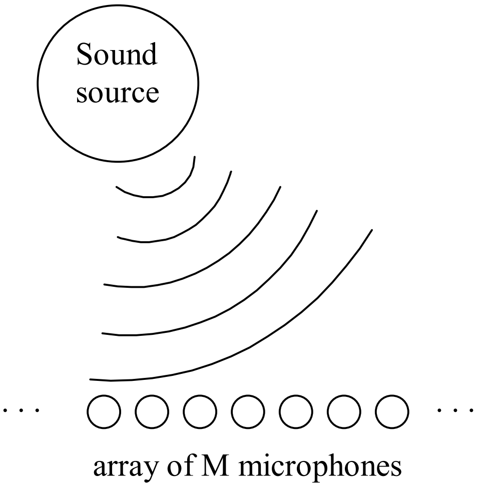


Figure 4: Nearfield spherical sound wave propagation

When far-field assumptions are valid, the complexity of the setup is reduced to one parameter, the angle of wave propagation, whereas when using nearfield wave propagation theory another parameter, the radial distance from the source to the array, is introduced.

The determination of whether the planar assumption is valid is found in the relationship between the spacing of the microphones and the distance of the sound source to the microphone array. The planar assumption becomes justified as the sound source distance to the array increases and as the spacing of the microphones and overall aperture length decreases. To quantify this, the valid nearfield region is given by (Steinberg, 1976):

$$r_{nf} = \frac{L^2}{2\lambda} \qquad [3]$$

where the nearfield radius is r_{nf}, the microphone spacing is $\lambda/2$, and the overall length of the array is L. Far-field planar assumptions are generally accepted to be valid for sound sources outside of this region.

There has been some interest in microphone array signal processing in areas where the array geometry and source location produce a nearfield model. In (Ryan, 1997), nearfield and far-field wavefront differences from a sound source and its respective reverberations were used to optimize an arbitrary microphone array design to decrease the noise from the reverberations. The reverberation noise created from the sound source was modeled as far-field, planar waves. The optimized nearfield algorithms outperformed traditional delay and sum beamformers in the experimental setups. Similarly, in (Tager, 1998) the sound source of interest was considered to be in the nearfield whereas interfering noise sources were placed in the far-field. Tager exploited the differences of these wavefronts to produce a nearfield superdirectivity algorithm that outperformed the traditional delay and sum beamforming algorithm, especially for low frequencies.

The microphone arrays used in this research with respect to the speech source locations allow for the use of far-field models in that the microphone array overall lengths are small as compared to the distances to the sound sources.

2.2 Beamformer Fundamentals

Beamforming algorithms, used in conjunction with an array of sensors, take advantage of the time differentials between incoming signals among the sensors in the array. This is due to the fact that a signal emitted from a source, located at specific position in space, will arrive at a unique time for each sensor in an array according to the relation between the sensors and the source. Using this spatial information, source location and primary signal extraction beamforming are possible. These tasks represent the two main fields of microphone array signal processing research being performed today.

Using beamformers to determine source locations has applications to teleconferencing and videoconferencing, in addition to radar and sonar. In (Rabinkin, 1996), a microphone array for a lecture room was created where a source location beamformer was used to determine the position of the current speaker. This microphone array was implemented using two sets of four microphones in a square geometry with application to source location in an auditorium setting. This research utilized the time delay of arrivals (TDOA) between microphone pairs as inputs to a correlation based beamforming algorithm to determine the source location. With the position of the speaker determined, a video camera was integrated into the array system to automatically point the camera at that speaker. This system, however, could handle only a one source environment.

Most source location estimators have been extended into a multiple source environment through determining the direction of signal propagation using a modification of the MUSIC algorithm (Rao, 1985), a spectral estimation method based on sub-space decomposition. Multiple source location algorithms require a higher resolution estimator with increased computational load typically through the use of cross correlation metrics to estimate the source locations (Rao, 1985; Wang, 1985; Friedlander, 1993).

For signal extraction, beamformers use time lags between sensors to reduce noise effects and improve the quality of the primary signal. Research has shown that beamforming algorithms outperform traditional, single channel enhancement methods (Bitzer, 2001; Saruwatari, 2000; Brandstein, 2001). To reduce noise in an environment, beamformers act as spatial filters through "steering" the array of sensors towards a "look" direction where the primary signal of interest is located, thus emphasizing the primary signal features while negating the noise signal features.

In (Widrow, 2001), beamforming algorithms were integrated into a hearing aid design, which increased the user's ability to understand speech up to 70 percent compared to traditional hearing aid designs. Widrow's design consists of a necklace microphone array as shown in Figure 5. Widrow assumed the look direction to be directly in front of the wearer or at zero degrees to the array.

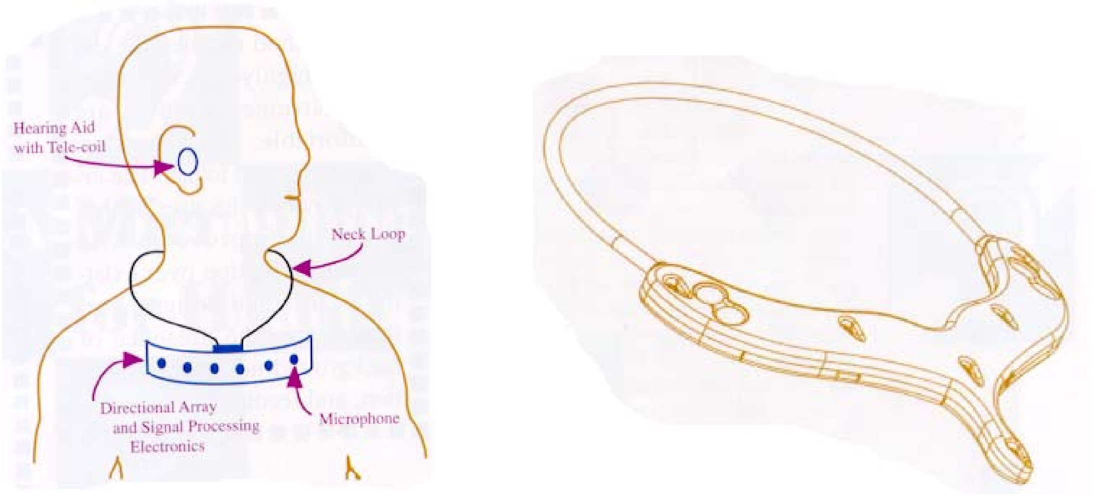


Figure 5: Hearing aid microphone necklace array (Widrow, 2001)

2.2.1 Delay and Sum Beamformer

The most fundamental of the beamforming algorithms is the delay and sum (DS) beamformer. Given a signal of interest in a certain location in space, the signal will arrive at the sensors, or microphones, at times determined by each microphone's location. For a linear, equally spaced array and a far-field model, those time differentials are as given previously in equation [1]. As mentioned, Figure 1 shows the graphical representation of the propagating signal.

Once the time difference of each microphone relative to the others is determined, each microphone signal is shifted in time to align the signal of interest, without aligning the noise. This is accomplished only when the noise is not propagating in the same direction as the signal if interest. As shown in Figure 6, the signal of interest is increased in magnitude by the number of microphones in the array, while the noise is linearly combined. The flow graph of the DS beamformer is shown in Figure 7. Once the signals are time shifted and summed, dividing by the number of microphones normalizes the signal of interest.

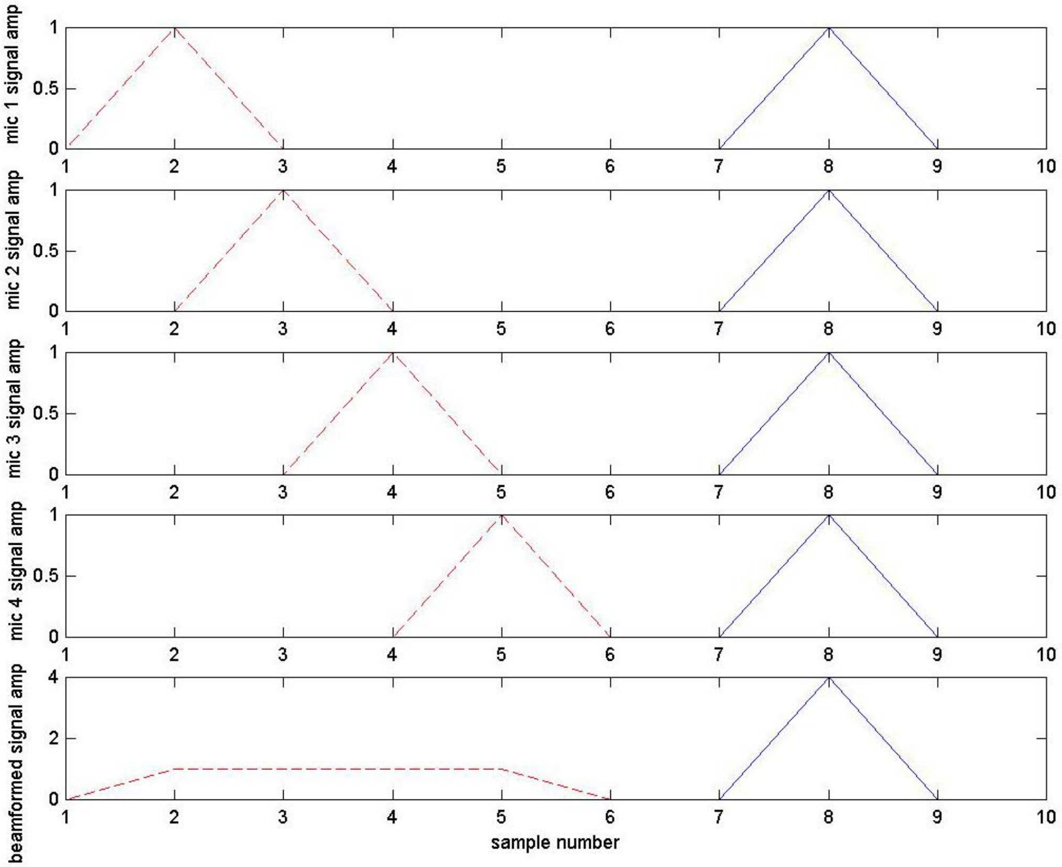


Figure 6: A graphic of a DS beamformer

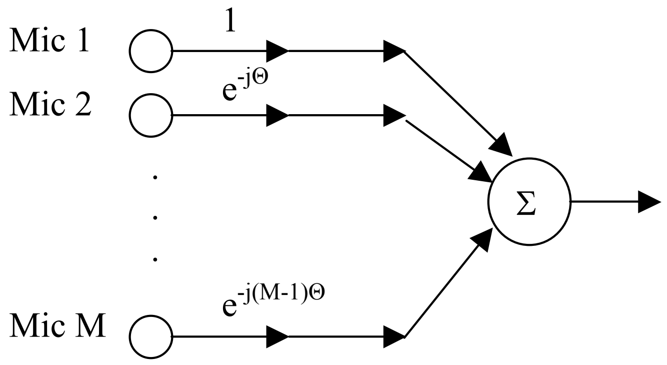


Figure 7: DS beamformer flow graph

The equation for the DS beamformer response, **z**, is given by:

$$\mathbf{z}_{\phi,f}[n] = \sum_{m=1}^{M} w_m(\phi, f) y_m[n] \quad \leftrightarrow \quad \mathbf{z}_{\phi,f}[n] = \mathbf{w}^{\mathbf{H}}(\phi, f)\mathbf{y}[n] \qquad [4]$$

$$\mathbf{w} = \frac{\mathbf{d}}{M} \qquad [5]$$

where *M* is the number of microphones in the array and where each narrowband microphone signal y_m or **y** has a center frequency *f* and arrival angle ϕ. The “$^{\mathbf{H}}$” denotes the hermitian transpose of a vector or matrix and a “$^{\mathbf{T}}$” denotes the transpose of a vector or matrix. The filter weights w_m or **w** are a function of the delay vector **d** that is normalized by dividing by the number of microphones. With a far-field model and a linear, equally spaced array, the filter weights' delay vector **d** is defined by:

$$\mathbf{d}^{\mathbf{T}} = [1, e^{(-j\Theta)}, e^{(-2j\Theta)}, ..., e^{(-j(M-1)\Theta)}] = [1, e^{(\frac{-j2\pi f \mathrm{d}\cos\phi}{v})}, e^{(\frac{-j2\pi f\, 2\mathrm{d}\cos\phi}{v})}, ... , e^{(\frac{-j2\pi f\,(M-1)\mathrm{d}\cos\phi}{v})}] \qquad [6]$$

Although the DS beamformer is extremely simple in design, this simplicity has a significant advantage over other more complicated beamformers through its fast computational abilities. This characteristic allows the use of the DS for real time implementation as is required in many applications like hearing aid design and teleconferencing.

2.2.2 MVDR Beamformer

The minimum variance distortionless response (MVDR) beamformer improves upon the DS beamformer through utilizing the correlations between microphone pair signals in addition to the time differentials. The MVDR is normally implemented as an adaptive algorithm because it computes the correlation matrix of the array signals for each segmented frame. Typically, speech signals are broken into frames to approximate stationarity of the signal characteristics.

The MVDR works by minimizing signals propagating from directions other than the look direction of the beamformer while constraining the signal response in the look direction to unity:

$$\mathbf{w}^{\mathbf{H}}\mathbf{d} = 1 \qquad [7]$$

A solution to this problem is found in the MVDR beamformer and can be derived using Lagrange multipliers (Frost, 1972). The resulting equation for the MVDR beamformer filter weights is given by:

$$\mathbf{w} = \frac{\mathbf{R}^{-1}\mathbf{d}}{\mathbf{d}^{\mathbf{H}}\mathbf{R}^{-1}\mathbf{d}} \qquad [8]$$

where **d** is the delay vector as previously defined on page 21 and **R** is the autocorrelation matrix of the array signals at a sample point. Note that the DS beamformer can be viewed as a sub-case of the MVDR beamformer, with **R** reduced to the identity matrix.

2.3 Implementation of Beamformers

2.3.1 Delay and Sum Beamformer

From equation [6], the DS beamformer output has, in general, both real and imaginary components. This is due to the narrowband assumption, where $e^{j\Theta}$ is intended as a delay element at a specific frequency, which does not generalize to real data. To deal with this, the relative time delays are calculated from equation [1] and the signal from each microphone is then shifted by the respective amount of points, rather than implementing the theoretical filter equation. For this research, the sampling frequency is always 16,000 Hz so that each point in the signal represents 62.5 microseconds. Once each of the microphone signals is appropriately shifted, they are summed together to create the beamformer output signal.

2.3.2 MVDR Beamformer

The theoretical MVDR filter weights in equation [8] also give both real and imaginary components. In a method similar to that above, the microphone signals are time shifted by the appropriate delay value. The correlation matrix, **R**, is calculated and applied to the microphone signals to create the beamformer output as detailed in the following steps outlined in Table 1:

Theoretical Approach	Practical Implementation
1. Calculate the M by M inverse correlation matrix of the signal array $\mathbf{R}^{-1}$	1. Calculate the time/sample delay given the signal's angle of approach using $\Delta t = \frac{(m-1)\mathrm{d}\sin\phi}{v}$
2. Calculate the theoretical M by 1 delay vector given by equation [6]: $\mathbf{d}^{\mathbf{T}} = [1, e^{(-j\Theta)}, e^{(-2j\Theta)}, \ldots, e^{(-j(M-1)\Theta)}]$	2. Time align each microphone signal so that the delay weights given by equation [6] reduce to a vector of unity values. Calculate the M by M inverse correlation matrix of the time aligned microphone signal array $\mathbf{R}^{-1}$
3. Multiply the M by M inverse correlation matrix by the M by 1 delay vector	3. Multiply the M by M inverse correlation matrix by the M by 1 unity vector of delays.
4. The resulting M by 1 vector is then calculated and divided by the scalar value $\mathbf{d}^{\mathbf{H}}\mathbf{R}^{-1}\mathbf{d}$ to create the filter weight vector, **w**	4. Divide the resulting M by 1 vector by the scalar value resulting from the 1 by M unity vector multiplied by the inverse correlation matrix and then again multiplied by the M by 1 unity vector.
5. Apply the filter weight vector to the signal array with: $\mathbf{z}_{\phi,f}[n] = \mathbf{w}^{\mathbf{H}}(\phi, f)\mathbf{y}[n]$ A sample point in the beamformer output is then produced	5. Apply the resulting M by 1filter weight vector to the signal array as in: $\mathbf{z}_{\phi,f}[n] = \mathbf{w}^{\mathbf{H}}(\phi, f)\mathbf{y}[n]$ A sample point in the beamformer output is then produced

Table 1: Theoretical versus practical MVDR beamformer process

2.4 Speech Enhancement Fundamentals

Speech enhancement is a large research area in speech signal processing. The goal of many enhancement algorithms is to suppress the noise in a noisy speech signal. In general, noise can be additive, multiplicative, or convolutional, narrowband or broadband, and stationary or nonstationary. The majority of research in speech enhancement addresses additive, broadband, stationary noise.

Speech enhancement algorithms have many applications in speech signal processing. Signal enhancement can be invaluable to hearing impaired persons because the ability to generate clean signals is critical to their comprehension of speech. Enhancement algorithms are also used in conjunction with speech recognizers and speech coders as front end processing. It has been shown that enhancing the noisy speech signal before running the signal through a recognizer can increase the recognition rate and thus create a more robust recognizer (Kajita, 1996; Bitzer, 2001; McCowan, 2001). Similarly, front end enhancing to speech coding has been shown to decrease the number of bits necessary to code the signal (Carnero, 1999).

2.4.1 Spectral Subtraction

One of the most simple and widely used enhancement methods is the power spectral subtraction algorithm. This algorithm's basis is in estimating the noise and subtracting it in the power spectral domain (Boll, 1979). The basic equations are given by:

$$y = s + n \qquad [9]$$

where *s* is the clean signal, *n* is the uncorrelated noise signal, and *y* is the noise corrupted input signal, and

$$\Gamma_{\hat{s}} = \Gamma_y - \Gamma_n \qquad [10]$$

where Γ_y is the power density spectrum (PDS) of the noise corrupted signal found by taking the Discrete Fourier Transform (DFT) of the noisy signal and Γ_n is the PDS of the noise signal estimate. In the above equation, the PDS of the noise estimate is subtracted from the PDS of the noise corrupted signal, yielding a PDS estimate for the clean signal. An inverse DFT is then applied to obtain the clean signal estimate. Some initial knowledge of the noise signal must be known in order to obtain a noise signal estimate. Because an *a priori* noise signal estimate is often difficult to find, iterative improvements are often performed on the algorithm by replacing the noise corrupted PDS by the resulting PDS clean signal estimate.

This algorithm has many adaptations and improvements (Hu 2002; Deller 2000). One particular example is the power spectral subtraction method given by:

$$\hat{\mathbf{S}}(\omega) = \left[\left|\mathbf{S}(\omega)\right|^2 - \left|\mathbf{N}(\omega)\right|^2 \right]^{1/2} e^{j\varphi_s} \qquad [11]$$

where $\hat{\mathbf{S}}$ is the resulting clean signal PDS estimate, **S** is the DFT of the noise corrupted input signal, **N** is the DFT of the noise signal estimate, and φ_S is the phase spectrum of the input signal. Capturing the phase information from the noise corrupted signal is a valid approximation because human perception places little importance on the phase information in speech signals (Wang, 1982). A similar algorithm is the generalized power spectral subtraction method given by (Berouti, 1979):

$$\hat{\mathbf{S}}(\omega) = \left[\left|\mathbf{S}(\omega)\right|^{a} - k\left|\mathbf{N}(\omega)\right|^{a} \right]^{1/a} e^{j\varphi_s} \qquad [12]$$

A noise correlation constant *k* and a power constant *a* are the differences between equation [11] and equation [12]. Integrating a noise correlation constant allows the generalized spectral subtraction method to further adjust how much noise power is subtracted.

From the above equations, imaginary values can result if the estimate of the noise PDS is greater than the PDS of the noise corrupted signal. Because speech signals are real valued signals, these imaginary values can be dealt with through spectral flooring. Spectral flooring sets negative PDS estimates to zero.

2.4.2 Wiener Filtering

Another widely utilized algorithm in speech enhancement research is the Wiener filter. If both the signal and the noise estimates are exactly true, this algorithm will yield the optimal estimate of the clean signal. Through minimizing the mean squared error between the estimated and clean speech signals, the Wiener filter is developed and given by:

$$\mathbf{H}(\omega) = \left[\frac{\left|\hat{\mathbf{S}}(\omega)\right|^{2}}{\left|\hat{\mathbf{S}}(\omega)\right|^{2} + \left|\mathbf{N}(\omega)\right|^{2}} \right] \qquad [13]$$

$$\hat{\mathbf{S}}(\omega) = \mathbf{H}(\omega)\mathbf{S}(\omega) \qquad [14]$$

where **H** is the Wiener filter, and **S** and **N** are the noise corrupted speech and noise spectra, respectively. Because the Wiener filter has a zero phase spectrum, the phase from the noisy signal is the output phase for the estimation of the PDS of the clean signal. This was similar to the spectral subtraction algorithms.

It should be noted that the Wiener filter assumes that the noise and the signal of interest are ergodic and stationary random processes and thus not correlated to each other. To accommodate the nonstationarity of speech signals, the signals can be broken into frames to assume stationarity, as is commonly done in speech signal processing research.

Another generalization to the Wiener filter is found through incorporating a noise correlation constant *k* and a power constant *a* to the filter:

$$\mathbf{H}(\omega) = \left[\frac{\left|\hat{\mathbf{S}}(\omega)\right|^2}{\left|\hat{\mathbf{S}}(\omega)\right|^2 + k\left|\mathbf{N}(\omega)\right|^2} \right]^a \qquad 15]$$

Again, similar to spectral subtraction, *a priori* knowledge of the noise signal is required, but is often difficult to obtain. Incorporating iterative techniques and methods of estimating the noise are therefore important to the Wiener filter algorithm (Hansen, 1987; Lim, 1978). The iterative techniques re-estimate the Wiener filter with each iteration.

2.4.3 Single Channel Systems

The traditional speech enhancement techniques of Wiener filtering and spectral subtraction have been based upon a single channel system given *a priori* knowledge of the noise characteristics. In most real world situations, however, *a priori* knowledge is not available. To obtain an estimation of the noise, detection methods were created to determine when speech silence regions occur. During these speech silent regions, it is assumed that only noise is present in the input signal, therefore allowing the extraction of a noise estimate. With this information, the noise characteristics can be determined and used in the enhancement algorithms listed above. In order to determine silence regions, most methods utilize an energy based calculation where a threshold is set. If the energy reaches a certain limit, a decision is made to flag a silence region and obtain a noise estimate at that time (Ris, 2001). It follows that these enhancement algorithms will perform less efficiently given a noise corrupted signal with a large signal to noise ratio. Given a small noise signal, obtaining a high-quality estimation of the noise signal is more difficult.

2.5 Speech Enhancement Measurement Fundamentals

Because the focus of this research is on speech signal enhancement, it is important to introduce the methods used to determine the amount of enhancement the algorithms developed here perform. Previous research of enhancement metrics has been unable to find a quantifier directly correlated to that of human perception. This is challenging because human perception varies from person to person and science has yet to unlock all of the secrets of human cognitive function.

2.5.1 Objective and Subjective Metrics

Objective metrics can be calculated given an equation whereas subjective metrics require human subjects and individual opinions to score them. Objective quantifications of enhancement are created through arithmetic algorithms such as the signal to noise ratio

(SNR) or Itakura distance measure (Itakura, 1975). Because subjective testing is extremely laborious to conduct compared to that of objective metrics, much research has been performed to try to create an objective measure that correlates well to human subjective testing, however this has so far been unsuccessful (Deller, 2000). Therefore, using subjective tests with people remains the best metric for speech enhancement.

Some of the more common performance measures are the SNR, the segmental SNR, the Itakura metric, and the accuracy rate of speech recognition engine. Although none of these metrics is a direct measure of perceived speech signal quality, it has been established that the segmental SNR is more correlated to perception than SNR (Quackenbush, 1988). This research utilizes the SNR and the segmental SNR as objective enhancement quantifications. The SNR is given by:

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_{n} s^2(n)}{\sum_{n}\left[s(n)-\hat{s}(n)\right]^2} \qquad [16]$$

where the clean signal is *s* and the enhanced signal is $\hat{s}$. The segmental SNR simply creates stationarity to the speech signal through dividing the signal into *i* frames each with *N* points and also helps give equal weight to softer-spoken speech segments. The final segmental SNR value is the average of the *i* segmental SNR frame values.

$$\mathrm{SNR}_{\mathrm{seg}} = \frac{1}{N}\sum_{j=0}^{N-1} 10\log_{10}\left[\frac{\sum_{n} s^2(n,i)}{\sum_{n}\left[s(n,i)-\hat{s}(n,i)\right]^2}\right] \qquad [17]$$

Using a speech recognition engine allows for comparison of the noisy signal and enhanced signal through comparing accuracy values. The noisy signal is first run through the recognizer and then the enhanced signal is put through it. Recognition accuracy is used as a measure of signal intelligibility.

2.5.2 Quality and Intelligibility

There are two separate issues to address with respect to enhancement: quality and intelligibility. Intelligibility is the capability of a person to understand what is being spoken whereas improving the speech signal quality is based more upon the naturalness and clarity of the signal. Although a listener may be able to understand words spoken in a signal, the signal may not sound "natural", and may be perceived as poor quality. This is true in robotic-like synthesized speech. As mentioned previously, telephone signals have a limited bandwidth and thus have a degraded quality compared to the same signal if no band limitations occurred. This degraded quality, however, does little to compromise the intelligibility of the signal.

Quantifying intelligibility is more definitive because listeners can be asked to write down what they hear or circle words that they heard on a questionnaire. This type of testing can have explicit quantities because the words written or circled are either correct or not. A commonly used test for intelligibility is the diagnostic rhyme test (DRT) that requires listeners to circle the word spoken among a pair of rhyming words. Although intelligibility is simple and definitive to score among listeners, the objective algorithms discussed in the previous section cannot quantify intelligibility.

The objective algorithms used to measure signal enhancement can only estimate relative change in quality of the signal. Quality testing is subjective among listeners because the basis of quality is rooted in the opinions of each individual's perception of quality. One person may be a more critical judge of quality compared to another; thus, creating a bias in quality rating among individuals. Typically, testing of quality is done with a rating system. A mean opinion score (MOS) is a common quality test that asks a listener to rate the speech on a scale of one to five, with five being the best quality. These tests can attempt to reduce individual biases through normalizing the means of each listener with test signals.

Although intelligibility testing with listeners is more easily and precisely quantified compared to quality testing, the implementation of intelligibility tests like the DRT is more difficult. Quality testing using a rating system on a signal, as in the MOS, is simple for a listener and allows for more types of speech to be utilized.

Chapter 3 Iterative Multiple Source Enhancement Method

When enhancing speech signals in a multiple speaker environment, the traditional enhancement methods have shortcomings and must be adapted. They are not able to cope with nonstationary noise with similar spectral characteristics, as is the situation with multiple speech signals. In addition, they are not designed for the multiple channel systems available with microphone arrays.

The traditional method of obtaining a noise estimate from a silent region of the primary speaker works well for noise that has stationary spectral characteristics; however, a speech signal corrupted with interfering speakers has nonstationary speech as noise. Interfering talkers may have characteristics changing at a rate faster than the primary speaker, and the silence region noise estimators will not perform well in the multiple speaker environment. Additionally, if the interfering talkers have similar energy characteristics, as is often the case in multiple speaker environments, the silence detector will not be able separate the energies of the primary speaker and interfering speakers. This will render a detector unable to separate the primary speaker’s silent regions. The silent region detectors cannot be used in multiple speaker environments because they are unable differentiate between the primary and interfering talker signals.

To integrate the spectral subtraction and Wiener filtering enhancement methods to a multiple channel system, post beamformer filtering has been performed in previous research (Marro, 1998; Bitzer, 1999; McCowan, 2000). The general block diagram is shown in Figure 8. The signals from each of the microphones, x_1 through x_N, are first time aligned into x'_1 through x'_N given *a priori* knowledge of the signal's location. Then, each signal is broken into *i* frequency bins where the data is processed through the beamformer's weighting function, *g*, and the noise filter function, *d*. Typically, *g* and *d* are identical functions. A noise post filter estimation is created and applied to the signal generating the post filtered beamformed output **Z**. To synthesize back to the time domain, an inverse transform is performed on each of the frequency bins. The noise post filter adapts itself based upon the output SNR (Brandstein, 2001).

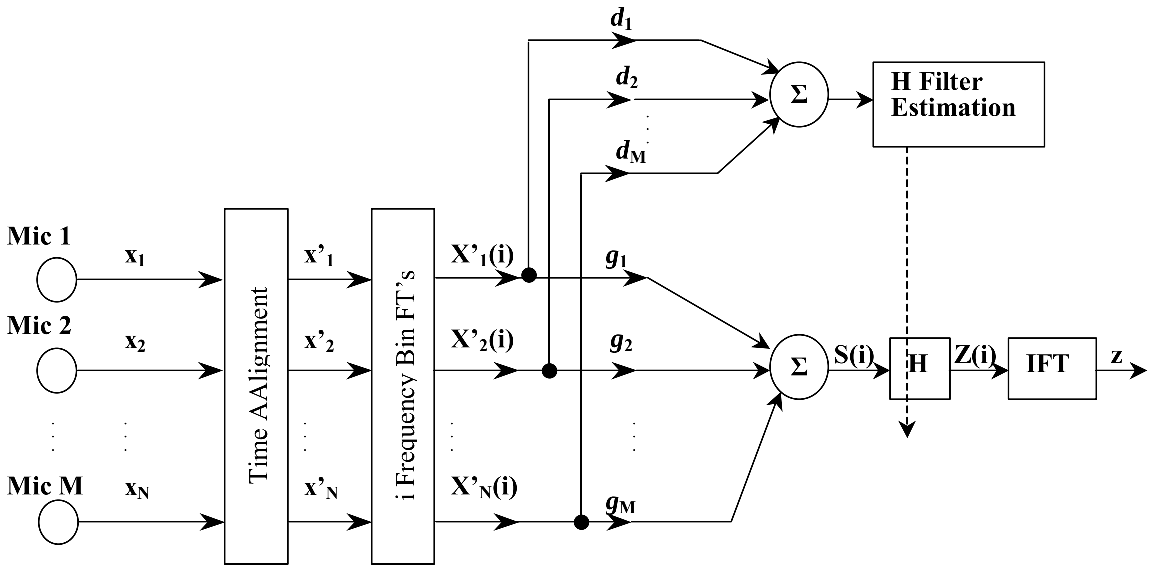


Figure 8: Block diagram of post filtering enhancement algorithms integrated with a microphone array

A generalized estimation of the noise post filter based upon the MVDR beamformer is derived in (Marro, 1998) to be:

$$H = \frac{(\mathbf{w^H R w - w^H R^H w) w^H w}}{\mathbf{(1 - w^H w) w^H R^D w}} \qquad [18]$$

where $\mathbf{R^D}$ is the diagonal matrix of the autocorrelation matrix.

Although equation [18] integrates noise filtering enhancement techniques with a multiple channel beamformer, it remains dependent upon *a priori* knowledge of the noise signal in order to first estimate the noise spectral characteristics. In multiple speaker environments using microphone arrays, as discussed previously in this section, it is not possible to predict or estimate the noise from an interfering talker. Further, the adaptive ability of the post filtering techniques relies on a prior knowledge of the primary signal in order to calculate an SNR.

To contend with the multiple speaker environment using a microphone array, this research uses multiple, parallel beamformers with a prior knowledge of source locations to acquire noise and signal estimates. In estimating the primary speaker, an initial beamforming algorithm is performed using either the DS or MVDR beamformer steered toward the primary speaker's direction. After the beamformer is used to extract the primary source signal, artifacts of the non-primary signals may still remain. To further extract the primary signal, the use of multiple beamformers obtains each noise source's estimate, and multiple source alterations of the traditional enhancement methods can then be utilized. A block diagram is shown in Figure 9.

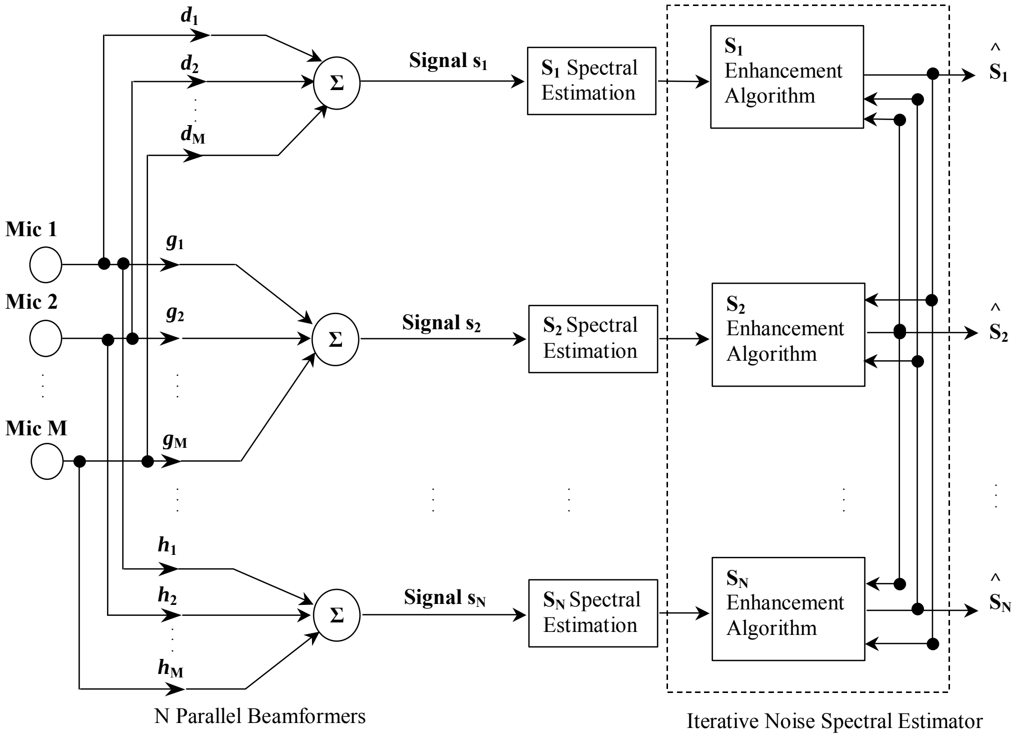

d_1
d_2
d_M
Mic 1
Mic 2
Mic M
g_1
g_2
g_M
h_1
h_2
h_M
Σ
Σ
Σ
Signal s_1
Signal s_2
Signal s_N
S_1 Spectral Estimation
S_2 Spectral Estimation
S_N Spectral Estimation
S_1 Enhancement Algorithm
S_2 Enhancement Algorithm
S_N Enhancement Algorithm
$\hat{S}_1$
$\hat{S}_2$
$\hat{S}_N$
N Parallel Beamformers
Iterative Noise Spectral Estimator

Figure 9: Multiple source enhancement algorithm flow graph

3.1 Multiple Source Spectral Subtraction Enhancement

To develop the multiple speaker spectral subtraction enhancement algorithm, the *N* noise source beamformer outputs are used as the initial noise estimates, $\hat{\mathbf{N}}_{\mathbf{i}}$, while the noise corrupted signal, **S**, is set to be the primary source beamformer output as shown in:

$$\hat{\mathbf{S}}(\omega)=\left[\left|\mathbf{S}(\omega)\right|^{a}-k_{1}\left|\hat{\mathbf{N}}_{1}(\omega)\right|^{a}-\ldots-k_{N}\left|\hat{\mathbf{N}}_{N}(\omega)\right|^{a}\right]^{1/a}e^{j\varphi_{s}(\omega)} \qquad [19]$$

Like the traditional generalized algorithm, the phase is added back in using the noisy signal phase and the signals are windowed to approximate the speech signal as stationary.

In this research, the power constant *a* is set to be two. This creates a power spectral subtraction so that there are only positive values in the noise spectral characteristics used in the algorithm. If the noise power estimates multiplied by their respective *k* constant factors are larger than the noise corrupted signal power, a negative new power estimate is created. Spectral flooring is used so that no negative values are established. The constant factors *k* are related to the coupling between the sources as discussed below in Section 3.3

The algorithm is iterated to maximize enhancement. As shown in the block diagram in Figure 8, the enhanced signal estimates can be looped back into the algorithm as an updated noise signal estimate for the other source signals. It is important to note that the original beamformed signal is always used as the noise corrupted signal **S**. The amount of improvement in the noise estimate will approach a limit as the number of iterations increases. This limit is dependent upon the unique multiple source environment, and in this research, the number of iterations is set to be five.

The iterative approach to the algorithms led to some investigation into the convergence of the noise signal estimates. As a result, a smoothing function is integrated into the implementation of the algorithms. The new estimate of the noise signal is simply averaged with the previous noise estimate signal upon each iteration process. This allows for faster convergence of the noise signal estimates and thus reduces the required iterations to five.

The rate of convergence for the noise estimates is highly dependent upon the initial noise estimates. For the multiple source situations presented here, a more spectrally dominant source will yield a high-quality estimate of that signal while at the same time creating a poor estimate for the less powerful source. A large difference between the estimates causes a longer convergence time, and incorporating the smoothing function helps minimize that convergence time.

3.2 Multiple Source Wiener Filtering Enhancement

Like the multiple source spectral subtraction algorithm, the multiple source Wiener filter utilizes the *N* noise source beamformer outputs as the initial noise estimates, $\hat{\mathbf{N}}_{\mathbf{i}}$. Again, the noise corrupted signal, **S**, is set to be the primary source beamformer output, and the signals are divided into frames. The resulting filter is:

$$\mathbf{H}(\omega) = \left[\frac{\left|\hat{\mathbf{S}}(\omega)\right|^2}{\left|\hat{\mathbf{S}}(\omega)\right|^2 + \left|k_1 \hat{\mathbf{N}}_1(\omega)\right|^2 + \ldots + \left|k_N \hat{\mathbf{N}}_N(\omega)\right|^2} \right] \qquad [20]$$

As shown, the noise signal estimates are the key factor to the amount of enhancement produced. Therefore, the algorithm is improved through the creation of an iterative noise estimation as shown in Figure 8.

3.3 Coupling function, k

The noise of the original signal is composed of multiple interfering speakers and is initially passed through a beamformer to de-emphasize the interfering talkers' noise signals; however, some level of the interfering talkers remain even after beamforming.

This is especially true for the small aperture array that is used in this research because the resolution of the beamformer to separate signals in space decreases with decreasing microphones. The separation resolution of the beamformer is also dependent on frequency and the separation of the signal sources. The lower the frequency, the less the beamformer is able to separate the signal at that frequency. Similarly, the closer the sources, the less the beamformer is able to separate the source signals.

It is therefore necessary to incorporate a function that is related to the beamformer response when filtering or subtracting the remaining interfering talker spectral information. Like the beamformer response, this function is dependent upon frequency and the talkers' physical separation. In the multiple source spectral subtraction and Wiener filtering techniques, a coupling factor k is applied. This coupling function determines the amount of the interfering signal's spectral energy that is filtered or subtracted from the beamformed signal.

The coupling function is calculated using an estimate of the amount of interfering talker noise that is passed through the beamformer using the beamformer lobe function. This function determines the amount of noise coupled in the beamformed signal, as shown in Figure 10 and given by:

$$W(f,\phi) = \frac{\sin\left[\dfrac{\pi f M d}{c}(\sin\phi_o - \sin\phi)\right]}{\sin\left[\dfrac{\pi f d}{c}(\sin\phi_o - \sin\phi)\right]} \qquad [21]$$

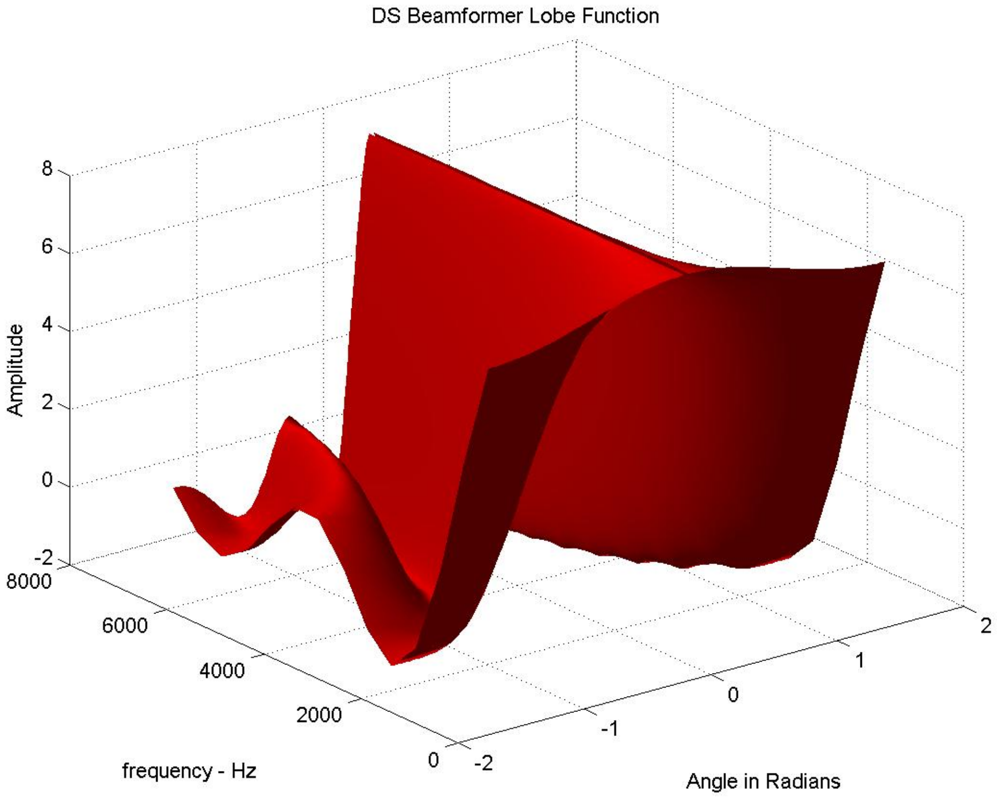


Figure 10: DS beamformer lobe for an array with eight microphones and 2.5 cm spacings and a changing ϕ in radians.

This equation is taken from the DS beamformer response and is a function of frequency and source direction. Given the beamformer function and a specific angular separation, it is possible to evaluate the coupling function across frequencies that can directly determine the spectral characteristics of the interfering sources passed through the beamformer. Incorporating this coupling function into the post filtering enhancement algorithms will ensure that the interfering talkers' signal residual spectrum is subtracted or filtered with the appropriate level across the frequency range.

Although the beamformer function is a sinc function in theory, it is judicious in practice to define the coupling function through using the sinc function envelope in order to make it more robust to slight discrepancies in the source locations. The coupling function based on this envelope is shown in Figure 11 and given by:

$$k = \begin{cases} \dfrac{\sin\left[\dfrac{Mf\pi d}{c}(\sin\phi_o - \sin\phi)\right]}{\dfrac{f\pi d}{c}(\sin\phi_o - \sin\phi)} & low\ f \\ \\ \dfrac{1}{sin\left[\dfrac{f\pi d}{c}(sin\phi_o - \sin\phi)\right]} & high\ f \end{cases} \qquad [22]$$

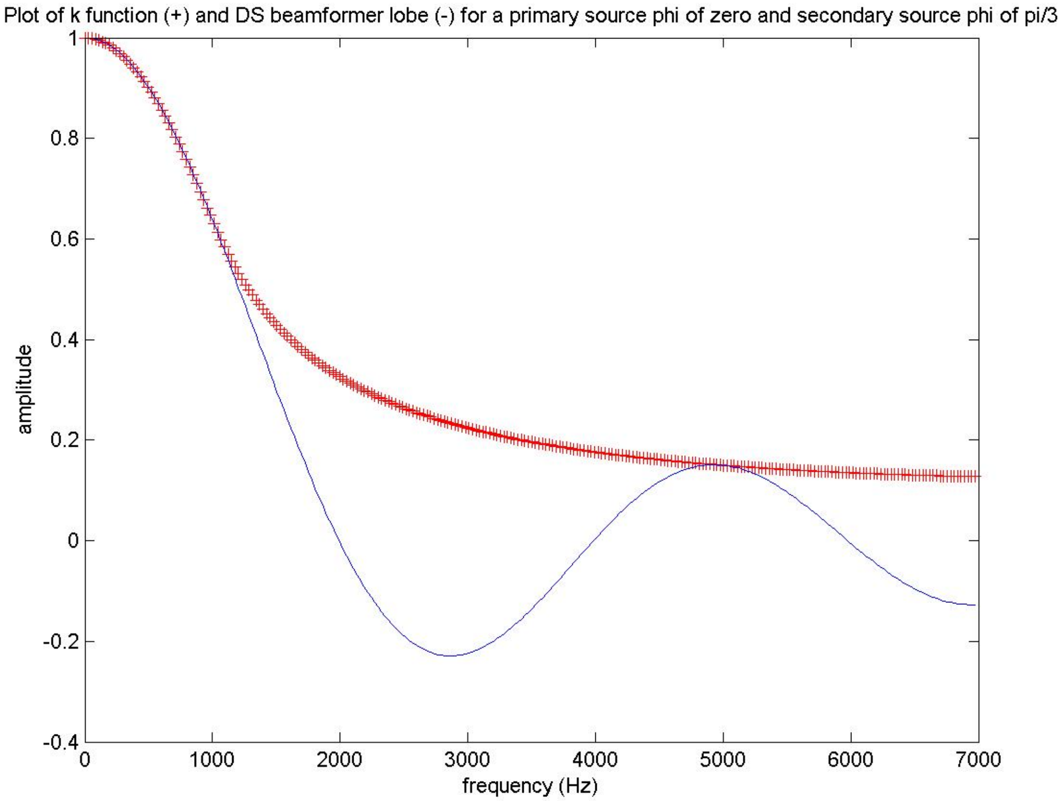


Figure 11: Coupling-function, *k*, as envelope of the beamformer sinc function

Chapter 4 Experimental Setup

4.1 Overall Setup

The experiments can be broken into two main types: simulated experiments and sound booth experiments. The sound booth experimental hardware and data acquisition system is discussed in Chapter 5. The algorithms are executed in the same manner for both the simulated and sound booth experiments.

4.2 Multiple Speaker Input Signals

To simulate the multiple speaker environment, equation [1] was used to determine the appropriate time shift for each source signal given the angle of direction for each source and the microphone signal being created.

4.2.1 Simulated geometries

First, a two source experiment was run with the first speaker placed at a constant location while the second speaker's location was varied as shown in Table 2 and Figure 12. For each geometry, the second speaker's signal was varied in magnitude ten times, thus creating ten SNR's per geometry. The speech signals used were the same signals for this entire geometry and SNR variation. This experimental set up was run five times or for five different speech signal combinations.

Source 1	Source 2
–π/3	π/3
–π/3	π/4
–π/3	π/5
–π/3	0

Table 2: Simulated two source geometries

Next, a three source experiment was run with the first speaker placed at a constant location while the second and third speakers' locations were varied as shown in Table 3 and Figure 12. Again, ten variations in the SNR were created by magnitude changes of speakers two and three for each geometry, and the speech signals used were the same signals for this entire geometry and SNR variation. The entire experiment of all of these geometries and SNR's was run five times or for five different speech signal combinations.

Source 1	Source 2	Source 3
0	–π/3	π/3
0	–π/4	π/4
0	–π/5	π/5

Table 3: Simulated three source geometries

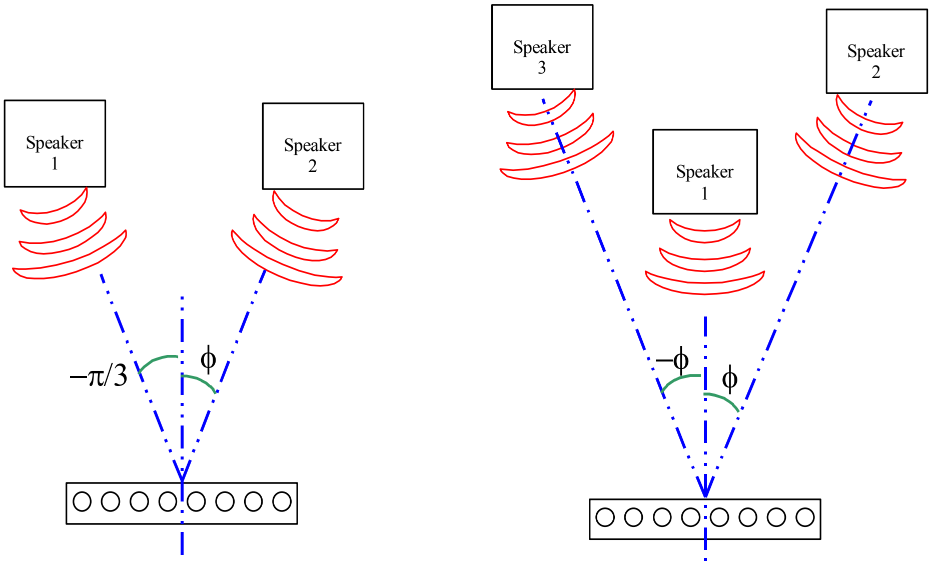


Figure 12: Experiment setups

4.2.2 Sound booth geometry

In the sound booth experiments, there is a primary source and one competing noise source. Both sources remain stationary in adjacent corners of the room, located at -π/7 and π/7 as shown in Figure 13. The amplitudes of the noise source were amplified differently to create five different SNR's.

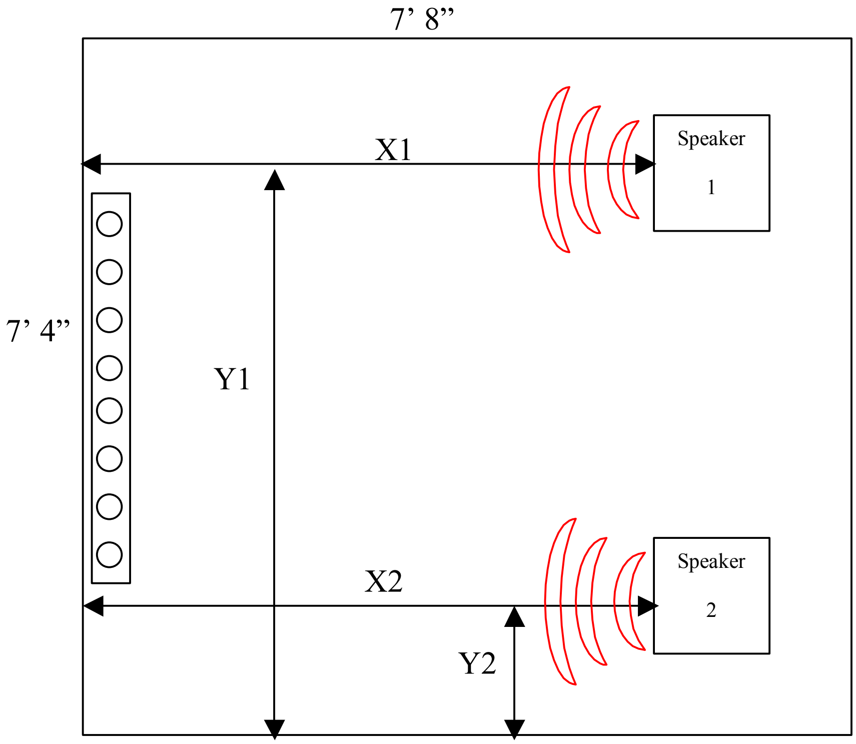


Figure 13: Sound booth multiple source experiment layout

4.2.3 Data

The data used to create the multiple speaker environments in this research is obtained from the DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus database (Garofolo, 1993). The training waveform files can be used to build automatic speech recognition systems, and the testing files can then be used to evaluate the systems and yield a percent recognition rate.

Other than the requirement that different sentences and different speakers be used for each source for any one experimental setup, the speakers and sentence waveforms in this subdivision were chosen at random from the North Midland dialect region and include both men and women speakers. These waveforms are then combined to create a multiple speaker signal for the simulated experiments and are independently output to speakers in the sound booth experiments.

4.3 Processing detail

A band pass filter from 300 to 6800 Hz is applied to each of the microphone signals to assure that only sounds within the capability of the sampling frequency and array geometry are present. Next, the input speech signal is divided into 512 point, 32 millisecond, triangular windowed frames. The 32 millisecond frames are commonly used in speech signal processing to approximate stationarity of the speech signal.

Each of the frames is run through a filter bank to produce ten bins of equal frequency bandwidths across the given range. Ten bins were chosen because it was determined that this was the fewest number of frequency bins that still resulted in a significant improvement in the beamformer algorithms. The filters used are twelfth order FIR band pass filters. Given the filtered and framed data for each of the microphone signals and given the *a priori* knowledge of each signal source direction, each source is beamformed using the DS or MVDR algorithm for every separate filter bin. After beamforming, the signal is resynthesized from the frames. The 50% overlapping triangular windows are useful because they allow for simple additive resynthesis without introducing distortion.

Finally, using the resynthesized beamformer output signals, the enhancement algorithms are implemented.

Similarly, the enhancement algorithms divide the signals into 512 point, 32 millisecond, triangular windowed frames. Each frame is processed and the enhanced signal is resynthesized by overlapping and adding the frames once again.

Chapter 5 Data Acquisition System Setup

A National Instruments data acquisition system is used to create the multiple speaker output system. It interfaces with the NI input and NI output cards using LabView software. Through the use of the LabView software, input signals for each microphone record the multiple output scenarios, and Matlab Version 6.1 software is used for developing the algorithms and completing the analysis on those signals. The LabView block diagram of the data acquisition system is shown in Figure 14.

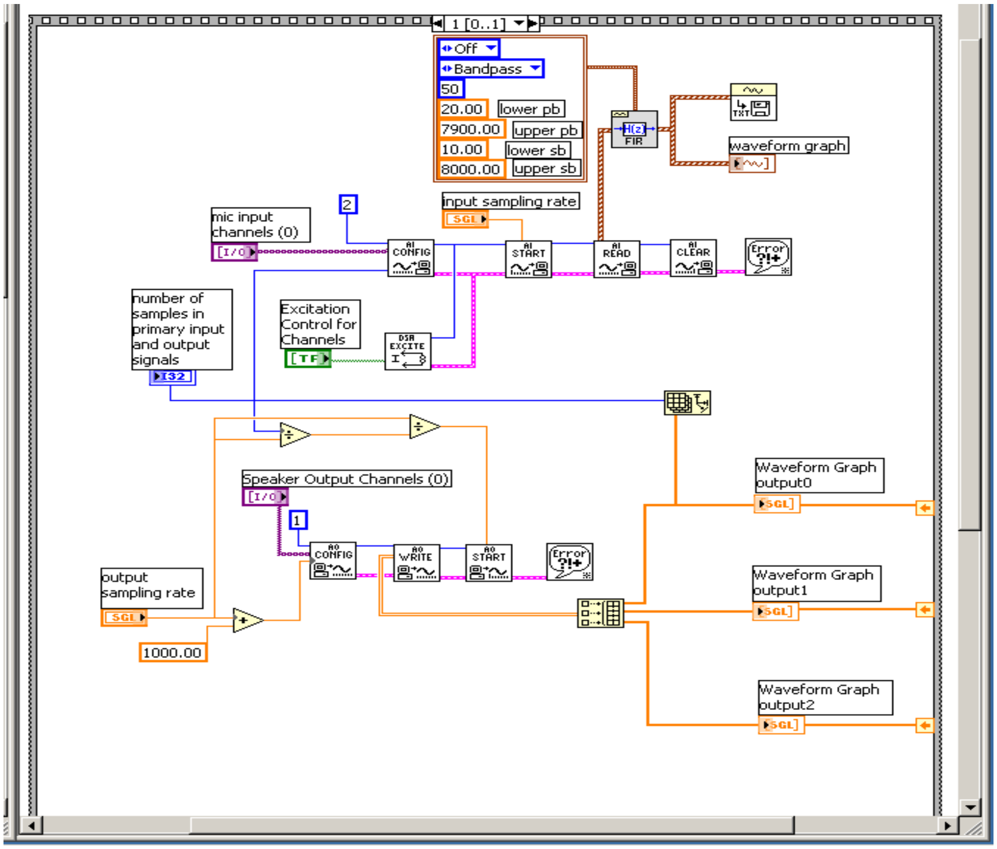


Figure 14: LabView block diagram of data acquisition system

5.1 Multiple Speaker Output System

5.1.1 Output Card

The NI-6731 output card is used to simultaneously convert up to four acoustic digital files to analog voltage signals that are then routed to an amplifier. The card uses a 16 bit resolution and spans ±10V with an accuracy of ±1.0mV. The sampling frequency is set to 16,000 Hz.

5.1.2 Speakers

Two satellite speakers are used to output the two separate speech signals from the output card, depending upon the experiment being performed. These speakers are placed at different locations in the sound booth with each of them facing the microphone array. The speakers and the microphone array are at the same elevations to simplify the setup to a two dimensional analysis. To acquire the speech signals, the TIMIT speech corpus is utilized. As discussed in Section 4.2.3, a sentence from a randomly chosen speaker in the TIMIT corpus is used as an output speech signal for each speaker. This multiple source signal is then recorded on each of the microphones. Two different people, each speaking an independent sentence from each other, are used from the TIMIT corpus for the speech signal output from the speakers. Thus, a multiple source signal, similar to the one shown in Figure 15, is recorded on each of the eight microphones.

5.2 Multiple Input System

All of the algorithms in this research are created and implemented on a Pentium IV OmniTech PC using MATLAB 6.1 software. The data acquisition system replicates the speech sources using a digital sound file with a digital to analog output board in series with amplifiers and speakers. An array of microphones collects the speaker outputs. The input signals are first amplified and then sent to an analog to digital data acquisition board. The board records the data from the microphone channels using LabView 6.1 software.

5.2.1 Microphones

Eight omnidirectional ICP® microphone/preamp modules, model number 130D10 array microphone with 130P series amplifier, are used to create the microphone array. A constant current power supply of two to 20 mA is required to power the modules while creating a 45 mV/Pa sensitivity where one Pa is equivalent to 94 dB.

The microphones have approximately a flat frequency response from 100 to 7,000 Hz. The microphone/preamp modules are linearly spaced in the array at 2.5 cm with the diameter of each microphone head at 6.99 mm.

5.2.2 Input card

BNC to SMB cables are used to connect the microphone/preamp modules to the input card. The National Instruments (NI) 4472 for PCI is used as the input card and is integrated into a Pentium IV processor OmniTech PC. The NI 4472 is an eight channel dynamic signal acquisition module for PCI. This input card supplies a 4 mA constant ICP current supply, which is necessary to power the microphone/preamp modules. An example of the microphone signals is shown in Figure 15.

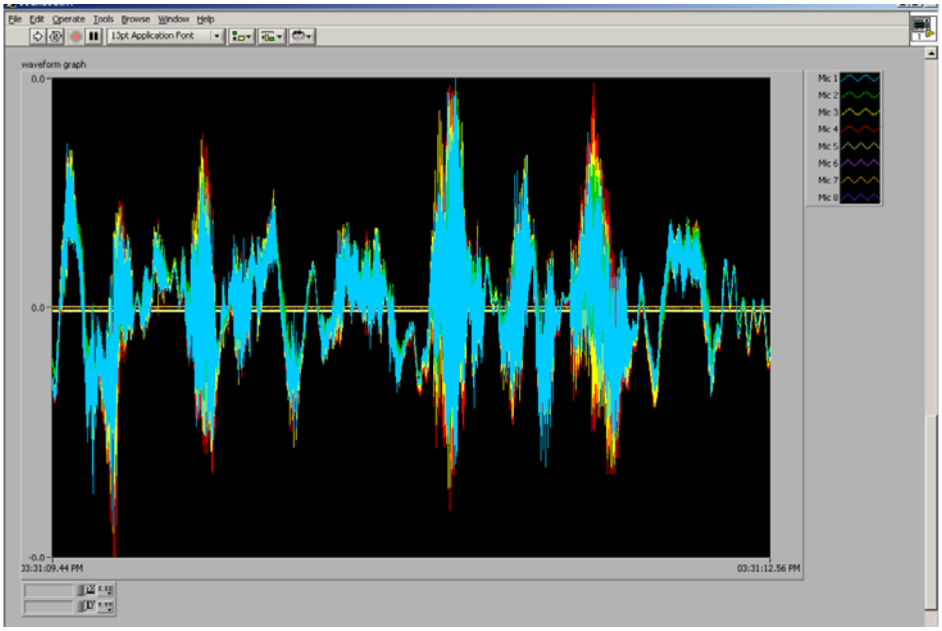


Figure 15: Microphone data from data acquisition system

5.3 Sound Booth Setup

Experiments were performed within an acoustically treated sound booth that is approximately 7.5 feet by 7.5 feet by 7 feet in height. The sound booth has 4 inch insulation and is also equipped with 2 inch anti-reverberation Sonex Classic Polyurethane Acoustical Foam wall treatments. The setup of the room was shown in Figure 13.

Chapter 6 Experimental Results

The majority of the results from this research are produced from the simulated experiments. The sound booth experiments were used as a demonstration of the enhancement algorithms. SNR and segmental SNR for both the simulated and sound booth experiments are presented here as performance metrics for the enhancement algorithms. To verify that the algorithms used enhance the simulated multiple source signal, listener tests to evaluate the signal quality before and after analysis are performed using the mean opinion score test (MOS). In this test, the listener ranks the quality of the speech on a scale of one to five with one being least favorable and five being the best. These tests entailed 15 people that listened to the multiple source signals with and without enhancements.

The specific signals tested with the listeners were a slice of each of the simulated three source experiments. Of the five experiments, only one geometry of one interferer at $\pi/3$ and another at $-\pi/3$ and at one magnitude level was presented to the listeners. The magnitude level of the interferers corresponds to the fourth data point in the SNR and sSNR plots. This slice of data resulted in seven signals per experiment: the noisy signal, the DS beamformed signal, the multiple source spectral subtraction using the DS beamformer, the multiple source Wiener filter using the DS beamformer, the MVDR beamformed signal, the multiple source spectral subtraction using the MVDR beamformer, and the multiple source Wiener filter using the MVDR beamformer. This totals to 35 specific signals that were randomly presented to the listeners.

Once the testing was completed, each listener's scores were tabulated and compared against other listener's scores. An overall quality score was then calculated and tabulated as shown in Table 5. The beamformed signal was used as a basis to score the algorithms. This yields a change in improvement rather than simply comparing the raw scores assigned to the enhanced signals. It also helps to normalize the listeners relative to one another.

When reviewing the MOS testing forms, it was evident that one listener did not score the sentences in the correct order. This listener had turned in the test sheets shuffled and as a result, these test scores are not incorporated in the analysis or in Table 5. The test forms used in the analysis are found in Appendix B.

	Noisy	DS	DS with Spectral Subtraction	DS with Wiener Filter	MVDR	MVDR with Spectral Subtraction	MVDR with Wiener Filter
Experiment 1	1.29	2.14	2.64	2.14	2.14	2.50	2.14
Experiment 2	2.50	3.71	4.43	4.00	3.29	4.07	4.07
Experiment 3	1.5	2.71	3.29	3.93	2.57	3.29	2.50
Experiment 4	2.21	3.57	4.21	3.71	3.79	4.14	3.86
Experiment 5	1.93	3.50	3.86	3.64	3.64	4.29	3.86

Table 4: MOS test results

	DS with Spectral Subtraction – DS alone	DS with Wiener Filter – DS alone	MVDR with Spectral Subtraction – MVDR alone	MVDR with Wiener Filter – MVDR alone
Experiment 1	0.500	0.000	0.357	0.000
Experiment 2	0.714	0.286	0.786	0.786
Experiment 3	0.571	1.214	0.714	-0.070
Experiment 4	0.643	0.143	0.357	0.071
Experiment 5	0.357	0.143	0.643	0.214
Overall Average	0.557±0.14	0.357±0.49	0.571±0.20	0.200±0.34

Table 5: MOS test results of average improvement

From the MOS scores, the multiple source spectral enhancement using the MVDR beamformer as a basis received the most improved scores while the same algorithm with the DS beamformer as a basis received the second highest improvement. The multiple source Wiener filter using the DS and using the MVDR also showed a net improvement, but with more variation in listener scores.

In addition to the MOS tests, comparisons of the enhanced primary source signal to the noise free primary source signal are performed with the quality measures of SNR and segmental SNR. With the simulated experiments, the clean reference signal is obtained by using the primary source signal TIMIT file as the clean signal. With the sound booth experiments, this is done through recording the primary signal alone. Recording the primary signal alone in the sound booth set up as opposed to analyzing the signal directly from the corpus compensates for channel effects imposed by the microphones.

The SNR and segmental SNR values for each simulated experiment and each algorithm are given in the plots in Appendix A. The plots show each enhancement algorithm with the corresponding beamformer results so that a comparison of how much more enhancement produced above the beamformer can be addressed. Examples of the data results from experiment 2 for both two and three speakers are shown in Figures 16 through 19. The overall average improvement for the algorithms across all experiments is tabulated in Tables 6 through 9.

	DS	DS with Spectral Subtraction	DS with Wiener Filter	MVDR	MVDR with Spectral Subtraction	MVDR with Wiener Filter
Experiment 1	2.58	6.28	5.68	9.21	10.96	10.56
Experiment 2	1.89	5.78	4.65	5.96	7.49	7.39
Experiment 3	1.13	3.79	3.48	7.18	9.32	9.10
Experiment 4	2.11	5.48	4.61	8.14	10.04	9.87
Experiment 5	2.38	5.36	4.35	5.06	6.38	6.24
Overall Average	2.02 ±0.56	5.34 ±0.94	4.55 ±0.79	7.11 ±1.66	8.84 ±1.87	8.63 ±1.78

Table 6: Average SNR improvements for two source experiments

	DS	DS with Spectral Subtraction	DS with Wiener Filter	MVDR	MVDR with Spectral Subtraction	MVDR with Wiener Filter
Experiment 1	2.45	4.59	4.07	6.60	6.87	5.04
Experiment 2	3.33	5.69	4.83	4.24	5.12	5.00
Experiment 3	1.77	3.05	2.94	3.89	4.89	4.76
Experiment 4	3.5	5.19	4.77	4.33	5.49	5.29
Experiment 5	3.14	5.39	4.67	4.80	5.63	5.45
Overall Average	2.84 ±0.72	4.78 ±1.05	4.26 ±0.80	4.77 ±1.07	5.60 ±0.77	5.10 ±0.27

Table 7: Average segmental SNR improvements for two source experiments

	DS	DS with Spectral Subtraction	DS with Wiener Filter	MVDR	MVDR with Spectral Subtraction	MVDR with Wiener Filter
Experiment 1	-0.39	2.50	1.34	4.62	7.23	6.72
Experiment 2	0.79	2.64	1.72	4.47	5.90	5.58
Experiment 3	-0.10	0.86	0.29	2.18	5.89	4.97
Experiment 4	-0.74	0.70	0.63	3.28	5.50	5.30
Experiment 5	-0.05	2.01	1.03	3.78	6.04	5.41
Overall Average	-0.10 ±0.57	1.74 ±0.91	1.00 ±0.56	3.67 ±0.99	6.11 ±0.65	5.60 ±0.67

Table 8: Average SNR improvements for three source experiments

	DS	DS with Spectral Subtraction	DS with Wiener Filter	MVDR	MVDR with Spectral Subtraction	MVDR with Wiener Filter
Experiment 1	0.39	1.97	0.98	2.80	4.96	4.49
Experiment 2	1.28	2.78	1.82	3.92	5.23	4.84
Experiment 3	0.60	1.24	0.58	1.54	3.14	2.70
Experiment 4	0.81	1.78	1.21	2.86	4.50	4.16
Experiment 5	0.44	2.15	1.01	3.96	4.65	2.53
Overall Average	0.70 ±0.36	1.98 ±0.56	1.12 ±0.45	3.02 ±0.99	4.50 ±0.81	3.74 ±1.06

Table 9: Average segmental SNR improvements for three source experiments

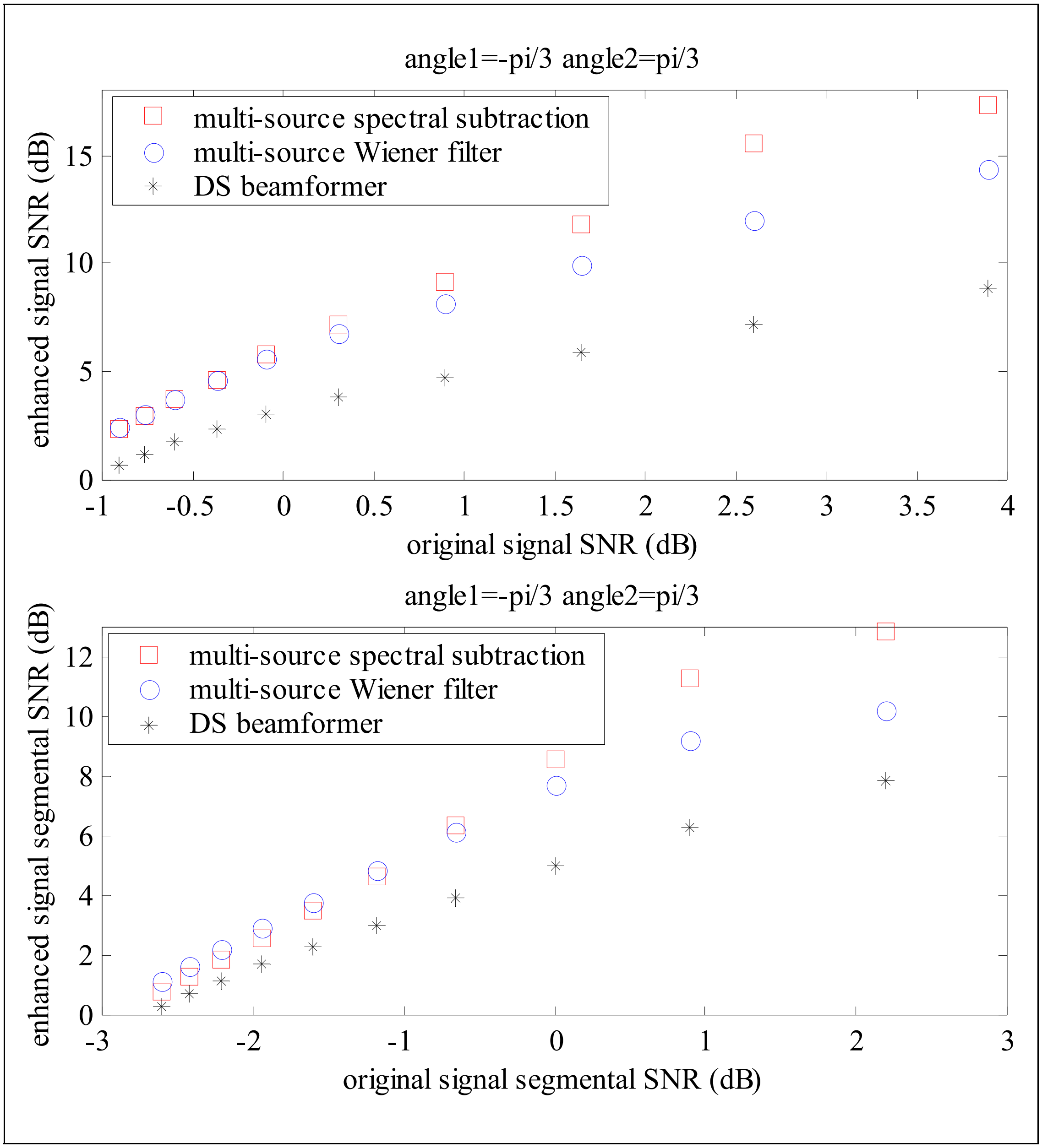


Figure 16: Results example of DS based enhancement algorithms for two sources in experiment 2

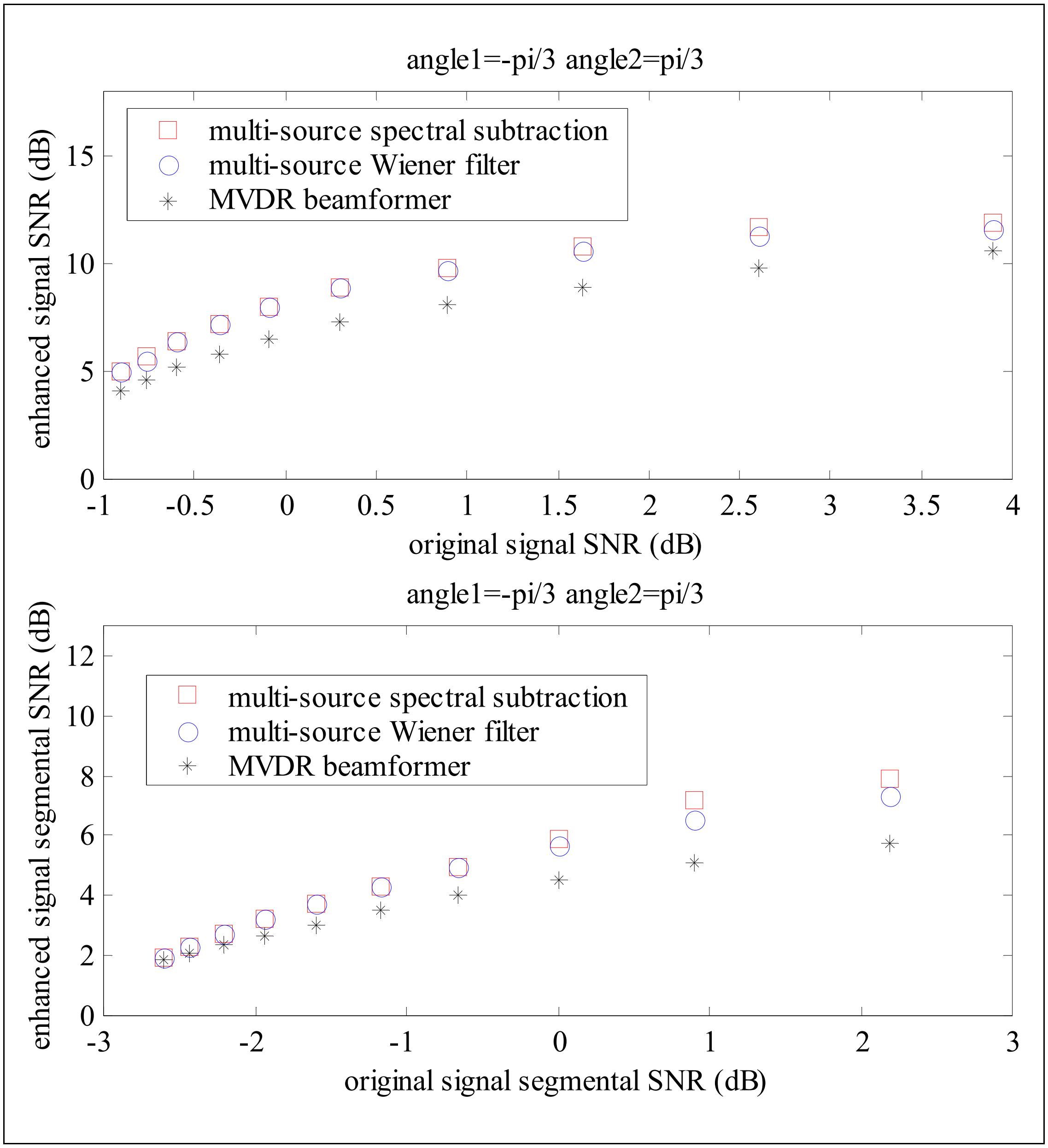


Figure 17: Experiment 2 MVDR based enhancement algorithms for two sources

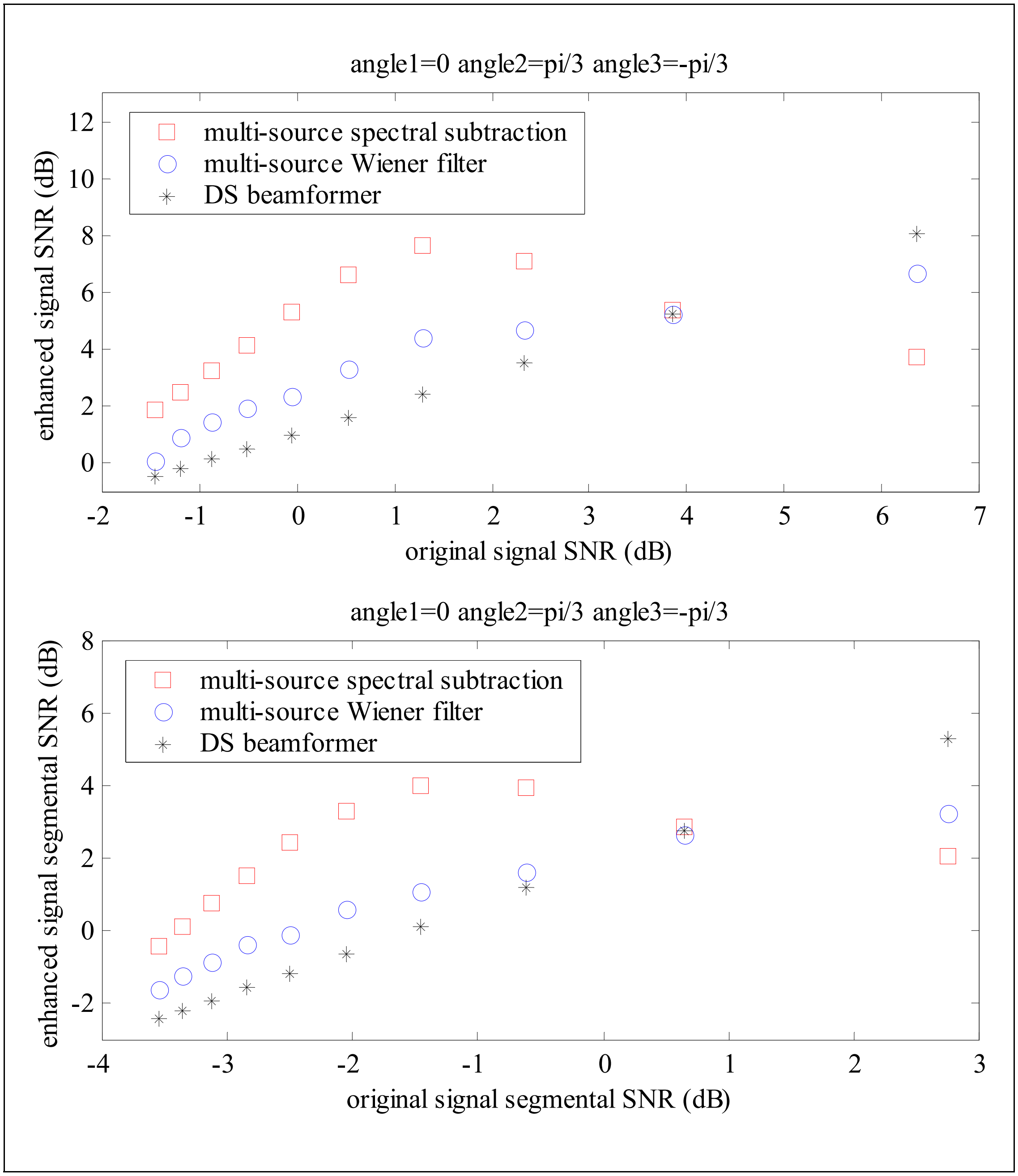


Figure 18: Experiment 2 DS based enhancement algorithms for three sources

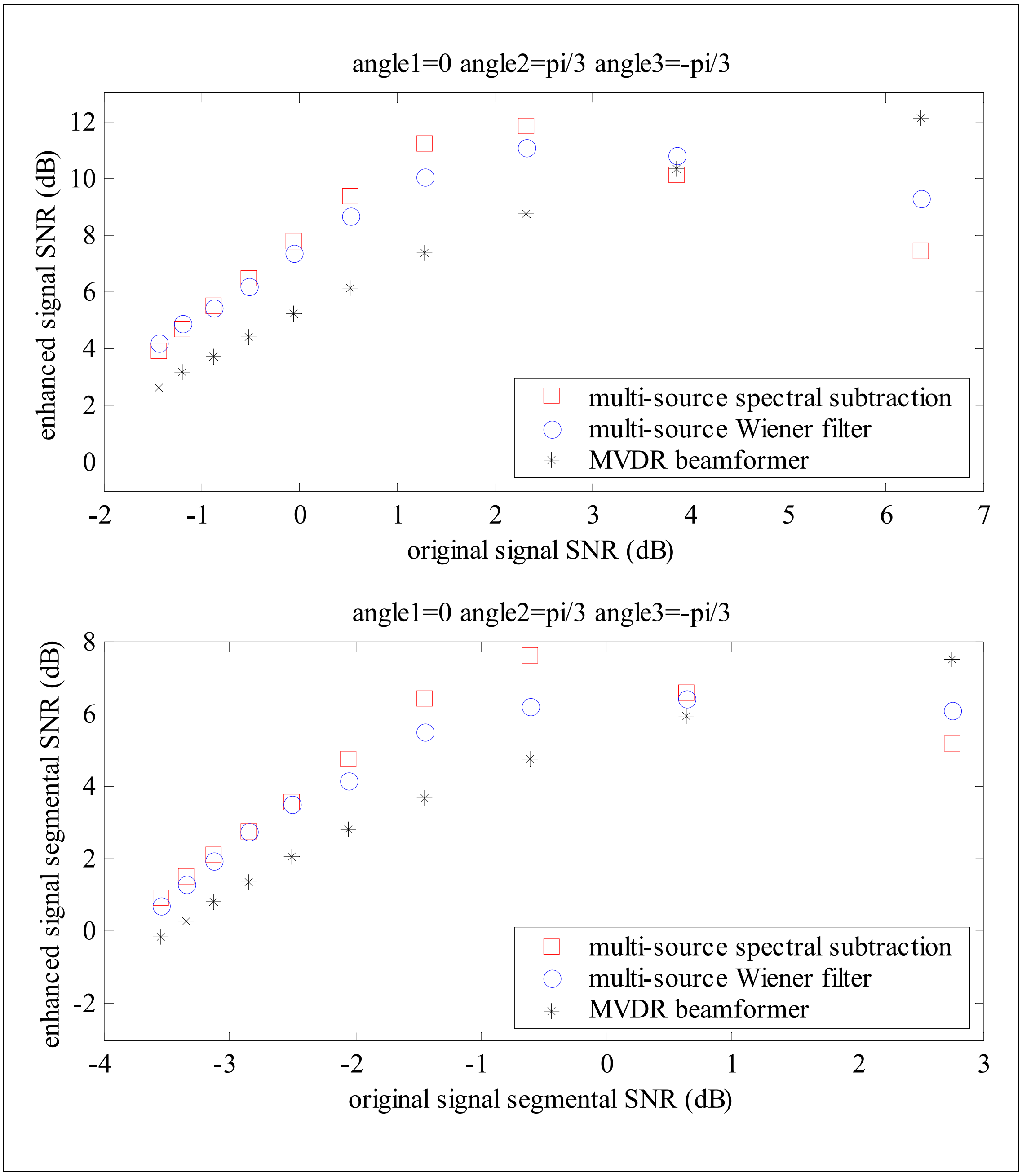


Figure 19: Experiment 2 MVDR based enhancement algorithms for three sources

After reviewing the SNR and sSNR data, overall, the best improvements were found for the experiments where the separation between sources was the greatest and for a limited SNR and sSNR range. This maximum improvement range varied from experiment to experiment but remained generally the same range across geometries for a specific experiment.

The results of the sound booth experiment are shown in Figure 20 and Figure 21. It is shown that there is an improvement in SNR and segmental SNR for four of the five data points associated with the MVDR based enhancement algorithms and an improvement in all the data points for the DS based ones. The DS based algorithms also had a larger increase in SNR and segmental SNR.

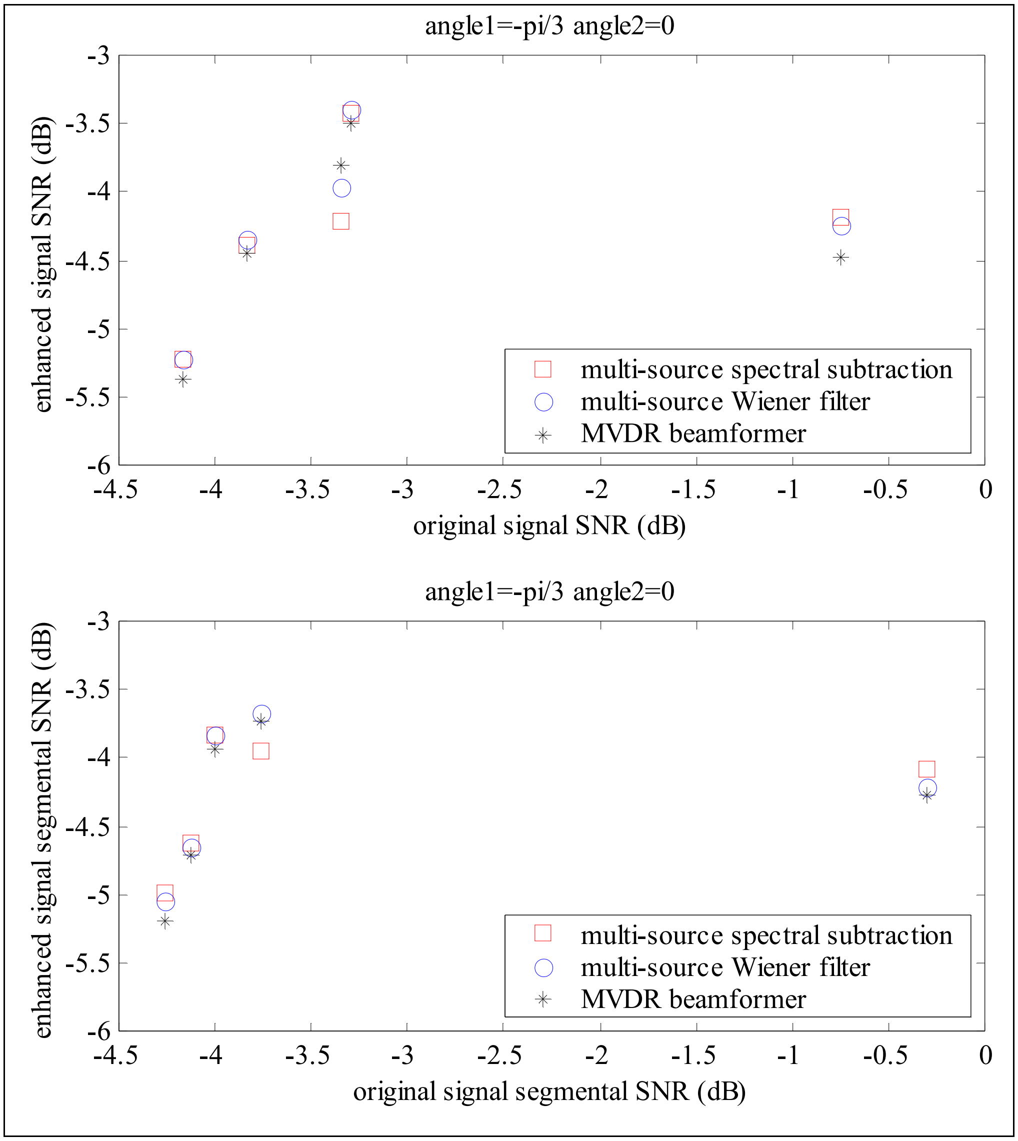


Figure 20: Sound booth experiment results for MVDR based enhancement algorithms

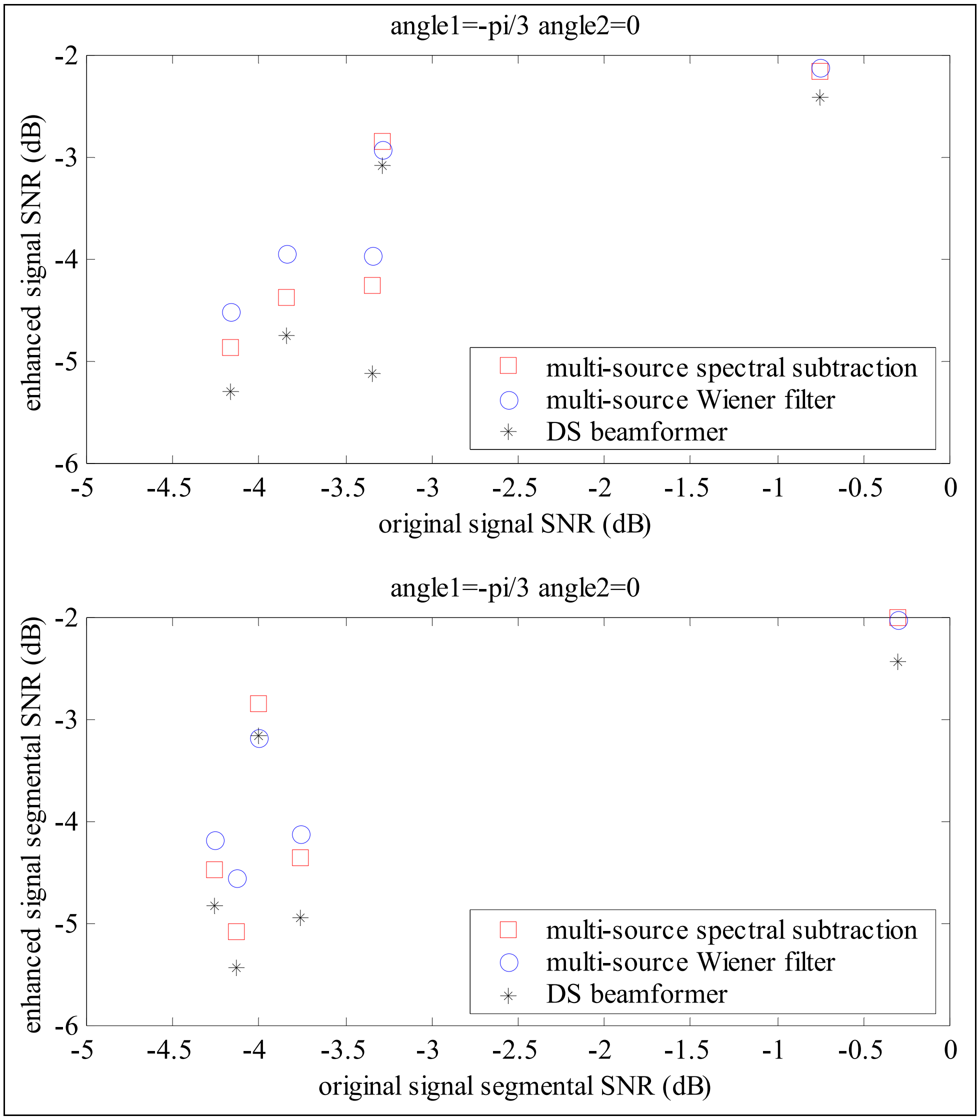


Figure 21: Sound booth experiment results for DS based enhancement algorithms

Chapter 7 Discussion

From the results of the simulated experimental SNR and sSNR data, it can be seen that the multiple source enhancement algorithms worked most efficiently for a specific SNR and sSNR range. As shown in the results, there is significant overall improvement by the enhancement algorithms above the beamformer alone for both the two source and three source experiments and across all geometries. However, it is important to note that the multiple source enhancement algorithms perform best for an intermediate range of SNR and sSNR's. The amount of this improvement decreases when the interfering source power is minimal, as the SNR and segmental SNR increases. This is because the algorithm is dependent upon the interfering talkers' signal estimation. When the interfering talkers' signal power increases, the parallel beamformers are able to obtain better noise signal estimates. However, as the interfering signals' power become much larger, as is true for extremely small SNR's, the algorithm does not perform well due to the inability of the beamformer to obtain an adequate estimate of the primary signal. These characteristics are summarized in Figure 22 and reflected in the results in Appendix A.

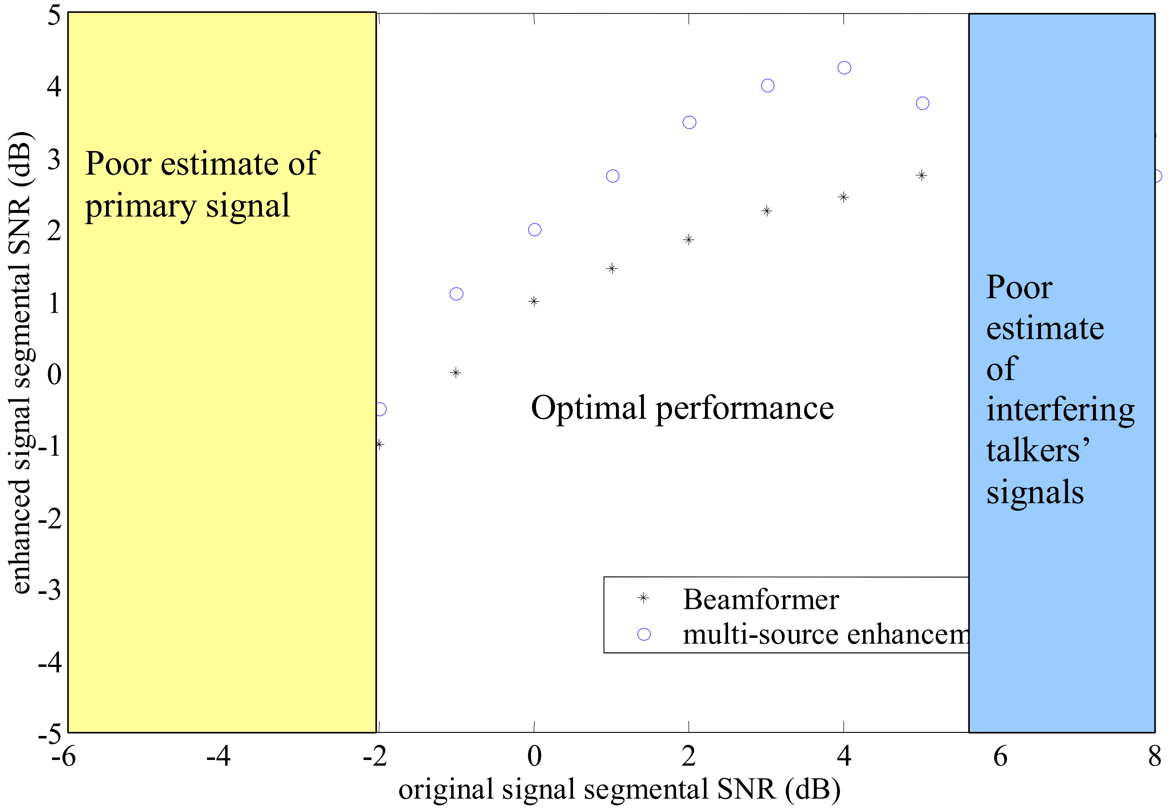


Figure 22: Illustration example of optimal performance sSNR range

The MOS results show that the multiple source algorithms increase signal quality in a multiple speaker environment. The importance of the results is that the listeners perceived an increase in signal quality for all of the algorithms evaluated. In general, the data points used in the MOS testing were within the range that the SNR and sSNR results showed the algorithms performed well. The multiple source Wiener filter algorithm showed less of an improvement compared with the spectral subtraction algorithms.

Although the Wiener filter suppresses the noise spectral characteristics, it also introduces "musical" tones or distortions into the signal. These tones may have been the reason the listeners had more variation in scores and did not feel the quality of the signal increased as much as the multiple source spectral subtraction methods. "Musical" tones are also present in the spectral subtraction method due to the spectral smoothing, but in the results presented here, they are not as noticeable compared with the Wiener filter output signals.

The sound booth experiment is used as a demonstration of the enhancement algorithms' capabilities. In Figure 20 and Figure 21, it is shown that there is a slight overall improvement in both the SNR and sSNR by the multiple source enhancement algorithms above the beamformer alone. Because some of the beamformer data points do not achieve an overall improved SNR and segmental SNR, the results also show that the beamformers can be further optimized for the given microphone array.

Chapter 8 Conclusion

The use of multiple, parallel beamformers integrated with a multiple source Wiener filter and multiple source spectral subtraction algorithm, as presented in this research, shows substantial improvement in the SNR and segmental SNR for a range consistent with a multiple interfering speaker environment. In the multiple speaker environment, interfering talkers will have enough power to generate an acceptable beamformed signal estimate for the enhancement algorithms. As a result, the speech enhancement methods presented in this paper are able to contend with nonstationary, broadband noise that occurs in a multiple speaker environment.

Further work is being conducted to research the multiple speaker environment. In particular, the integration of other traditional enhancement methods into the multiple source domain using parallel beamformers and the utilization of a more powerful beamformer as the basis of the algorithms are being researched. To create a more robust algorithm, the use of *a priori* knowledge of the source directions would be eliminated with the integration of a source location algorithm like the Root-MUSIC algorithm. Additionally, reverberation signal adaptations could be incorporated into the iterative analysis of the algorithms, creating the ability to work well in an acoustically imperfect room, as is true for most “real world” rooms. More experimental runs using the sound booth and a conference room can be performed to test these improvements.

References

Berouti, M. G., R. Schwartz, and J. Makhoul; "Enhancement of speech corrupted by acoustic noise;" IEEE Transactions on Acoustics, Speech, and Signal Processing, pp. 208-211, 1979.

Boll, S. F.; "Suppression of acoustic noise in speech using spectral subtraction;" IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 27, pp. 113-120, April 1979.

Bitzer, Joerg, K. U. Simmer, and K. D. Kammeyer; "Multi-microphone noise reduction by post-filter and superdirective beamformer;" Proceedings of the International Workshop on Acoustic Echo and Noise Control, pp. 100-103, September 1999.

Bitzer, Joerg, Klaus Uwe Simmer, Karl-Dirk Kammeyer; "Multi-microphone noise reduction techniques as front-end devices for speech recognition;" Speech Communication, vol. 34, pp. 3-12, 2001.

Brandsein, M., J. E. Adcock, and H. F. Silverman; "A practical time-delay estimator for localizing speech sources with a microphone array;" Computer Speech and Language, vol. 9, pp. 153-169, 1995.

Brandstein, M. and D. Ward, eds.; Microphone Arrays: Signal Processing Techniques and Applications; Springer, New York: 2001.

Carnero, Benito and Andrzej Drygajlo; "Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms;" IEEE Transactions on Signal Processing, vol. 47, pp. 1622-1635, 1999.

Deller, J.R., J. Hansen, and J. Proakis; Discrete-Time Processing of Speech Signals; IEEE Press, New York: 2000.

Dundgeon, Dan E. and Don H. Johnson; Array Signal Processing: Concepts and Techniques; PRT Prentice Hall, Englewood Cliffs, NJ: 1993.

Elledge, Mark, et al; "Real-time implementation of a frequency-domain beamformer on the TI C62X EVM;" Texas instruments DSPSFest 2000.

Friedlander, B. and A. J.Weiss; "Direction finding for wide-band signals using an interpolated array;" IEEE Trans. Signal Processing, vol. 41, pp. 1618–1634, April 1993.

Frost III, O.L.; "An algorithm for linearly constrained adaptive array processing;" Proceedings of the IEEE, vol. 60, pp. 916-935, 1972.

Garofolo, J., L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue; "TIMIT Acoustic-Phonetic Continuous Speech Corpus;" Linguistic Data Consortium, 1993.

Hansen, J. H. L.; and M. A. Clements; "Iterative speech enhancement with spectral constraints;" IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 1, pp. 189-192, 1987.

Hu, H. T., F. J. Kuo, and H. J. Wang; "Supplementary schemes to spectral subtraction for speech enhancement;" Speech Communication, vol. 36, pp. 205-218, 2002.

Itakura; F.; "Minimum prediction residual applied to speech recognition;" IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 23, pp. 67-72, 1975.

Kajita, Shoji, Kazuya Takeda, and Fumitada Itakura; "Subband-crosscorrelation analysis for robust speech recognition;" International Conference on Spoken Language Processing, ICSLP, Proceedings, vol. 1, pp. 422-425, 1996.

Kellermann, W.; "Analysis and design of multirate systems for canceling of acoustic echoes;" Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 2570-2573, 1988.

Lim, J. S. and A. V. Oppenheim; "All-pole modeling of degraded speech;" IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 26, pp. 197-210, June 1978.

Marro, C., Y.Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering;" IEEE Transactions on Speech and Audio Processing, vol. 6, pp. 240-259, May 1998.

McCowan, I.A., C. Marro, and L. Mauuary; "Robust speech recognition using near-field superdirective beamforming with post-filtering;" Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1723-1726, 2000.

McCowan, I.A. and S. Sridharan; "Microphone array sub-band speech recognition;" Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 185-188, 2001.

Omologo, M., et al; "Microphone array based speech recognition with different talker-array positions;" Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 227-230, 1997.

Plomp, R.; “A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired;” Journal of Speech and Hearing Research, vol. 29, pp. 146-154, 1986.

Quackenbush, S.R., T. P. Barnwell, and M. A. Clements; Objective Measures of Speech Quality; Englewood Cliffs, J.J.: Prentice Hall, 1998.

Rabinkin, Daniel, et al.; “A DSP implementation of source location using microphone arrays;” Proceedings of the SPIE, vol. 2846, pp. 88-99, Denver, Colorado, August 1996.

Rabinkin, Daniel, R. Renomeron, J. French, and J. Flanagan; "Optimum microphone placement for array sound capture;" Proceedings of SPIE, vol. 3162, pp. 227-239, 1997.

Rao, Bhaskar D. Hari, K V S.; “Performance analysis of root-music;” Asilomar Conference on Circuits, Systems & Computers, vol. 2, pp. 578-582, 1985.

Ris, Christophe and Stephane Dupont; “Assessing local noise level estimation methods: Application to noise robust ASR;” Speech Communication, vol. 34, pp. 141-158, 2001.

Ryan, James G. and R. Goubran; “Near-field beamforming for microphone arrays;” Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 363-366, 1997.

Saruwatari, H., et al; “Speech enhancement using nonlinear microphone array with noise adaptive complementary beamforming;” Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp.1049-1052, 2000.

Steinberg, B.D.; Principles of Apertre and Array System Design, John Wiley and Sons, 1976.

Svaizer, Piergiorgio, Marco Matassoni, and Maurizio Omologo; “Acoustic source location in a three dimensional space using crosspower spectrum phase;” Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 231-234, 1997.

Tager, W.; “Near field superdirectivity (NFSD);” IEEE Transactions on Acoustics, Speech, and Signal Processing, pp. 2045-2048, 1998.

Wang, A., et al.; “Calibration, optimization, and DSP implementation of microphone array for speech processing;” IEEE, pp. 221-230, 1996.

Wang, D. L. and J. S. Lim; “The unimportance of phase in speech enhancement;” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, pp. 679-681, August 1982.

Wang, Fang-Ming, et al.; “Frequency domain adaptive postfiltering for enhancement of noisy speech;” Speech Communication; vol. 12, pp. 41-56, 1993.

Wang, H. and M. Kaveh; “Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources;” IEEE Transactions on Acoustics, Speech, and Signal Processing; vol. 33, pp. 823–831, Aug 1985.

Ward, D.B., Rodney A. Kennedy, and Robert C. Williamson; “Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns;” Journal of Acoustical Society of America; vol. 97, pp. 1023-1034, February 1995.

Weiss, S., M. Harteneck, and R. W. Stewart; “On implementation and design offFilter banks for subband adaptive systems;” IEEE Workshop on Signal Processing Systems, pp. 172-181, October 1998.

Widrow, Bernard; “A microphone array for hearing aids;” IEEE Circuits and Systems, vol. 1, no. 2, pp. 26-32, Second Quarter 2001.

Wu, Wen-Rong and Po-Cheng Chen; “Subband kalman filtering for speech enhancement;” IEEE Transactions on Circuits and Systems II-Analog and Digital Signal Processing; vol. 45, pp. 1072-1083, August 1998.

Appendix A: Simulated Data Experimental Results

Two source experiments

Experiment 1:

Primary speaker TIMIT ID = dr3 mkch0 SI2008

Second speaker TIMIT ID= dr3 fcmh0 SX194

Experiment 2:

Primary speaker TIMIT ID = dr3 fkms0 SX410.WAV

Second speaker TIMIT ID= dr3 mkch0 SX298.WAV

Experiment 3:

Primary speaker TIMIT ID = dr3 mctw0 SI2003.WAV

Second speaker TIMIT ID= dr3 mglb0 SI904.WAV

Experiment 4:

Primary speaker TIMIT ID = dr3 mjvw0 SX293.WAV

Second speaker TIMIT ID= dr3 fcmh0 SI1454.WAV

Experiment 5:

Primary speaker TIMIT ID = dr3 fcmh0 SX374.WAV

Second speaker TIMIT ID= dr3 mctw0 SI743.WAV

Three source experiments

Experiment 1:

Primary speaker TIMIT ID = dr3 mkch0 SI2008

Second speaker TIMIT ID= dr3 fcmh0 SX194

Third speaker TIMIT ID= dr3 mglb0 SA2

Experiment 2:

Primary speaker TIMIT ID = dr3 fkms0 SX410.WAV

Second speaker TIMIT ID= dr3 mkch0 SX298.WAV

Third speaker TIMIT ID= dr3 mctw0 SI743.WAV

Experiment 3:

Primary speaker TIMIT ID = dr3 mctw0 SI2003.WAV

Second speaker TIMIT ID= dr3 mglb0 SI904.WAV

Third speaker TIMIT ID= dr3 mjvw0 SX383.WAV

Experiment 4:

Primary speaker TIMIT ID = dr3 mjvw0 SX293.WAV

Second speaker TIMIT ID= dr3 fcmh0 SI1454.WAV

Third speaker TIMIT ID= dr3 mglb0 SI1534.WAV

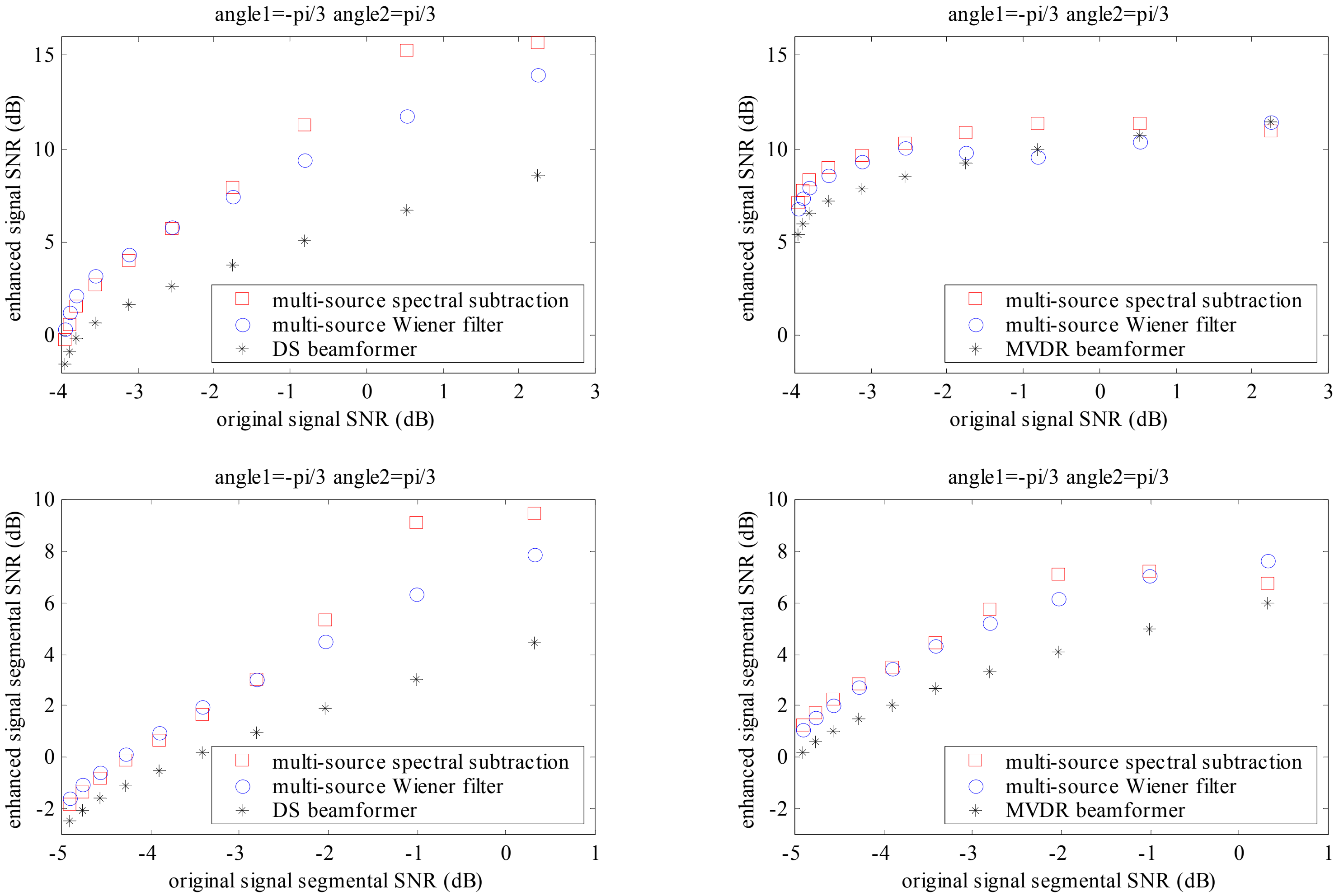
Experiment 5:

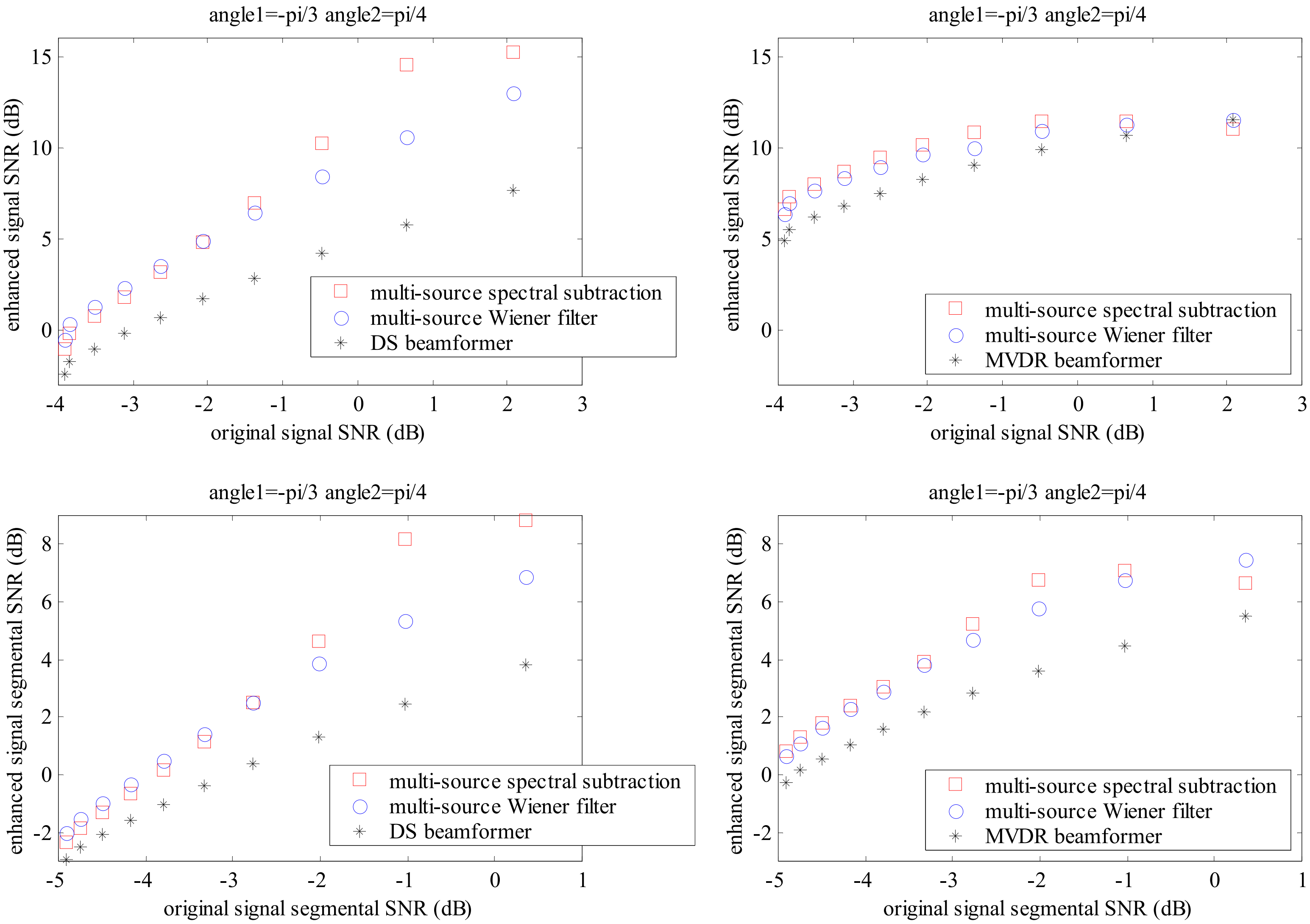
Primary speaker TIMIT ID = dr3 fcmh0 SX374.WAV

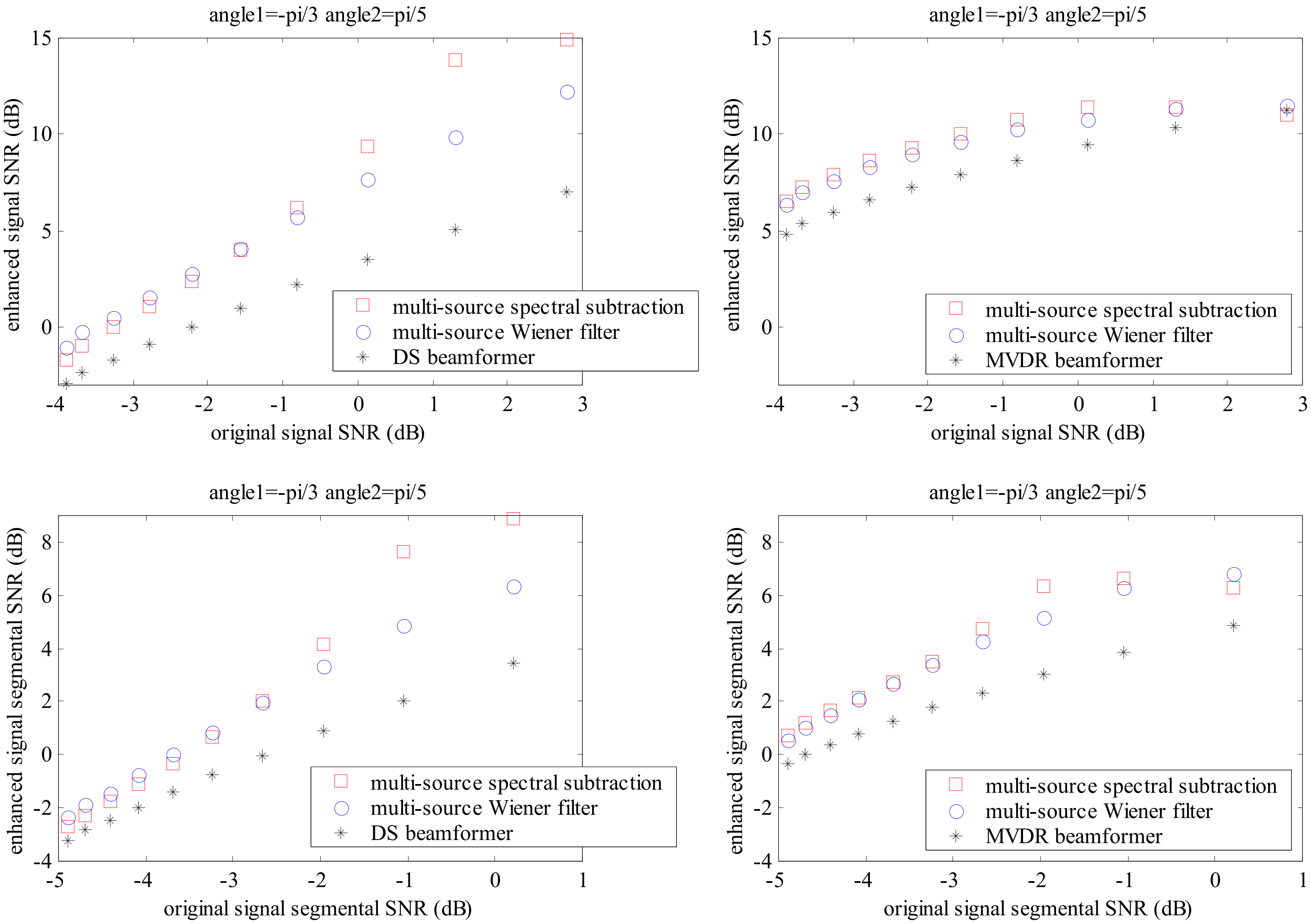
Second speaker TIMIT ID= dr3 mctw0 SI743.WAV

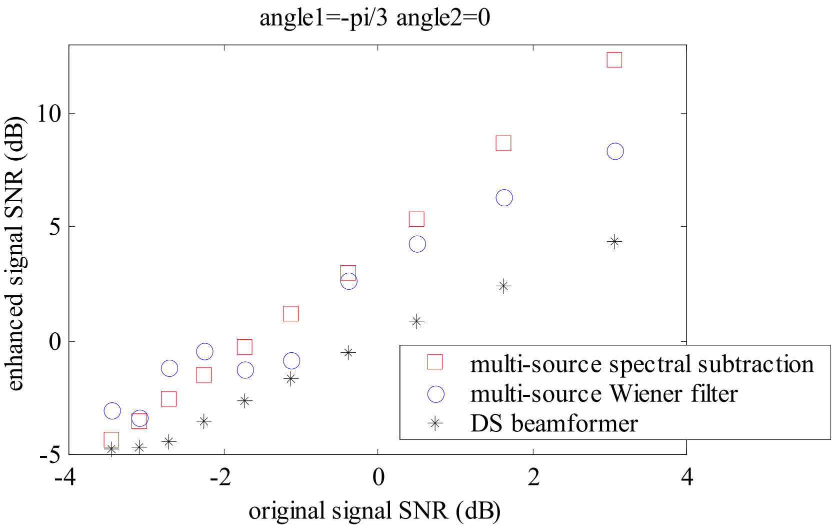
Third speaker TIMIT ID= dr3 mjvw0 SA2.WAV

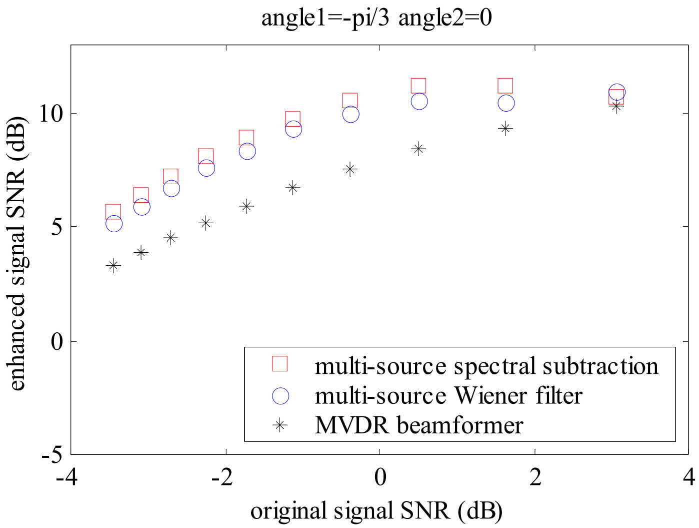
Two source experiments: Experiment 1:

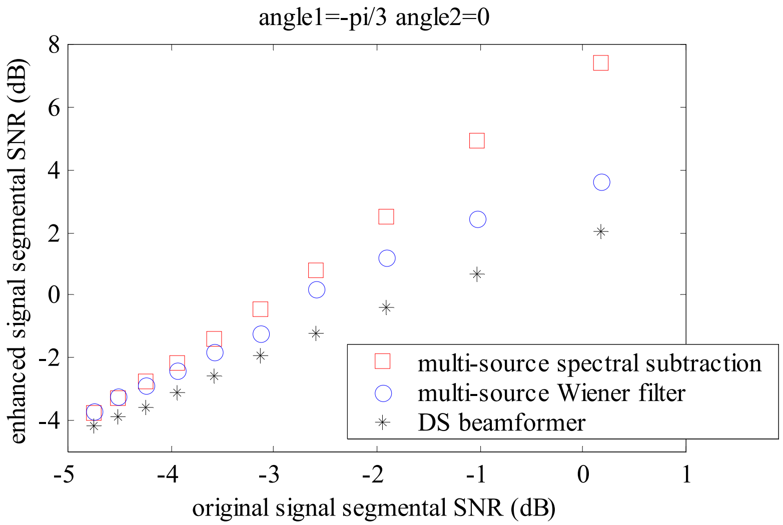


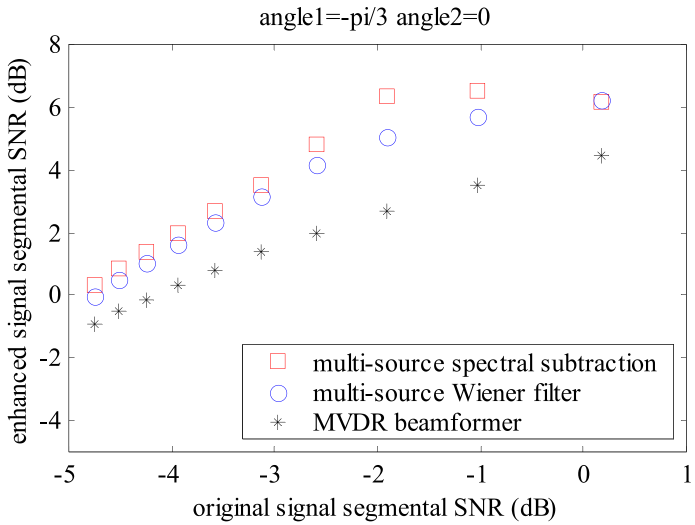

angle1=-pi/3 angle2=pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
15
10
5
0
-4
-3
-2
-1
0
1
2
3
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
15
10
5
0
-4
-3
-2
-1
0
1
2
3
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
8
6
4
2
0
-2
-5
-4
-3
-2
-1
0
1
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
8
6
4
2
0
-2
-5
-4
-3
-2
-1
0
1
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer


angle1=-pi/3 angle2=pi/5
enhanced signal SNR (dB)
15
10
5
0
-4
-3
-2
-1
0
1
2
3
original signal SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal SNR (dB)
15
10
5
0
-4
-3
-2
-1
0
1
2
3
original signal SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal segmental SNR (dB)
8
6
4
2
0
-2
-4
-5
-4
-3
-2
-1
0
1
original signal segmental SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal segmental SNR (dB)
8
6
4
2
0
-2
-4
-5
-4
-3
-2
-1
0
1
original signal segmental SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer

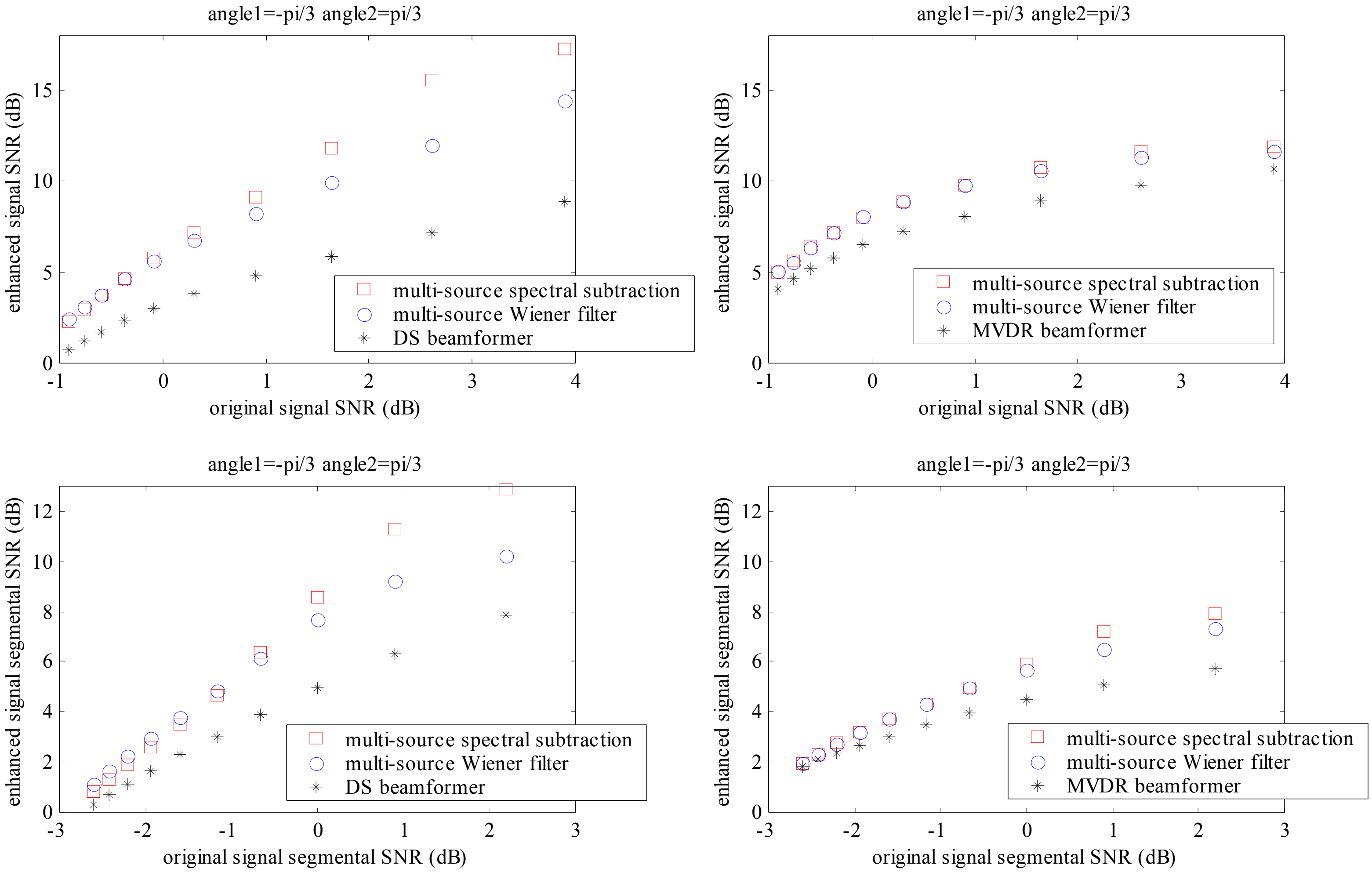

angle1=-pi/3 angle2=0
enhanced signal SNR (dB)
original signal SNR (dB)
10
5
0
-5
-4
-2
0
2
4
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer

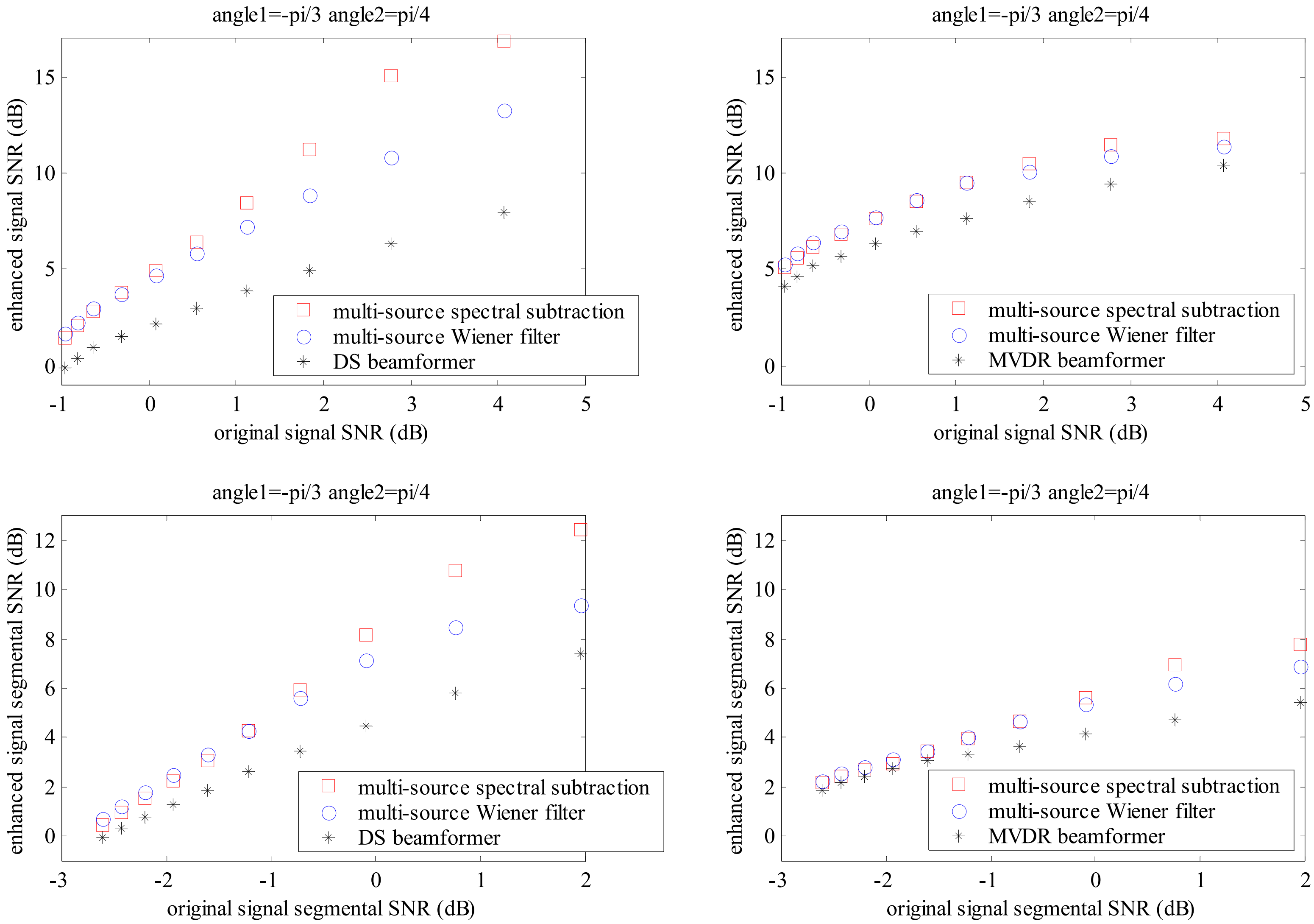

angle1=-pi/3 angle2=0
enhanced signal SNR (dB)
original signal SNR (dB)
10
5
0
-5
-4
-2
0
2
4
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer

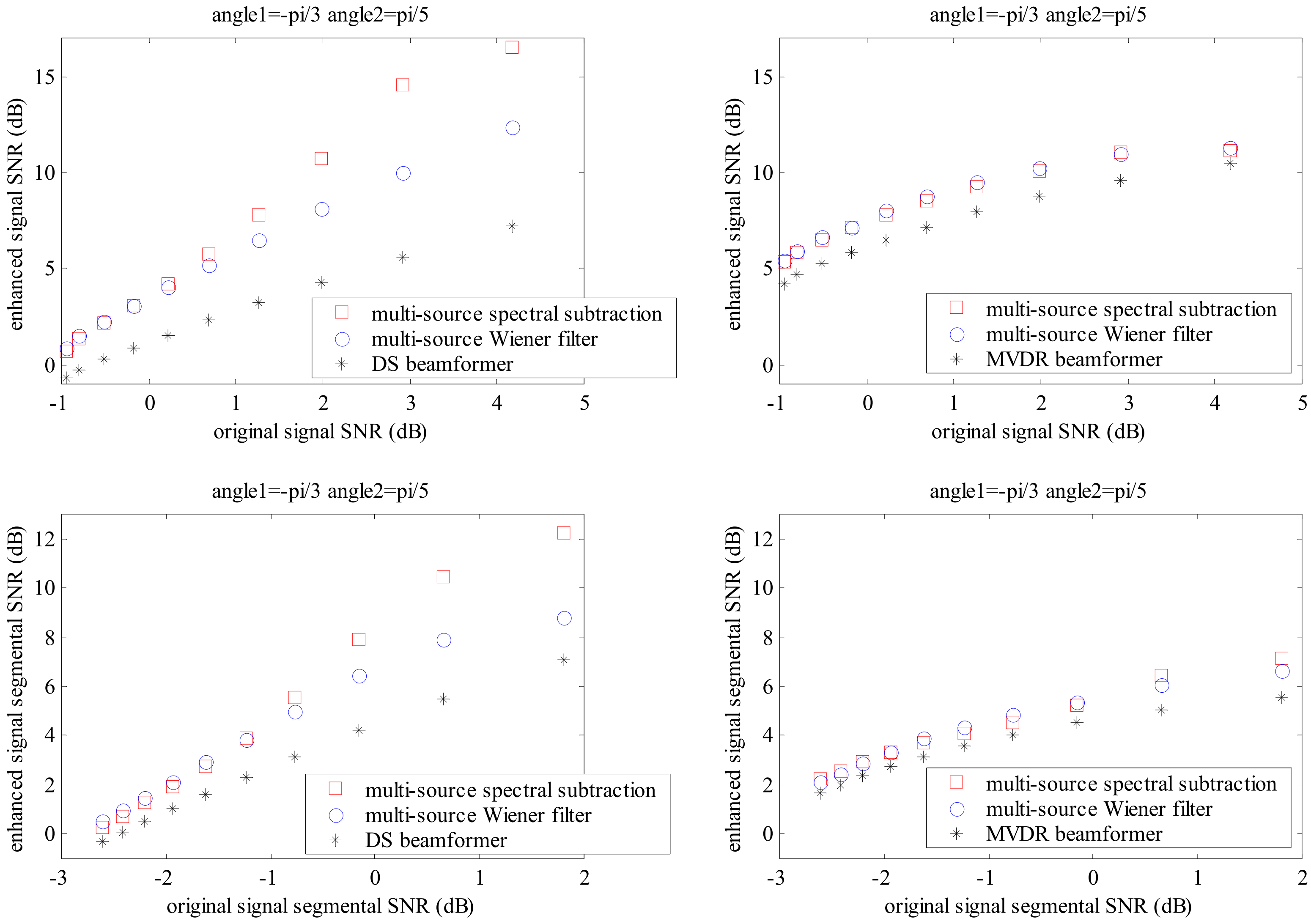

angle1=-pi/3 angle2=0
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
8
6
4
2
0
-2
-4
-5
-4
-3
-2
-1
0
1
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer

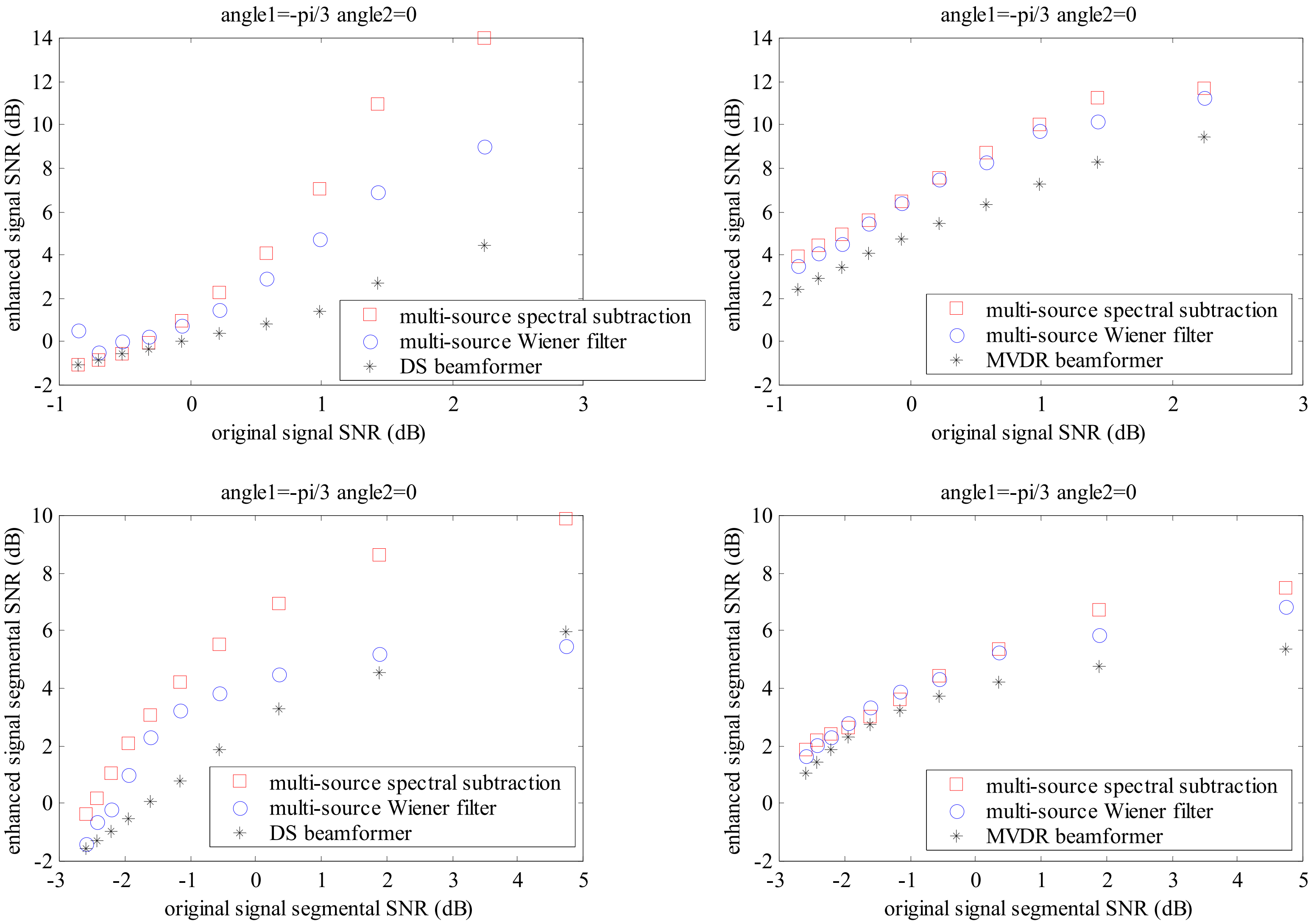

angle1=-pi/3 angle2=0
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
8
6
4
2
0
-2
-4
-5
-4
-3
-2
-1
0
1
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer

Experiment 2:

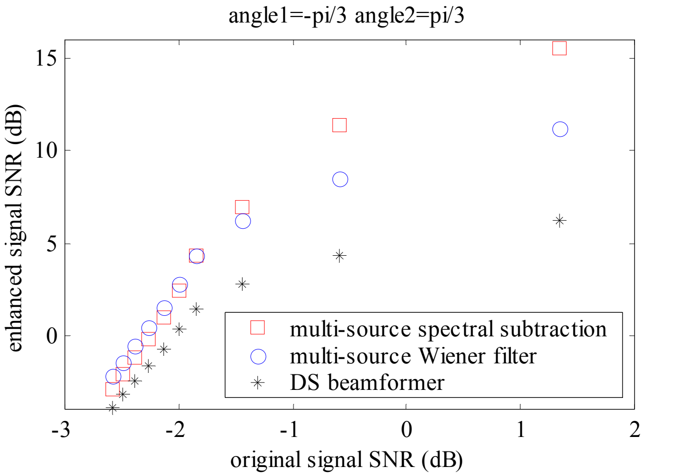


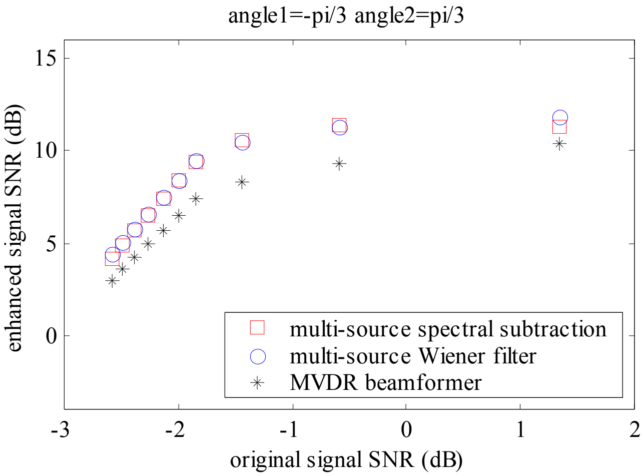

angle1=-pi/3 angle2=pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
0
5
10
15
-1
0
1
2
3
4
5
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
0
5
10
15
-1
0
1
2
3
4
5
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
0
2
4
6
8
10
12
-3
-2
-1
0
1
2
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
0
2
4
6
8
10
12
-3
-2
-1
0
1
2
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer

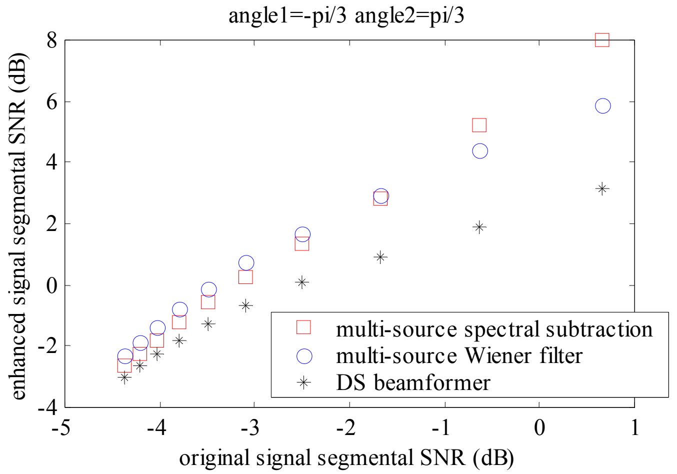

angle1=-pi/3 angle2=pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
0
5
10
15
-1
0
1
2
3
4
5
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
0
2
4
6
8
10
12
-3
-2
-1
0
1
2
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer

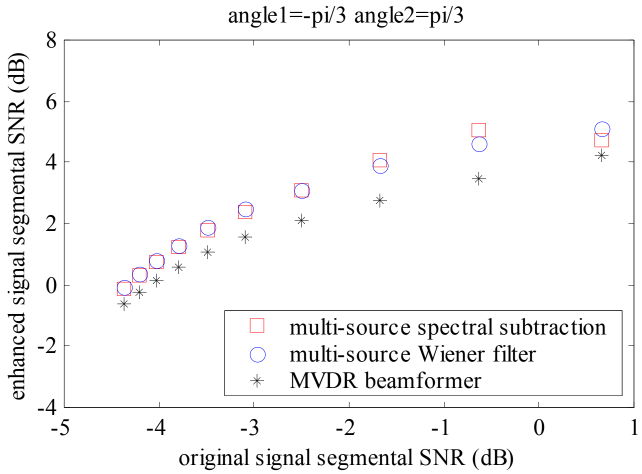

angle1=-pi/3 angle2=0
enhanced signal SNR (dB)
14
12
10
8
6
4
2
0
-2
-1
0
1
2
3
original signal SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
angle1=-pi/3 angle2=0
enhanced signal SNR (dB)
14
12
10
8
6
4
2
0
-2
-1
0
1
2
3
original signal SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer
angle1=-pi/3 angle2=0
enhanced signal segmental SNR (dB)
10
8
6
4
2
0
-2
-3
-2
-1
0
1
2
3
4
5
original signal segmental SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
angle1=-pi/3 angle2=0
enhanced signal segmental SNR (dB)
10
8
6
4
2
0
-2
-3
-2
-1
0
1
2
3
4
5
original signal segmental SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer

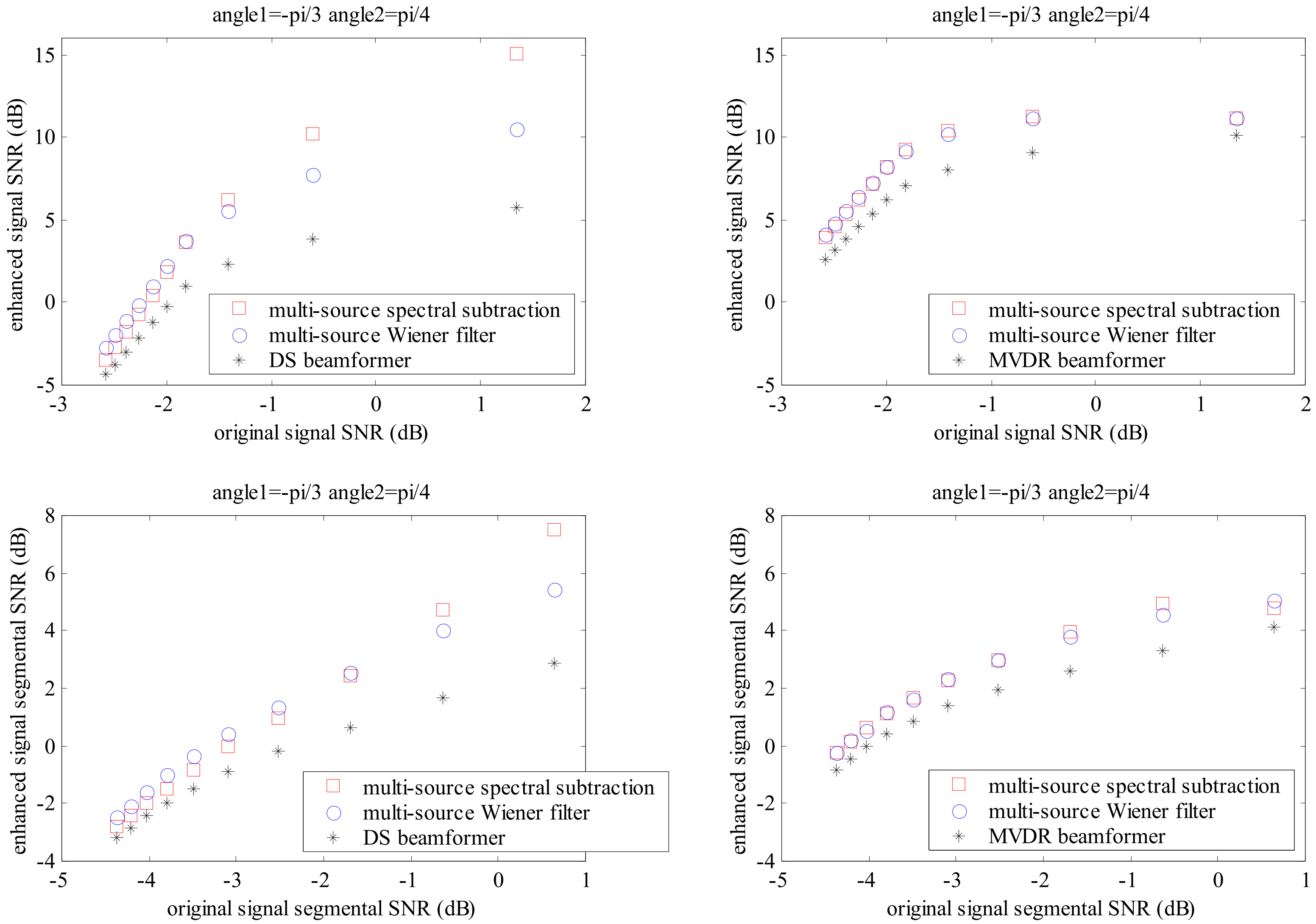
Experiment 3:

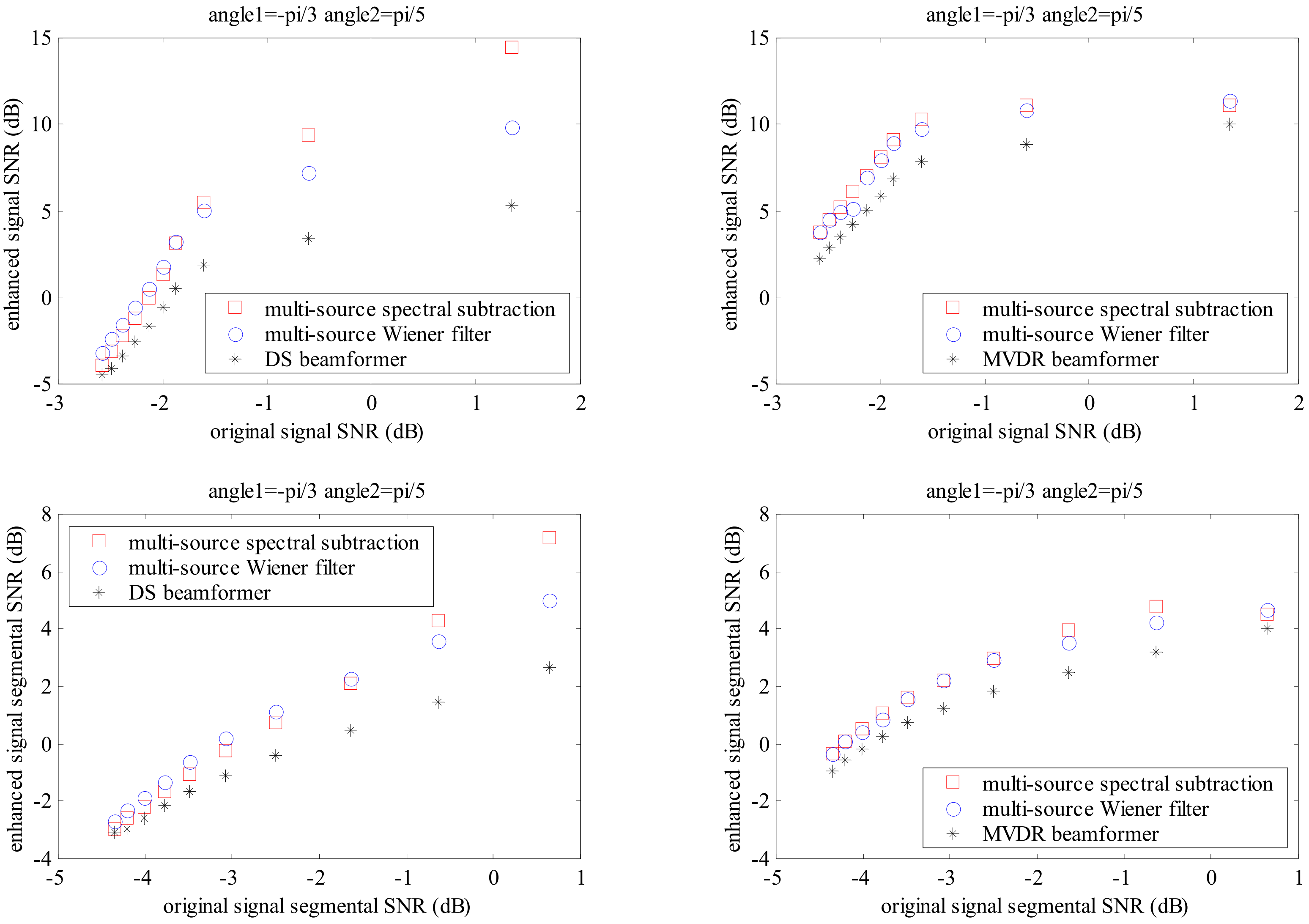

angle1=-pi/3 angle2=pi/3

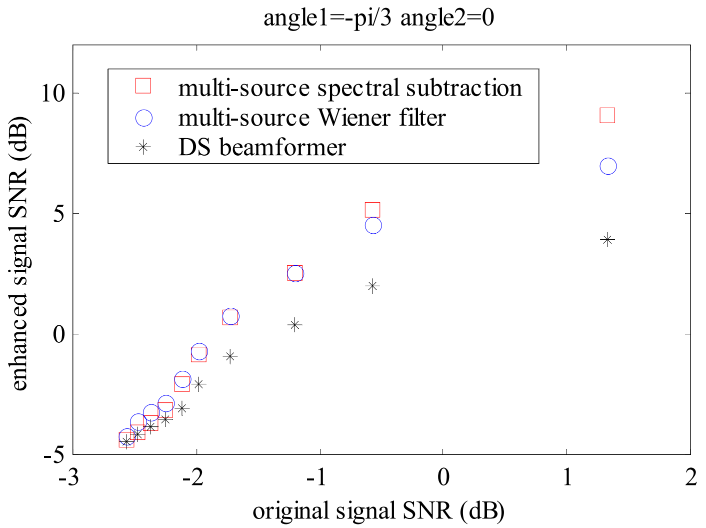

angle1=-pi/3 angle2=pi/3

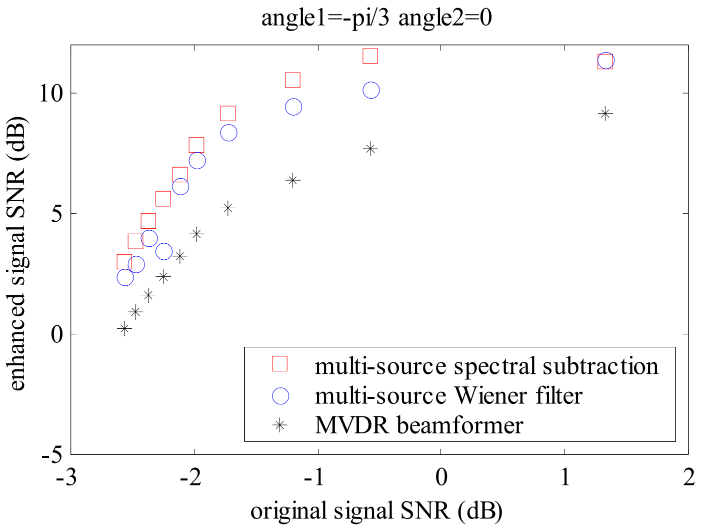

angle1=-pi/3 angle2=pi/3

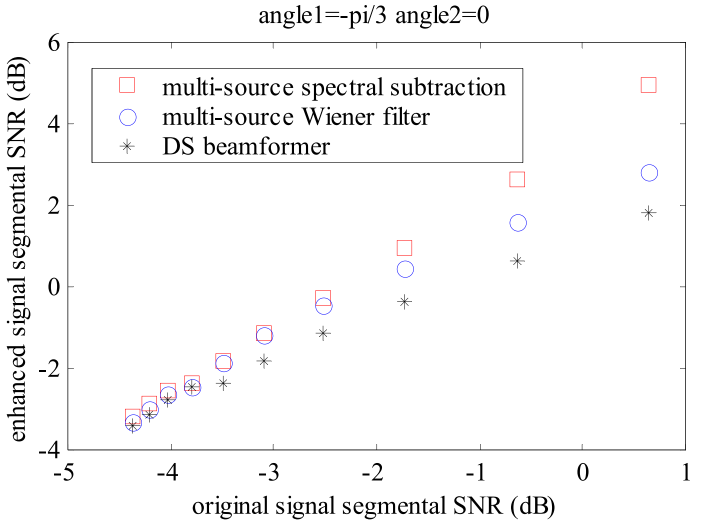

angle1=-pi/3 angle2=pi/3

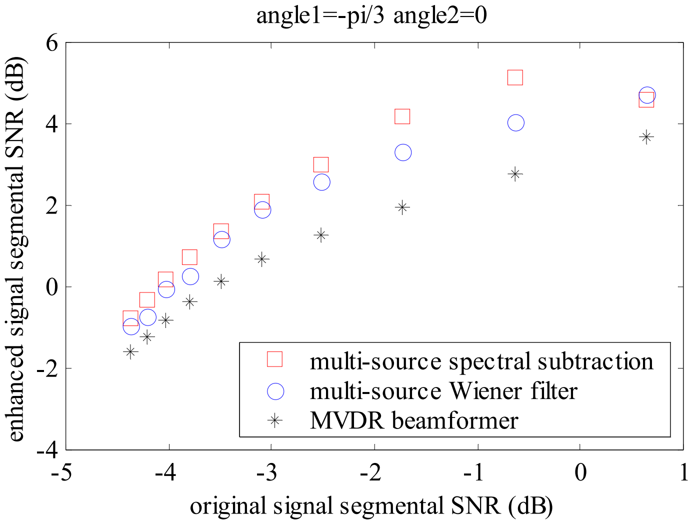

angle1=-pi/3 angle2=pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
15
10
5
0
-5
-3
-2
-1
0
1
2
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
8
6
4
2
0
-2
-4
-5
-4
-3
-2
-1
0
1
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer


angle1=-pi/3 angle2=pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
15
10
5
0
-5
-3
-2
-1
0
1
2
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
15
10
5
0
-5
-3
-2
-1
0
1
2
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
8
6
4
2
0
-2
-4
-5
-4
-3
-2
-1
0
1
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
8
6
4
2
0
-2
-4
-5
-4
-3
-2
-1
0
1
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer

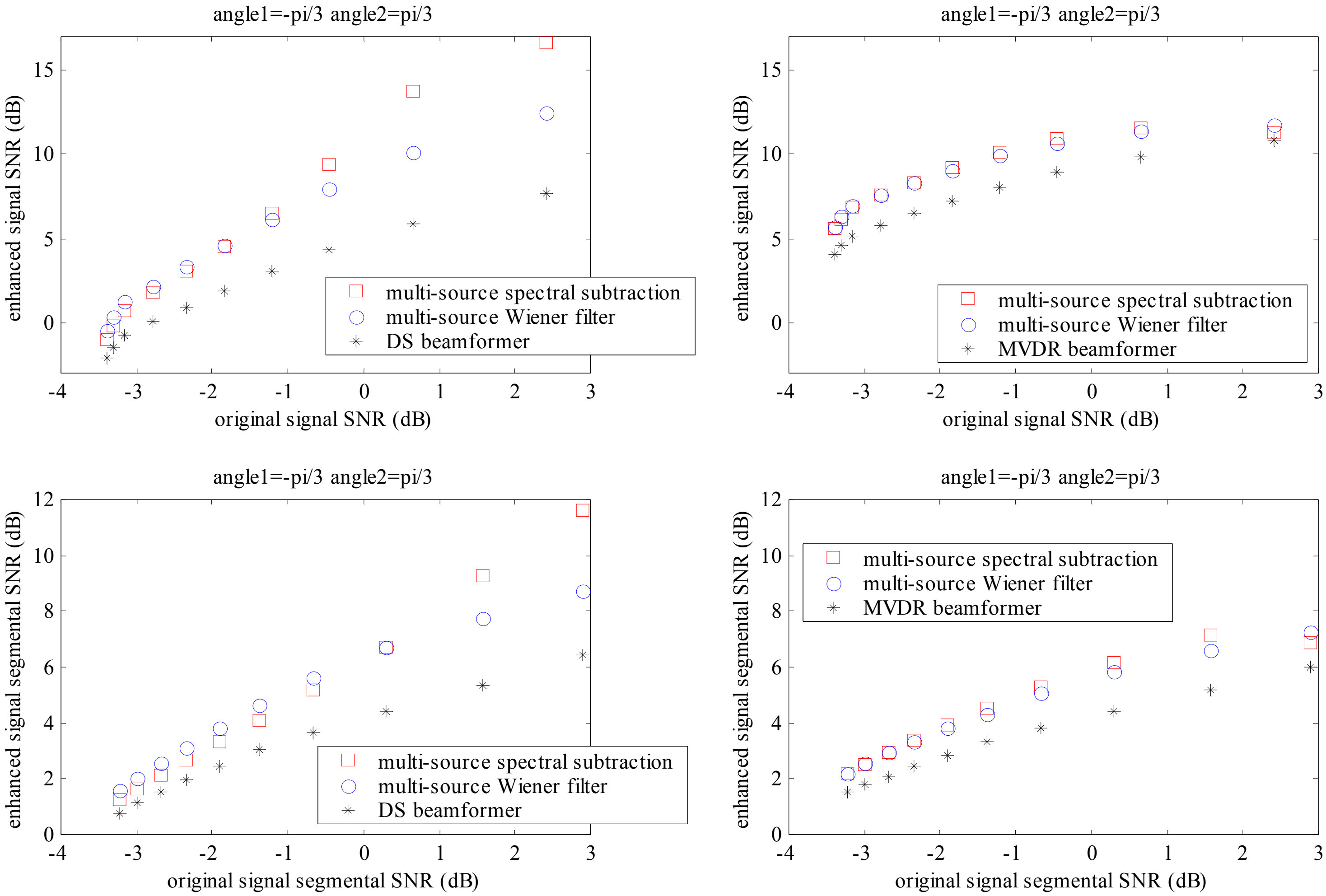

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
enhanced signal SNR (dB)
original signal SNR (dB)
10
5
0
-5
-3
-2
-1
0
1
2

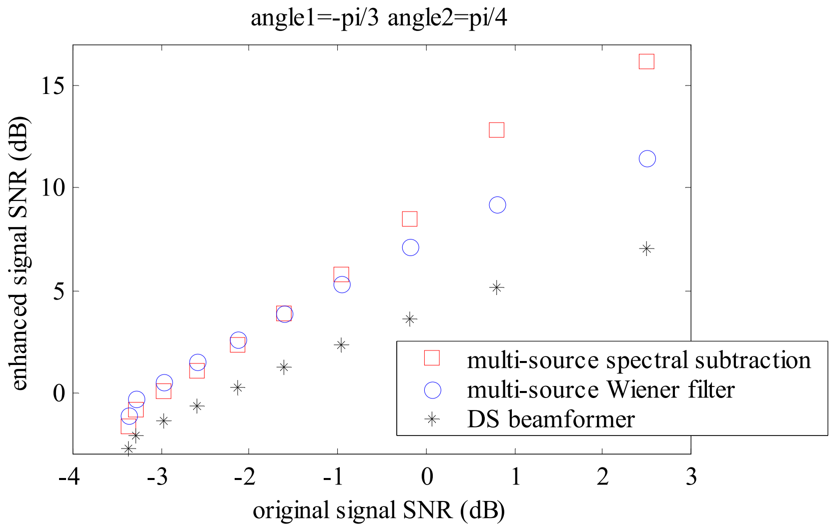

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
enhanced signal SNR (dB)
original signal SNR (dB)
10
5
0
-5
-3
-2
-1
0
1
2

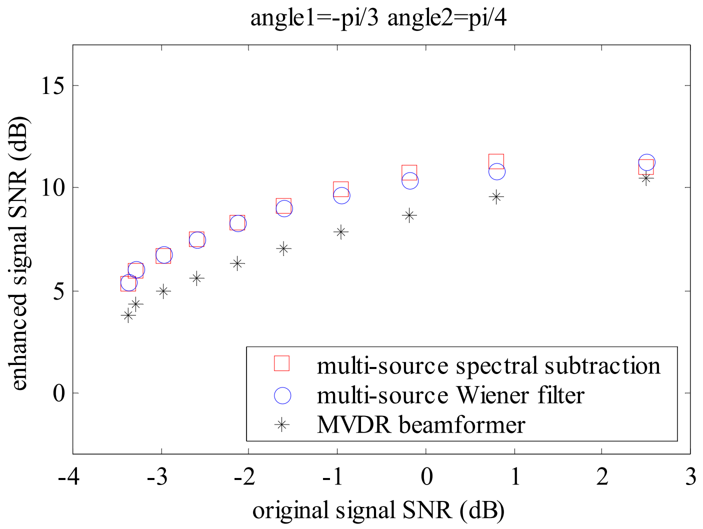

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
6
4
2
0
-2
-4
-5
-4
-3
-2
-1
0
1

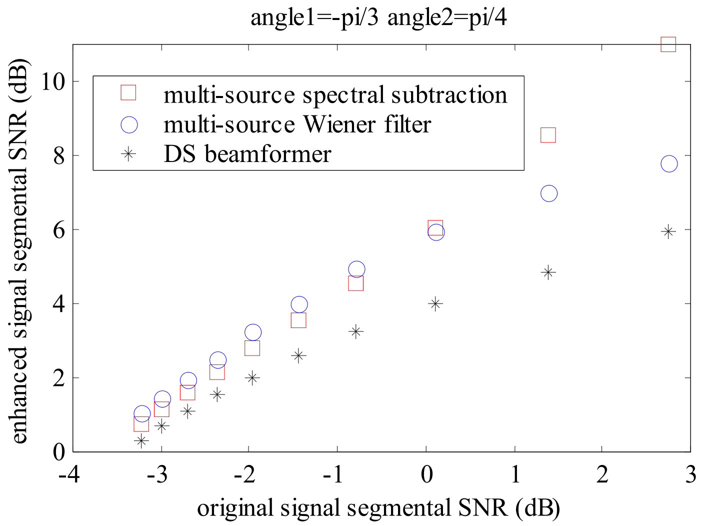

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
6
4
2
0
-2
-4
-5
-4
-3
-2
-1
0
1

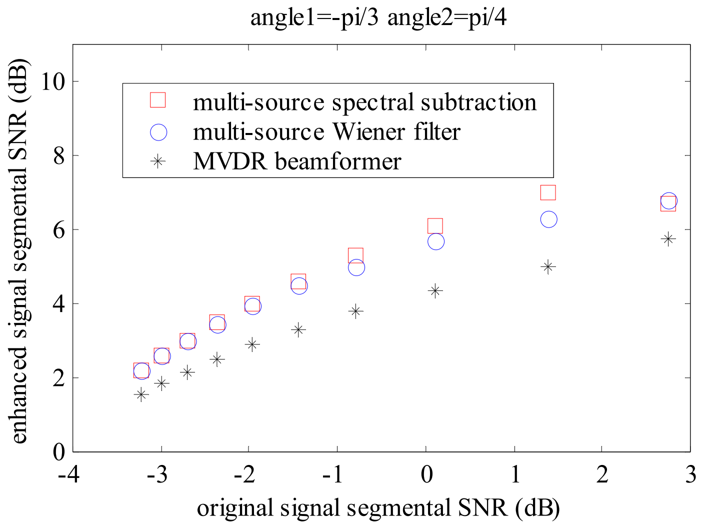
Experiment 4:

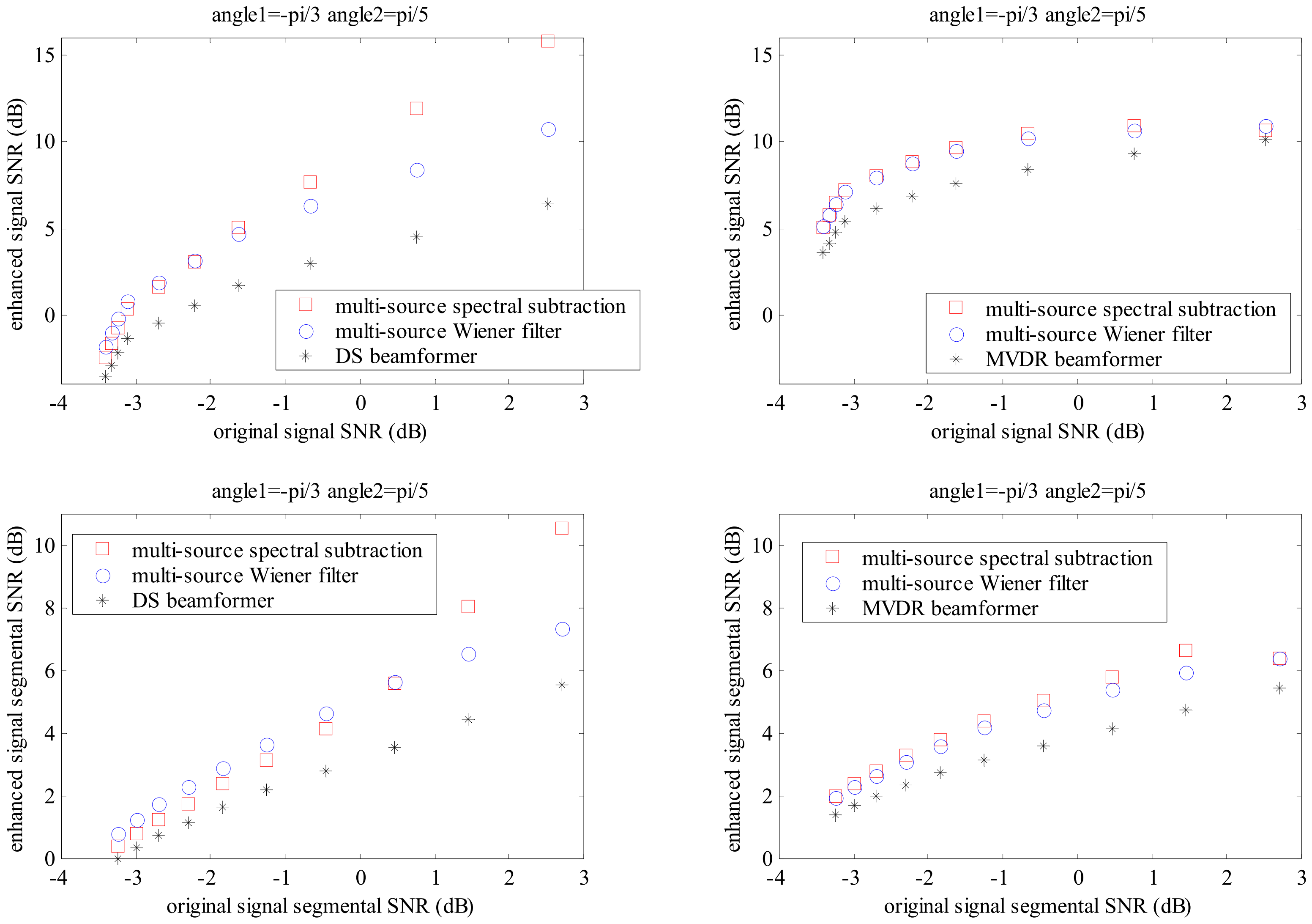


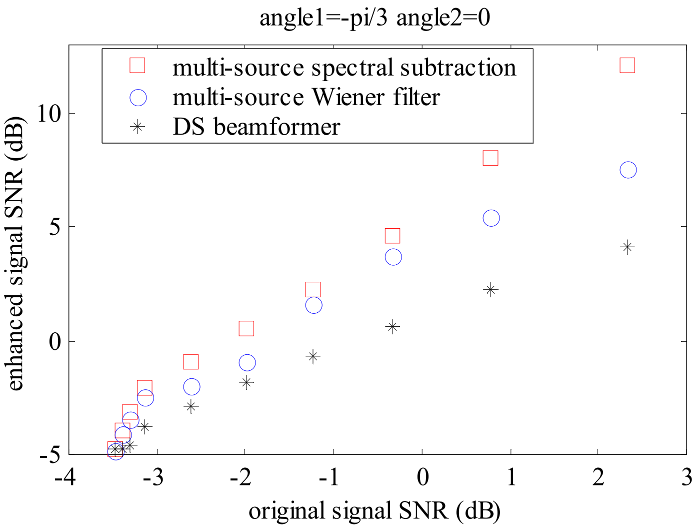

angle1=-pi/3 angle2=pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
15
10
5
0
-4
-3
-2
-1
0
1
2
3
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer

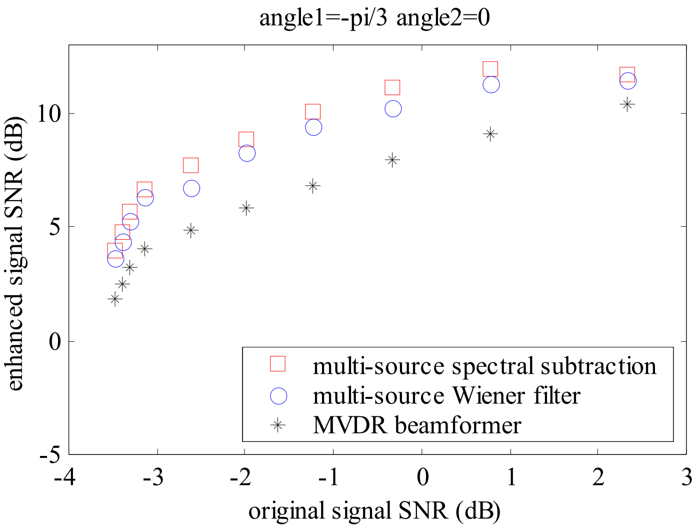

angle1=-pi/3 angle2=pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
15
10
5
0
-4
-3
-2
-1
0
1
2
3
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer

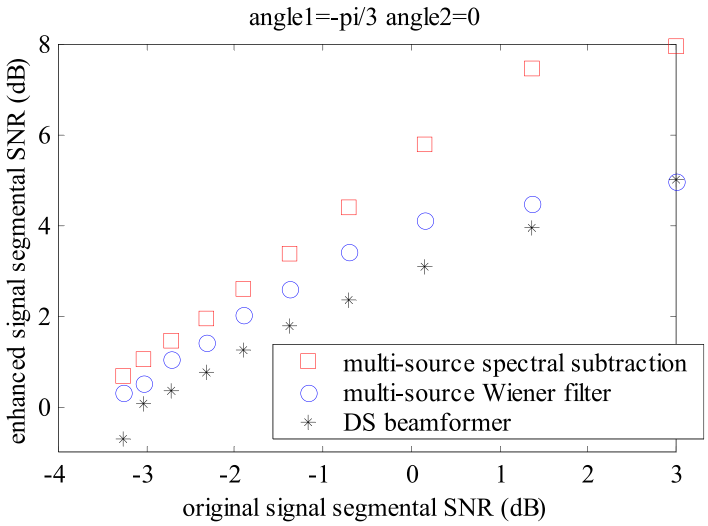

angle1=-pi/3 angle2=pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
10
8
6
4
2
0
-4
-3
-2
-1
0
1
2
3
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer

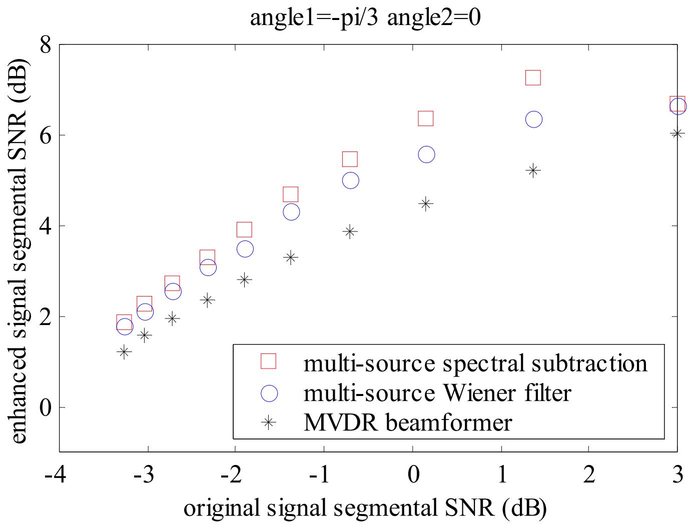

angle1=-pi/3 angle2=pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
10
8
6
4
2
0
-4
-3
-2
-1
0
1
2
3
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer


angle1=-pi/3 angle2=pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=-pi/3 angle2=pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
-4
-3
-2
-1
0
1
2
3
0
5
10
15
0
2
4
6
8
10

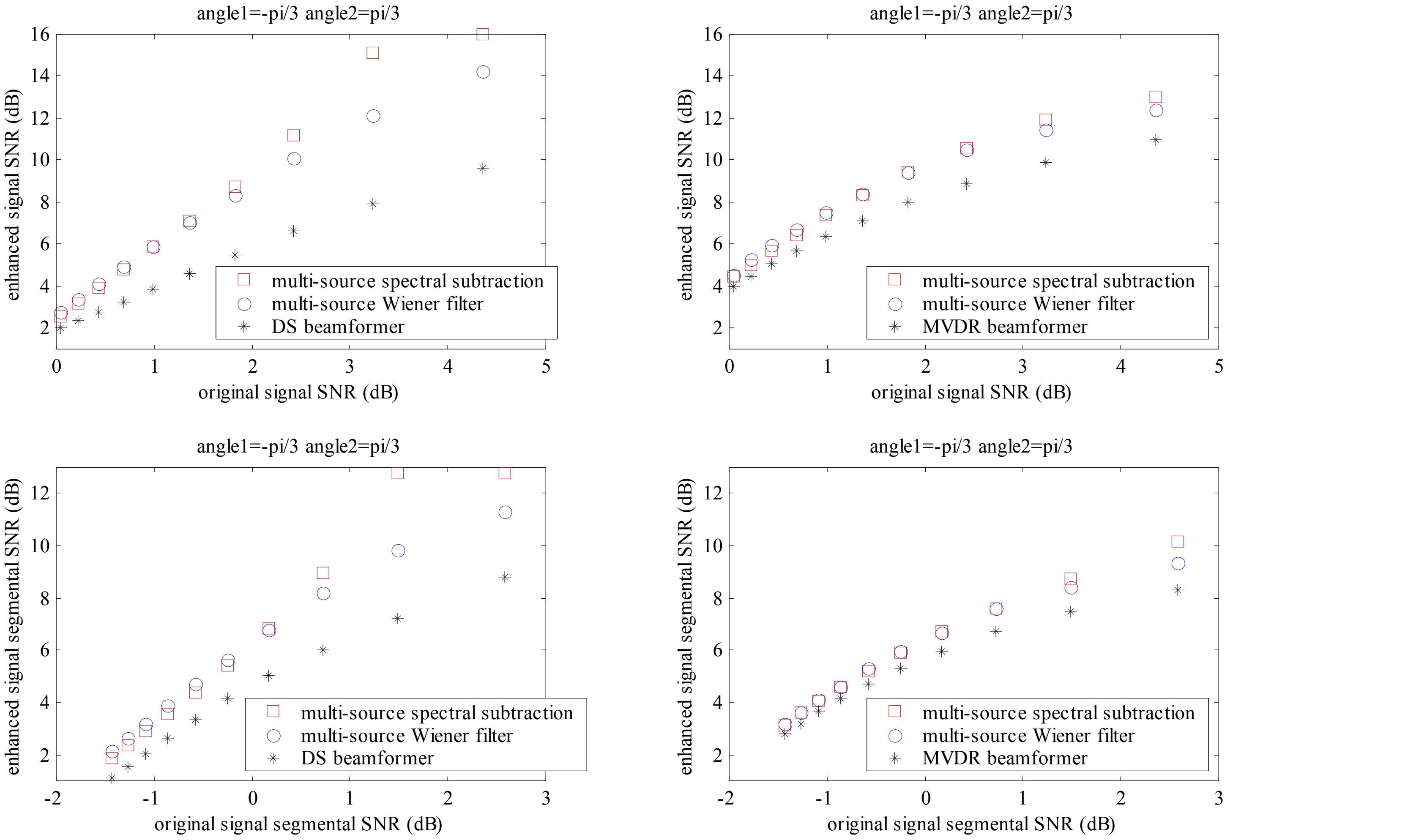

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
enhanced signal SNR (dB)
original signal SNR (dB)
10
5
0
-5
-4
-3
-2
-1
0
1
2
3

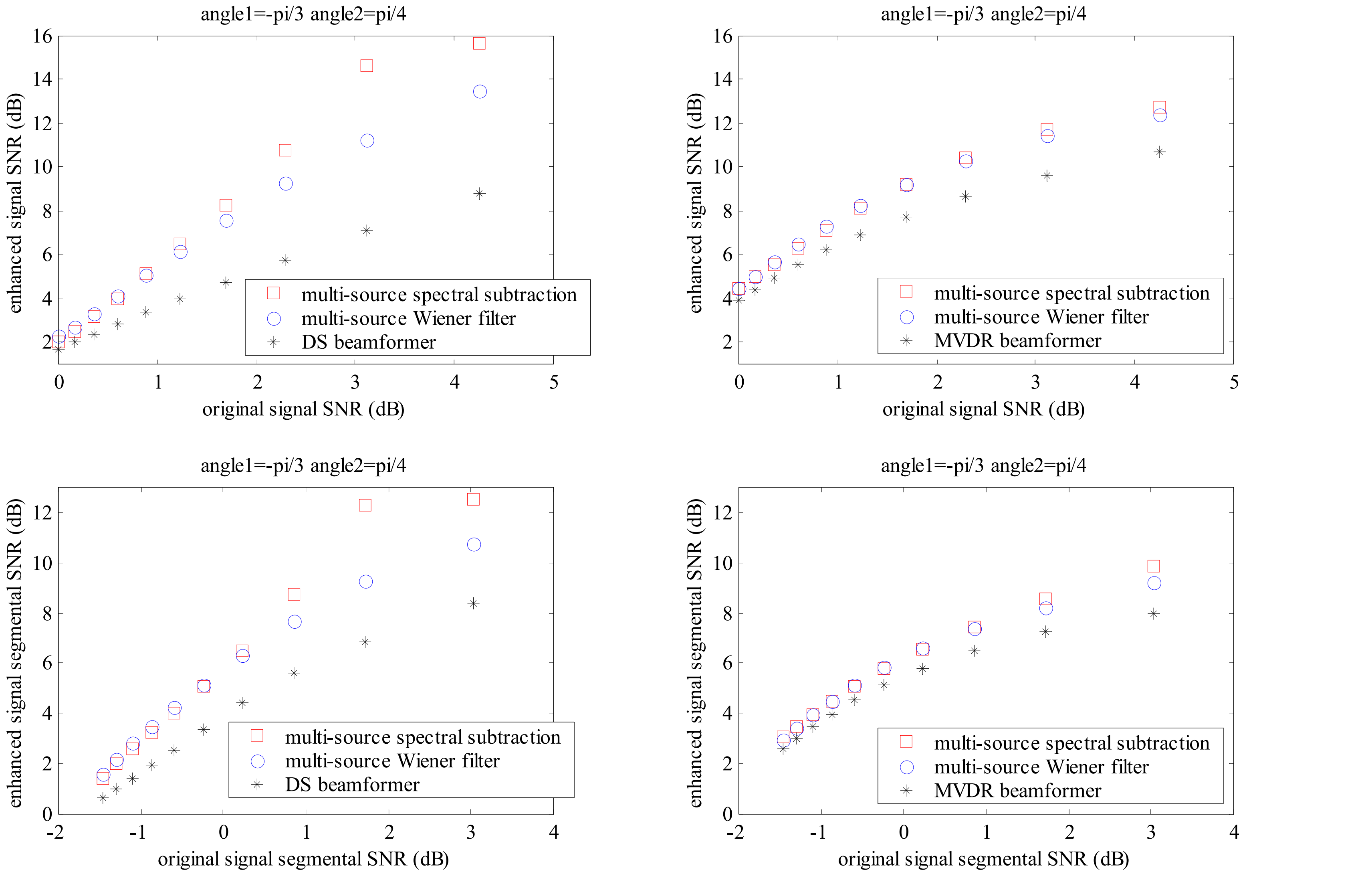

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
enhanced signal SNR (dB)
original signal SNR (dB)
10
5
0
-5
-4
-3
-2
-1
0
1
2
3

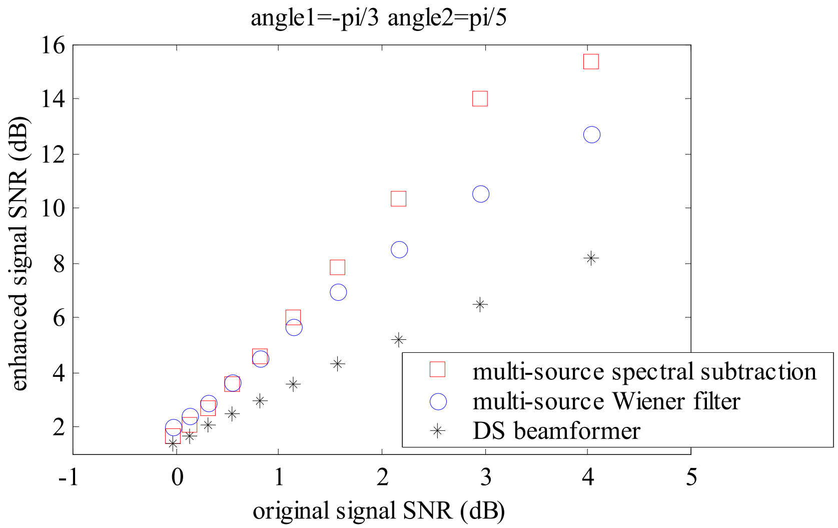

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
8
6
4
2
0
-4
-3
-2
-1
0
1
2
3

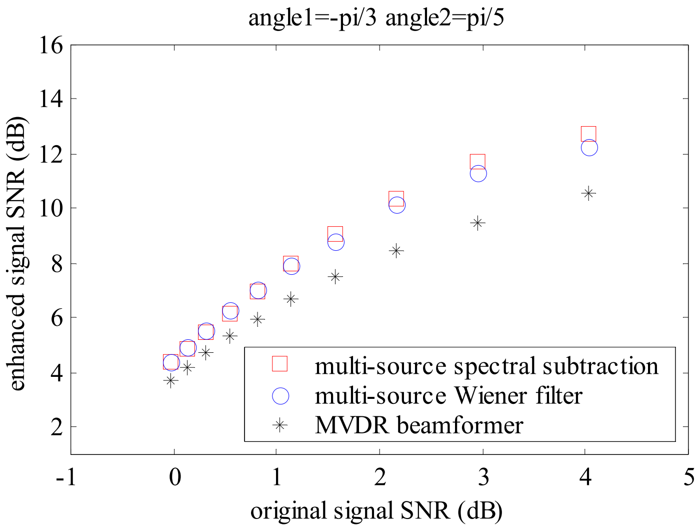

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
8
6
4
2
0
-4
-3
-2
-1
0
1
2
3

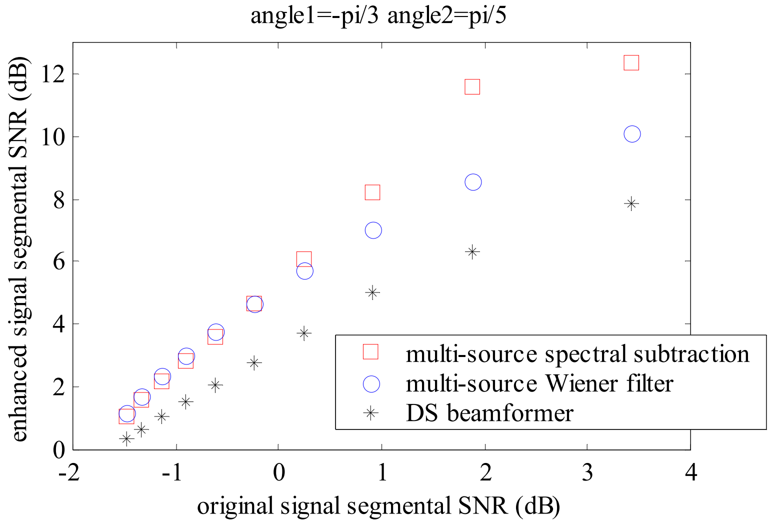
Experiment 5:

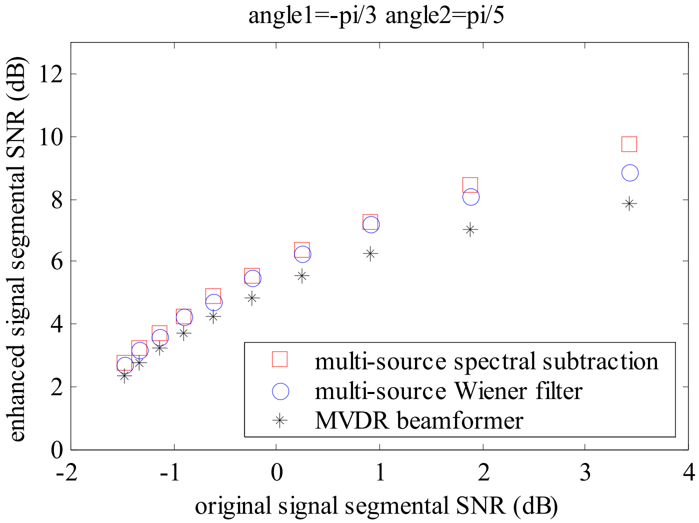


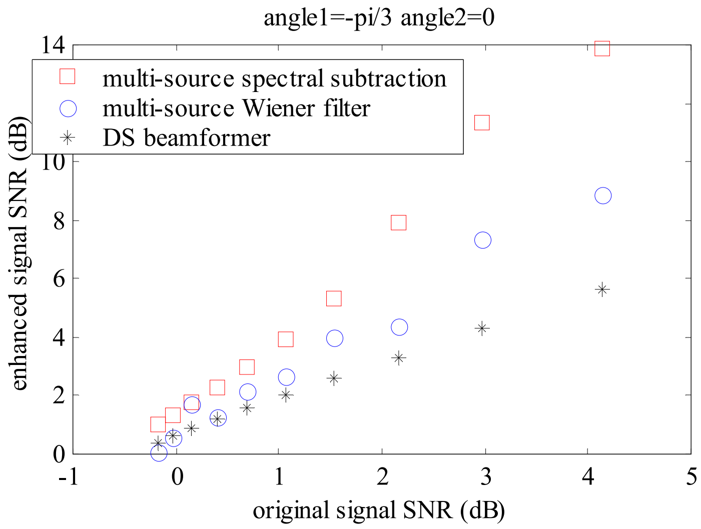

angle1=-pi/3 angle2=pi/4
enhanced signal SNR (dB)
16
14
12
10
8
6
4
2
0
1
2
3
4
5
original signal SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal segmental SNR (dB)
12
10
8
6
4
2
0
-2
-1
0
1
2
3
4
original signal segmental SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
angle1=-pi/3 angle2=pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer

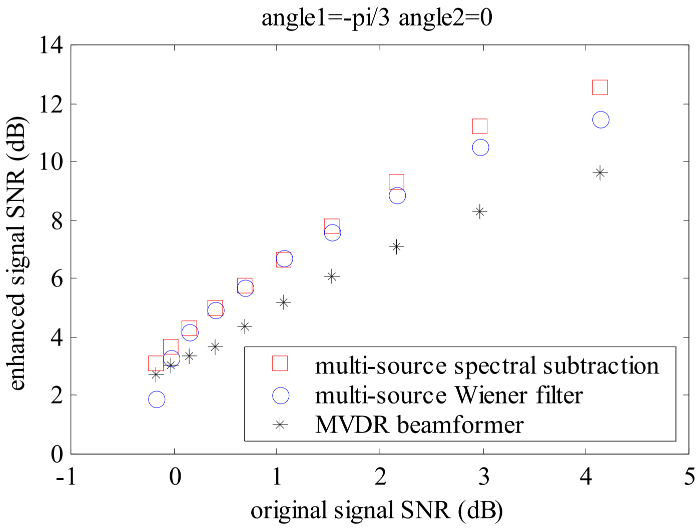

angle1=-pi/3 angle2=pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
16
14
12
10
8
6
4
2
-1
0
1
2
3
4
5
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer

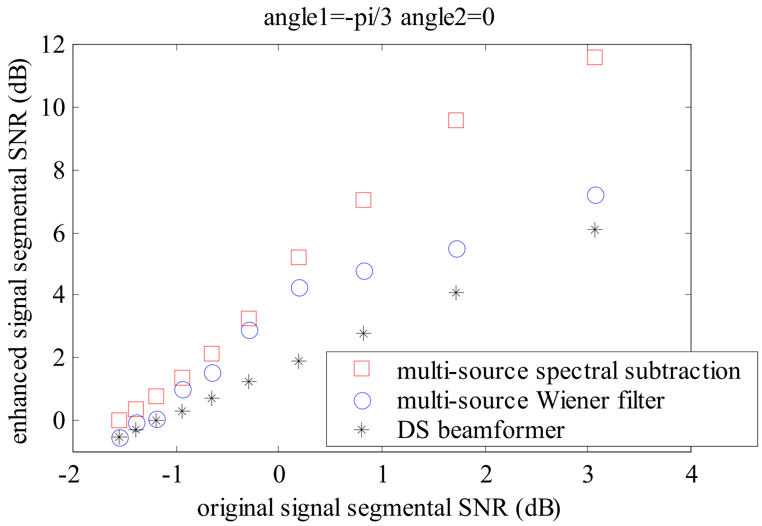

angle1=-pi/3 angle2=pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
16
14
12
10
8
6
4
2
-1
0
1
2
3
4
5
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer

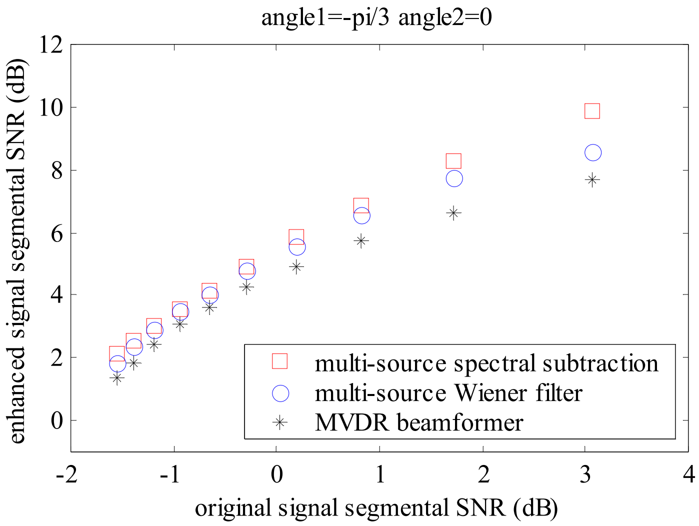

angle1=-pi/3 angle2=pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
12
10
8
6
4
2
0
-2
-1
0
1
2
3
4
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer


angle1=-pi/3 angle2=pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
12
10
8
6
4
2
0
-2
-1
0
1
2
3
4
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer

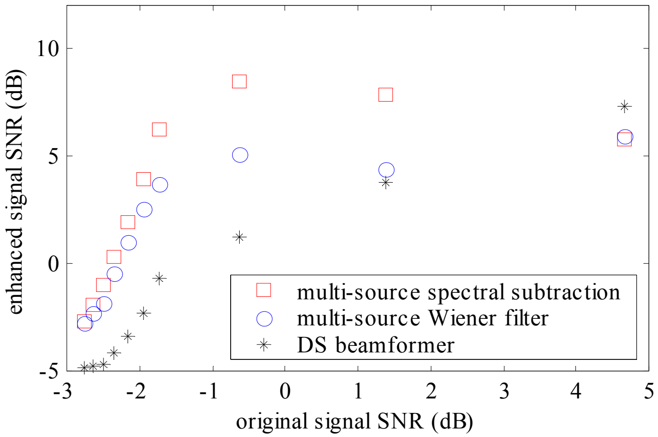

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
enhanced signal SNR (dB)
original signal SNR (dB)
14
10
8
6
4
2
0
-1
0
1
2
3
4
5

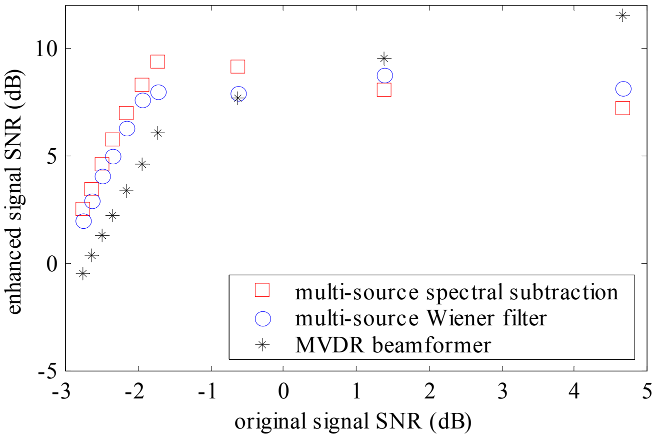

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
enhanced signal SNR (dB)
original signal SNR (dB)
14
12
10
8
6
4
2
0
-1
0
1
2
3
4
5

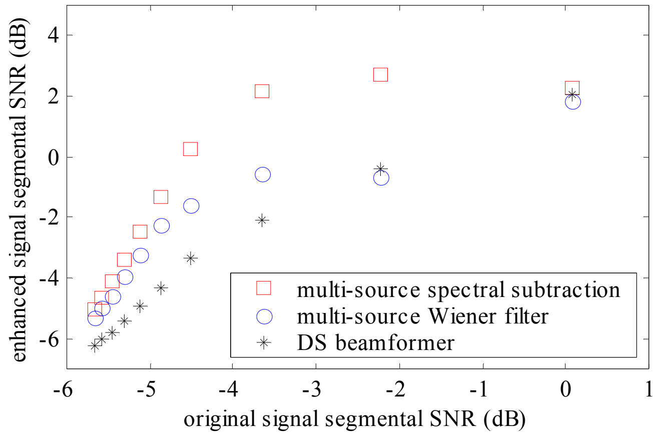

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
12
10
8
6
4
2
0
-2
-1
0
1
2
3
4

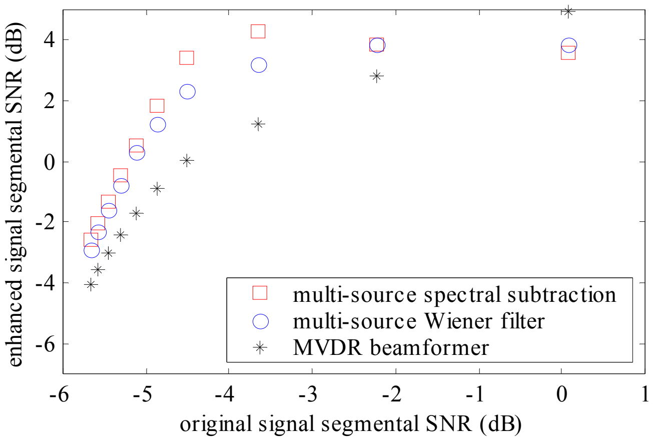

angle1=-pi/3 angle2=0
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
12
10
8
6
4
2
0
-2
-1
0
1
2
3
4

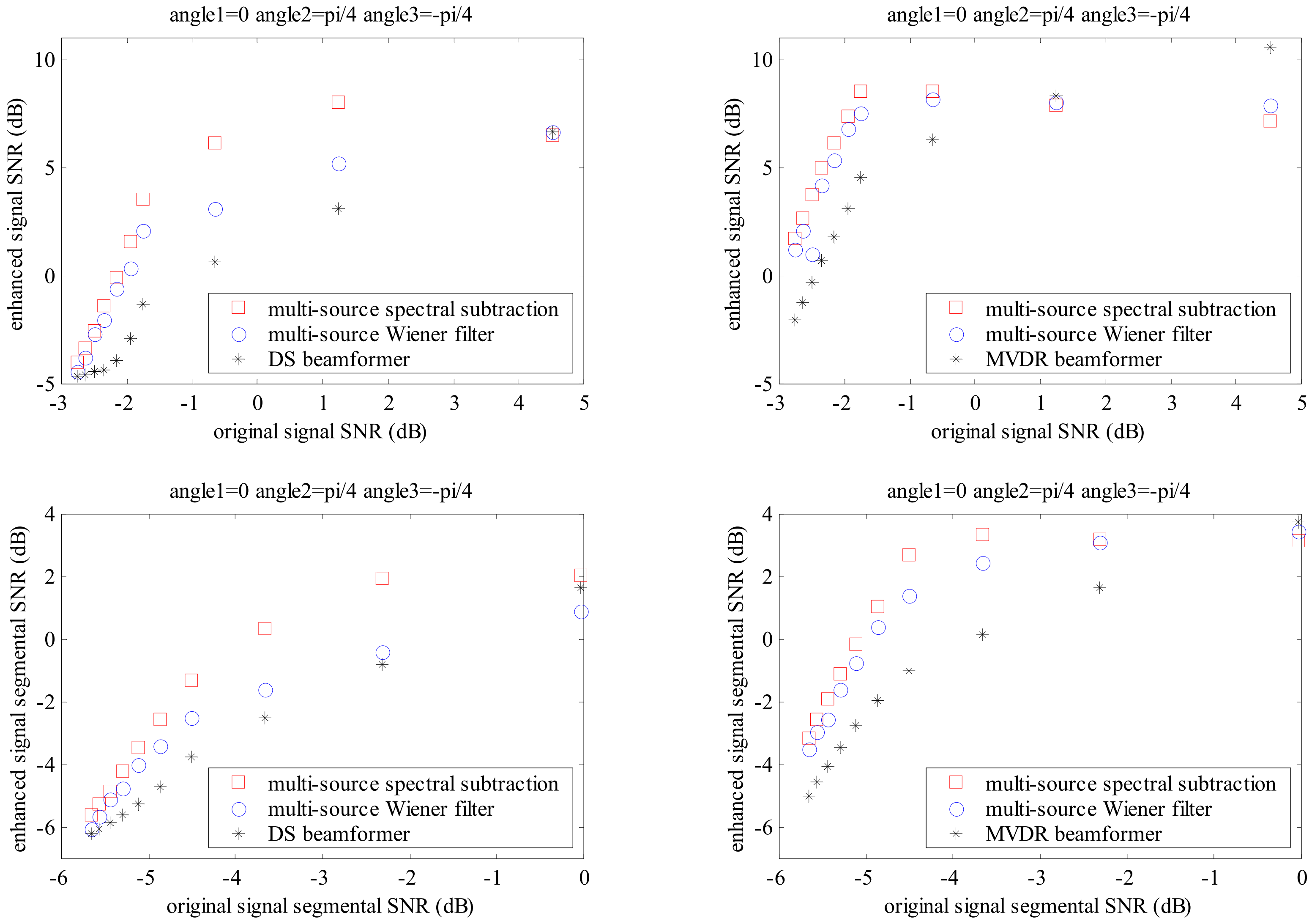
Three source experiments: Experiment 1:

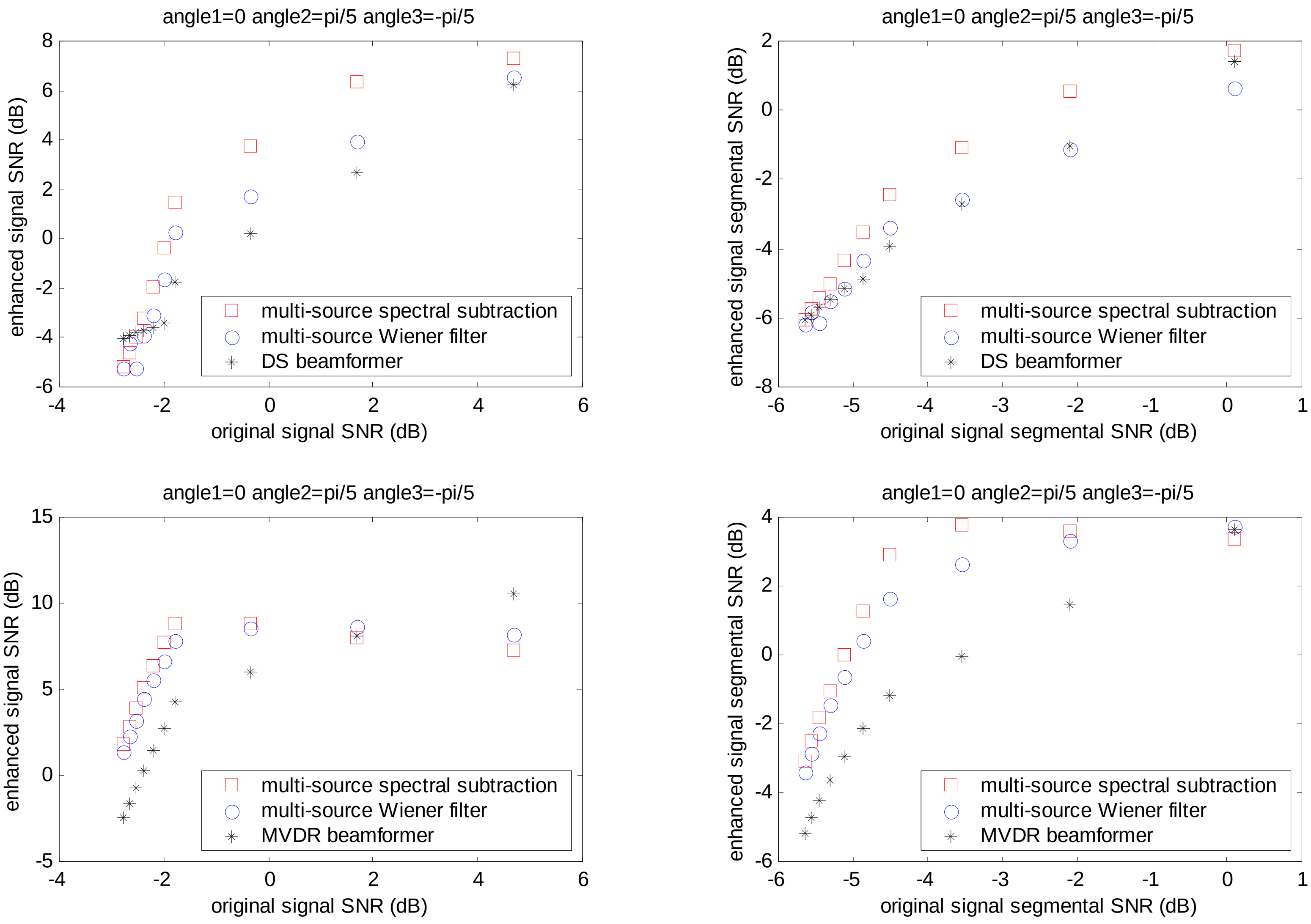

angle1=0 angle2=pi/3 angle3=-pi/3


angle1=0 angle2=pi/3 angle3=-pi/3

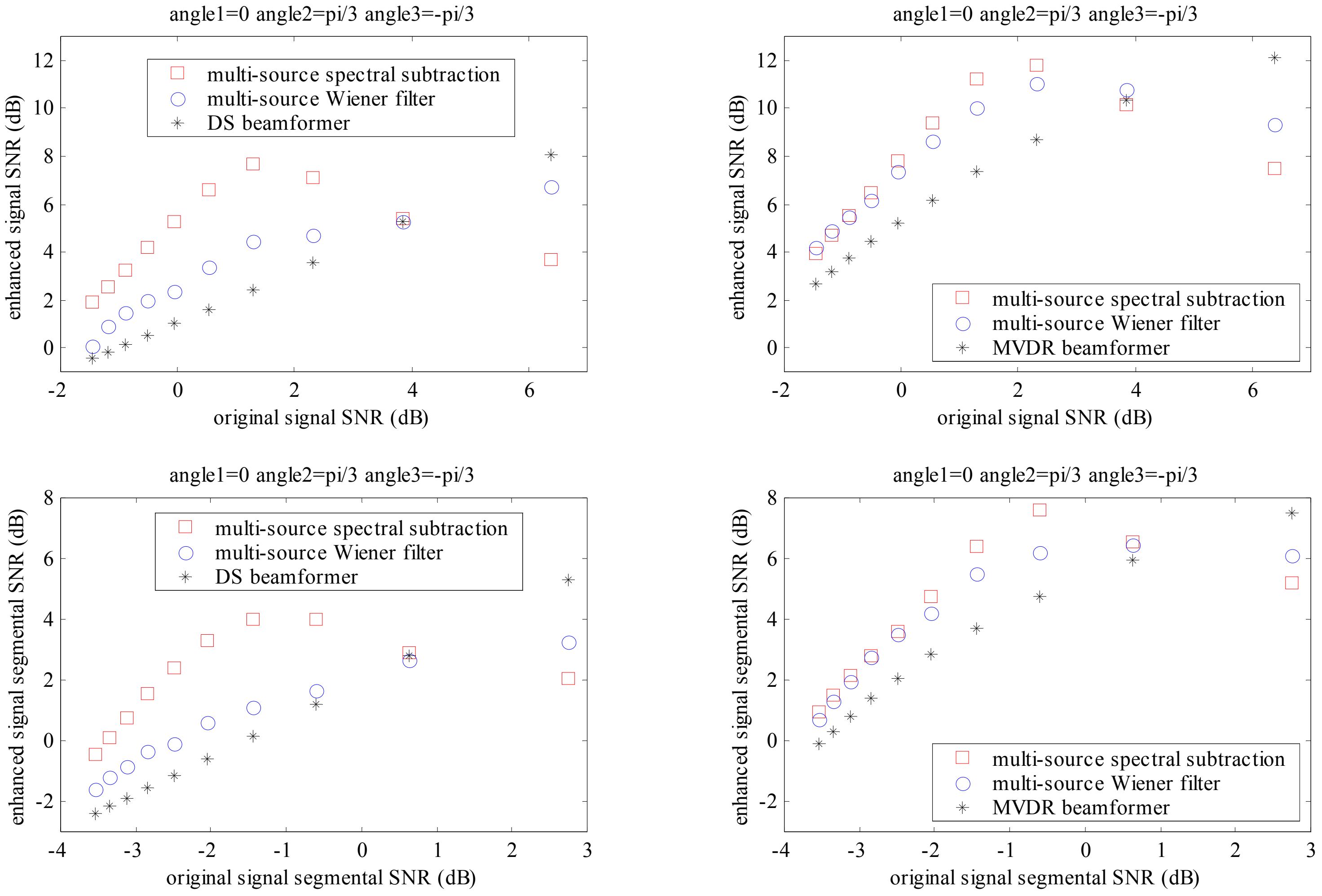

angle1=0 angle2=pi/3 angle3=-pi/3

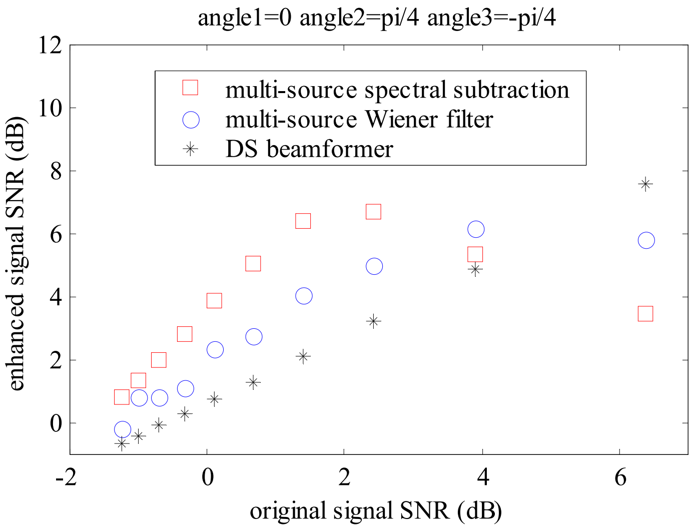

angle1=0 angle2=pi/3 angle3=-pi/3

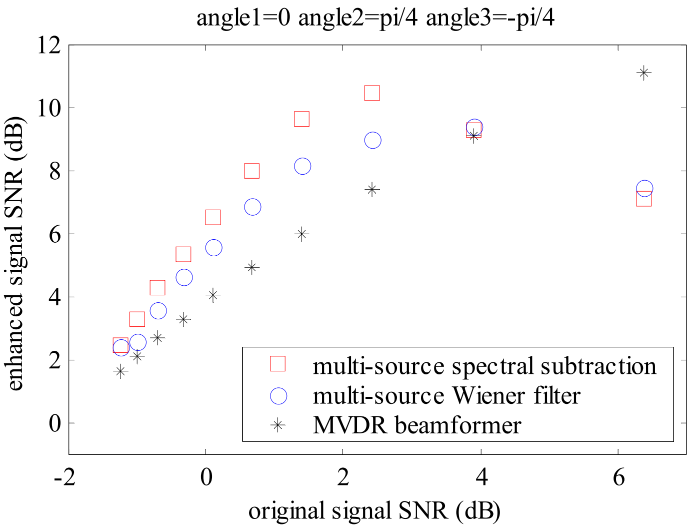

angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal SNR (dB)
10
5
0
-5
-3
-2
-1
0
1
2
3
4
5
original signal SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal SNR (dB)
10
5
0
-5
-3
-2
-1
0
1
2
3
4
5
original signal SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal segmental SNR (dB)
4
2
0
-2
-4
-6
-6
-5
-4
-3
-2
-1
0
original signal segmental SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal segmental SNR (dB)
4
2
0
-2
-4
-6
-6
-5
-4
-3
-2
-1
0
original signal segmental SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer

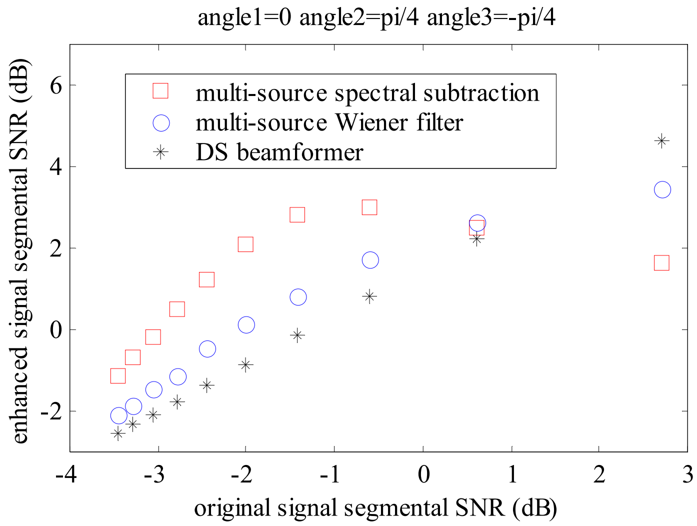

angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer

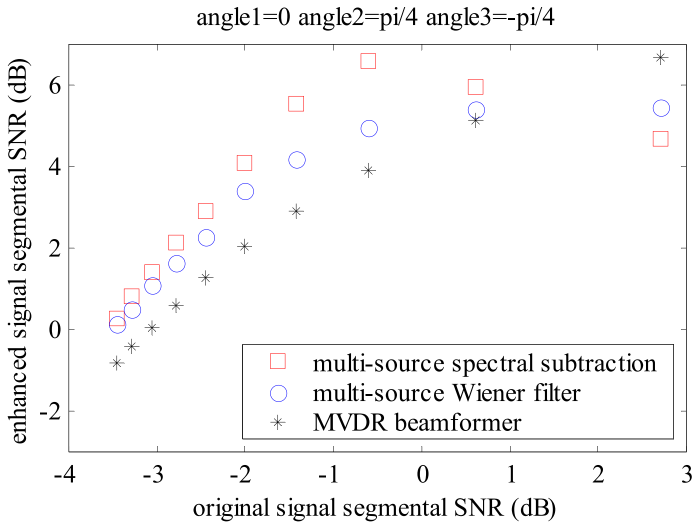
Experiment 2:

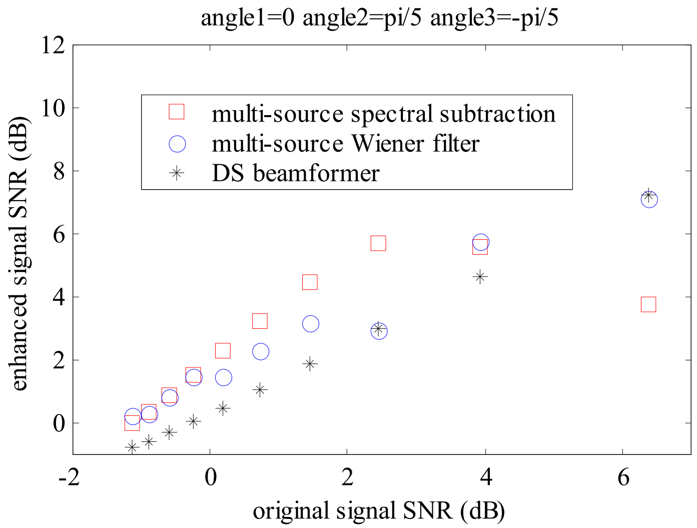


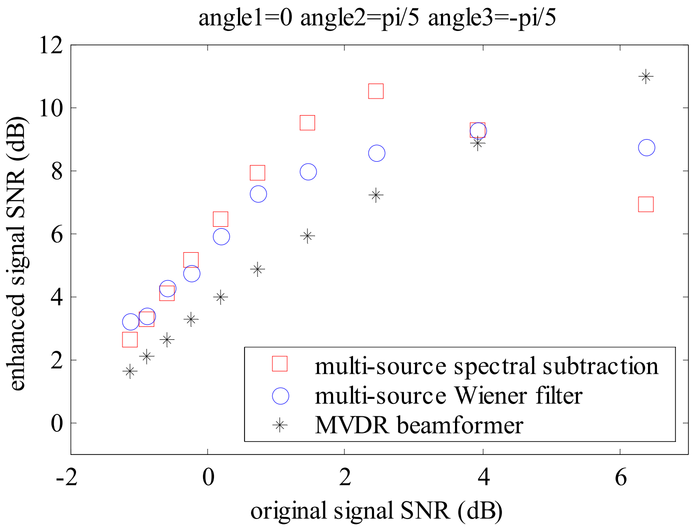

angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
12
10
8
6
4
2
0
-2
0
2
4
6
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer

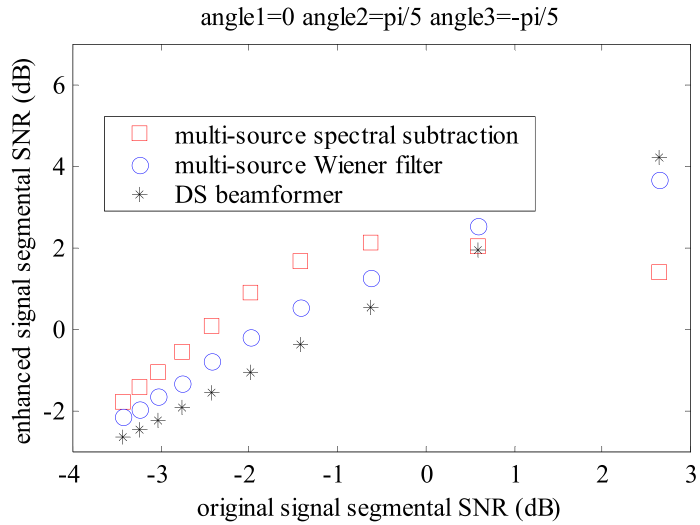

angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
12
10
8
6
4
2
0
-2
0
2
4
6
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer

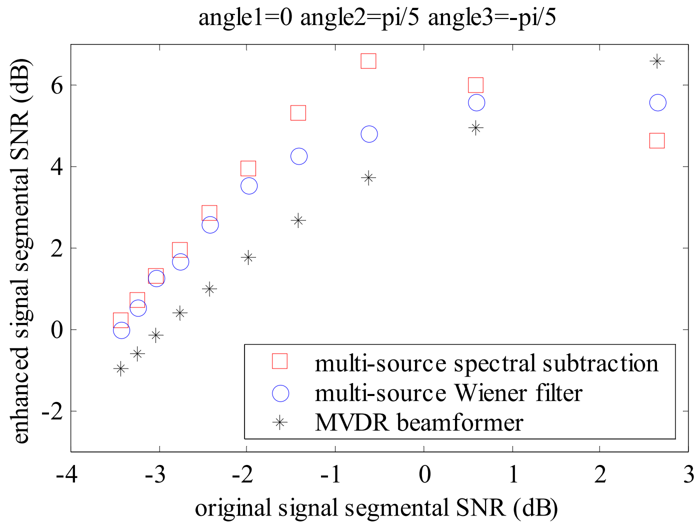

angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
6
4
2
0
-2
-4
-3
-2
-1
0
1
2
3
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer


angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
6
4
2
0
-2
-4
-3
-2
-1
0
1
2
3
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer

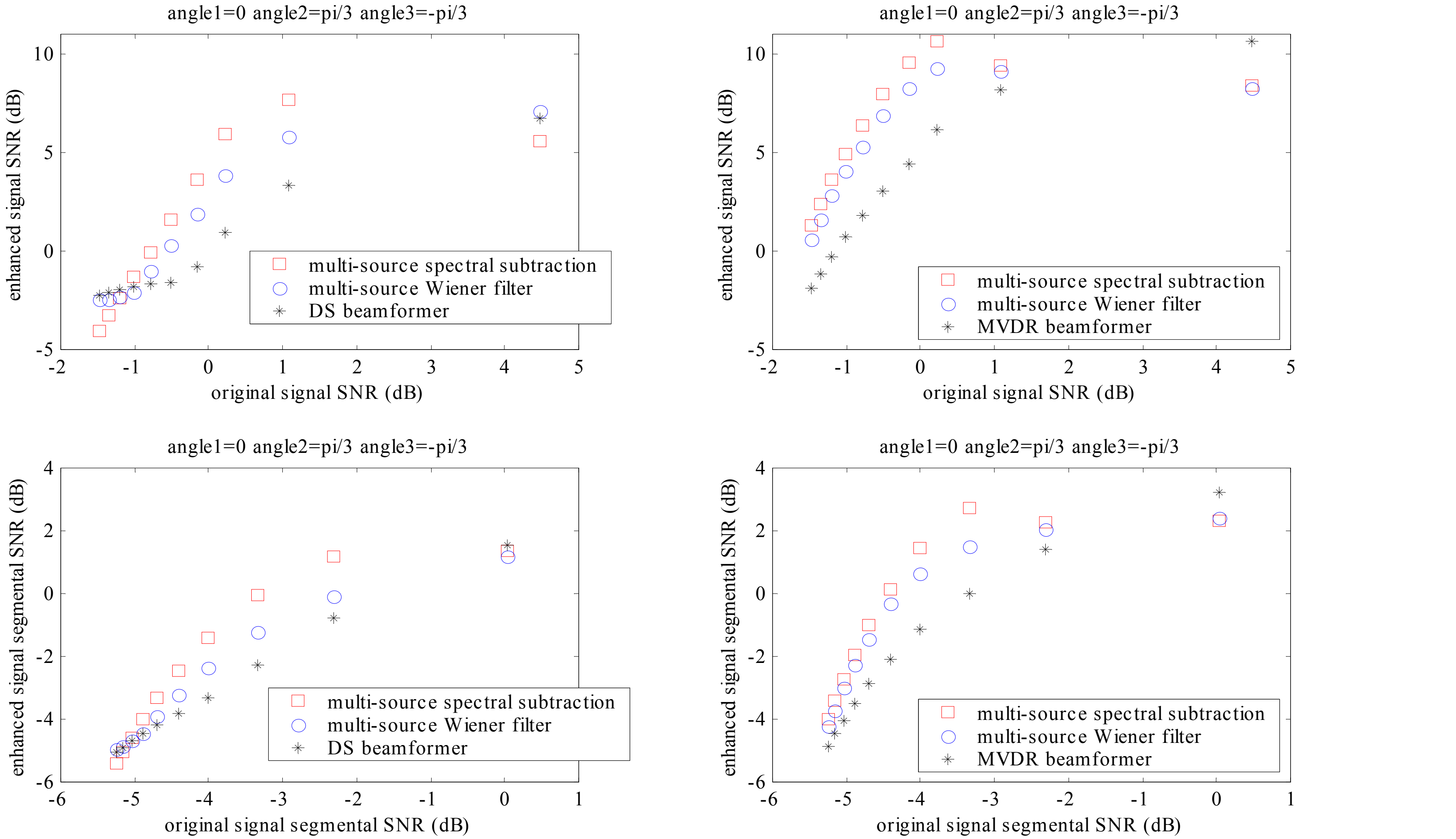

angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
12
10
8
6
4
2
0
-2
0
2
4
6
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer

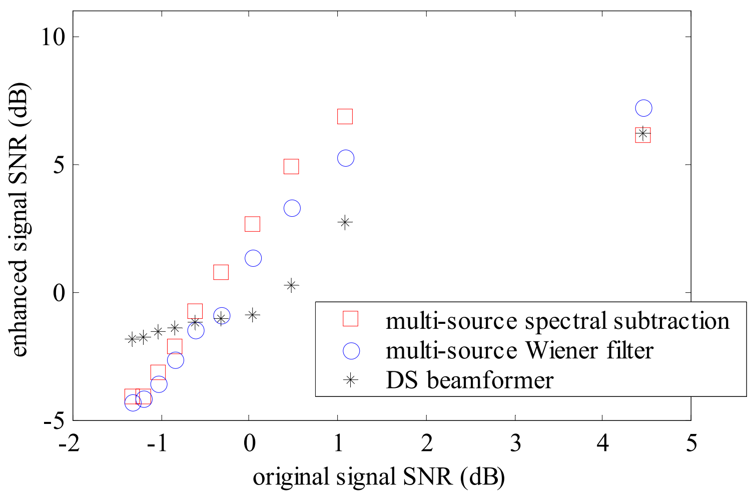

angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
12
10
8
6
4
2
0
-2
0
2
4
6
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer


angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
6
4
2
0
-2
-4
-3
-2
-1
0
1
2
3
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer

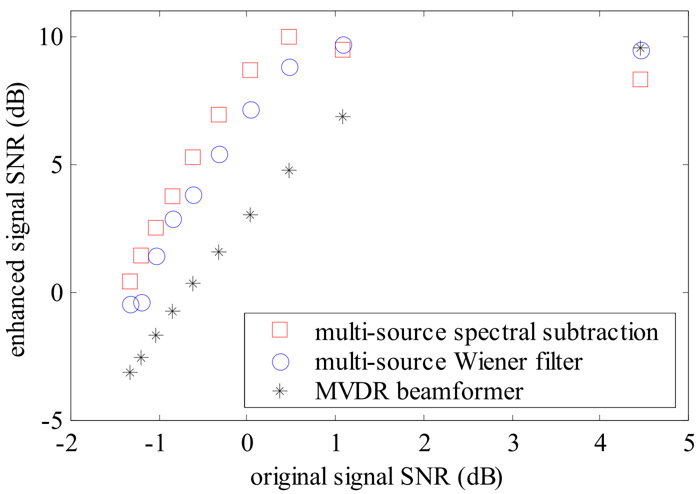

angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
6
4
2
0
-2
-4
-3
-2
-1
0
1
2
3
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer

Experiment 3:

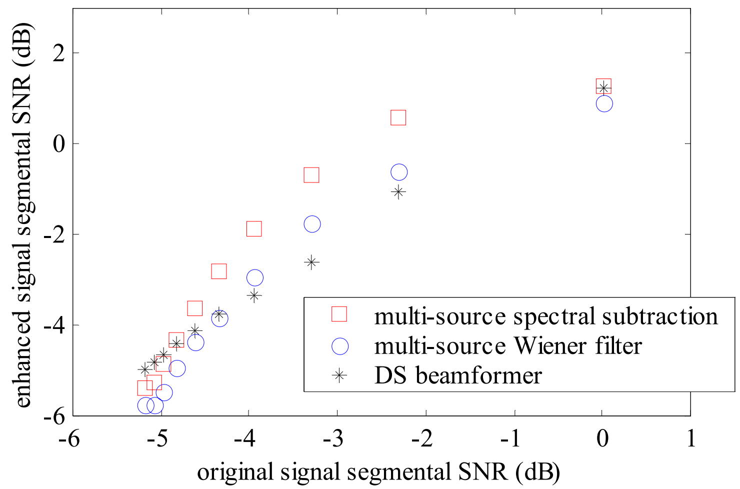




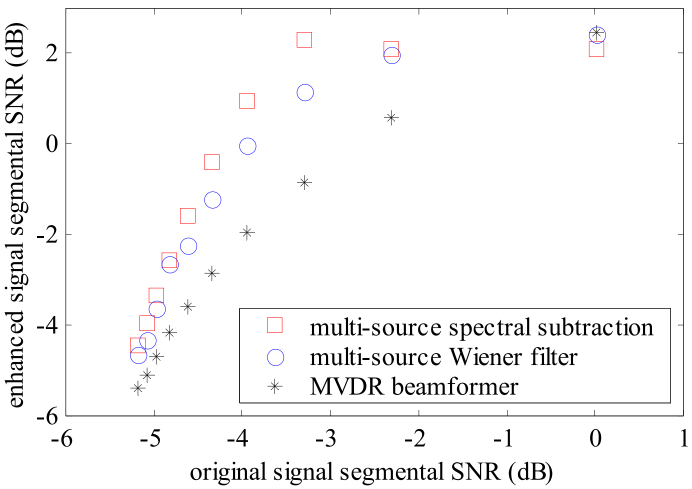
angle1=0 angle2=pi/4 angle3=-pi/4



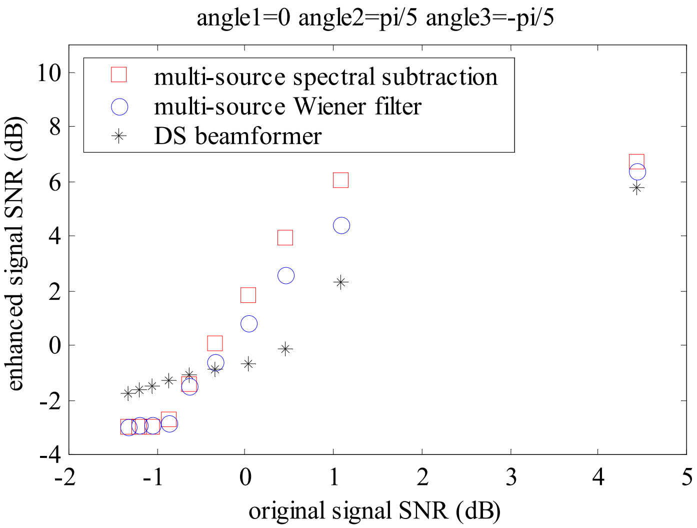
angle1=0 angle2=pi/4 angle3=-pi/4

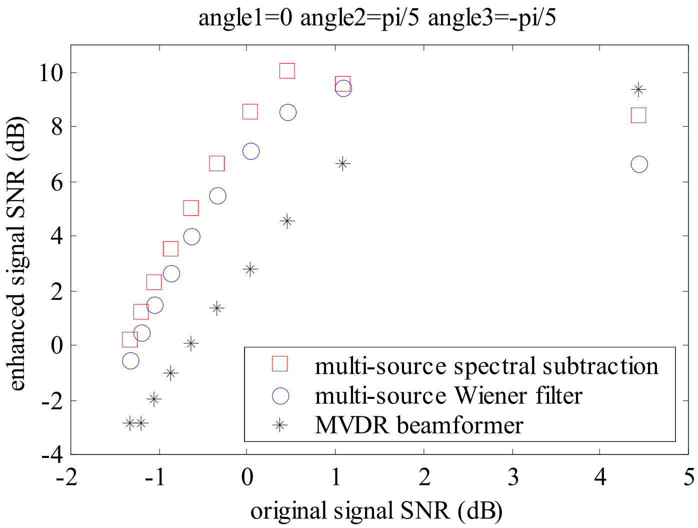


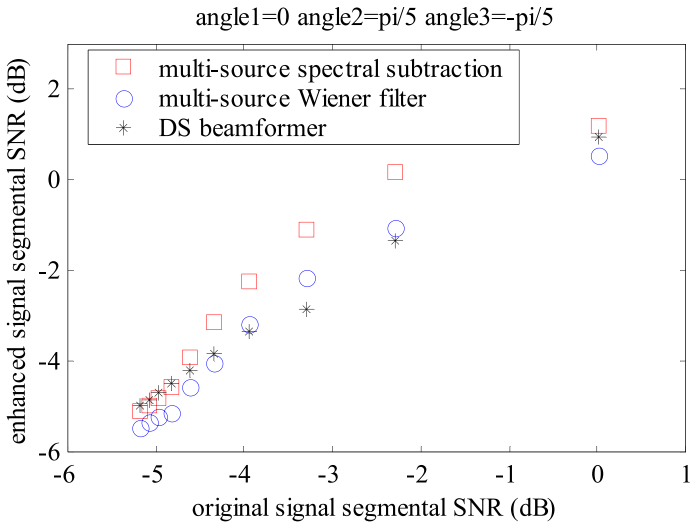
angle1=0 angle2=pi/4 angle3=-pi/4

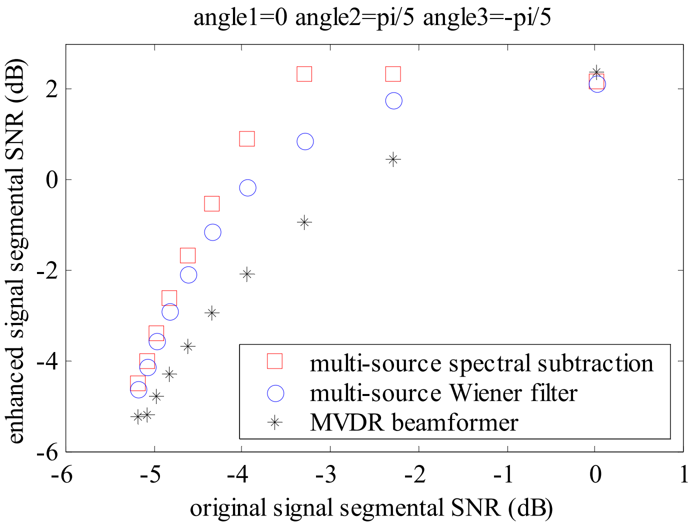


angle1=0 angle2=pi/4 angle3=-pi/4

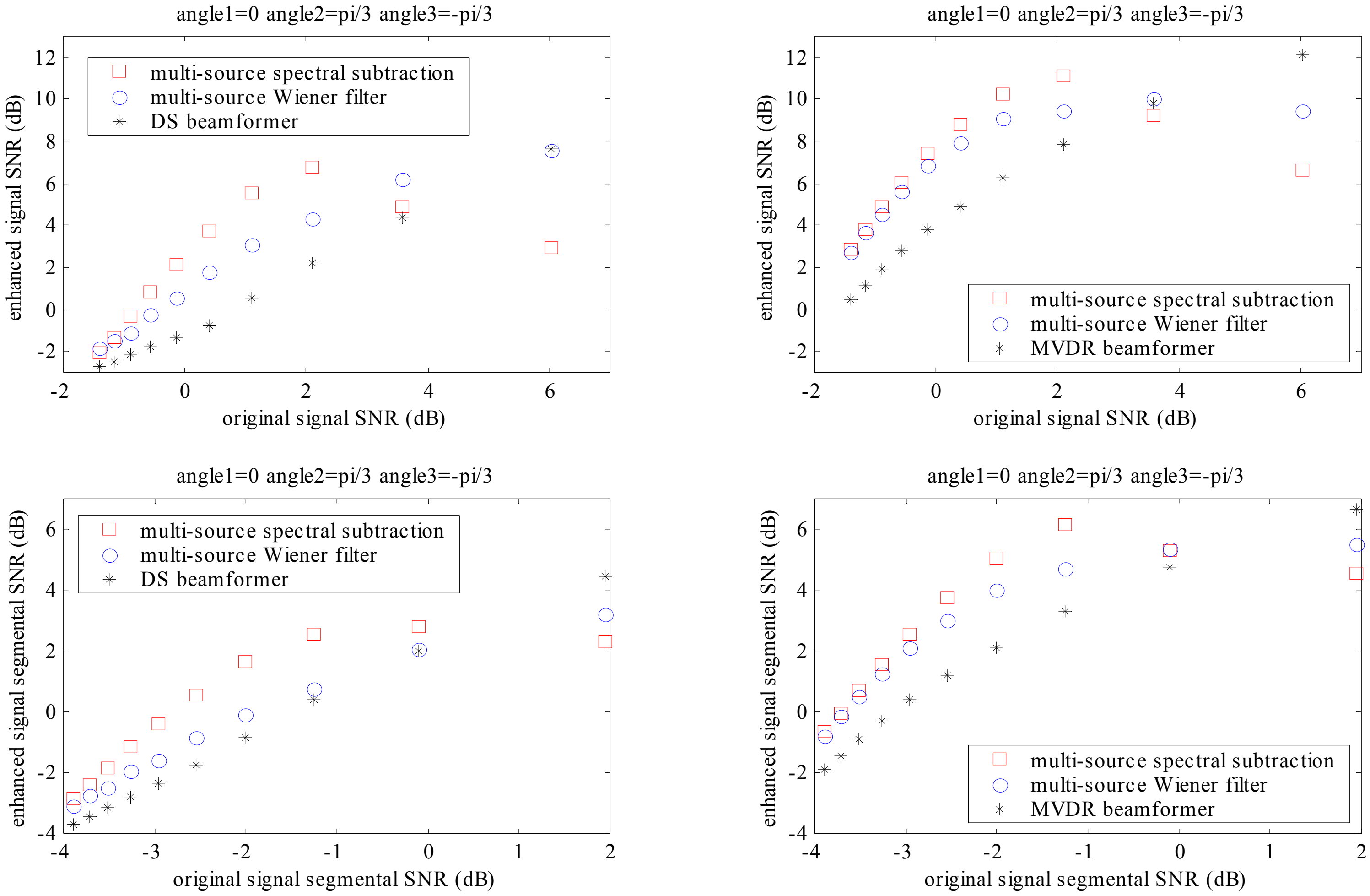

angle1=0 angle2=pi/5 angle3=-pi/5
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
enhanced signal SNR (dB)
original signal SNR (dB)
10
8
6
4
2
0
-2
-4
-2
-1
0
1
2
3
4
5

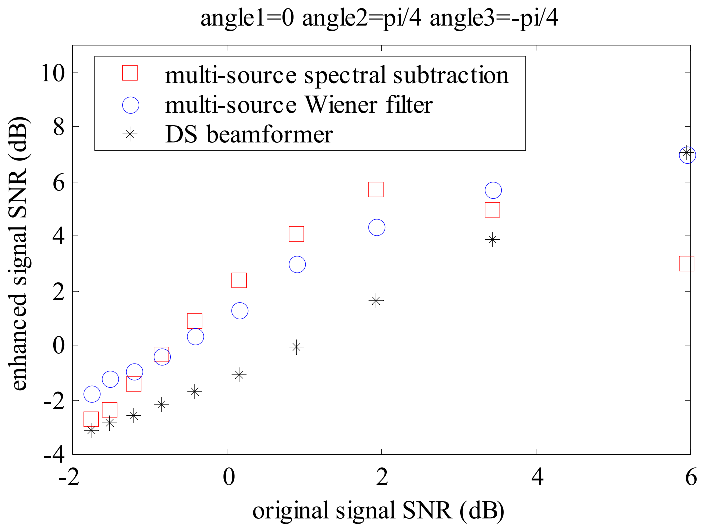

angle1=0 angle2=pi/5 angle3=-pi/5
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer
enhanced signal SNR (dB)
original signal SNR (dB)
10
8
6
4
2
0
-2
-4
-2
-1
0
1
2
3
4
5

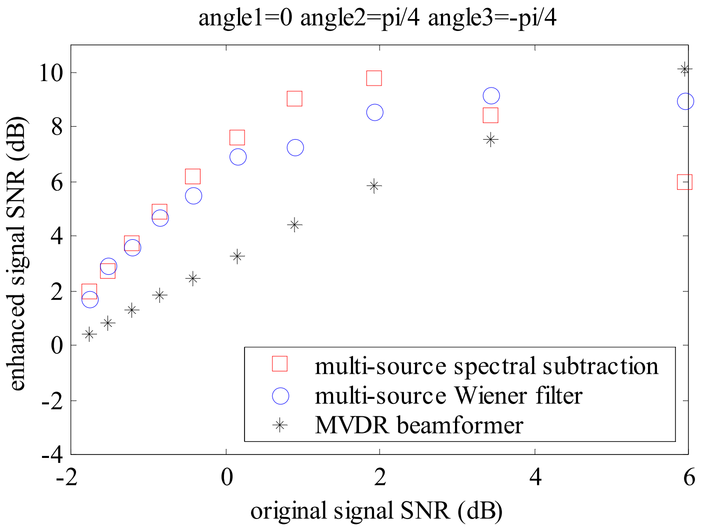

angle1=0 angle2=pi/5 angle3=-pi/5
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ DS beamformer
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
2
0
-2
-4
-6
-6
-5
-4
-3
-2
-1
0
1

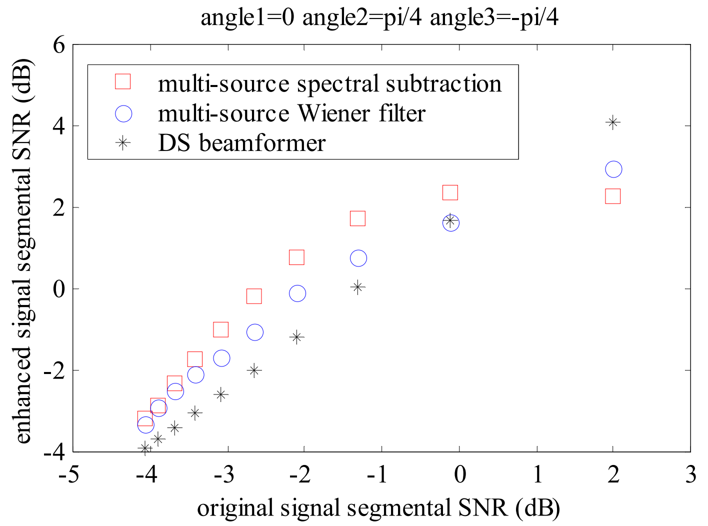

angle1=0 angle2=pi/5 angle3=-pi/5
□ multi-source spectral subtraction
○ multi-source Wiener filter
✳ MVDR beamformer
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
2
0
-2
-4
-6
-6
-5
-4
-3
-2
-1
0
1

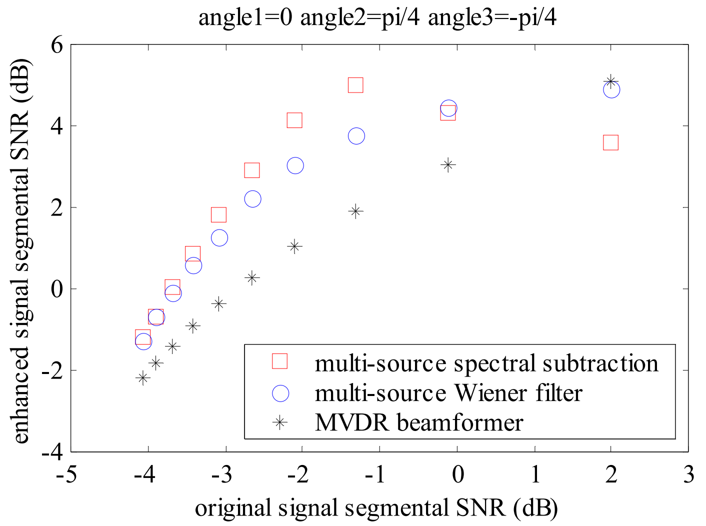
Experiment 4:

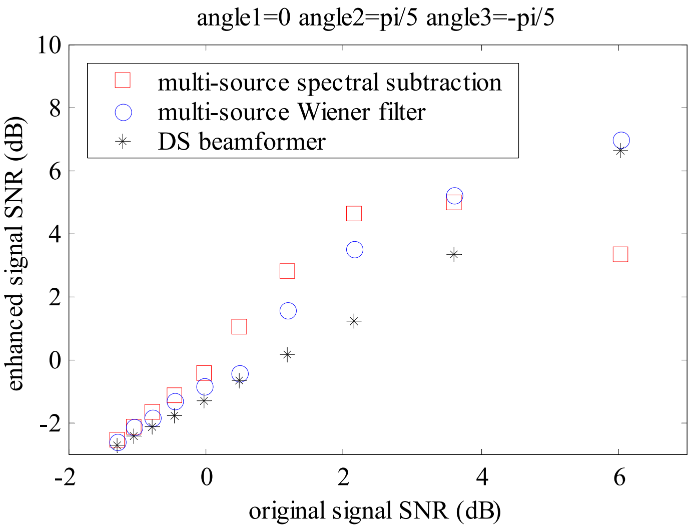


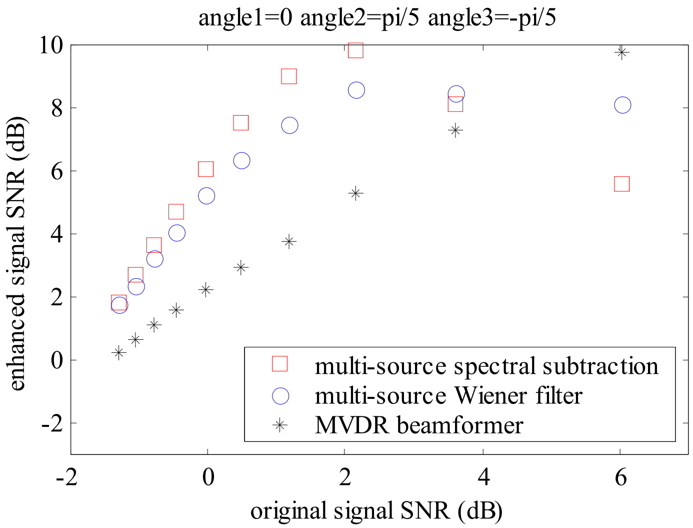

angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
10
8
6
4
2
0
-2
-4
-2
0
2
4
6
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer

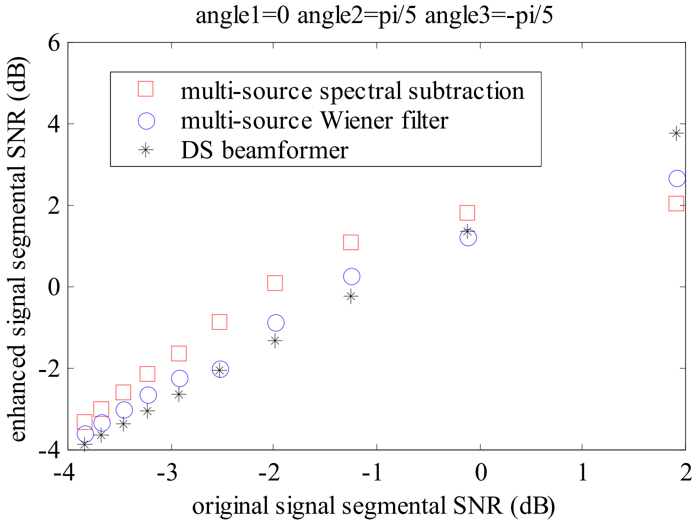

angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
10
8
6
4
2
0
-2
-4
-2
0
2
4
6
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer

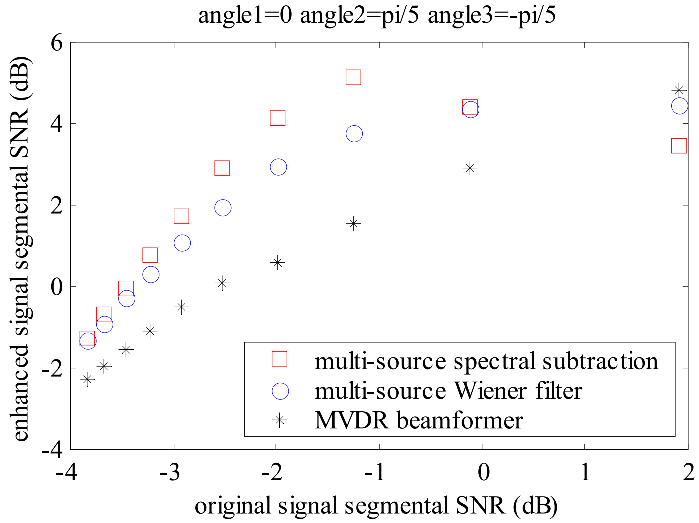

angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
6
4
2
0
-2
-4
-5
-4
-3
-2
-1
0
1
2
3
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer


angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
6
4
2
0
-2
-4
-5
-4
-3
-2
-1
0
1
2
3
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer

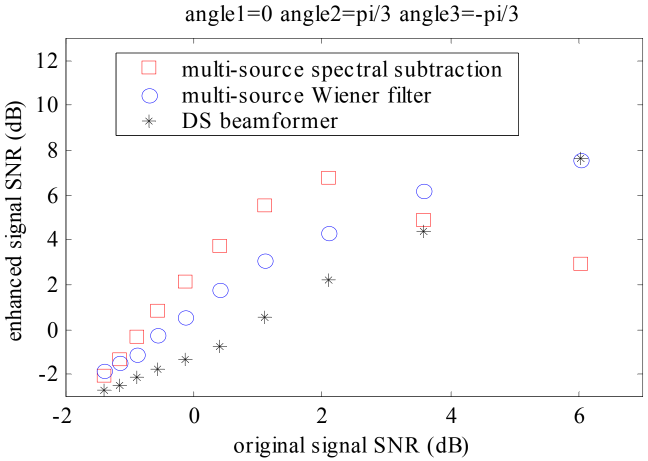

angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
* DS beamformer
10
8
6
4
2
0
-2
-2
0
2
4
6

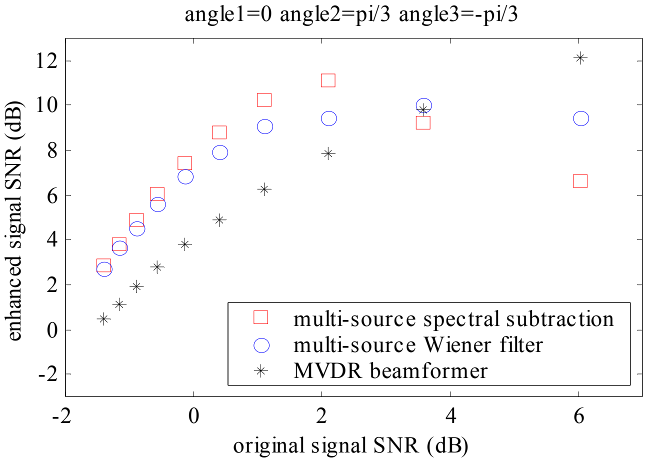

angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal SNR (dB)
original signal SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
* MVDR beamformer
10
8
6
4
2
0
-2
-2
0
2
4
6

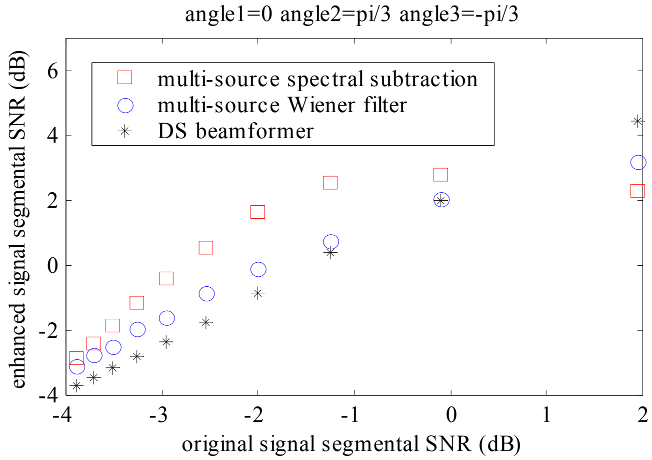

angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
* DS beamformer
6
4
2
0
-2
-4
-4
-3
-2
-1
0
1
2

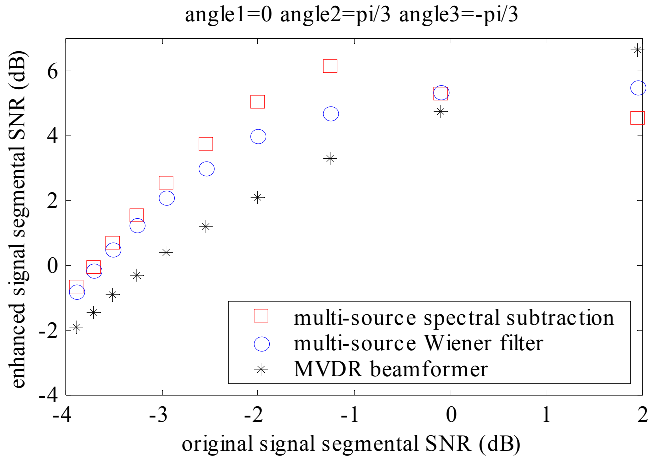

angle1=0 angle2=pi/5 angle3=-pi/5
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
□ multi-source spectral subtraction
○ multi-source Wiener filter
* MVDR beamformer
6
4
2
0
-2
-4
-4
-3
-2
-1
0
1
2

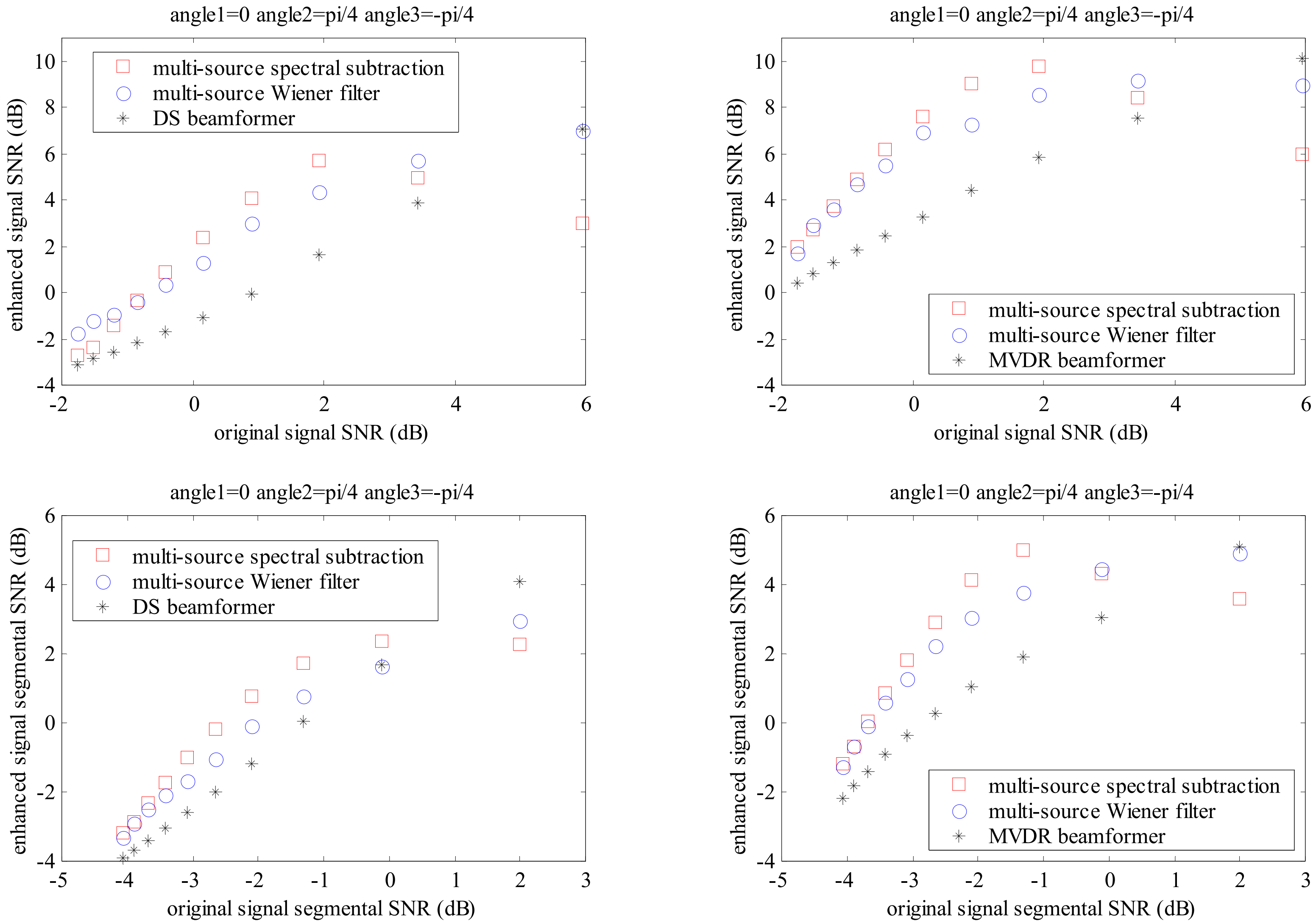
Experiment 5:

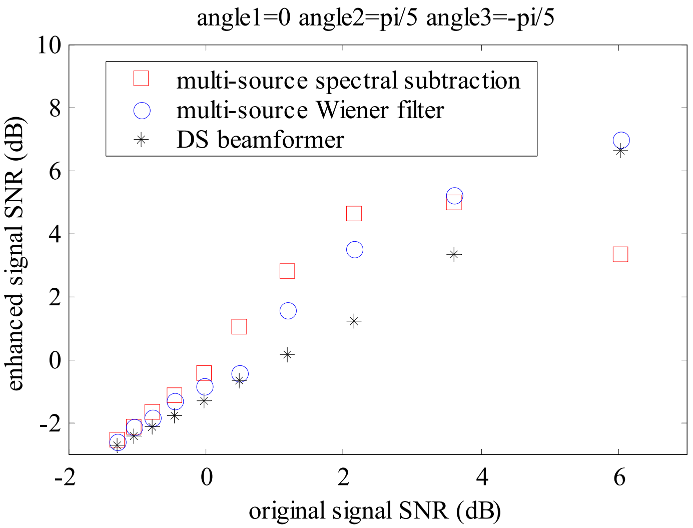


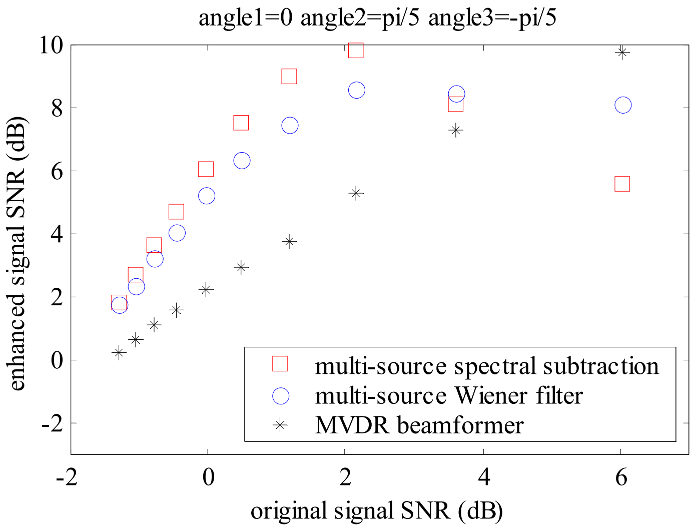


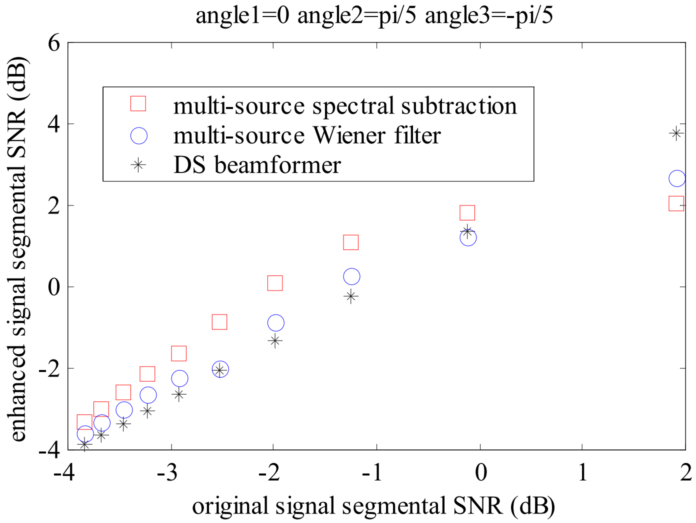


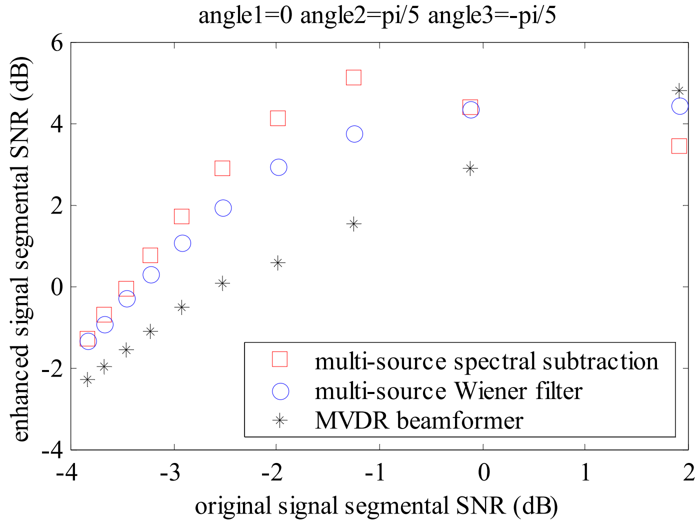



angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
-2
0
2
4
6
8
10
-4
angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal SNR (dB)
original signal SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
-5
-4
-3
-2
-1
0
1
2
3
angle1=0 angle2=pi/4 angle3=-pi/4
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer


angle1=0 angle2=pi/5 angle3=-pi/5
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
enhanced signal SNR (dB)
original signal SNR (dB)
10
8
6
4
2
0
-2
-2
0
2
4
6


angle1=0 angle2=pi/5 angle3=-pi/5
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
enhanced signal SNR (dB)
original signal SNR (dB)
10
8
6
4
2
0
-2
-2
0
2
4
6


angle1=0 angle2=pi/5 angle3=-pi/5
multi-source spectral subtraction
multi-source Wiener filter
DS beamformer
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
6
4
2
0
-2
-4
-4
-3
-2
-1
0
1
2


angle1=0 angle2=pi/5 angle3=-pi/5
multi-source spectral subtraction
multi-source Wiener filter
MVDR beamformer
enhanced signal segmental SNR (dB)
original signal segmental SNR (dB)
6
4
2
0
-2
-4
-4
-3
-2
-1
0
1
2

Appendix B: MOS Test Form

Mean Opinion Score Test

Rating	Speech Quality	Level of Distortion
5	Excellent	Imperceptible
4	Good	Just perceptible but not annoying
3	Fair	Perceptible and slightly annoying
2	Poor	Annoying but not objectionable
1	Unsatisfactory	Very annoying and objectionable

Circle the rating of the speech signal sample

MARQUETTE UNIVERSITY

This is to certify that we have examined this copy of the master's thesis by Heather E. Ewalt, B.A. and have found that it is complete and satisfactory in all respects.

The thesis has been approved by:

____

Michael T. Johnson, Ph.D.

Thesis Director, Department of Electrical and Computer Engineering

____

Kristina M. Ropella, Ph.D.

Thesis Committee Member, Department of Biomedical Engineering

____

James A. Heinen, Ph.D.

Thesis Committee Member, Department of Electrical and Computer Engineering

Approved on ____________________________