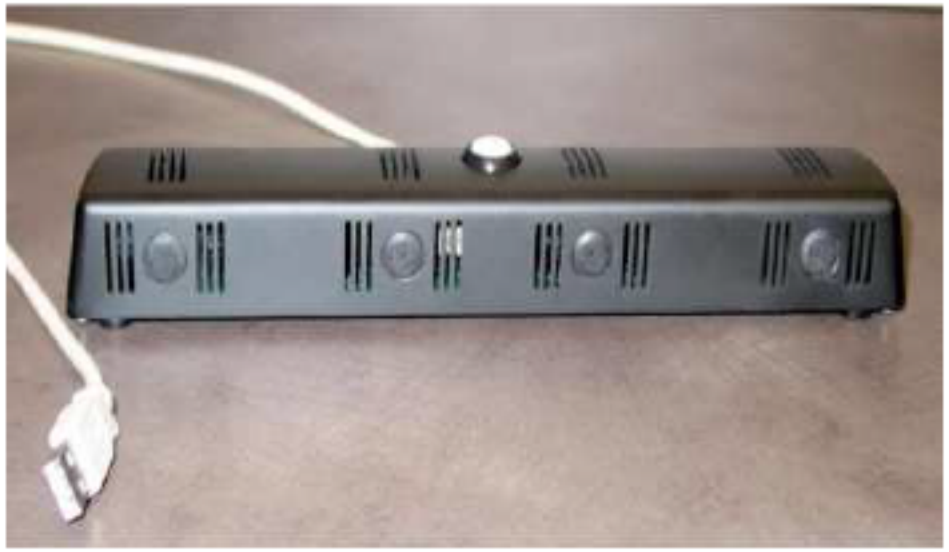
Microphone Array Processing for Speech Enhancement and Source Separation

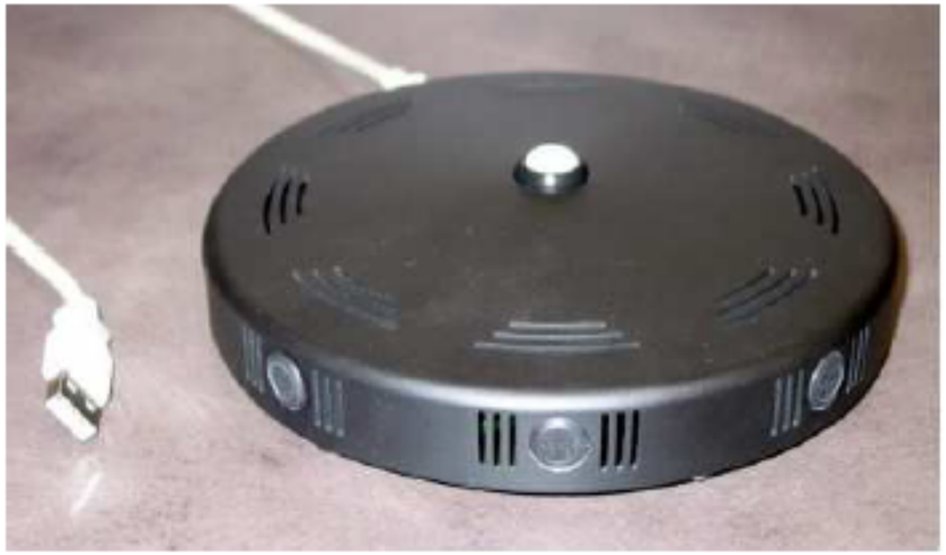
Rajesh Hegde
Professor
Dept. of EE, Kanpur

Tutorial @
National Conference on Communications (NCC 2017)
IIT Madras
Mar. 2 - 2017

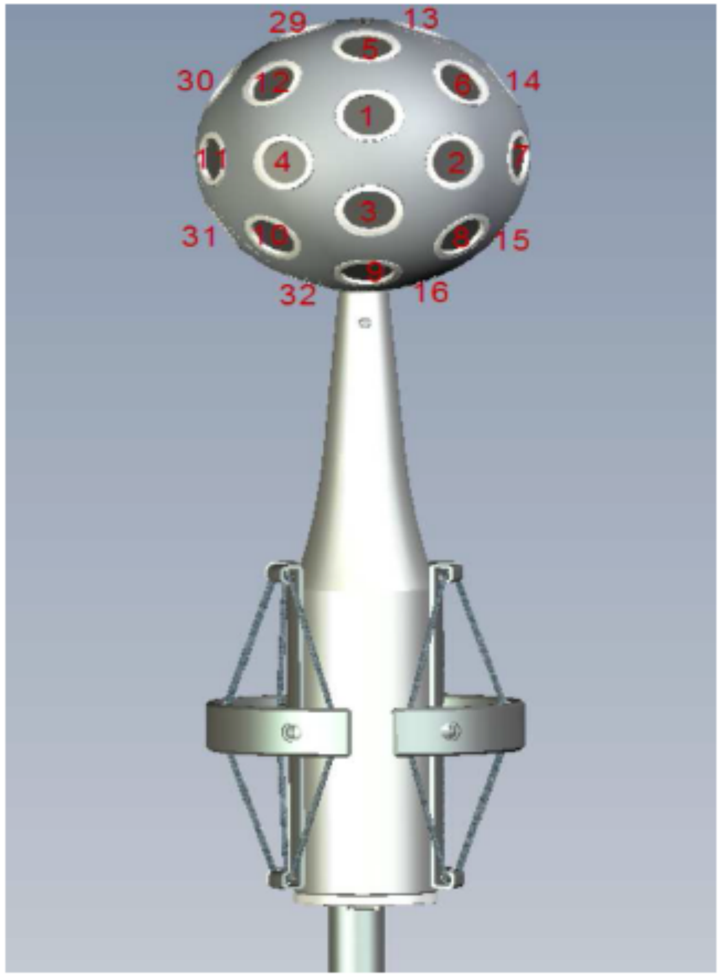
What is a Microphone Array?



Uniform Linear Array

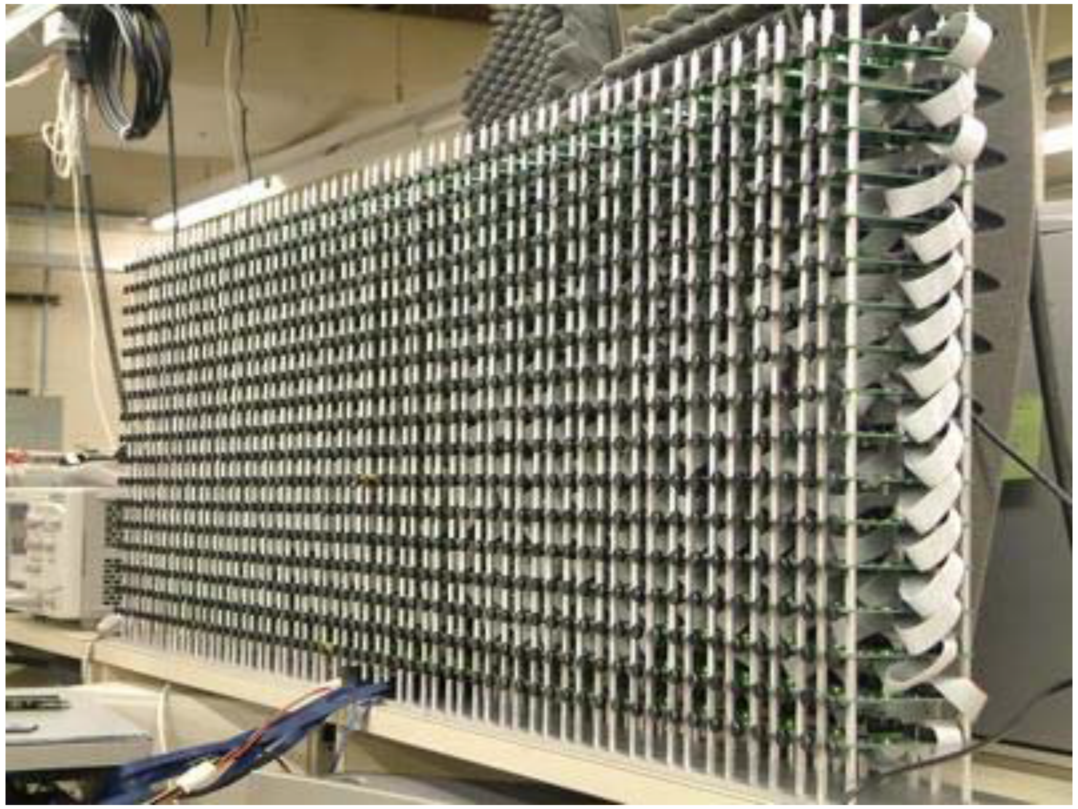


Uniform Circular Array

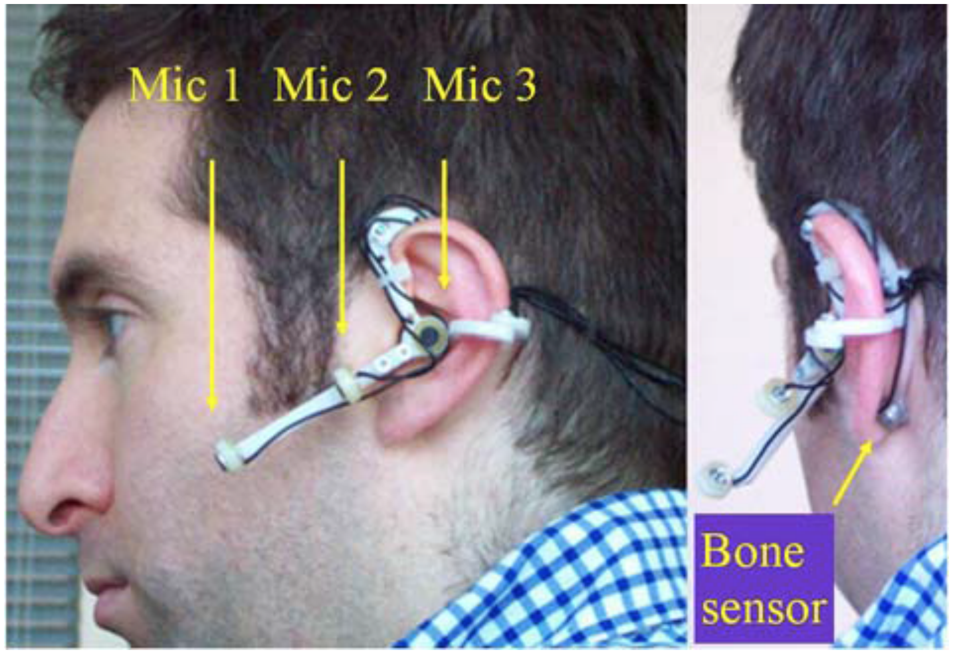


Spherical Microphone Array

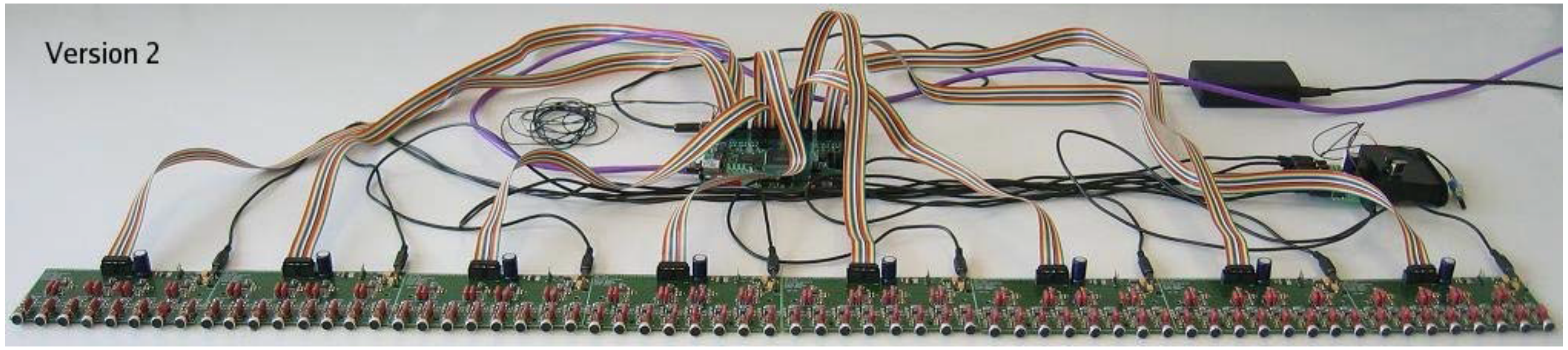
Some Interesting Microphone Arrays



MIT's LOUD array
(1020 mics)
Weinstein et al. 2005

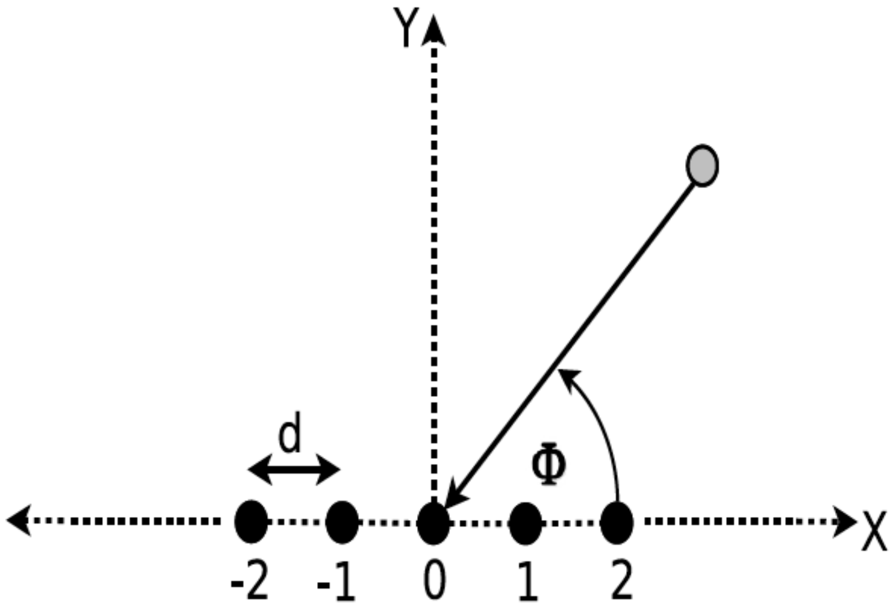


Multi-sensor headset from MSR
[Liu et al. 2005]

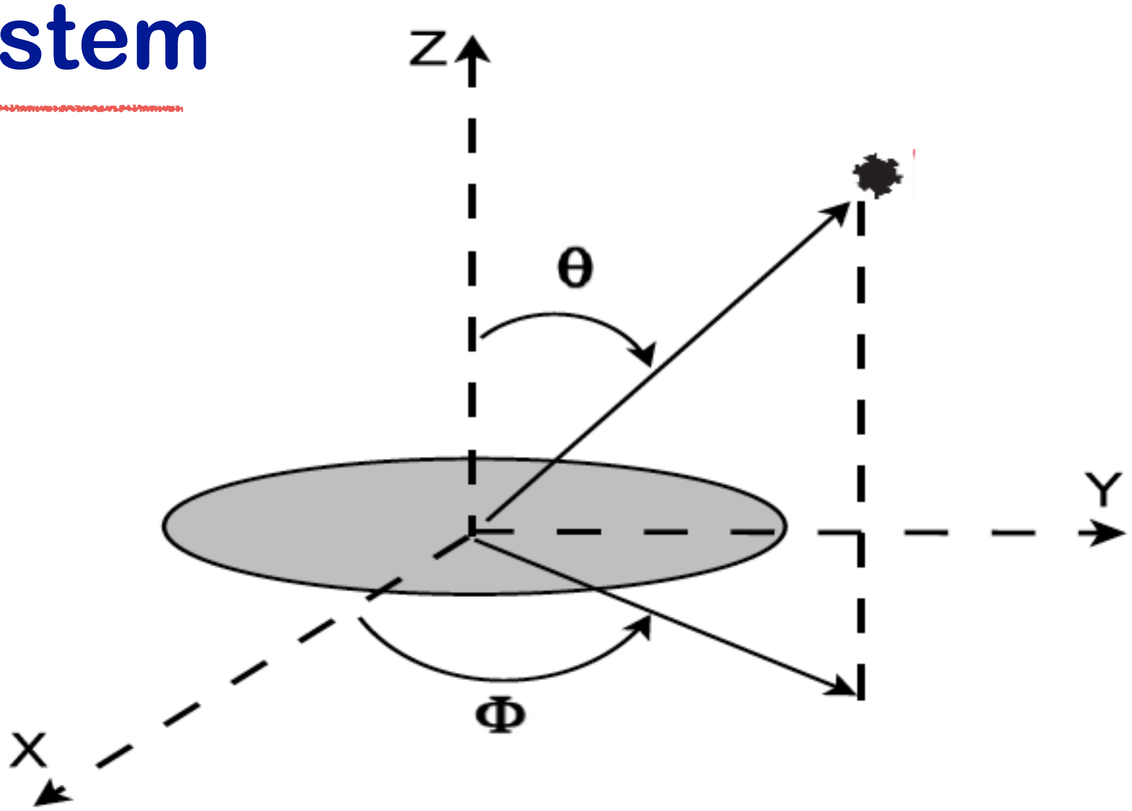


NIST's Mark III Array
(64 mics)
Stanford et al. 2004

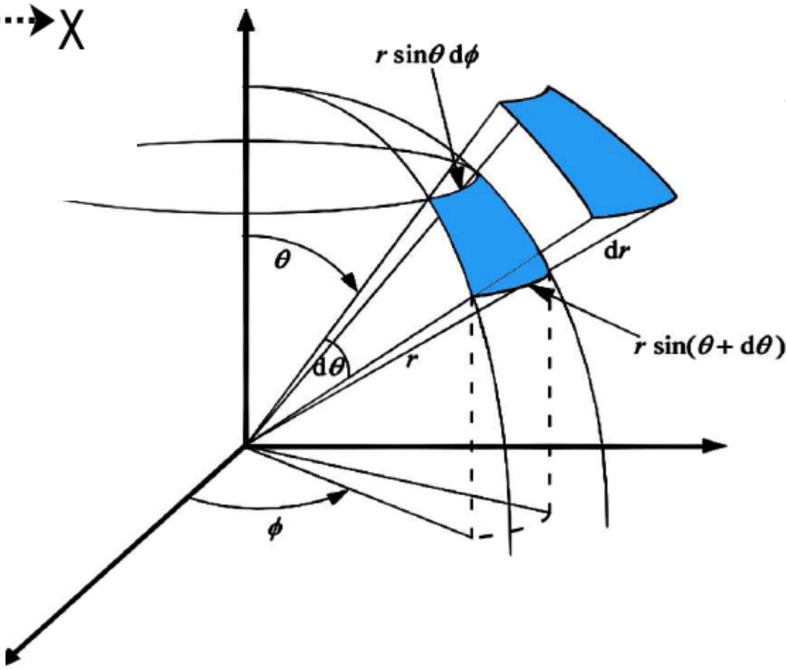
Microphone Arrays : Co-ordinate System



Uniform Linear Array



Uniform Circular Array



Spherical Microphone Array

- ▶ **ULA can measure azimuth (ϕ) only with front back ambiguity, where $\phi \in [0, \pi]$.**
- ▶ **UCA can measure azimuth (ϕ) and elevation (θ) both, where, $\phi \in [0, 2\pi]$ and $\theta \in [0, \pi/2]$.**
- ▶ **The spherical array can measure azimuth (ϕ) and elevation (θ) both, where, $\phi \in [0, 2\pi]$ and $\theta \in [0, \pi]$.**

Microphone Array Processing : Motivation

- Use of multiple microphones provides, at least in theory, exclusive advantages over a single microphone.
  - Signal enhancement : Suppressing the background and interfering signal, achieved by filter and sum - important in any sound rendering
  - Beamforming : Multiple microphones allow us to selectively capture sounds from particular direction.
  - Acoustic zoom effect : Direct sounds within the listening angular range are amplified, while other sounds are suppressed.
- Adoption of multi-microphone techniques in practical systems has not been popular until very recently.
- In real-life scenarios, the microphone array techniques provided insufficient improvement over single-microphone techniques, with increase in computational complexity, and manufacturing costs.
- Because of increased computational power and evolution of compact device technology, smartphone and hearing aid industries are utilizing microphone array, which has recently become a standard for these devices.

Microphone Array Processing (In) Parametric Spatial Sound Processing

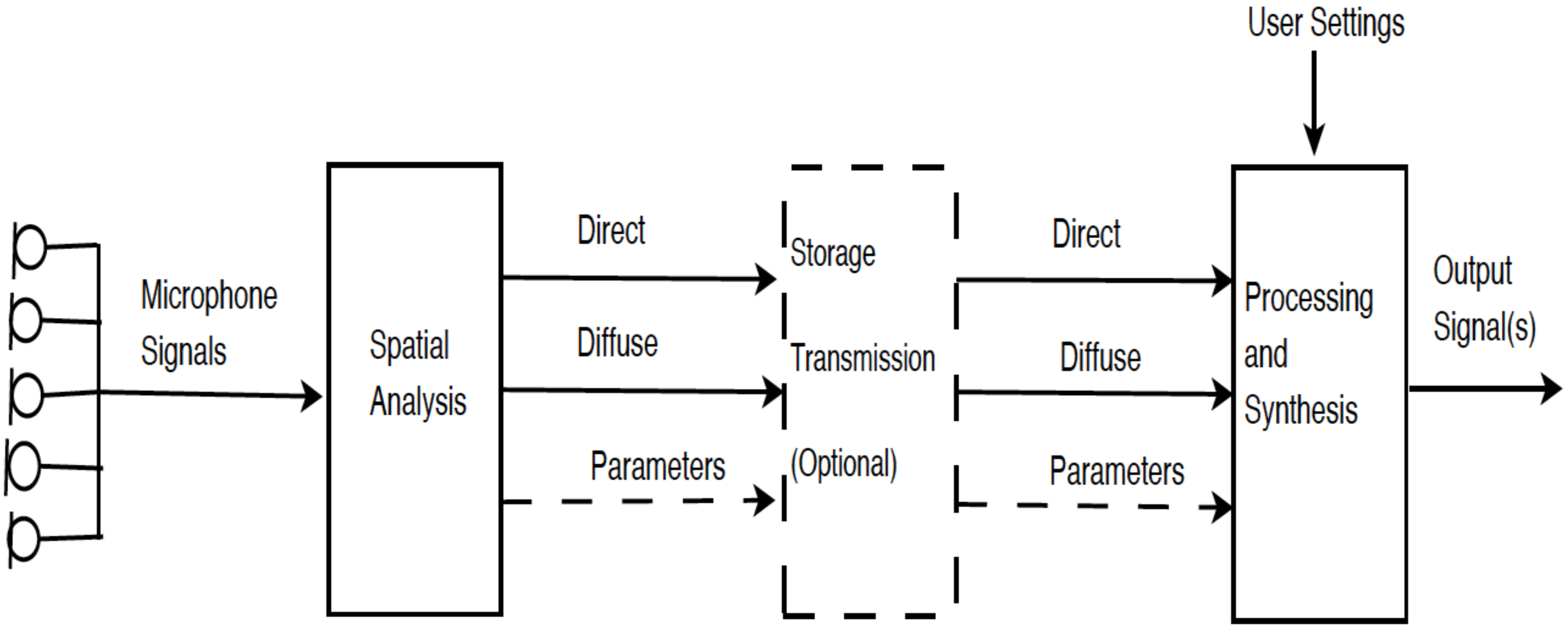
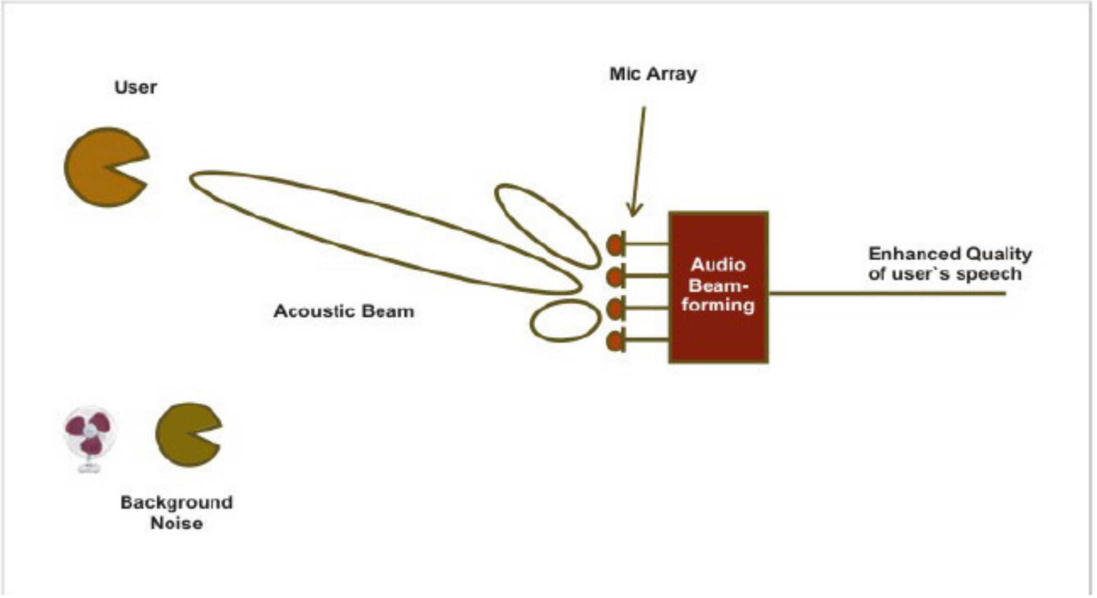


Figure: 1: An overview of the parametric spatial sound processing scheme

- Parametric processing is performed in two parts.
- In the first part, the sound field is analyzed in narrow frequency bands using multiple microphones to obtain direct and diffuse sound components and parametric information like DOAs and positions.
- In the second part, the input signals are modified, and one or more output signals are synthesized which is audio rendering.

Motivation : Socially Relevant Applications



(a)

(b)

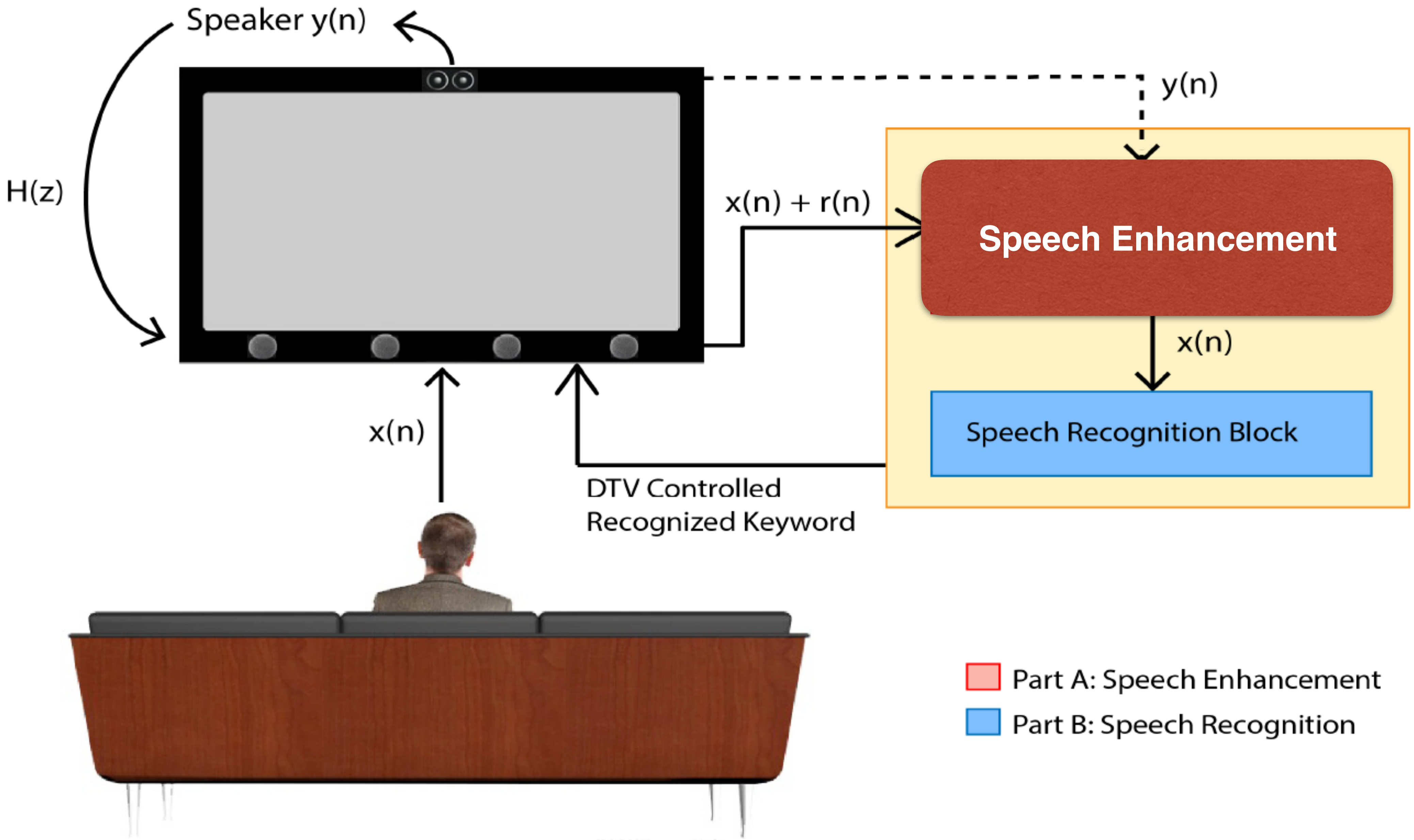
- Has daily life applications like localization and tracking of multiple sources, estimation of number of sources, noise reduction, echo cancellation, dereverberation, cocktail party, assistive living etc.
- SMA can localize sources anywhere in the space with no spatial ambiguity.
- A general approach for spherical harmonics signal processing was proposed in 2002. Most of source localization algorithms were proposed in last five years.

Motivation : Automotive Applications

Interior sound localization in car

Acoustic analysis of cavities in aircrafts during operation

Motivation : Speech Enhancement and RecognitionApplications



Speaker y(n)

H(z)

y(n)

x(n) + r(n)

Speech Enhancement

x(n)

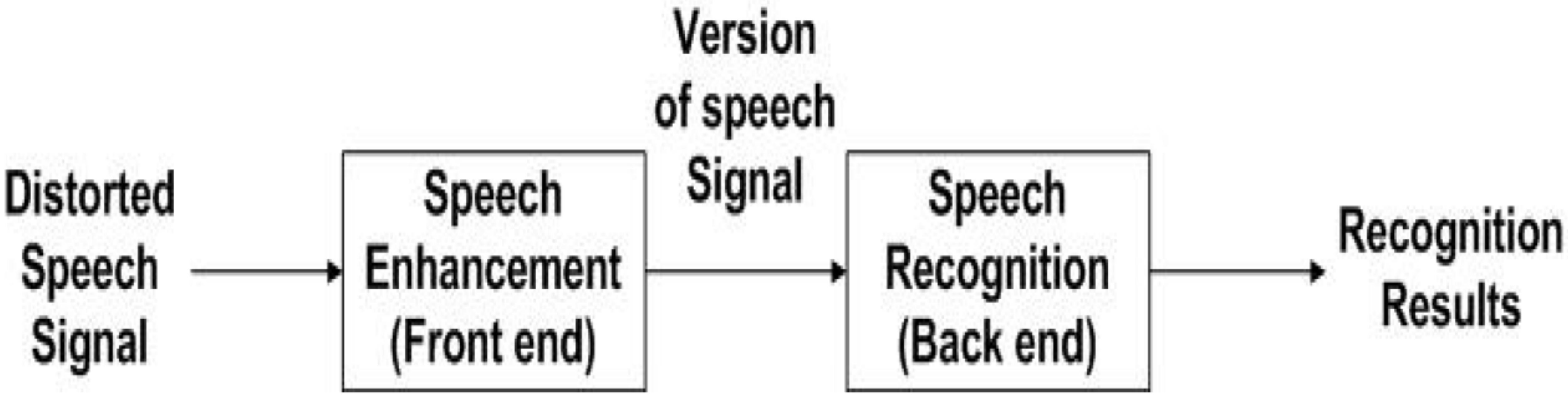
Speech Recognition Block

x(n)

DTV Controlled Recognized Keyword

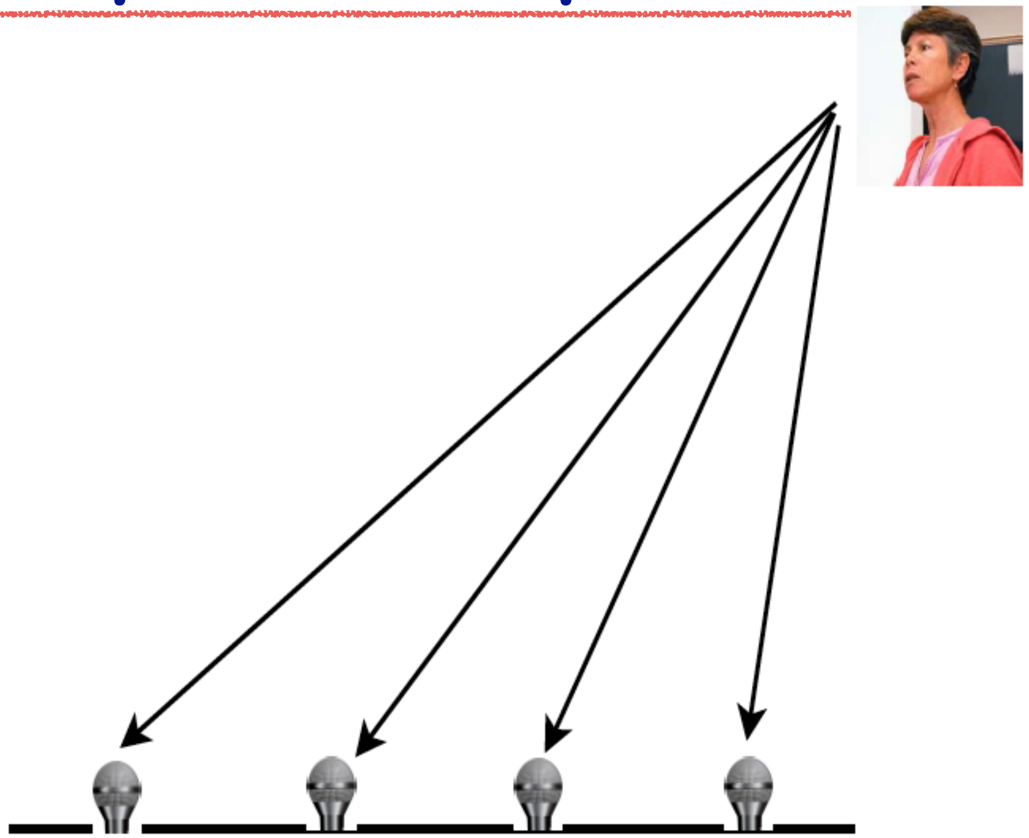
🟥 Part A: Speech Enhancement

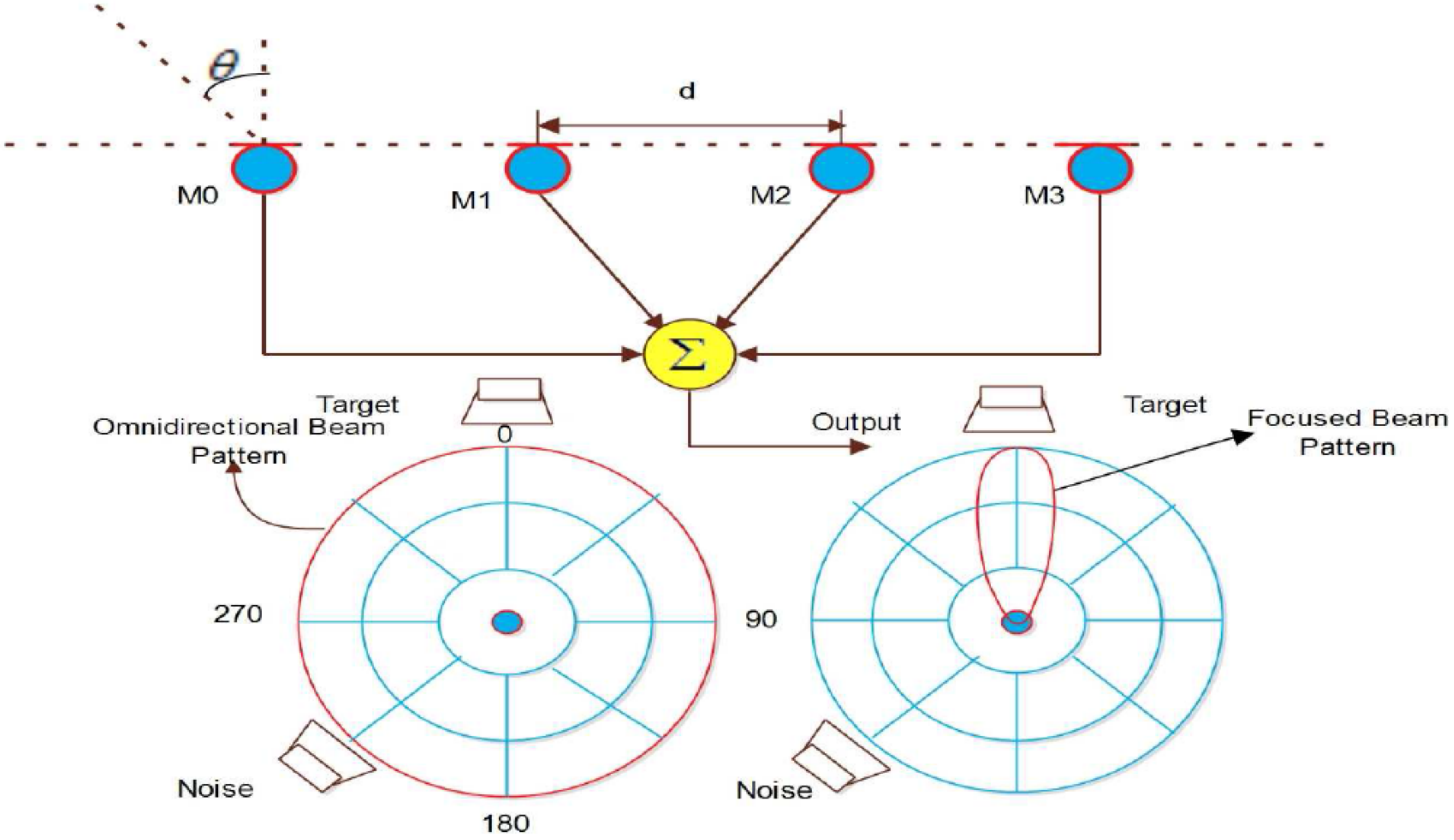
🟦 Part B: Speech Recognition



Distorted Speech Signal → Speech Enhancement (Front end) → Version of speech Signal → Speech Recognition (Back end) → Recognition Results

Close Talking and Distant Speech Acquisition





Far-field and Near-field Microphone Set-up[1]

- Far-field (Fraunhofer) region is defined by $r > \frac{2D^2}{\lambda}$. The Near Field (Fresnel) is given by $0.62\sqrt{\frac{D^3}{\lambda}} < r < \frac{2D^2}{\lambda}$.
- For far-field, the incident wavefront is planar while for near-field the incident wavefront is spherical.
- Far-field source localization refers estimating DOAs only, while near-field source localization refers estimating DOA and range both.

[1] Courtesy : http://www.clearone.com/

Basic Concepts of Spatial Filtering

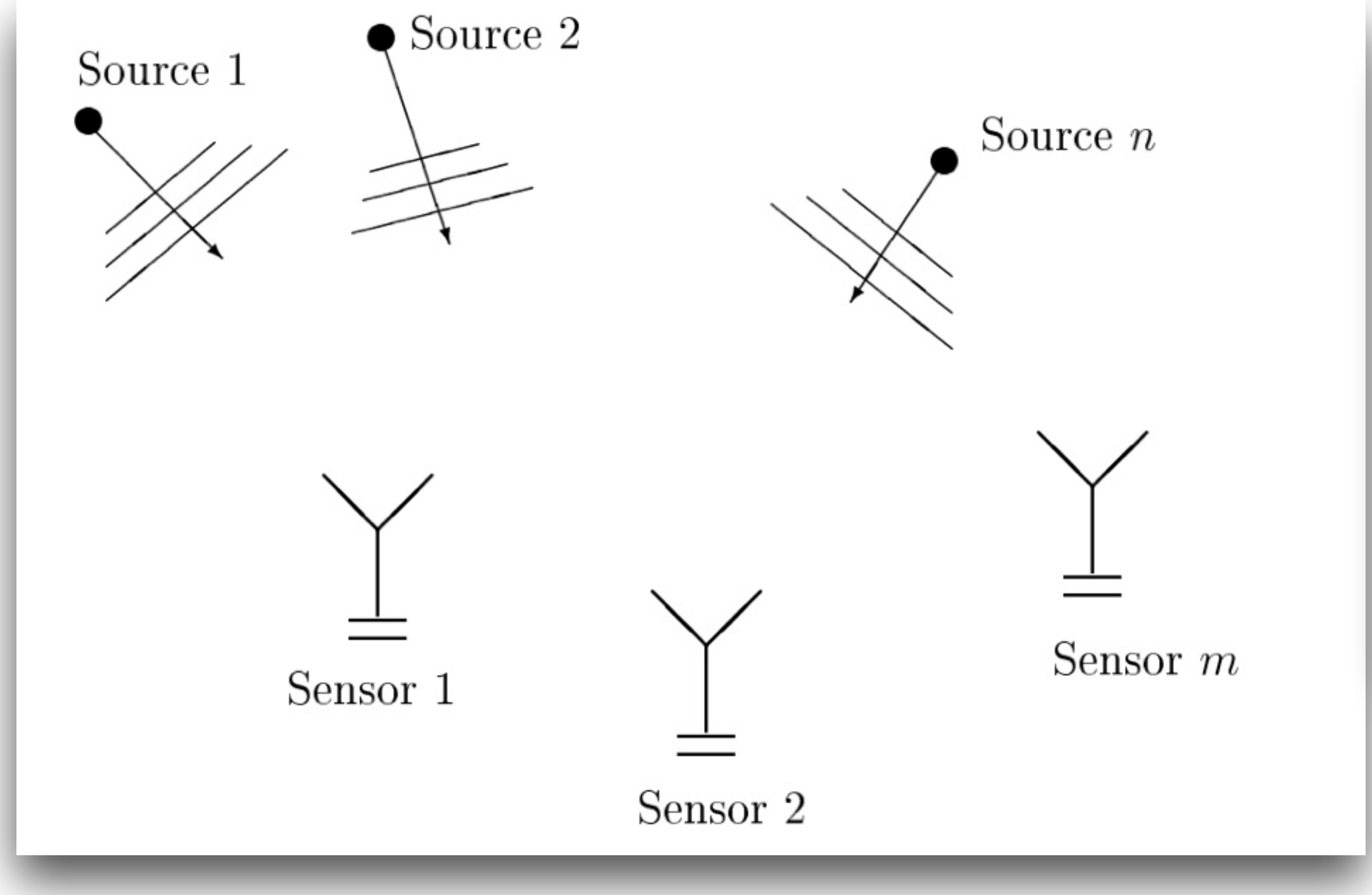
Material drawn from Stoica’s Slides (S13-S26)

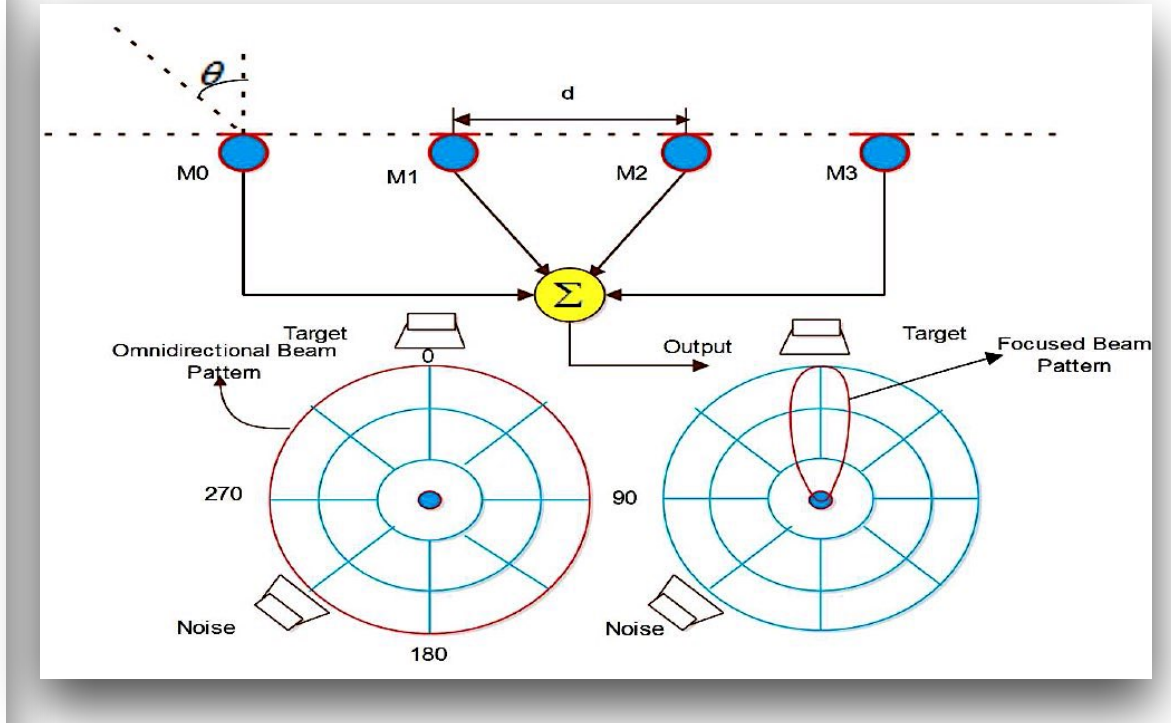
Spatial Filter

- To pass the signal of interest only, hence filtering out interferences located outside the filter's beam (but possibly having the same temporal characteristics as the signal).

- To locate an emitter in the field of view, by sweeping the filter through the DOA range of interest

Spatial Filtering as a Detection Problem

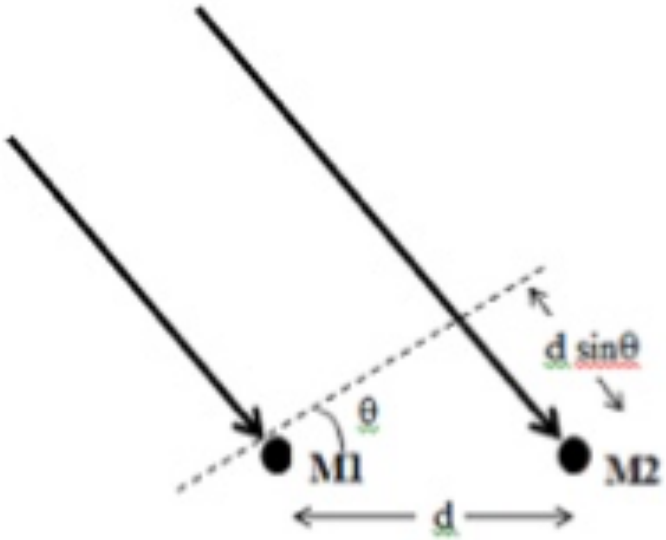




Problem: *Detect* and *locate* n radiating sources by using an array of m passive sensors.

- The number of sources n is known. (We do not treat the detection problem)
- Far-field sources in the same plane as the array of sensors

Time Delay Estimation in Spatial Filtering



$x(t) =$ the signal waveform as measured at a reference point (*e.g.*, at the "first" sensor)

$\tau_k =$ the delay between the reference point and the kth sensor

$h_k(t) =$ the impulse response (weighting function) of sensor k

$\bar{e}_k(t) =$ "noise" at the kth sensor (*e.g.*, thermal noise in sensor electronics; background noise, etc.)

Time Delay Estimation in Spatial Filtering

Then the output of sensor k is

$$\bar{y}_k(t) = h_k(t) * x(t - \tau_k) + \bar{e}_k(t)$$

Then the output of sensor k is

$$\bar{y}_k(t) = h_k(t) * x(t - \tau_k) + \bar{e}_k(t)$$

Basic Problem: Estimate the *time delays* $\{\tau_k\}$ with $h_k(t)$ known but $x(t)$ unknown.

we get the **complex representation**: (for $t \in \mathcal{Z}$)

$$\boxed{y_k(t) = s(t) H_k(\omega_c)\, e^{-i\omega_c \tau_k} + e_k(t)}$$

where $H_k(\omega) = \mathcal{F}\{h_k(t)\}$ is the kth sensor's transfer function

$s(t)$ is the *complex envelope* of $x(t)$

Time delay is now $\simeq$ to a *phase shift* $\omega_c \tau_k$:

Vector Data Model for a Single Narrow band Source

$$\begin{aligned} \theta &= \text{the emitter DOA} \\ m &= \text{the number of sensors} \\ a(\theta) &= \begin{bmatrix} H_1(\omega_c)\, e^{-i\omega_c\tau_1} \\ \vdots \\ H_m(\omega_c)\, e^{-i\omega_c\tau_m} \end{bmatrix} \end{aligned}$$

$$y(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_m(t) \end{bmatrix} \qquad e(t) = \begin{bmatrix} e_1(t) \\ \vdots \\ e_m(t) \end{bmatrix}$$

Then

$$\boxed{y(t) = a(\theta)s(t) + e(t)}$$

NOTE: θ enters $a(\theta)$ via both $\{\tau_k\}$ and $\{H_k(\omega_c)\}$. For *omnidirectional* sensors the $\{H_k(\omega_c)\}$ do not depend on θ.

Vector Data Model for Multiple Narrow band Sources

received signals: $\{s_k(t)\}_{k=1}^{n}$

DOAs: θ_k

$$y(t) = a(\theta_1)s_1(t) + \cdots + a(\theta_n)s_n(t) + e(t)$$

Let

$$A = [a(\theta_1) \ldots a(\theta_n)], \; (m \times n)$$

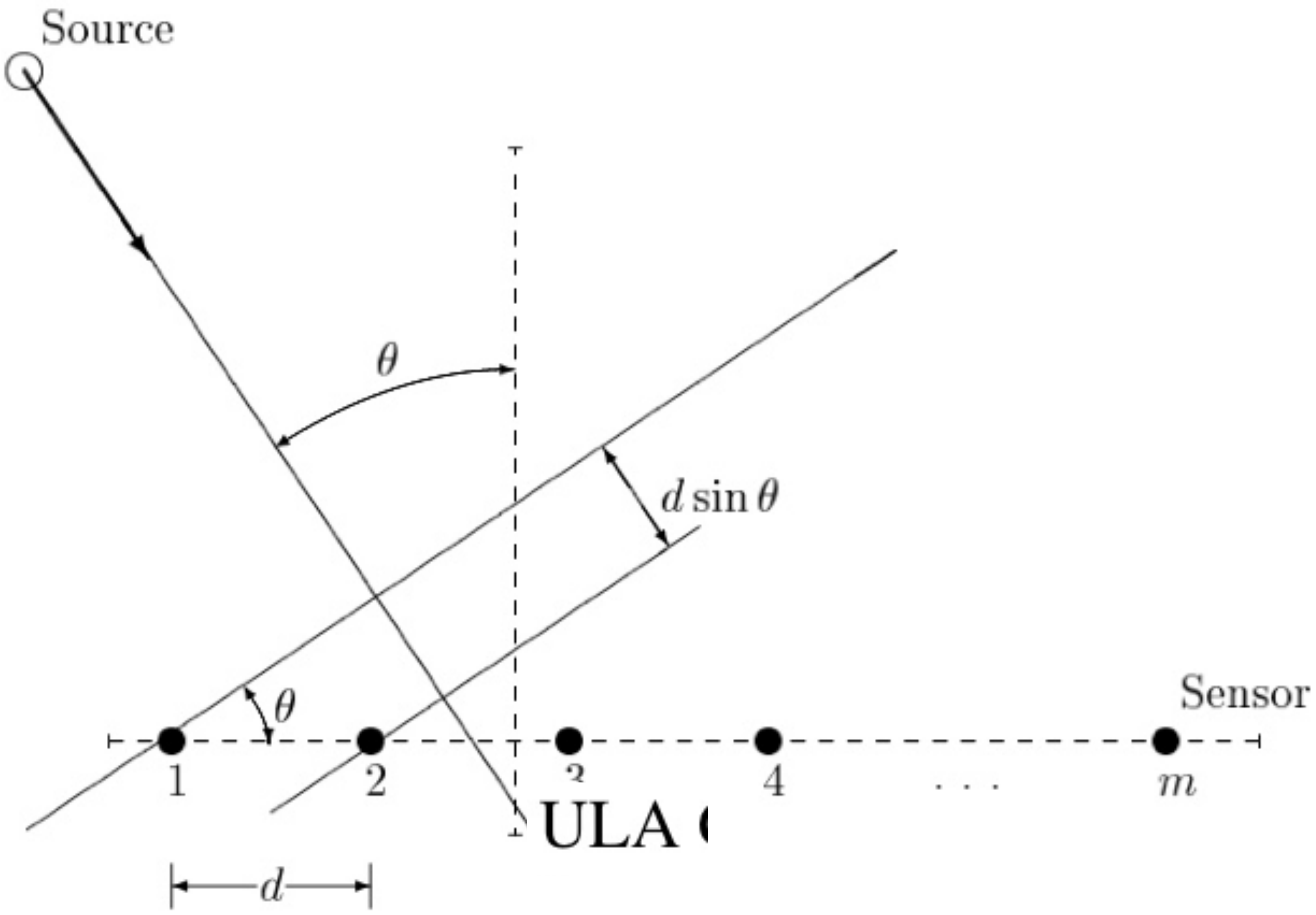
$$s(t) = [s_1(t) \ldots s_n(t)]^T, \; (n \times 1)$$

the **array equation** is:

$$\boxed{y(t) = As(t) + e(t)}$$

Use the *planar wave* assumption to find the dependence of τ_k on θ.

Computing the Time Delay



Time Delay for sensor k:

$$\tau_k = (k-1)\,\frac{d\sin\theta}{c}$$

where c = wave propagation speed

Spatial Sampling (Theorem)

Let:

$$\begin{aligned}\omega_s &\overset{\triangle}{=} \omega_c \frac{d \sin\theta}{c} = 2\pi \frac{d \sin\theta}{c/f_c} = 2\pi \frac{d \sin\theta}{\lambda} \\ \lambda &= c/f_c = \textit{signal wavelength} \\ a(\theta) &= [1, e^{-i\omega_s} \; \ldots \; e^{-i(m-1)\omega_s}]^T\end{aligned}$$

By direct analogy with the vector $a(\omega)$ made from uniform samples of a *sinusoidal time series*,

$$\omega_s = \text{spatial frequency}$$

The function $\omega_s \mapsto a(\theta)$ is *one-to-one* for

$$|\omega_s| \leq \pi \leftrightarrow \frac{d|\sin\theta|}{\lambda/2} \leq 1 \leftarrow \boxed{d \leq \lambda/2}$$

$$d = \text{spatial sampling period}$$

$d \leq \lambda/2$ is a **spatial** Shannon sampling theorem.

Spatial Filtering : Comparison with a Temporal Filter (FIR)

$$\begin{aligned} y_F(t) &= \sum_{k=0}^{m-1} h_k u(t-k) = h^* y(t) \\ h &= [h_o \ldots h_{m-1}]^* \\ y(t) &= [u(t) \ldots u(t-m+1)]^T \end{aligned}$$

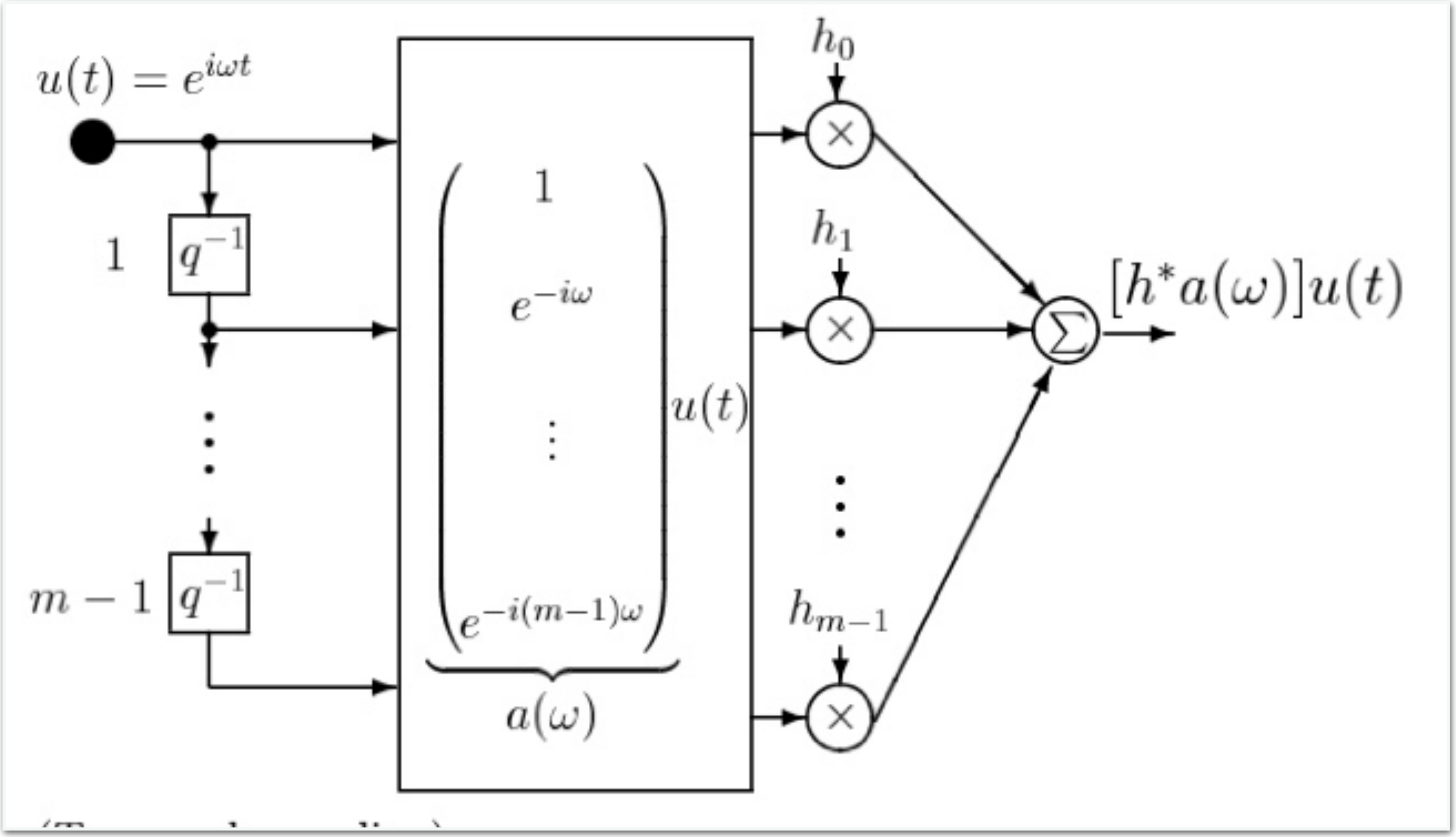
If $u(t) = e^{i\omega t}$ then

$$y_F(t) = \underbrace{[h^* a(\omega)]}_{\text{filter transfer function}} u(t)$$

$$a(\omega) = [1, e^{-i\omega} \ \ldots \ e^{-i(m-1)\omega}]^T$$

We can select h to enhance or attenuate signals with different frequencies ω.

Spatial Filtering : Comparison with a Temporal Filter (FIR)



Spatial Filtering (FIR)

$\{y_k(t)\}_{k=1}^{m}$ = the "spatial samples" obtained with a sensor array.

Spatial FIR Filter output:

$$y_F(t) = \sum_{k=1}^{m} h_k y_k(t) = h^* y(t)$$

Spatial Filtering (FIR) for Narrow Band Sources

Narrowband Wavefront: The array's (noise-free) response to a narrowband (∼ sinusoidal) wavefront with complex envelope $s(t)$ is:

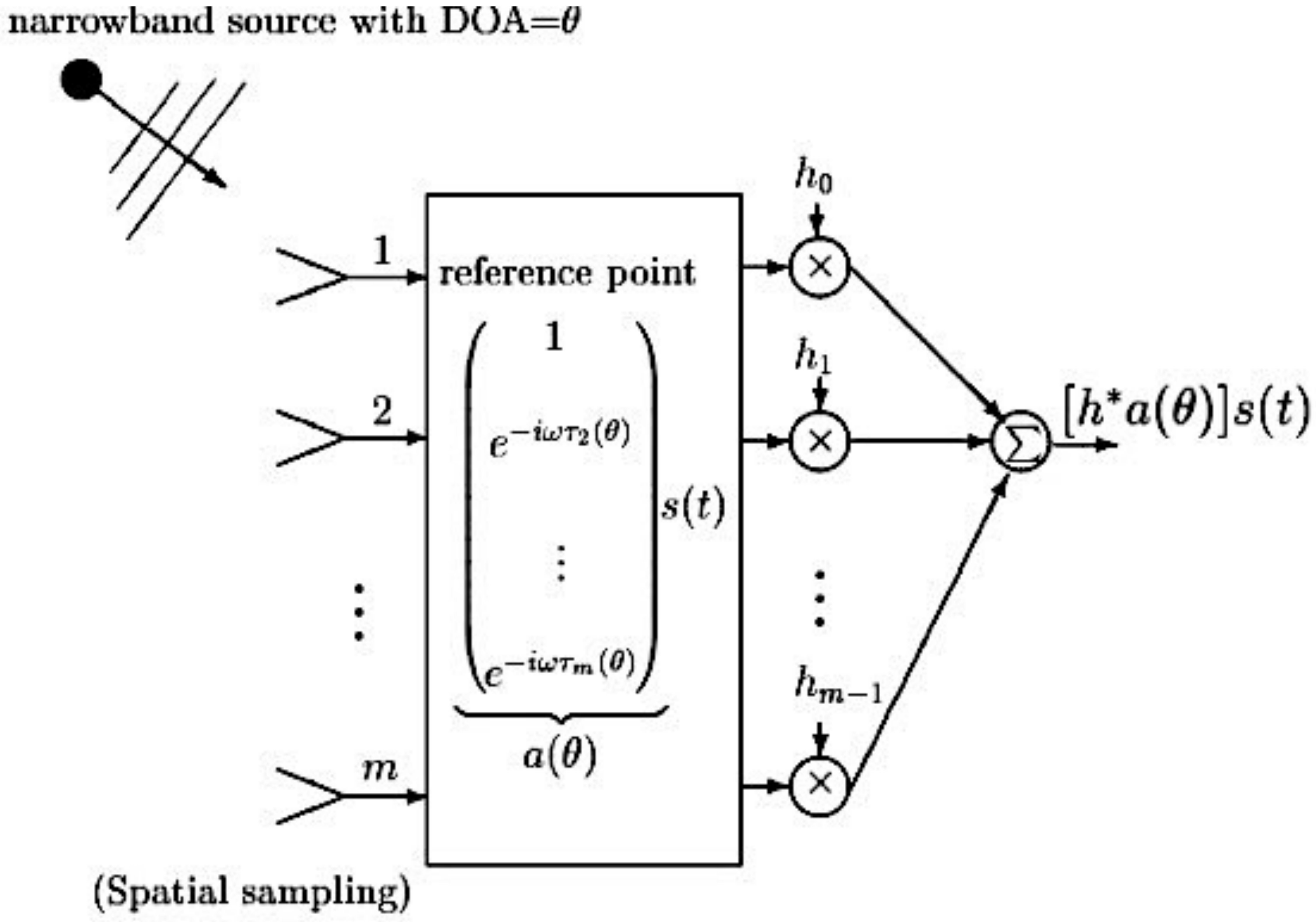
$$\begin{aligned} y(t) &= a(\theta)s(t) \\ a(\theta) &= [1, e^{-i\omega_c\tau_2} \ \ldots \ e^{-i\omega_c\tau_m}]^T \end{aligned}$$

The corresponding filter output is

$$\boxed{y_F(t) = \underbrace{[h^* a(\theta)]}_{\text{filter transfer function}} s(t)}$$

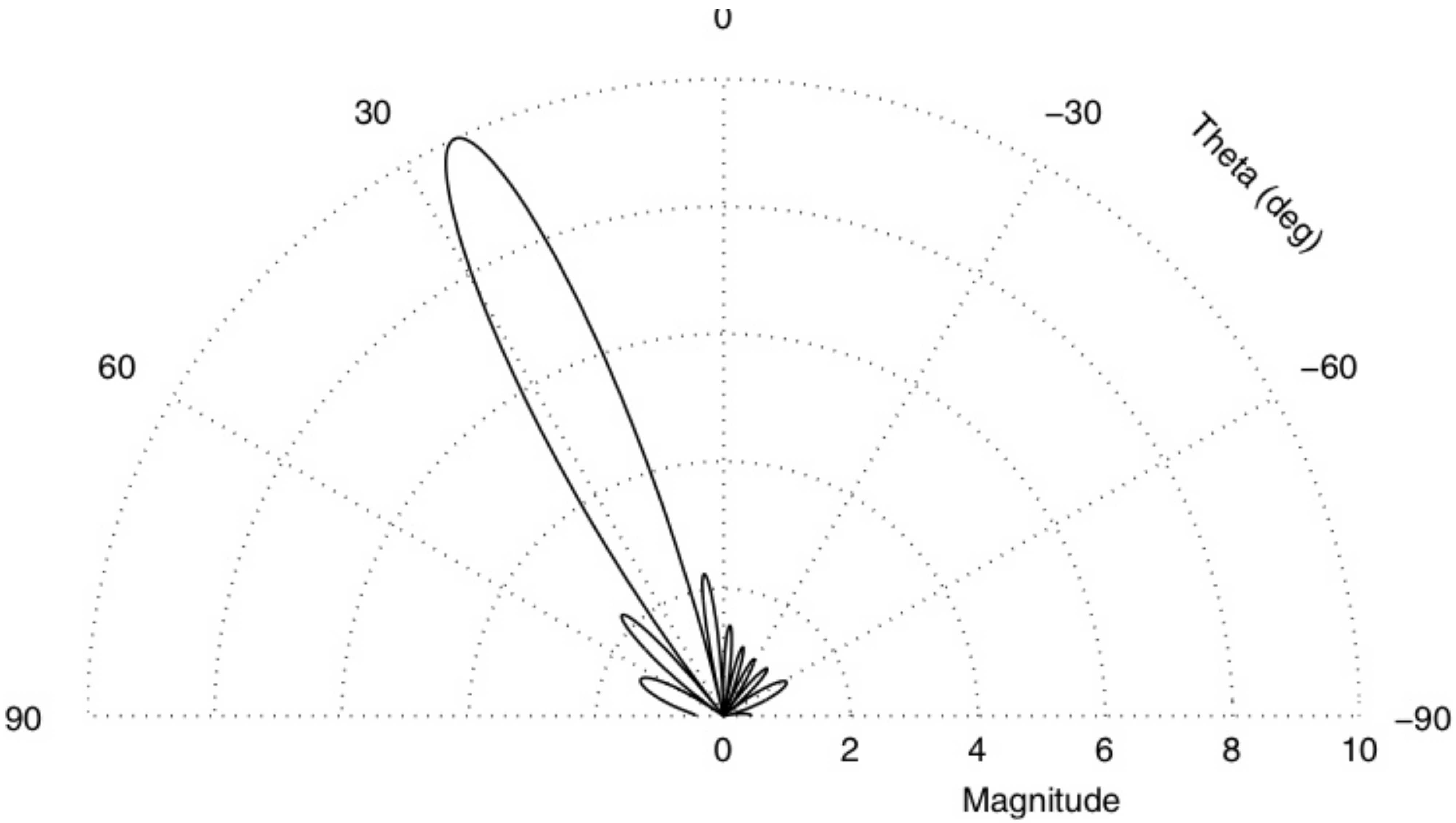
We can select h to enhance or attenuate signals coming from different DOAs.

Spatial Filtering (FIR) for Narrow Band Sources


narrowband source with DOA=θ
1
2
m
reference point
$\begin{pmatrix} 1 \\ e^{-i\omega\tau_2(\theta)} \\ \vdots \\ e^{-i\omega\tau_m(\theta)} \end{pmatrix}$
$s(t)$
$a(\theta)$
h_0
h_1
h_{m-1}
$\times$
Σ
$[h^* a(\theta)]s(t)$
(Spatial sampling)

Spatial Filtering Beam (Directivity) Pattern for Narrow Band Sources

Example: The response magnitude $|h^* a(\theta)|$ of a spatial filter (or beamformer) for a 10-element ULA. Here, $h = a(\theta_0)$, where $\theta_0 = 25°$



Time Delay/DOA Estimation Methods

Non Parametric Methods

Filter Sum Beamforming
Capon Beamforming
LCMV Beamforming

Parametric Methods

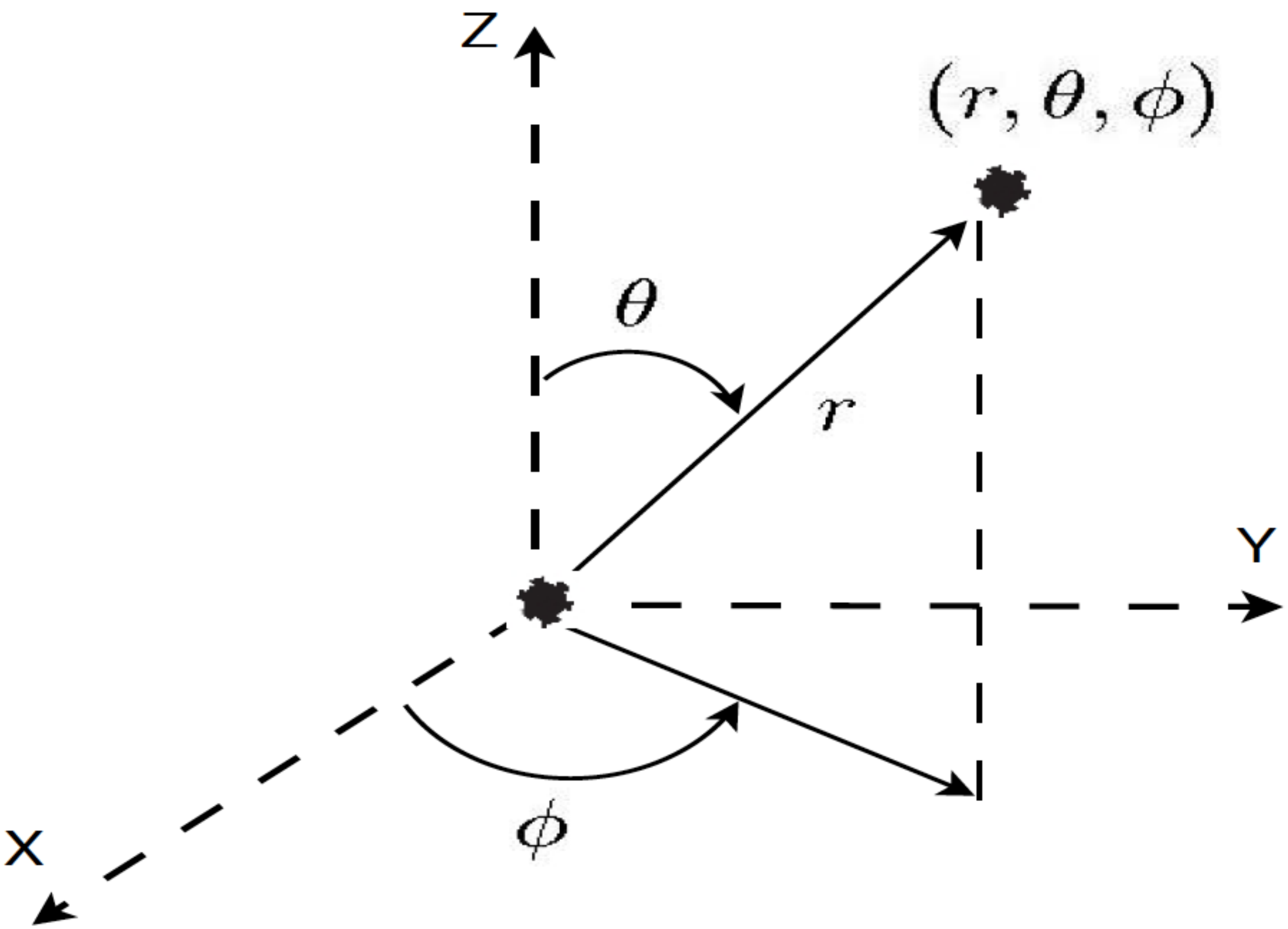
Non Linear Least Squares
MUSIC, root-MUSIC
Min Norm, Esprit

Correlation based Methods

GCC, GCC-PHAT, GCC-ROTH

Spherical Co-ordinate System

- Location of a source is given by $\mathbf{r}_l = (r_l, \Psi_l)$, with $\Psi_l = (\theta_l, \phi_l)$.
- Location of a receiver is denoted as $\mathbf{r}_i = (r_i, \mathbf{\Phi}_i)$, where $\mathbf{\Phi}_i = (\theta_i, \phi_i)$.
- The range (r), elevation (θ) and azimuth (ϕ) takes values as $r \in (0, \infty)$, $\theta \in [0, \pi]$, $\phi \in [0, 2\pi)$



Vector Data Model for Far Field Sources in Spherical Co-ordinates

- A sound field of L far-field sources with wavenumber $k = \frac{\omega_c}{c}$, is incident on an arbitrary microphone array of I microphones, $I > L$.
- In spatial domain, the sound pressure, $\mathbf{p}(t) = [p_1(t), p_2(t), \ldots, p_I(t)]^T$, is written as,

$$\mathbf{p}(t) = \mathbf{A}(\Psi, k)\mathbf{s}(t) + \mathbf{v}(t) \ , t = 1, 2, \cdots, N_s$$

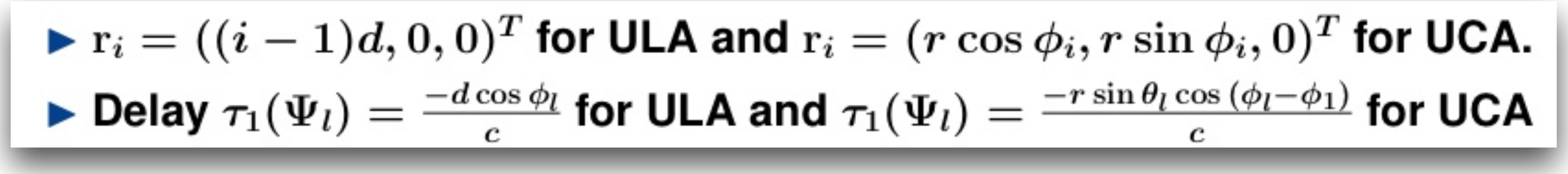
- $\mathbf{A}(\Psi, k)$ is $I \times L$ steering matrix, $\mathbf{s}(t)$ is $L \times N_s$ vector of signal amplitudes $\mathbf{v}(t)$ is $I \times N_s$ baseband additive white Gaussian sensor noise.

- The steering matrix $\mathbf{A}(\Psi, k)$ is expressed as

$$\mathbf{A}(\Psi, k) = [\mathbf{a}_1(\Psi_1, k), \mathbf{a}_2(\Psi, k), \ldots, \mathbf{a}_L(\Psi, k)], \text{ where}$$

◆ $\mathbf{a}_l(\Psi_l, k) = [e^{-j\mathbf{k}_l^T\mathbf{r}_1}, e^{-j\mathbf{k}_l^T\mathbf{r}_2}, \ldots, e^{-j\mathbf{k}_l^T\mathbf{r}_I}]^T$ and $\mathbf{k}_l^T\mathbf{r}_i = \omega_c\tau_i(\Psi_l)$

- $\mathbf{k}_l = -(k \sin\theta_l \cos\phi_l, k \sin\theta_l \sin\phi_l, k \cos\theta_l)^T$, with $\theta_l = \pi/2$ for ULA.



- ▶ $\mathrm{r}_i = ((i-1)d, 0, 0)^T$ **for ULA and** $\mathrm{r}_i = (r\cos\phi_i, r\sin\phi_i, 0)^T$ **for UCA.**
- ▶ **Delay** $\tau_1(\Psi_l) = \frac{-d\cos\phi_l}{c}$ **for ULA and** $\tau_1(\Psi_l) = \frac{-r\sin\theta_l\cos(\phi_l-\phi_1)}{c}$ **for UCA**

Computing the Time Delay

- The position vector of i^{th} sensor in ULA is

$$\mathbf{r}_i = \left[(i-1)d \quad 0 \quad 0\right]^T.$$

- The elevation angle θ is 90° for an ULA. Therefore, the expression for wavevector becomes

$$\mathbf{k}_l = -\left[k\cos\phi_l \quad k\sin\phi_l \quad 0\right]^T.$$

- The propagation delay at the i^{th} sensor can now be written as

$$\tau_i(\Psi_l) = \frac{-(i-1)d\cos\phi_l}{c}, \; i = 1, 2, \cdots, I.$$

the steering vector for ULA takes the form as

$$\mathbf{a}_l(\phi_l, k) = \left[1 \quad e^{jkd\cos\phi_l} \quad e^{j2kd\cos\phi_l} \quad \dots \quad e^{j(I-1)kd\cos\phi_l}\right]^T$$

where $k = \frac{\omega_c}{c}$

Spatial Sampling (Theorem)

- A minimum sampling frequency (called Nyquist frequency) is required to avoid aliasing in the time sampled signal, given by

$$fs = \frac{1}{T_s} \geq 2f_{max}$$

where f_{max} is maximum frequency component in frequency spectrum of signal.

- As signal is spatially sampled using antenna array, a similar condition on spatial sampling frequency for exists to avoid spatial aliasing.

$$f_{x_s} = \frac{1}{d} \geq 2f_{x_{max}}$$

where f_{x_s} is spatial sampling frequency in samples per meter, d is spatial sampling period and $f_{x_{max}}$ is the highest spatial frequency component present in spatial spectrum of the signal.

Spatial Sampling (Theorem)

- The wave vector $\mathbf{k} = -2\pi/\lambda \left[\sin\theta_l \cos\phi_l \quad \sin\theta_l \sin\phi_l \quad \cos\theta_l\right]'$, consists of spatial frequencies in x, y and z directions. Each spatial frequency denotes 2π times the number of cycles per meter.

- The spatial frequency (number of cycles per meter) in x-axis is given by

$$f_{x_s} = \frac{\sin\theta\cos\phi}{\lambda}.$$

- Maximizing numerator and minimizing denominator yields $f_{x_{max}}$ as below

$$f_{x_{max}} = \frac{1}{\lambda}.$$

- The Nyquist condition for alias free spatial sampling is given by

$$d \leq \frac{\lambda}{2}$$

Correlation Based Methods for DOA Estimation and Beamforming

Correlation based DOA Estimation

- The time delay of arrival (TDOA) is computed between a pair of sensors.
- The time delay corresponds to lag at which the cross-correlation is maximum [6].
- The DOA is estimated from time delay using

$$\tau_i(\Psi_l) = \frac{-(i-1)d\cos\phi_l}{c}, \quad i = 1, 2, \cdots, l.$$

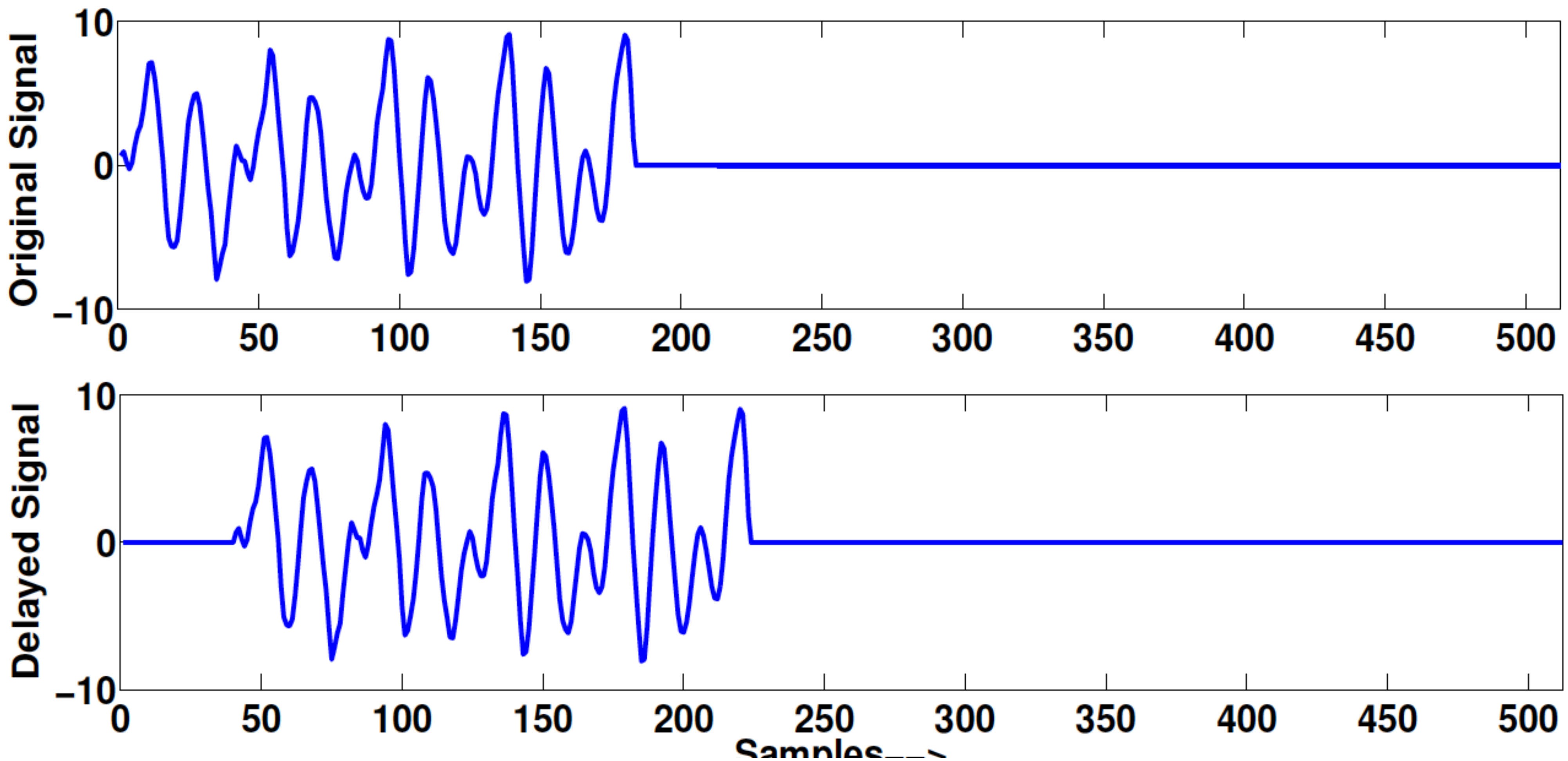


Figure: Signal with length 512 samples, original signal (top) and signal delayed by 40 samples (bottom).

Plain Time Correlation

- The plain time correlation between two observed signals, $p_1(t)$ and $p_2(t)$ is defined as
$$r^{PTC}_{p_1p_2}(l_g) = E[p_1(t)p_2^*(t - l_g)]$$
where l_g is the lag and $(.)^*$ denotes complex conjugate.
- In practice, the cross-correlation is estimated for any two finite signals as
$$\hat{r}^{PTC}_{p_1p_2}(l_g) = \sum_{t=-N_s}^{N_s} p_1(t)p_2^*(t - l_g).$$
- The TDOA can be estimated as
$$\hat{\tau}^{PTC} = \frac{1}{f_s} \operatorname*{argmax}_{l_g} \hat{r}^{PTC}_{p_1p_2}(l_g).$$
where f_s is the sampling rate.

Plain Time Correlation

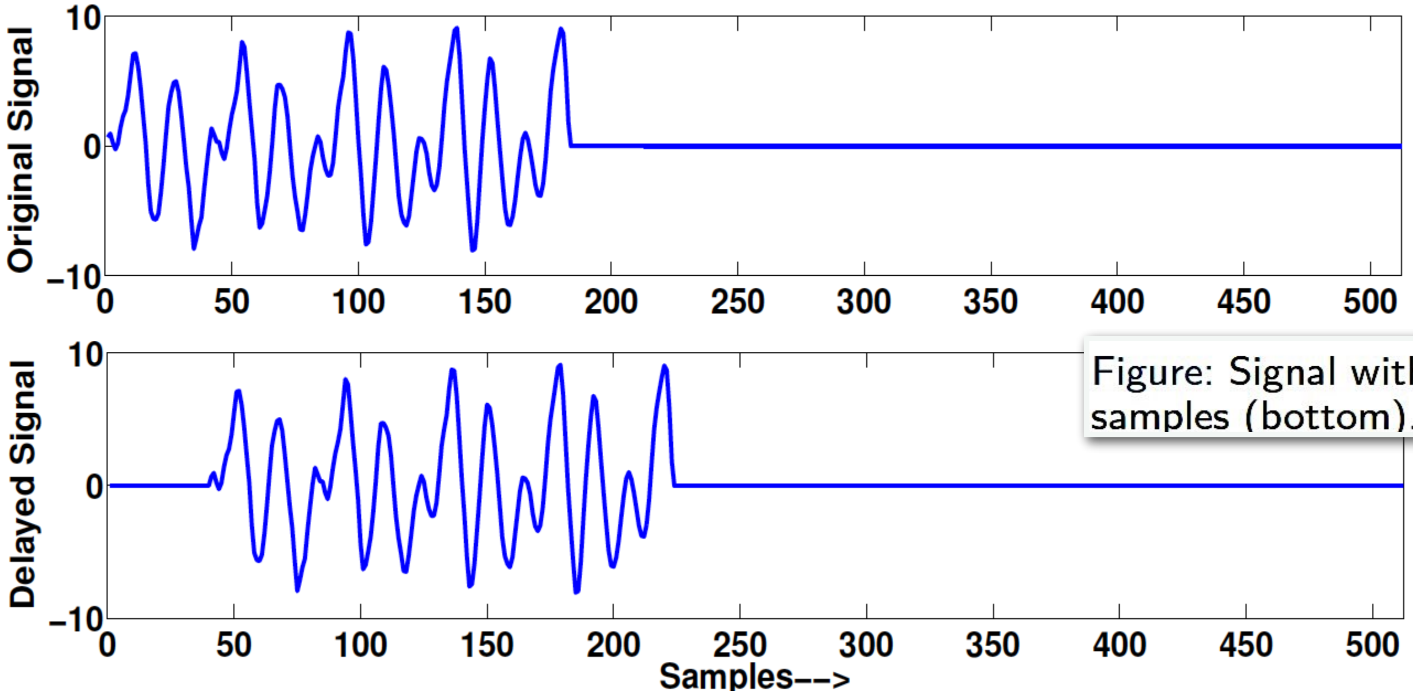
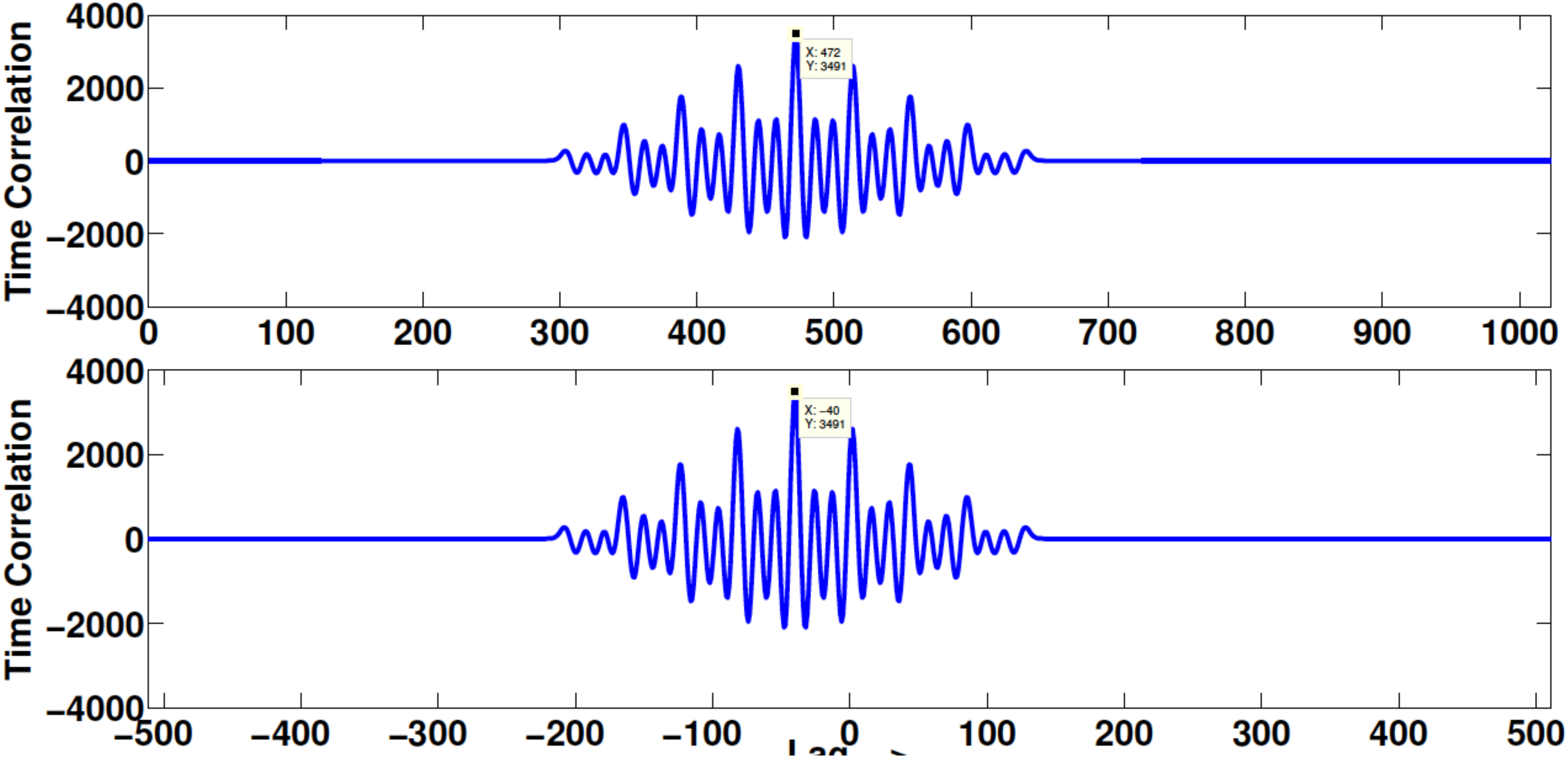


Figure: Signal with length 512 samples, original signal (top) and signal delayed by 40 samples (bottom).



Generalised Cross Correlation (GCC)

- Implements frequency domain cross-spectrum with a weighting function.
- Assuming DFT of signal output be represented by $p_1(k)$ and $p_2(k)$, the general expression for GCC is given by

$$r_{p_1 p_2}^{GCC}(l_g) = F^{-1}\{w(k)p_1(k)p_2^*(k)\}$$

  where F^{-1} stands for inverse discrete-time Fourier transform and $w(k)$ is weighting function.
- The term $w(k)p_1(k)p_2^*(k)$ is called generalized cross-spectrum.
- The TDOA estimate is obtained from the lag time that maximizes the generalized cross-correlation, as

$$\hat{\tau}^{GCC} = \frac{1}{f_s} \underset{l_g}{\operatorname{argmax}}\, r_{p_1 p_2}^{GCC}(l_g)$$

- Now the DOA estimate computed from τ ̂GCC

Generalised Cross Correlation (GCC)

- For $w(k) = 1$, GCC degenerates to cross-correlation with implementation through DFT and inverse DFT (IDFT).
- In GCC-Roth method, a Roth filter weighs GCC by a factor of auto-correlation of one of the signal.
- The Roth filter is given by

$$w_{ROTH}(k) = \frac{1}{p_1(k)p_1^*(k)}$$

- For reverberant environments, phase transform (PHAT) [7] weighting function is used for TDOA estimation using GCC.
- The PHAT weighting function is given by

$$w_{PHAT}(k) = \frac{1}{|p_1(k)p_2^*(k)|}$$

- The PHAT filter normalizes the amplitude of the spectral density of the two signal and utilizes only the phase information for computing the cross-correlation.

Generalised Cross Correlation (GCC) : Comparison

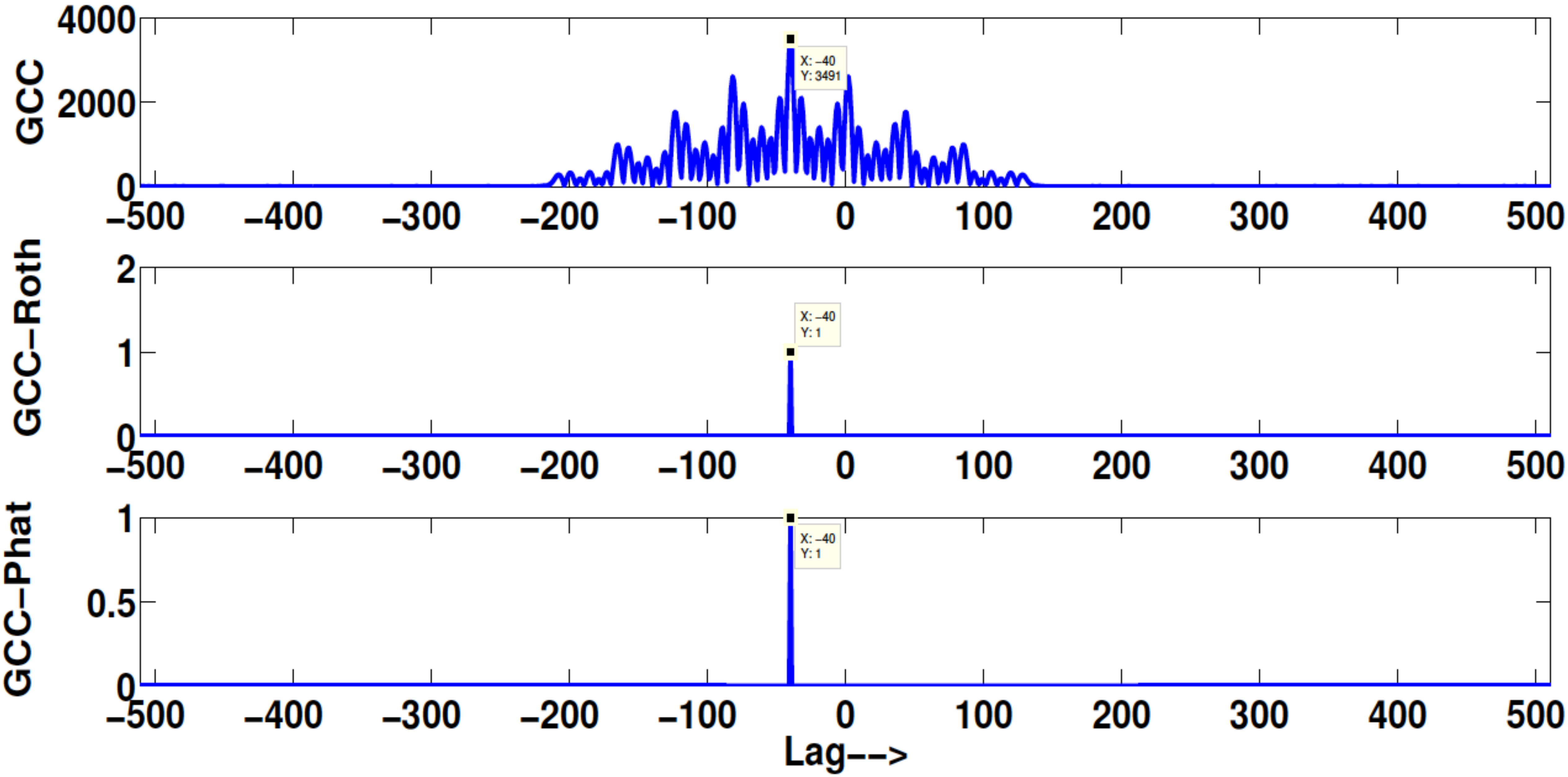
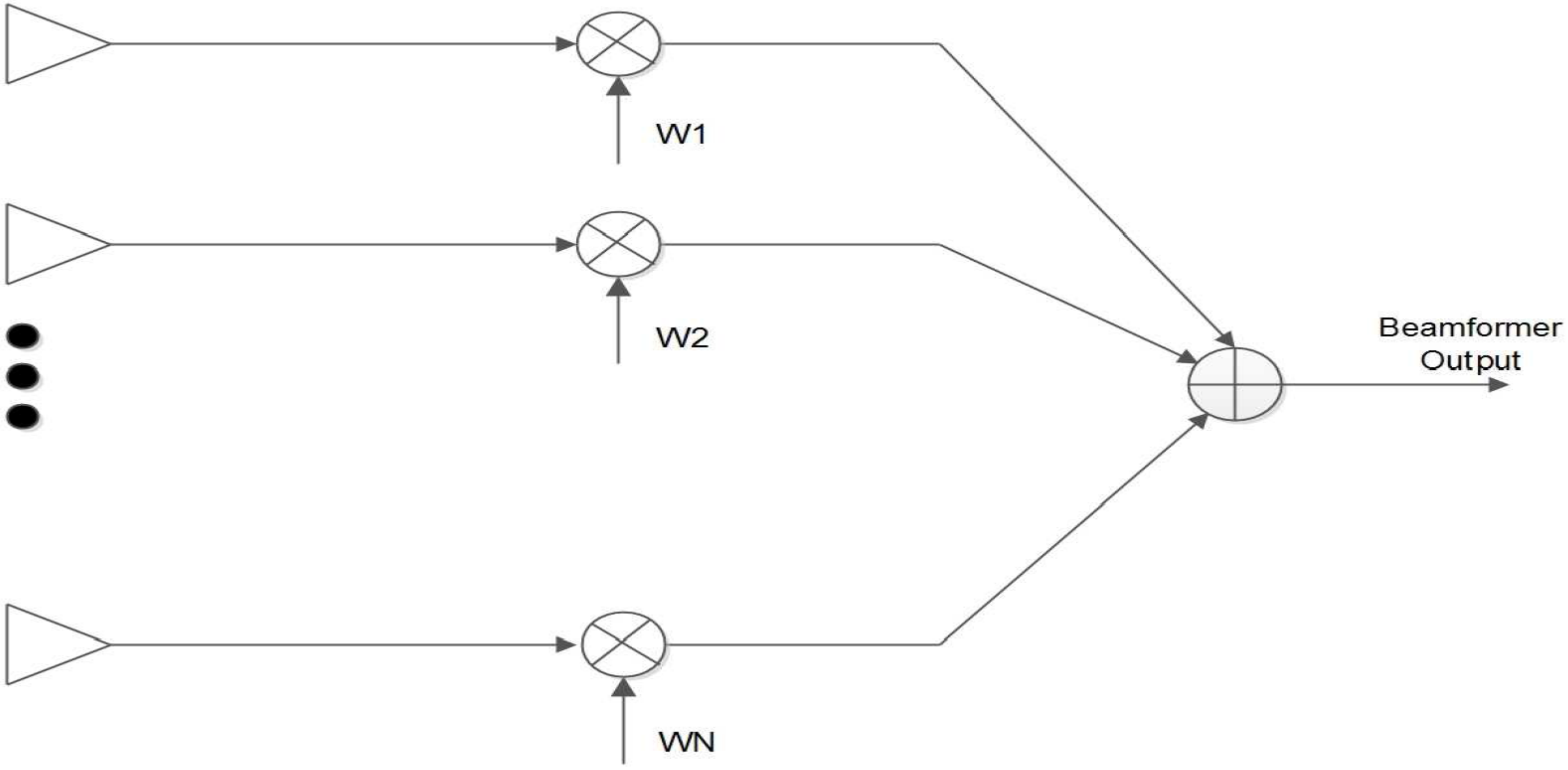


Figure: Generalized cross-correlation (GCC), GCC-Roth and GCC-PHAT plots (top to bottom)

Digital Beamforming using Spatial Filtering DSB, Filter and Sum, Capon Methods, and Beampattern Analysis

Digital Beamforming using a Spatial Filter

- Beamforming is a spatial filtering technique where signal from a given direction is passed undistorted, while signals from all other directions are attenuated.
- It is done by forming a beam in the look direction which is done by weighting and summing the array outputs.



Digital Beamforming using a Spatial Filter

- The beamformed array output is given by

$$\mathbf{p}_o(t) = \mathbf{w}^H \mathbf{p}(t)$$

where $\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_I \end{bmatrix}^T$ is beamforming weight vector and $(.)^H$ denotes conjugate transpose of $(.)$.

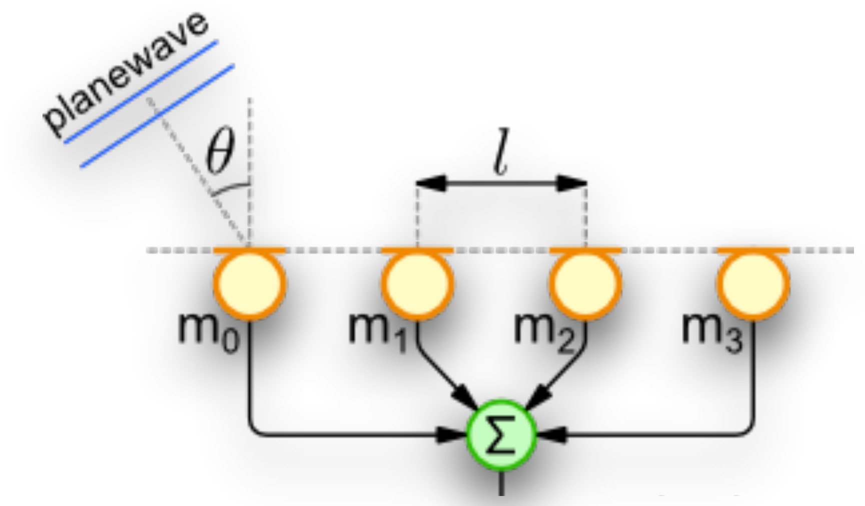
- Power spectrum of the spatially filtered signal

$$E\{|\mathbf{p}_o(t)|^2\} = \mathbf{w}^H \mathbf{R}_\mathbf{p} \mathbf{w}, \ \text{where}, \ \mathbf{R}_\mathbf{p} = E[\mathbf{p}(t)\mathbf{p}(t)^H]$$

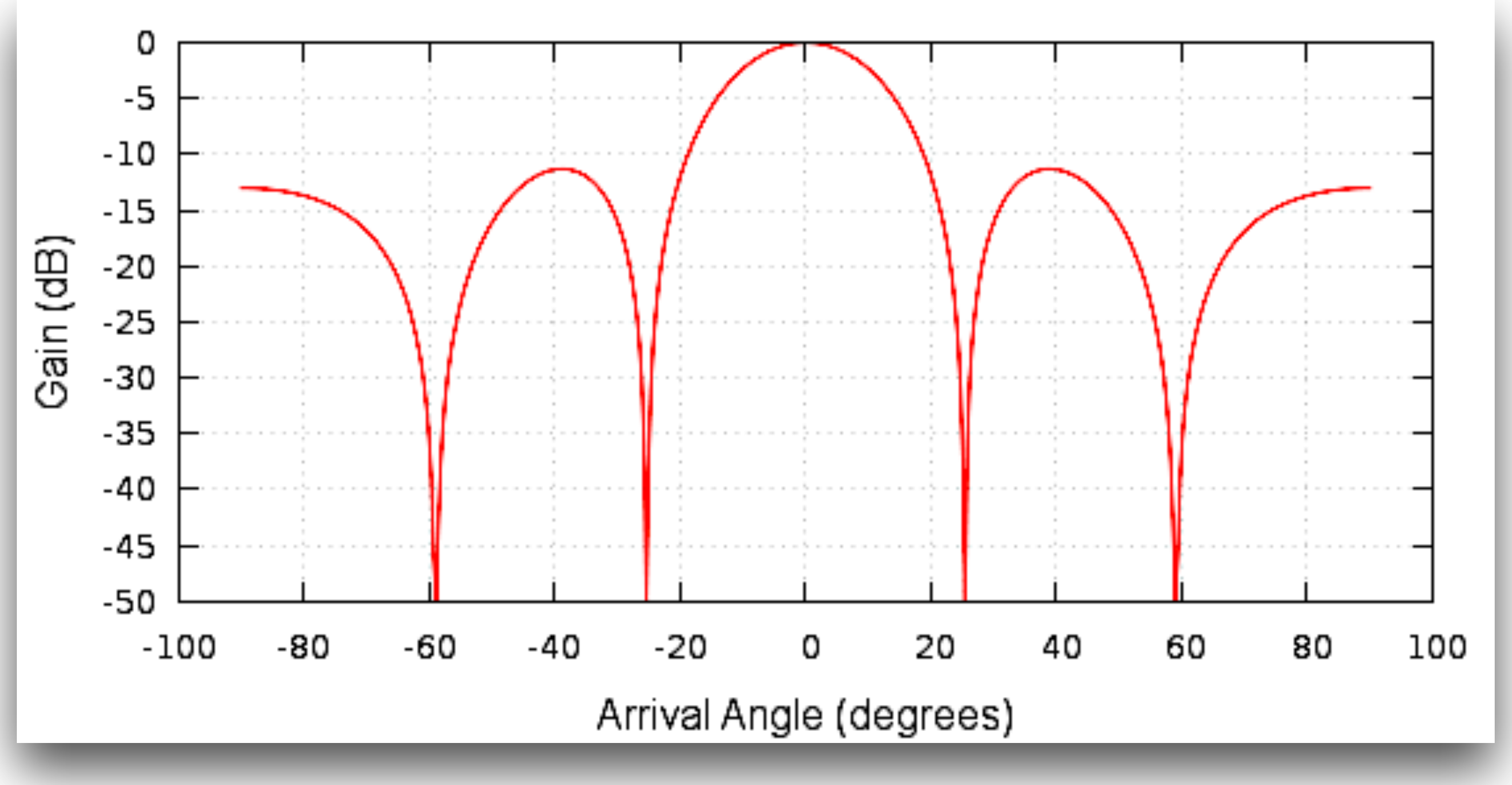
should give peak in DOAs.

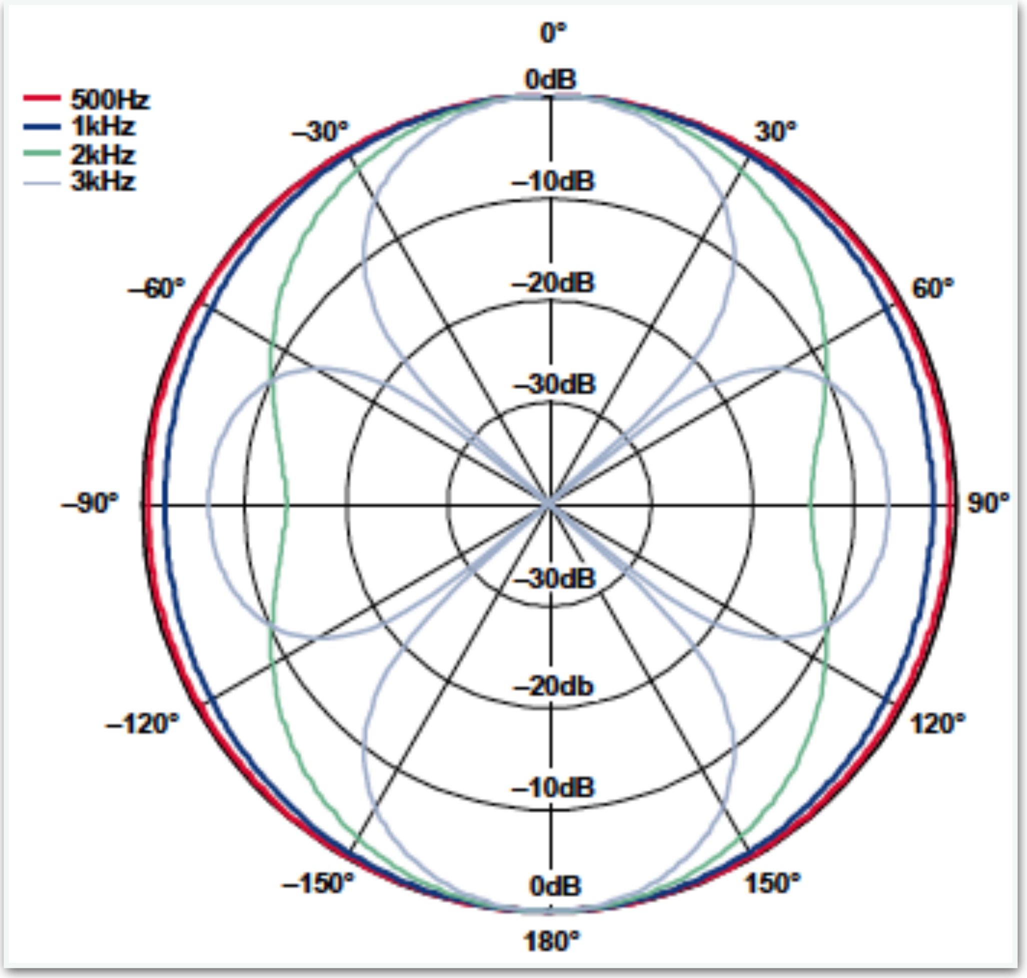
- Different choice of weights leads to different beamforming techniques.

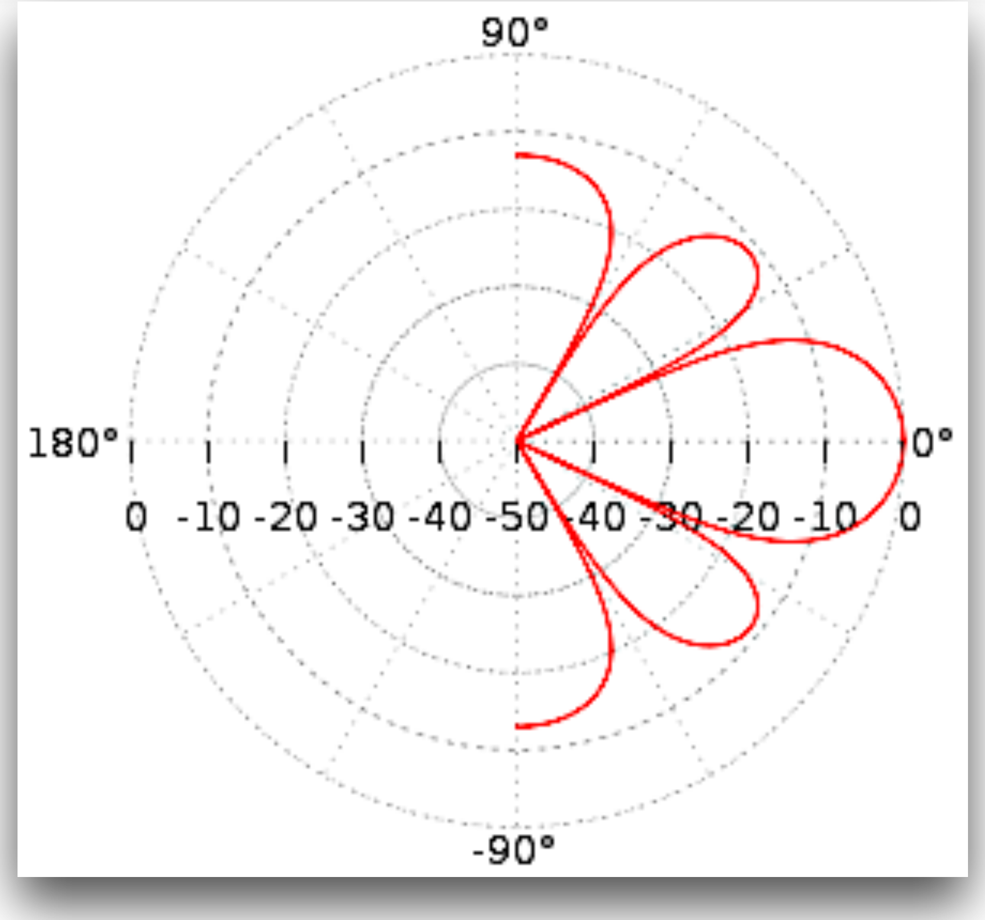
What is Beam (Directivity) Pattern of a Beamformer ?



$$20\log_{10}\left(\frac{1}{N}\sum_{i=0}^{N-1} e^{\frac{j\,2\pi f\, i\, l\, \sin(\theta)}{c}}\right)$$

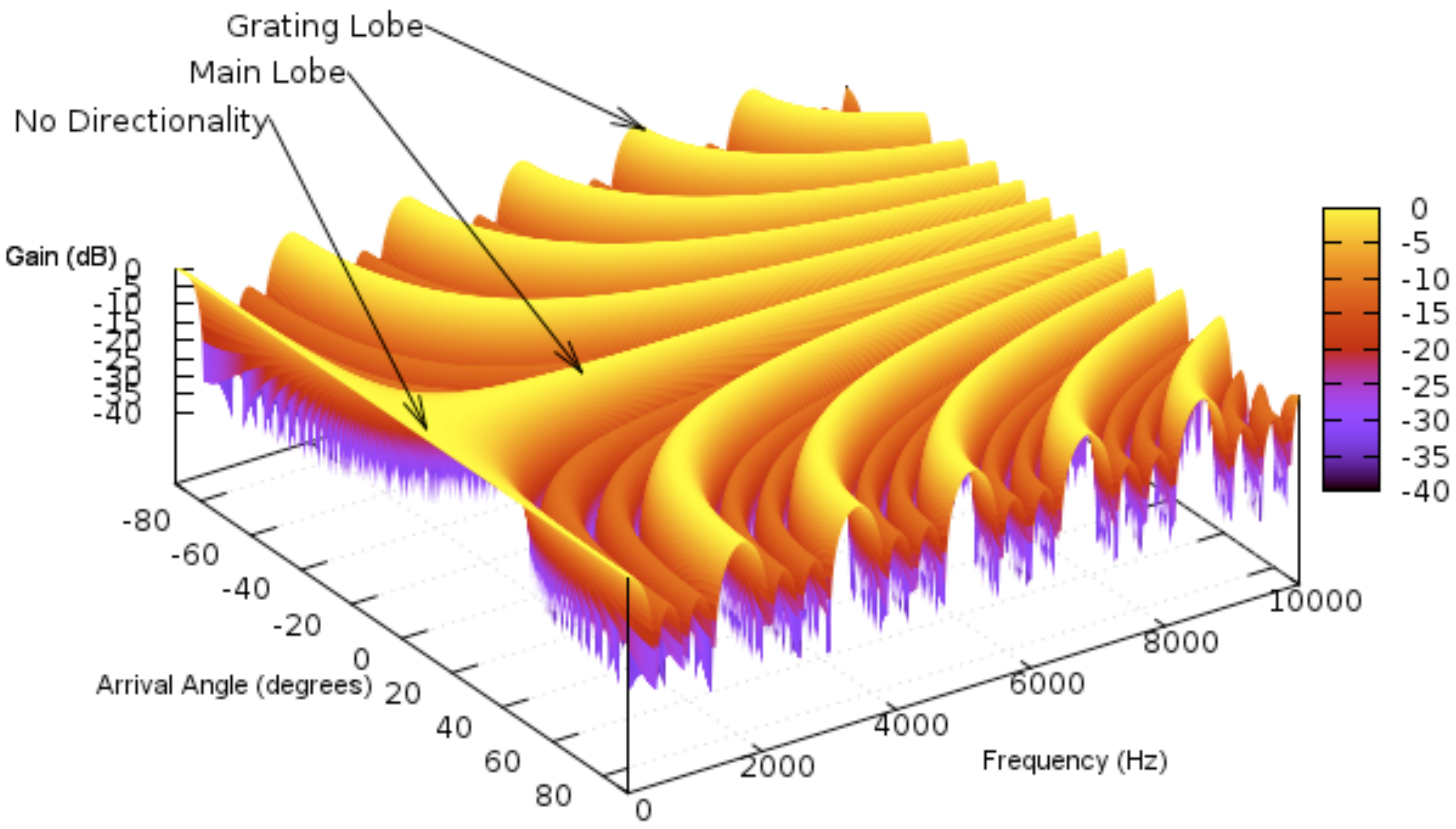






http://www.labbookpages.co.uk/audio/beamforming/delaySum.html

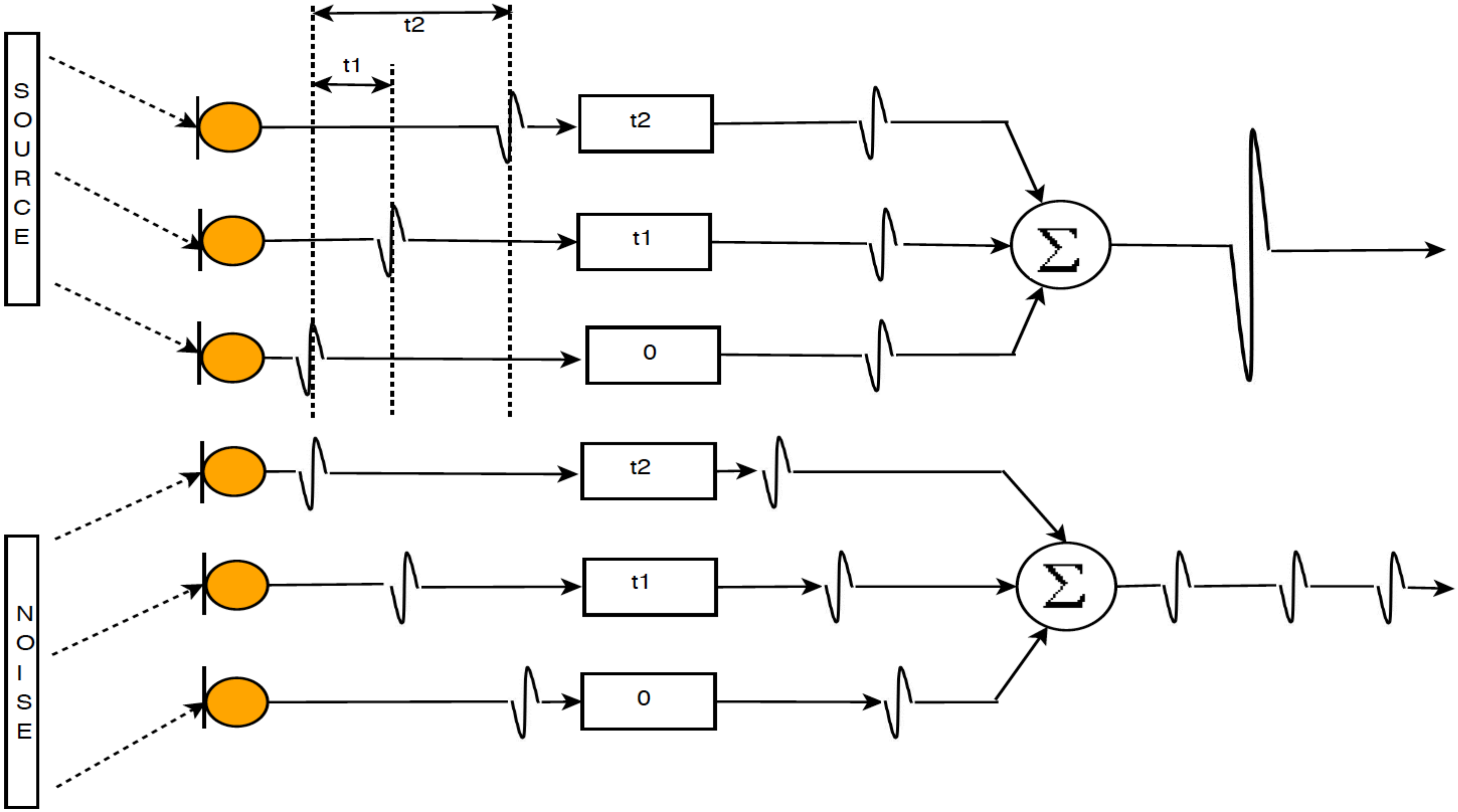
Direction - Frequency Beam patterns (Directional Frequency Response)



http://www.labbookpages.co.uk/audio/beamforming/delaySum.html

Delay Sum Beamforming

- Signal incident on array, suffers different delays at different sensors.
- The array output is delayed so that signal from desired direction is aligned.
- The aligned signals are summed, to realize a delay-and-sum beamformer (DSB).



Delay Sum Beamforming

- The Delay-and-sum beamformer design problem is formulated as :

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{w} \qquad \text{subject to} \qquad \mathbf{w}^H \mathbf{a}(\phi, k) = 1$$

- Solution to the optimization problem, results in DSB weights as

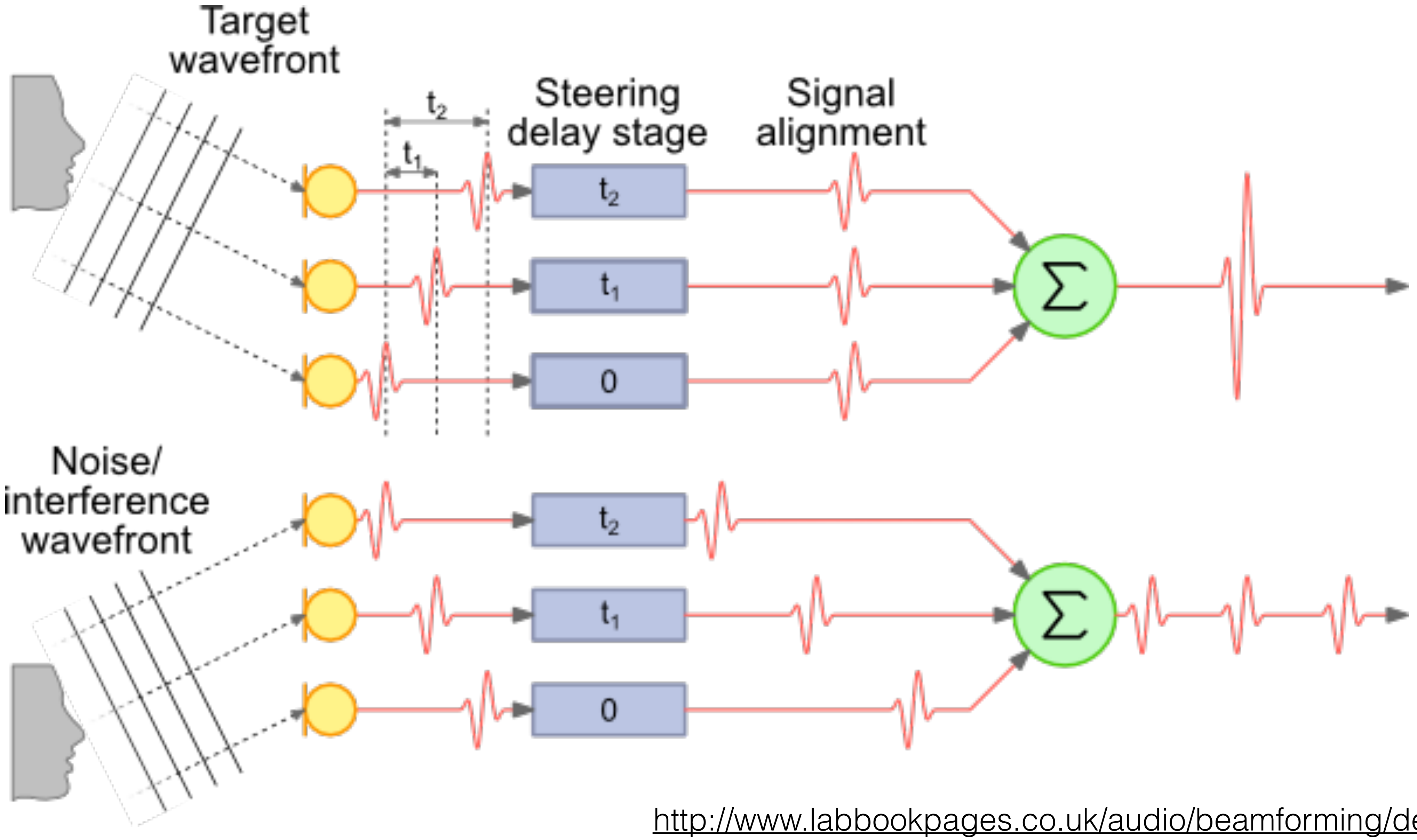
$$\mathbf{w} = \frac{\mathbf{a}(\phi, k)}{I}.$$

- The solution doesn't depend upon the input signal and only takes into consideration the steering vector of the signal of interest. Hence, the Delay-and-sum beamformers are not adaptive.
- The spatial power spectrum for DSB, can now be written as

$$E\{|\mathbf{p}_o(t)|^2\} = \mathbf{w}^H \mathbf{R}_\mathbf{p} \mathbf{w} \longrightarrow P_{DSB}(\phi) = \mathbf{a}^H(\phi) \mathbf{R}_\mathbf{p} \mathbf{a}(\phi).$$

Delay Sum Beamforming for Speech Source Localisation and Enhancement

- Delay-and-sum beamforming DOA estimates are given by the location of L highest peaks corresponding to L sources in DSB spatial power spectrum.
- DSB based soured localization is inconsistent when multiple sources are present. Bias of the estimates also become significant for closely spaced and correlated sources.



http://www.labbookpages.co.uk/audio/beamforming/delaySum.html

Beam Pattern Analysis for ULA

- Beampattern (directivity pattern, array pattern or spatial pattern) is defined as the magnitude of the spatial filter's directional response.
- For given weight vector **w** of a beamformer, beampattern specifies the response of the beamformer to a source arriving from the arbitrary direction in the field of view of the array.
- Beampattern is typically measured as the array response to a single plane wave.

- The beamformed output can be written as

$$\mathbf{p}_o(t) = \mathbf{w}^H\mathbf{p}(t) = \mathbf{w}^H\mathbf{a}(\Psi, k)s(t) + \mathbf{w}^H\mathbf{v}(t)$$

  where $\mathbf{w}^H\mathbf{a}(\Psi, k)$ is directional response of an array.
- Assuming ULA aperture, steered to direction ϕ_s, the delay-and-sum beampattern for a ULA can be written as

$$G(\phi, \phi_s) = |\mathbf{w}^H(\phi_s)\mathbf{a}(\phi, k)|$$

Beam Pattern Analysis for DSB

- Utilizing the DSB weight the beampattern for the ULA is given by

$$G(\phi, \phi_s) = \frac{1}{I}|\mathbf{a}^H(\phi_s, k)\mathbf{a}(\phi, k)|$$

where $|(.)|$ is absolute value of $(.)$.

- Substituting the expression for steering vector beampattern for a ULA can be written as

$$\begin{aligned} G(\phi, \phi_s) &= \frac{1}{I}\sum_{i=1}^{I} e^{j(i-1)kd(\cos(\phi)-\cos(\phi_s))} \\ &= \left| \frac{\sin\left(\frac{Ikd}{2}\left(\cos(\phi) - \cos(\phi_s)\right)\right)}{I\sin\left(\frac{kd}{2}\left(\cos(\phi) - \cos(\phi_s)\right)\right)} \right| \end{aligned}$$

- Narrowband beampatterns of a delay-and-sum beamformer, is illustrated without spatial aliasing and under aliasing.

Beam Pattern Analysis : Non Aliased case for DSB

Beam Pattern Analysis : Non Aliased Case for DSB

- A ULA with 10 sensors are used for this, with steering angle $\phi_s = 90°$.

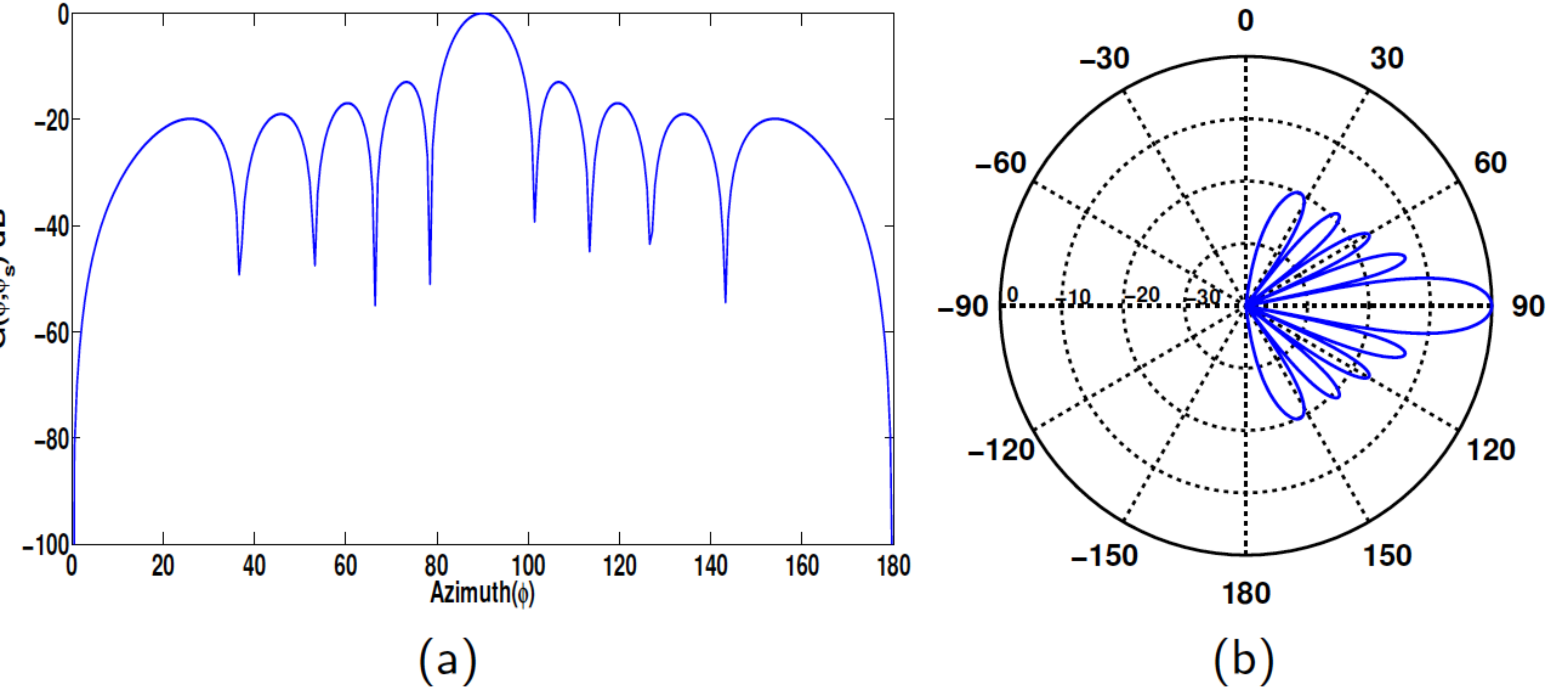
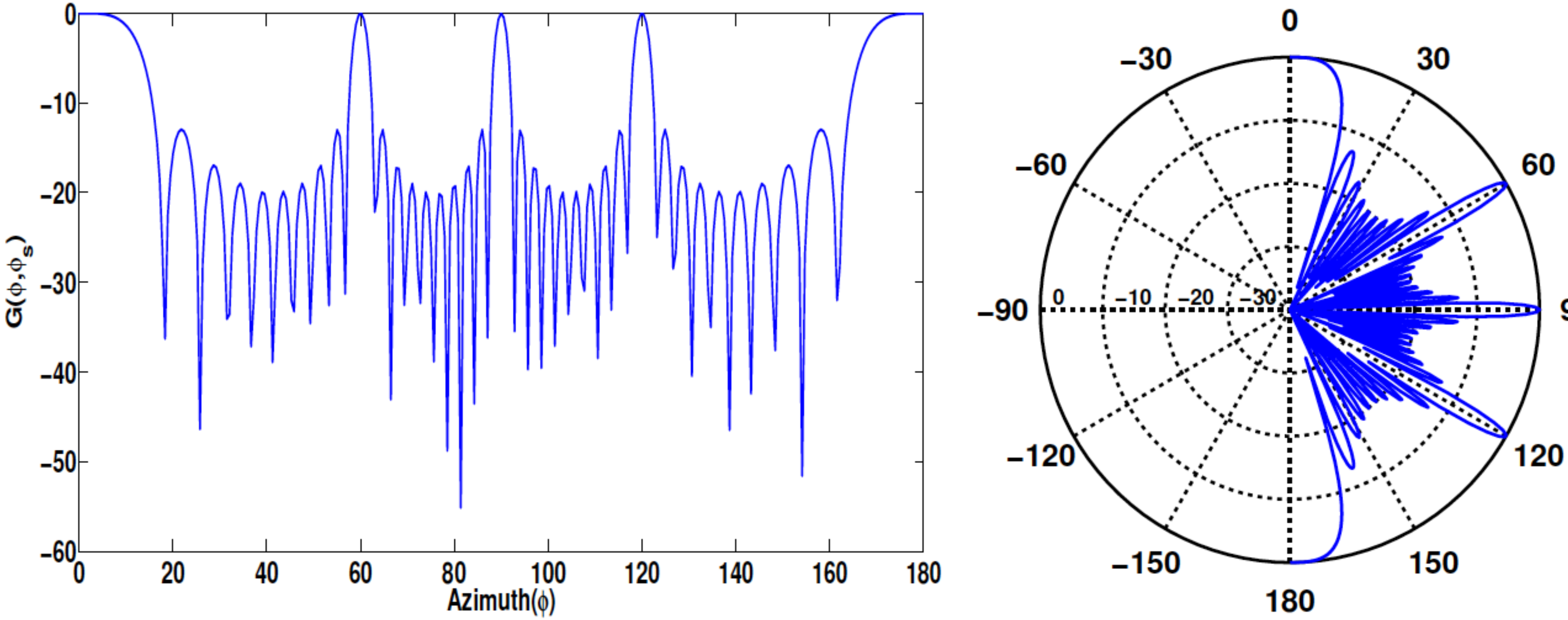


Figure: Delay-and-sum beampattern for ULA with no spatial aliasing for $I = 10$, $\phi_s = 90°$ and $d = 0.5\lambda$ (a) in Cartesian coordinates and (b) in polar coordinates

Beam Pattern Analysis : Aliased case for DSB

- A ULA with 10 sensors are used for this, with steering angle $\phi_s = 90°$.



(a)

(b)

Figure: Delay-and-sum beampattern for ULA under aliasing for $I = 10$, $\phi_s = 90°$ and $d = 2\lambda$ (a) in Cartesian coordinates and (b) in polar coordinates.

Filter And Sum Beamforming

Design a filter $h(\theta)$ such that for each θ

It passes undistorted the signal with DOA = θ

It attenuates all DOAs $\neq \theta$

Sweep the filter through the DOA range of interest, and evaluate the powers of the filtered signals:

$$\begin{aligned} E\left\{|y_F(t)|^2\right\} &= E\left\{|h^*(\theta)y(t)|^2\right\} \\ &= h^*(\theta)Rh(\theta) \end{aligned}$$

$$R = E\{y(t)y^*(t)\}$$

The (dominant) peaks of $h^*(\theta)Rh(\theta)$ give the DOAs of the sources.

Filter And Sum Beamforming

Assume the array output is *spatially white*:

$$R = E\left\{y(t)y^*(t)\right\} = I$$

Then: $$E\left\{|y_F(t)|^2\right\} = h^*h$$

Hence: In direct analogy with the temporally white assumption for filter bank methods, $y(t)$ can be considered as impinging on the array from *all* DOAs.

Filter And Sum Beamforming

Filter Design:

$$\boxed{\min_{h} \; (h^* h) \;\; \text{subject to} \;\; h^* a(\theta) = 1}$$

Solution:

$$h = a(\theta)/a^*(\theta)a(\theta) = a(\theta)/m$$

$$E\left\{ |y_F(t)|^2 \right\} = a^*(\theta) R a(\theta)/m^2$$

Implementation and DOA Estimation in Filter And Sum Beamforming

$$E\left\{|y_F(t)|^2\right\} = a^*(\theta)Ra(\theta)/m^2$$

$$\hat{R} = \frac{1}{N}\sum_{t=1}^{N} y(t)y^*(t)$$

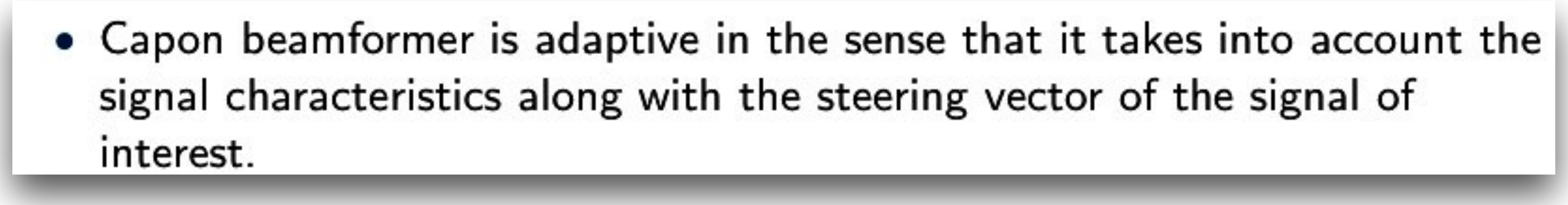
DOA estimates are $\{\hat{\theta}_k\} =$ the locations of the n largest peaks of $a^*(\theta)\hat{R}a(\theta)$.

Resolution Threshold:

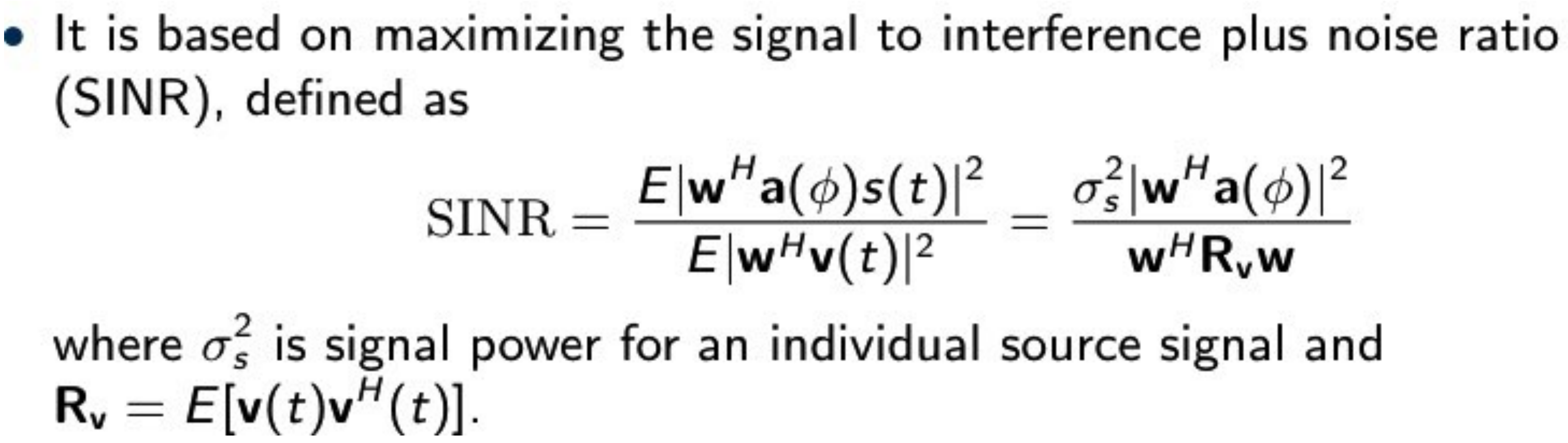
$$\begin{aligned} \inf\ |\theta_k - \theta_p| &> \frac{\text{wavelength}}{\text{array length}} \\ &= \text{array beamwidth} \end{aligned}$$

Beamforming DOA estimates are consistent if $n = 1$, but inconsistent if $n > 1$.

Capon/ MVDR Beamforming



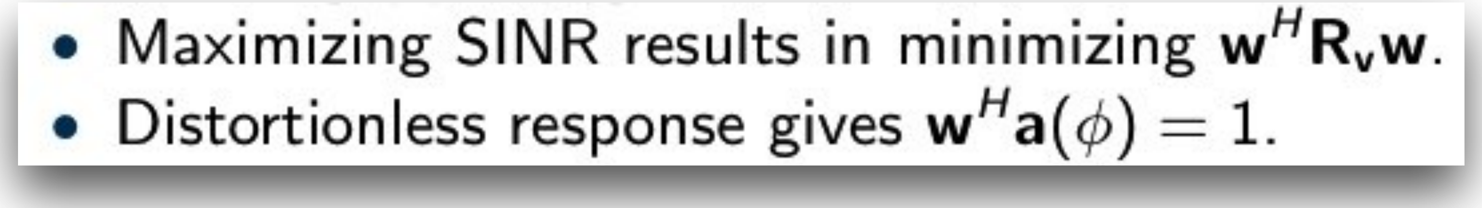
- Capon beamformer is adaptive in the sense that it takes into account the signal characteristics along with the steering vector of the signal of interest.



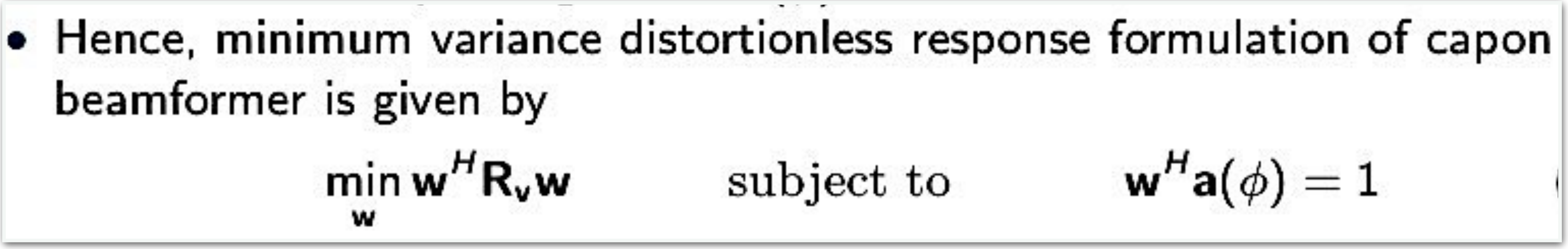
- It is based on maximizing the signal to interference plus noise ratio (SINR), defined as

$$\mathrm{SINR} = \frac{E|\mathbf{w}^H\mathbf{a}(\phi)s(t)|^2}{E|\mathbf{w}^H\mathbf{v}(t)|^2} = \frac{\sigma_s^2|\mathbf{w}^H\mathbf{a}(\phi)|^2}{\mathbf{w}^H\mathbf{R}_\mathbf{v}\mathbf{w}}$$

where σ_s^2 is signal power for an individual source signal and $\mathbf{R}_\mathbf{v} = E[\mathbf{v}(t)\mathbf{v}^H(t)]$.



- Maximizing SINR results in minimizing $\mathbf{w}^H\mathbf{R}_\mathbf{v}\mathbf{w}$.
- Distortionless response gives $\mathbf{w}^H\mathbf{a}(\phi) = 1$.



- Hence, minimum variance distortionless response formulation of capon beamformer is given by

$$\min_{\mathbf{w}} \mathbf{w}^H\mathbf{R}_\mathbf{v}\mathbf{w} \qquad \text{subject to} \qquad \mathbf{w}^H\mathbf{a}(\phi) = 1$$

Capon/ MVDR Beamforming

- Solution to the constrained problem

$$\mathbf{w} = \frac{\mathbf{R_v}^{-1}\mathbf{a}(\phi)}{\mathbf{a}^H(\phi)\mathbf{R_v}^{-1}\mathbf{a}(\phi)}$$

- However, $\mathbf{R_v}^{-1}$ is not available in practice. Therefore, $\mathbf{R_p}$ is used in place of $\mathbf{R_v}$.

- This results in final form of weight vector as

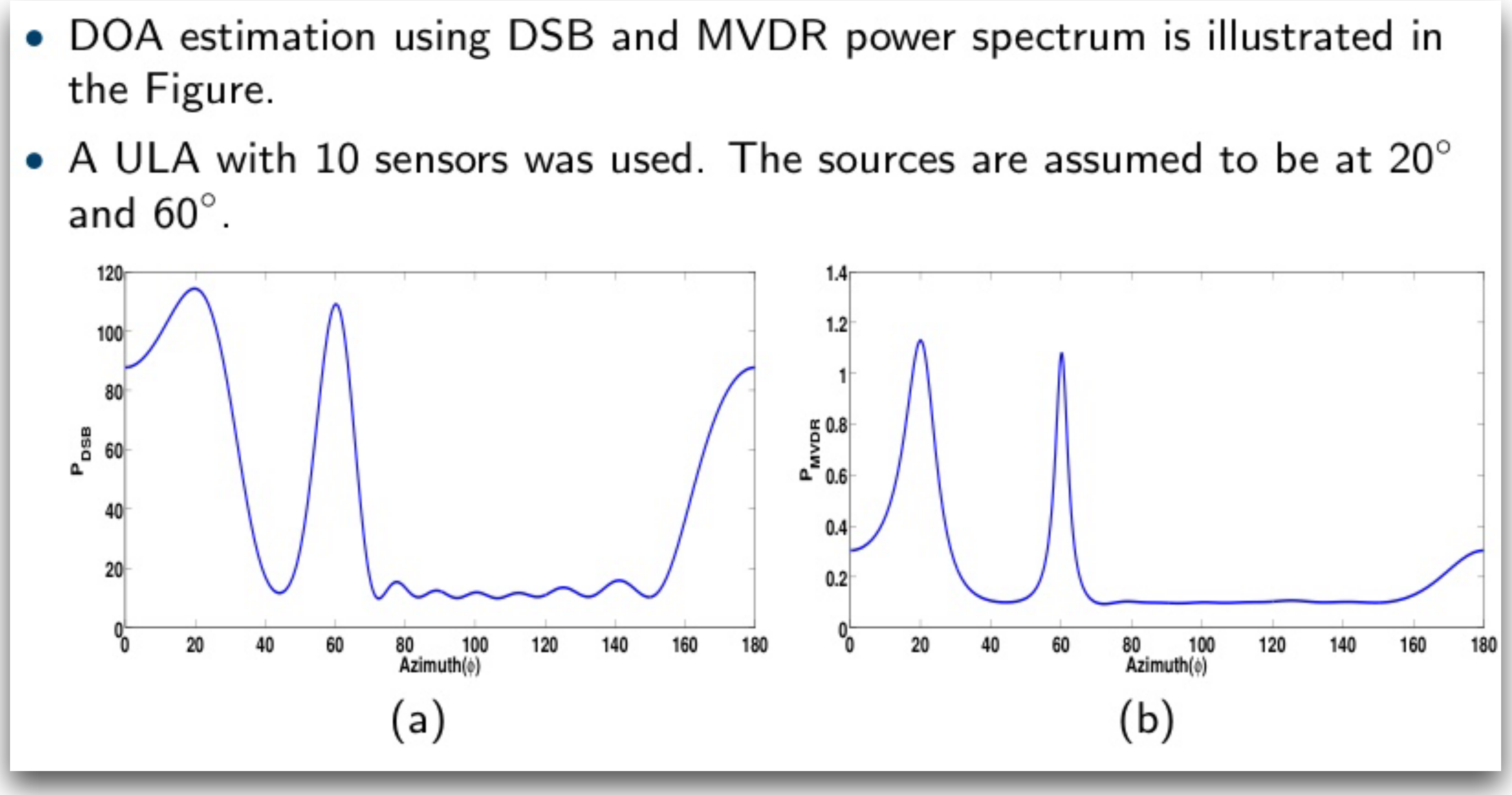
$$\mathbf{w} = \frac{\mathbf{R_p}^{-1}\mathbf{a}(\phi)}{\mathbf{a}^H(\phi)\mathbf{R_p}^{-1}\mathbf{a}(\phi)}$$

- Utilizing the expression for MVDR weights power spectrum for MVDR can written as

$$P_{MVDR} = \frac{1}{\mathbf{a}^H(\phi)\mathbf{R_p}^{-1}\mathbf{a}(\phi)}.$$

The MVDR DOA estimates can be given as L largest peaks in the MVDR power spectrum corresponding to L sources.

Capon/ MVDR Beamforming versus DSB

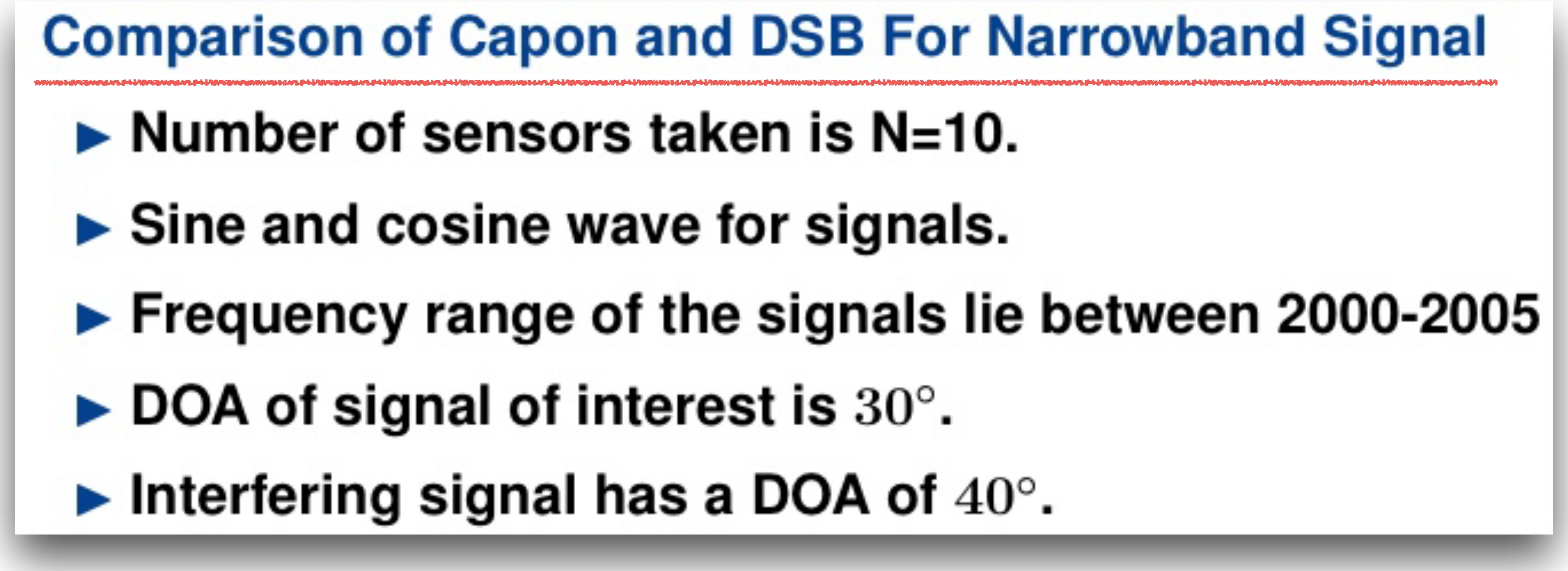


- DOA estimation using DSB and MVDR power spectrum is illustrated in the Figure.
- A ULA with 10 sensors was used. The sources are assumed to be at 20° and 60°.

(a)

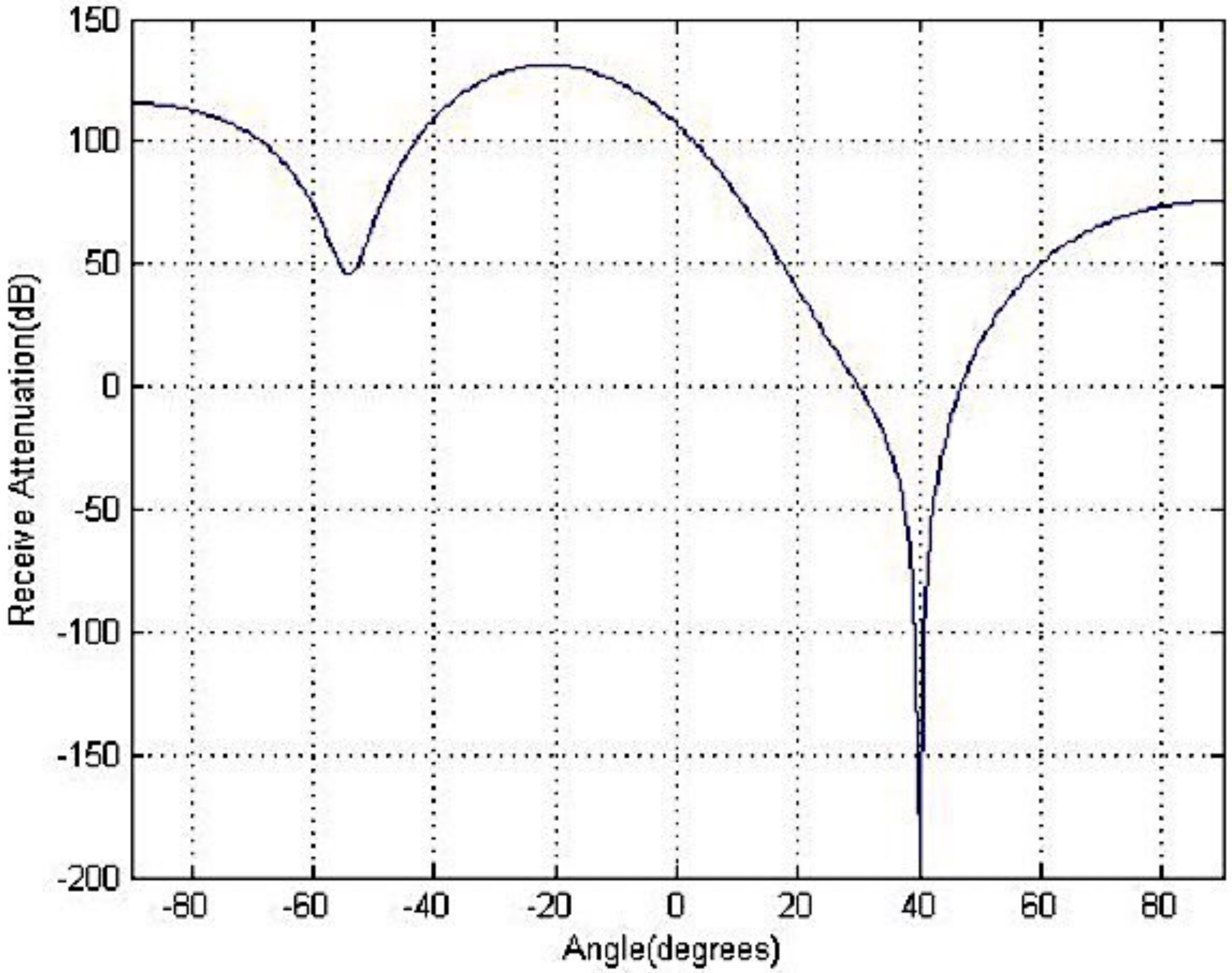
(b)

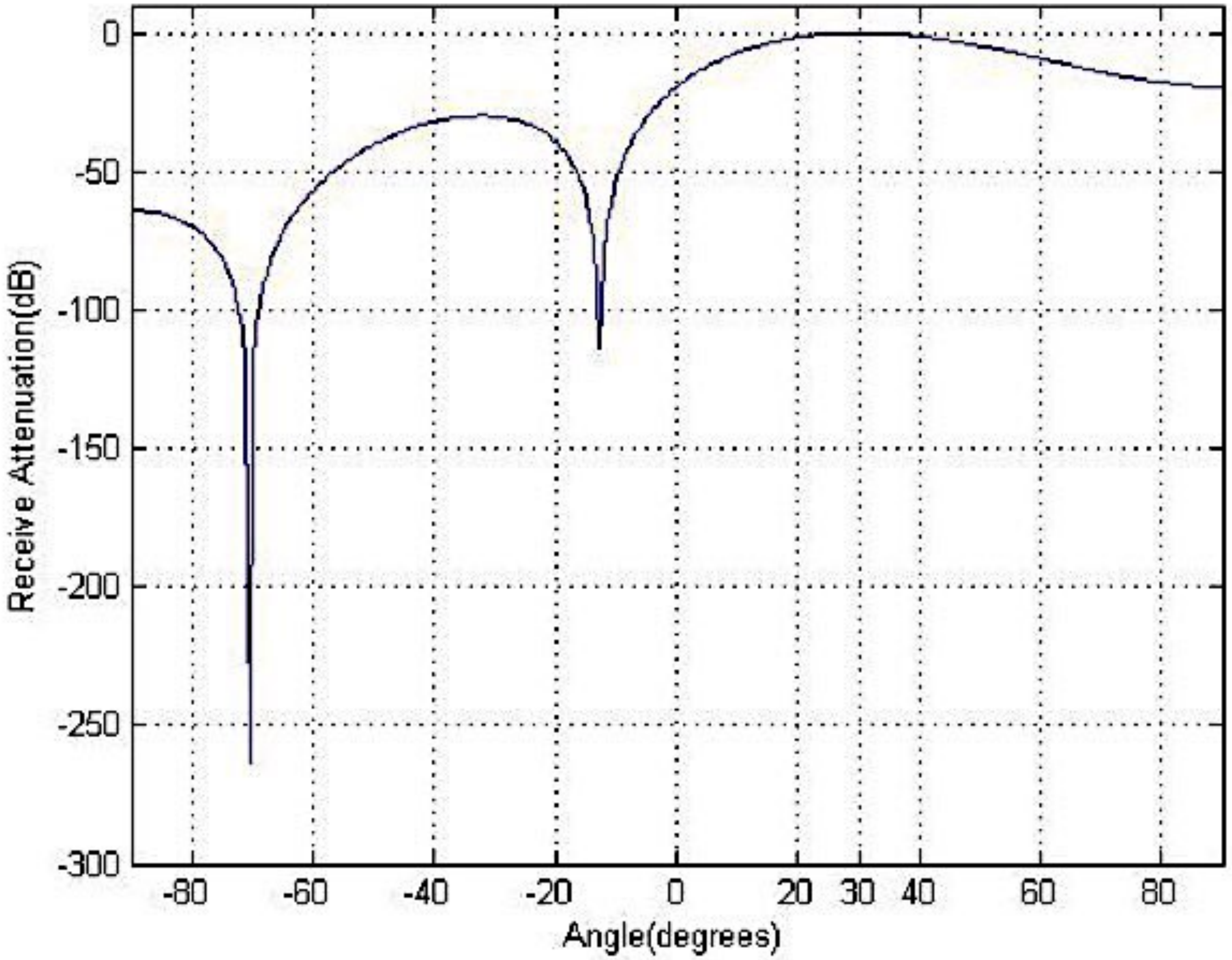
Figure: DOA estimation using (a) DSB and (b) MVDR method. A ULA with $I = 10$ sensors was used for sources located at 20° and 60°.



Comparison of Capon and DSB For Narrowband Signal

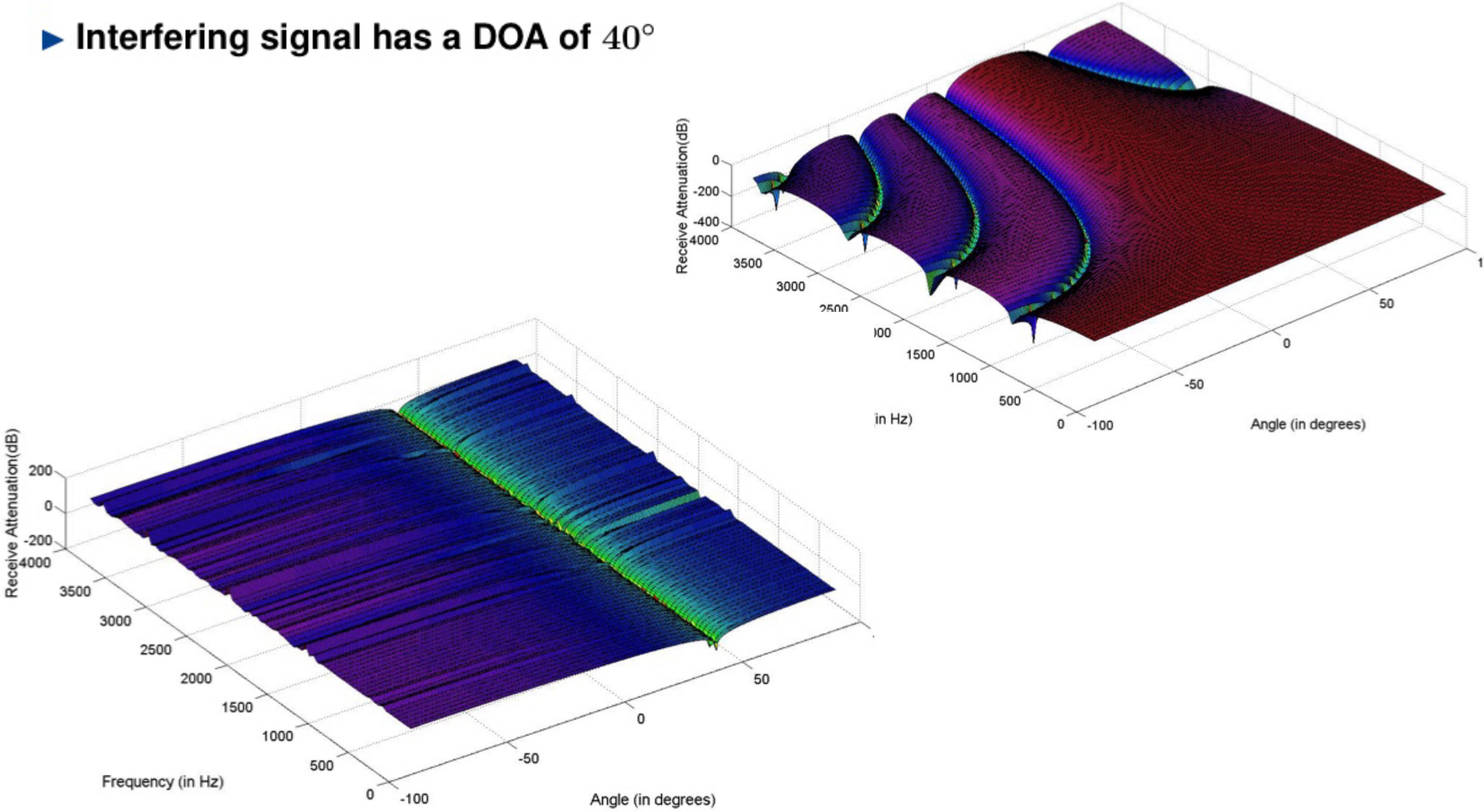
- Number of sensors taken is N=10.
- Sine and cosine wave for signals.
- Frequency range of the signals lie between 2000-2005
- DOA of signal of interest is 30°.
- Interfering signal has a DOA of 40°.





Comparison of Capon and DSB For Broadband Signal

- Number of sensors taken is N=10.
- Speech signals are taken for signal of interest and interfering signal.
- DOA of signal of interest is $30°$.
- Interfering signal has a DOA of $40°$



Capon/ MVDR Beamforming versus DSB : Comments

- Beampattern of Capon has severe attenuation in interfering direction at all the frequencies.
- Beampattern of Delay-and-sum beamformer does not have any null in interfering signal's DOA.
- In presence of interference Capon can reconstruct the signal of interest exactly whereas Delay-and-sum beamformer cannot completely filter interference.

Quadratically Constrained Capon Beamformer

- Additional Constraint on norm of weights for MVDR.

$$||\mathbf{w}||^2 \; = \; \mathbf{T}_1$$

where w is the weight vector and T_1 is a design parameter.

- The new constrained problem can be written as,

$$\begin{aligned} &\min_{\mathbf{w}} \mathbf{w}^T \mathbf{R}_y \mathbf{w} \\ \text{subject to} \quad &\mathbf{s}^T \mathbf{w} = 1 \\ &||\mathbf{w}||^2 \; = \; \mathbf{T}_1 \end{aligned}$$

where s corresponds to steering vector.

Quadratically Constrained Capon Beamformer

- Previous constrained problem is solved using Lagrange multiplier.
- The solution for the weight vector can be written as,

$$\mathbf{w}^{\mathrm{H}} = \frac{\mathbf{s}^{\mathrm{H}}(\mathbf{R}_y + \lambda_1 \mathbf{I})^{-1}}{\mathbf{s}^{\mathrm{H}}(\mathbf{R}_y + \lambda_1 \mathbf{I})^{-1}\mathbf{s}}$$

- Value of λ_1 depends on choice of T_1.

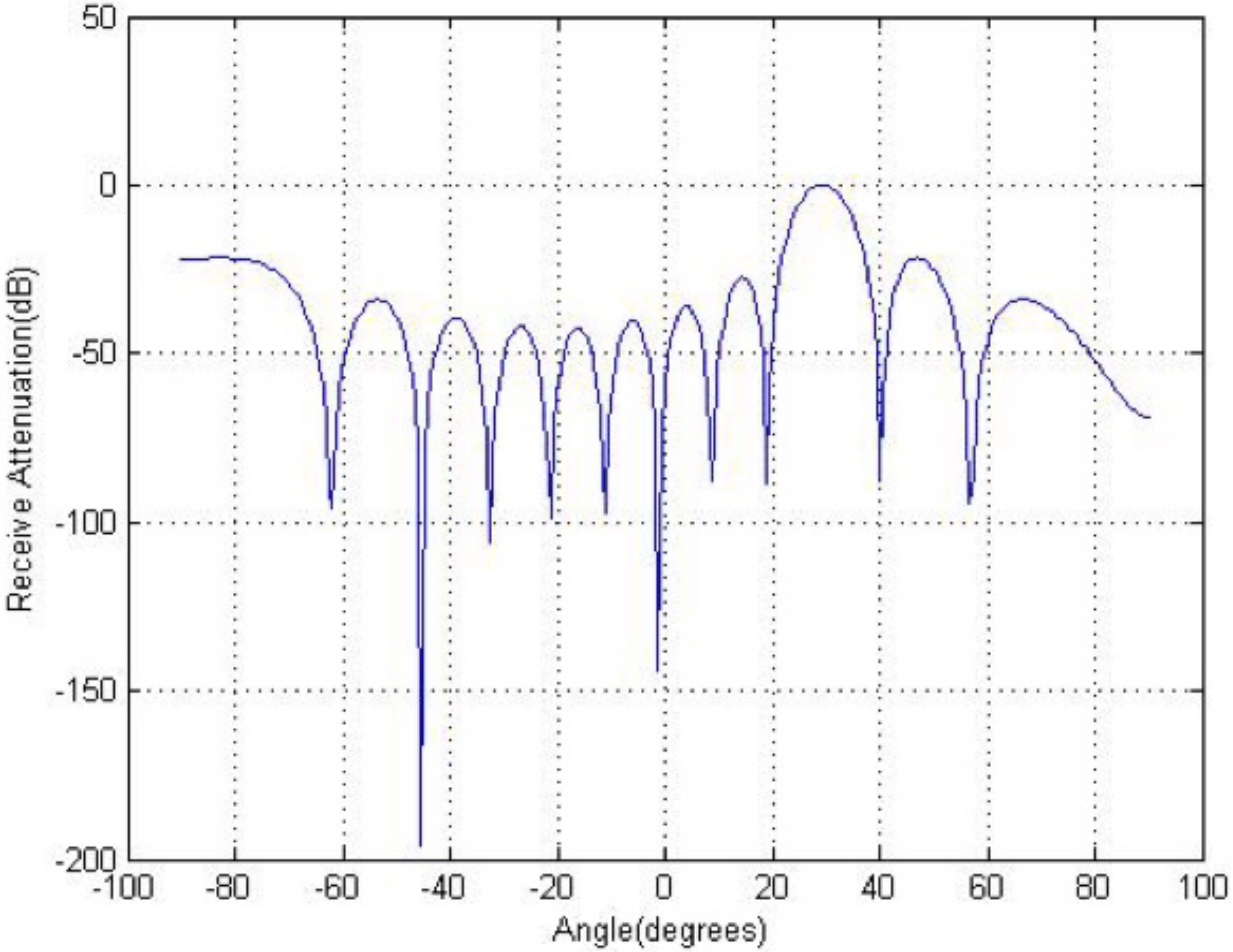
$$\frac{(\mathbf{R}_y + \lambda_1 \mathbf{I})^{-1}}{(\mathbf{R}_y + \lambda_1 \mathbf{I})^{-1}}$$

- Leads to diagonal loading.

Performance of Quadratically Constrained MVDR

For Narrowband Signals

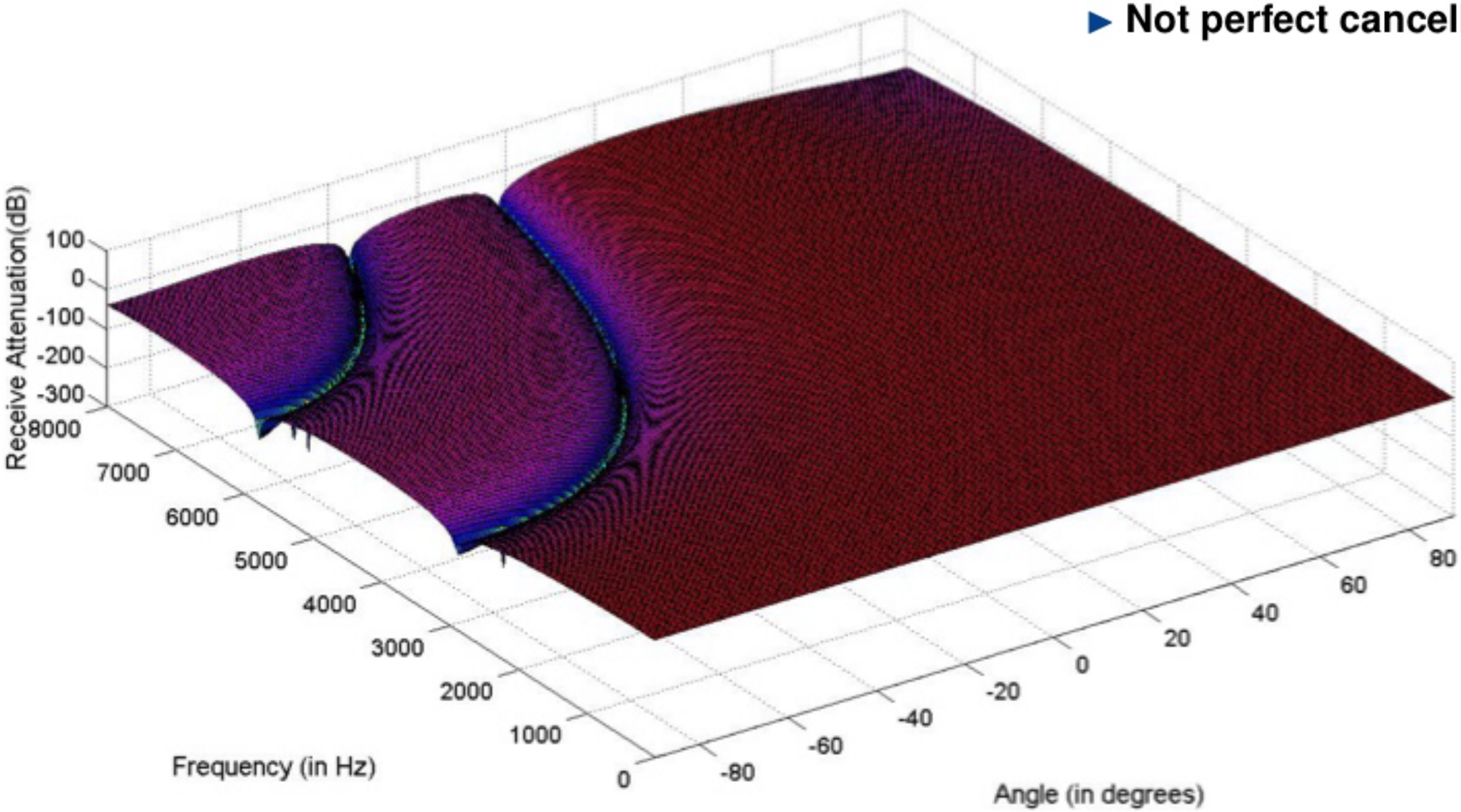
- Number of sensors taken is N=10.
- Sine and cosine wave for signals.
- Frequency range of the signals lie between 2000-2005 Hz.
- DOA of signal of interest is 30°.
- Interfering signal has a DOA of 40°.
- λ_1 is choosen as 0.5.



Performance of Quadratically Constrained MVDR For Broadband Signal

- Number of sensors taken is N=10.
- Speech signal is taken as signal of interest.
- Speech signal is taken as interfering signal.
- DOA of signal of interest is $30°$.
- Interfering signal has a DOA of $40°$.
- λ_1 is taken as 0.5.

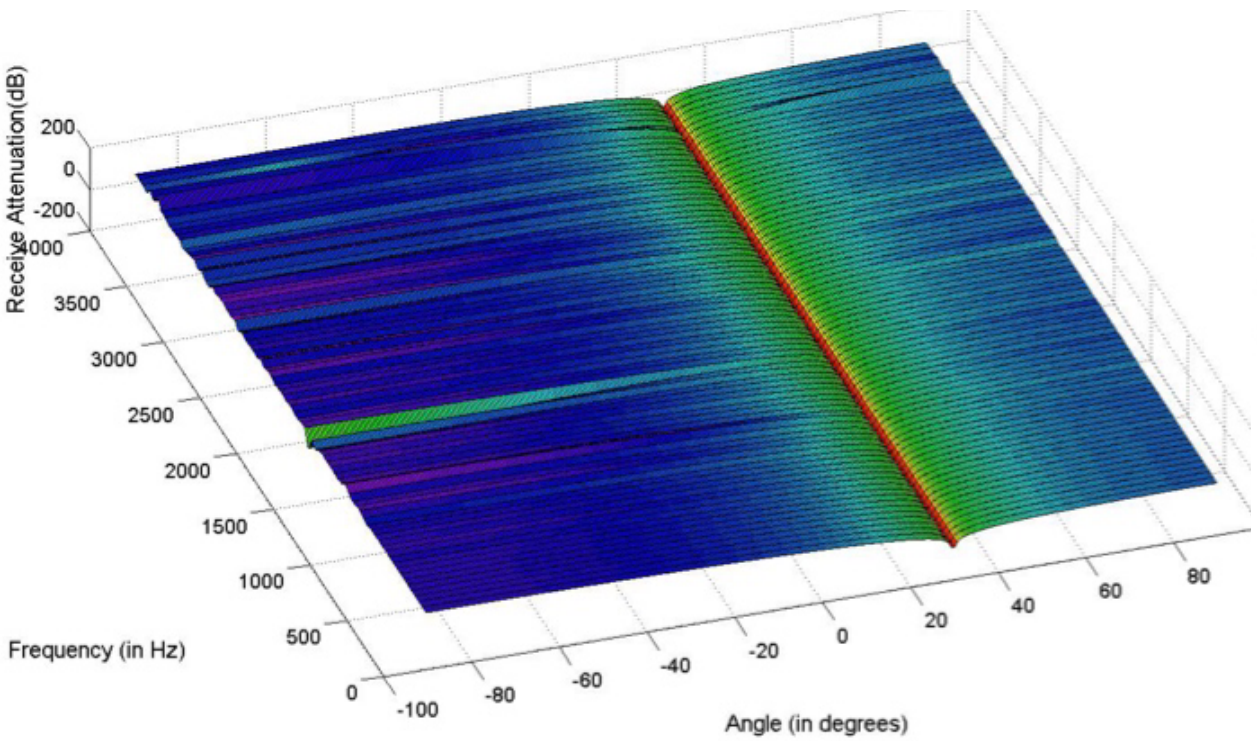
- Less attenuation of interfering signal.
- More emphasis on white noise.
- Not perfect cancellation of interfering signal.

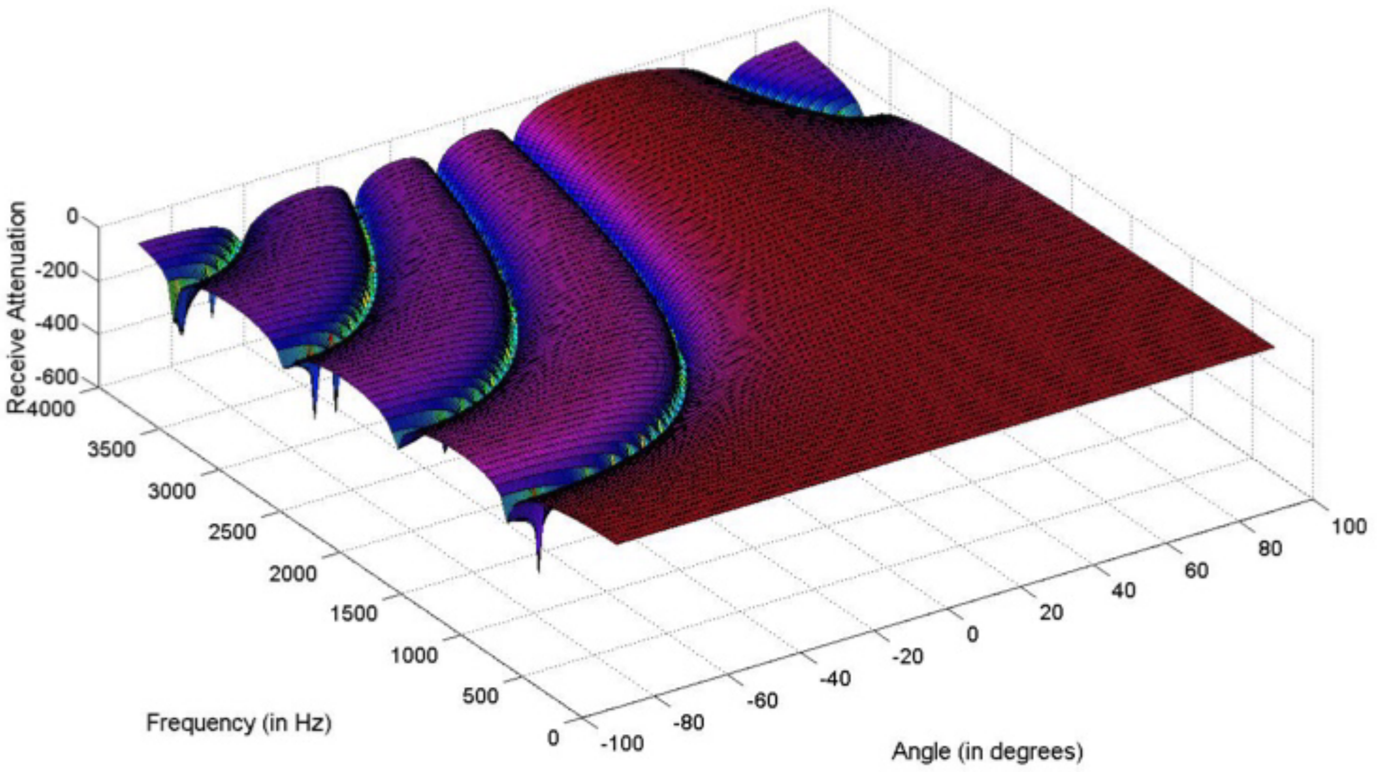


Performance of Beamformers Under DOA Estimation Error

- Comparison for Capon and Delay-and-sum beamforming method is provided.
- Number of sensors taken is N=10.
- Speech signal is taken as signal of interest.
- No interfering signal is taken.
- DOA of signal of interest is 35°.
- Estimated DOA is taken as 30°

- Delay-and-sum beamformer output is not affected much.
- Capon Output is deteriorated severely.
- Actual signal is taken as interfering signal.





Digital Beamforming using Linearly Constrained Minimum Variance (LCMV) Filter, Generalised Sidelobe Canceller (GSB)

Linearly Constrained Minimum Variance (LCMV) Beamforming

- MVDR beamformer imposes a linear constraint,

$$\boldsymbol{w}^H \boldsymbol{a}(\phi) = 1$$

- Lacks robustness to interference sources.
- Additional linear constraints are added to realize LCMV and LCMP beamformers.
- LCMV uses $\boldsymbol{R}_v$, while LCMP uses $\boldsymbol{R}_p$.
- LCMV beamformer minimizes,

$$\boldsymbol{P}_n = \boldsymbol{w}^H \boldsymbol{R}_v \boldsymbol{w}$$

  under the constraint, $\boldsymbol{w}^H \boldsymbol{C} = \boldsymbol{g}^H$, or $\boldsymbol{C}^H \boldsymbol{w} = \boldsymbol{g}$ where C is a $N\text{X}M_c$ constraint matrix where columns are linearly independent.
- LCMP beamformer minimizes,

$$\boldsymbol{P}_0 = \boldsymbol{w}^H \boldsymbol{R}_p \boldsymbol{w}$$

  under the constraint, $\boldsymbol{w}^H \boldsymbol{C} = \boldsymbol{g}^H$, or $\boldsymbol{C}^H \boldsymbol{w} = \boldsymbol{g}$

Linearly Constrained Minimum Variance (LCMV) Beamforming

Types of constraints in LCMV beamformers

- Distortionless Constraints,

$$\boldsymbol{w}^H\boldsymbol{a}(\boldsymbol{\phi}) = 1$$

It guarantees that signal from direction θ will pass through undistorted.

- Directional Constraints,

$$\boldsymbol{w}^H\boldsymbol{a}(\boldsymbol{\phi_i}) = g_i$$

where $i = 1, 2, .., M_0$ and $\boldsymbol{\phi}_i$ denotes DOA along which we want to impose the constraint g_i which is a complex number in general.

For example:

$$\boldsymbol{w}^H\boldsymbol{a}(\phi_i) = 1$$

and

$$\boldsymbol{w}^H\boldsymbol{a}(\phi_i + \Delta\phi_i) = 1$$
$$\boldsymbol{w}^H\boldsymbol{a}(\phi_i - \Delta\phi_i) = 1$$
$$\boldsymbol{C} = [\boldsymbol{a}(\phi_i)\vdots\boldsymbol{a}(\phi_i + \Delta\phi_i)\vdots\boldsymbol{a}(\phi_i - \Delta\phi_i)]$$
$$\boldsymbol{g} = [1\ 1\ 1]^T$$

and,

$$\boldsymbol{w}^H\boldsymbol{C} = \boldsymbol{g}^H$$

Linearly Constrained Minimum Variance (LCMV) Beamforming

Null Constraints, applied when interference coming from known direction,

$$\boldsymbol{w}^H \boldsymbol{a}(\boldsymbol{\phi_i}) = 0$$

where $i = 2, .., M_0$ Thus,

$$\boldsymbol{C} = [\boldsymbol{a}(\phi_m) \vdots \boldsymbol{a}(\phi_2) \vdots \boldsymbol{a}(\phi_3) \vdots \boldsymbol{a}(\phi_{M0})]$$

$$\boldsymbol{g}^T = [1\ 0\ 0\ ...\ 0]$$

- MVDR beamformer puts a perfect null on directional noise when σ_I^2/σ_w^2 is infinite. σ_I^2: variance of directional noise
  σ_w^2: variance of white noise
- A better constraint would be

$$\boldsymbol{w}^H \boldsymbol{a}(\boldsymbol{\phi_i}) = \epsilon_i$$

where $i = 2, 3, .., M_0$ and ϵ_i is related to σ_I^2/σ_w^2.

Linearly Constrained Minimum Variance (LCMV) Beamforming

Derivative constraints

used to maintain a particular shape of the

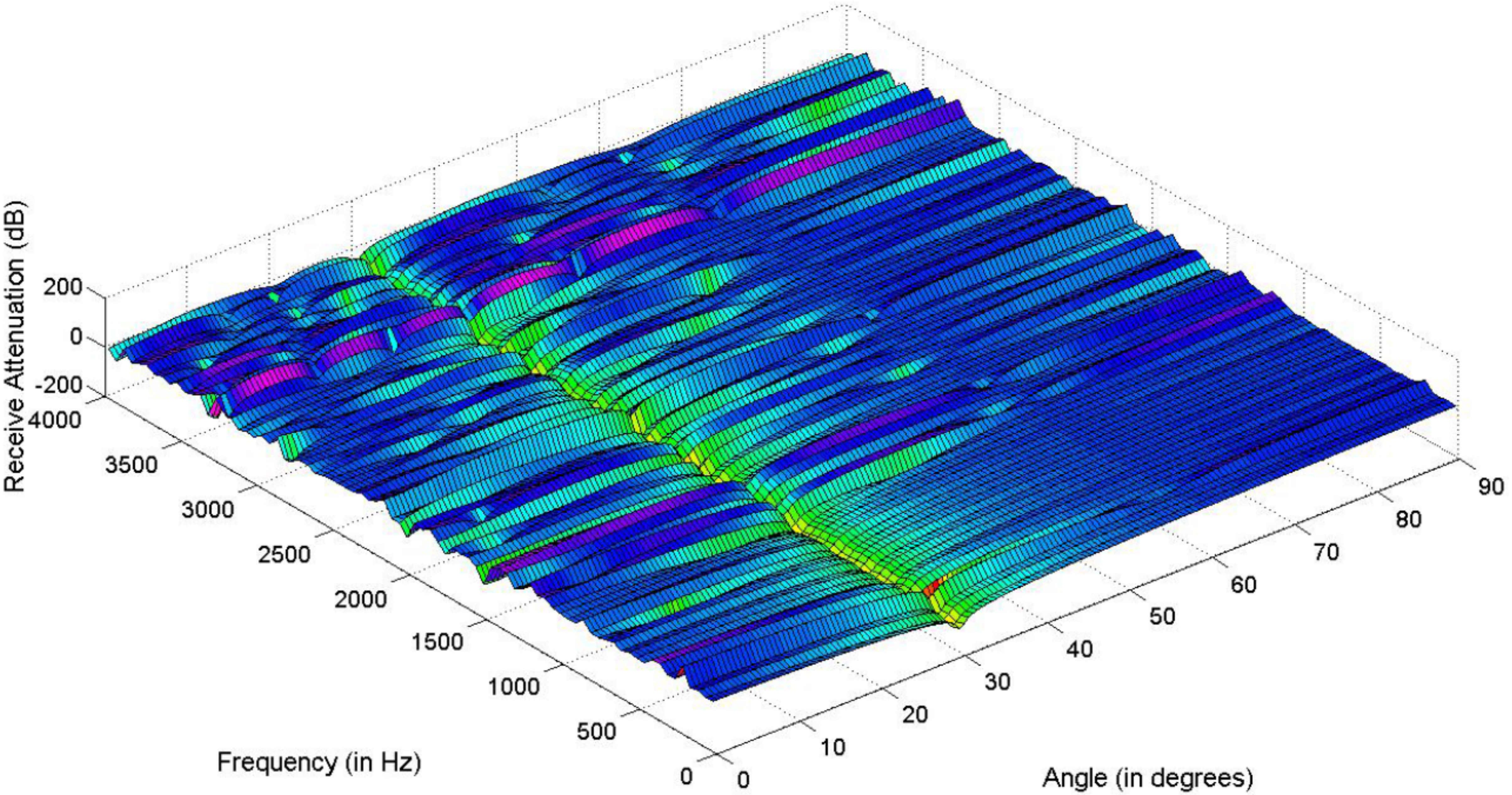
beampattern near the peak or the null

- set beam pattern derivatives with respect to θ and ϕ
- set power pattern derivatives $P(\omega, \theta, \phi)$ at θ and ϕ
- set frequency derivatives $P(\omega, \theta, \phi)$ with respect to ω

LCMV Beamformer : Beampattern Analysis

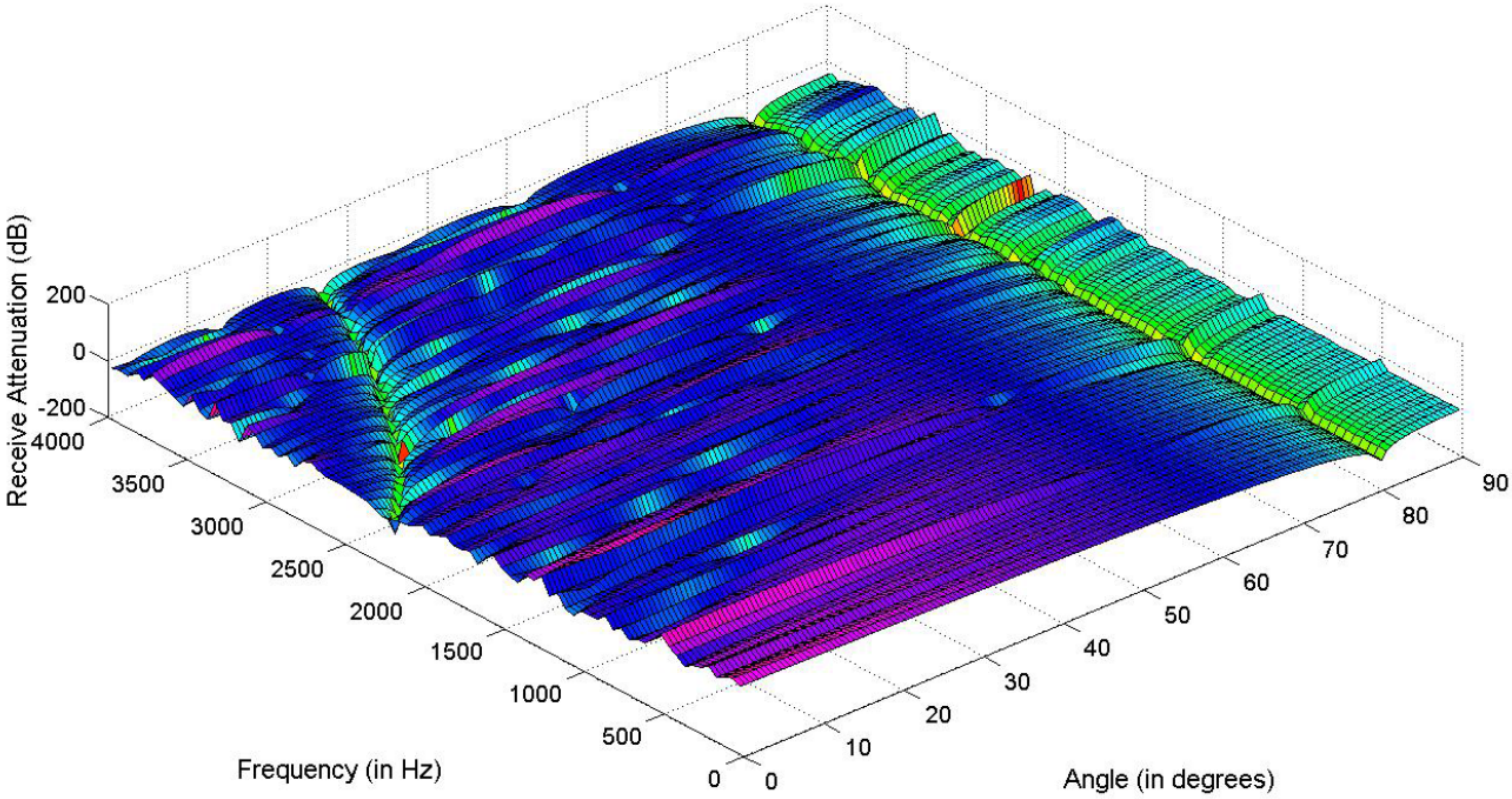
- Number of sensors taken is N=5.
- Speech signals are used as signal of interest and interfering signal.
- DOA of SOI is taken as $60°$.
- One interfering signal is taken with DOA as $30°$.
- Another interfering signal is taken with DOA as $80°$.

LCMV Beamformer : Beampattern Analysis



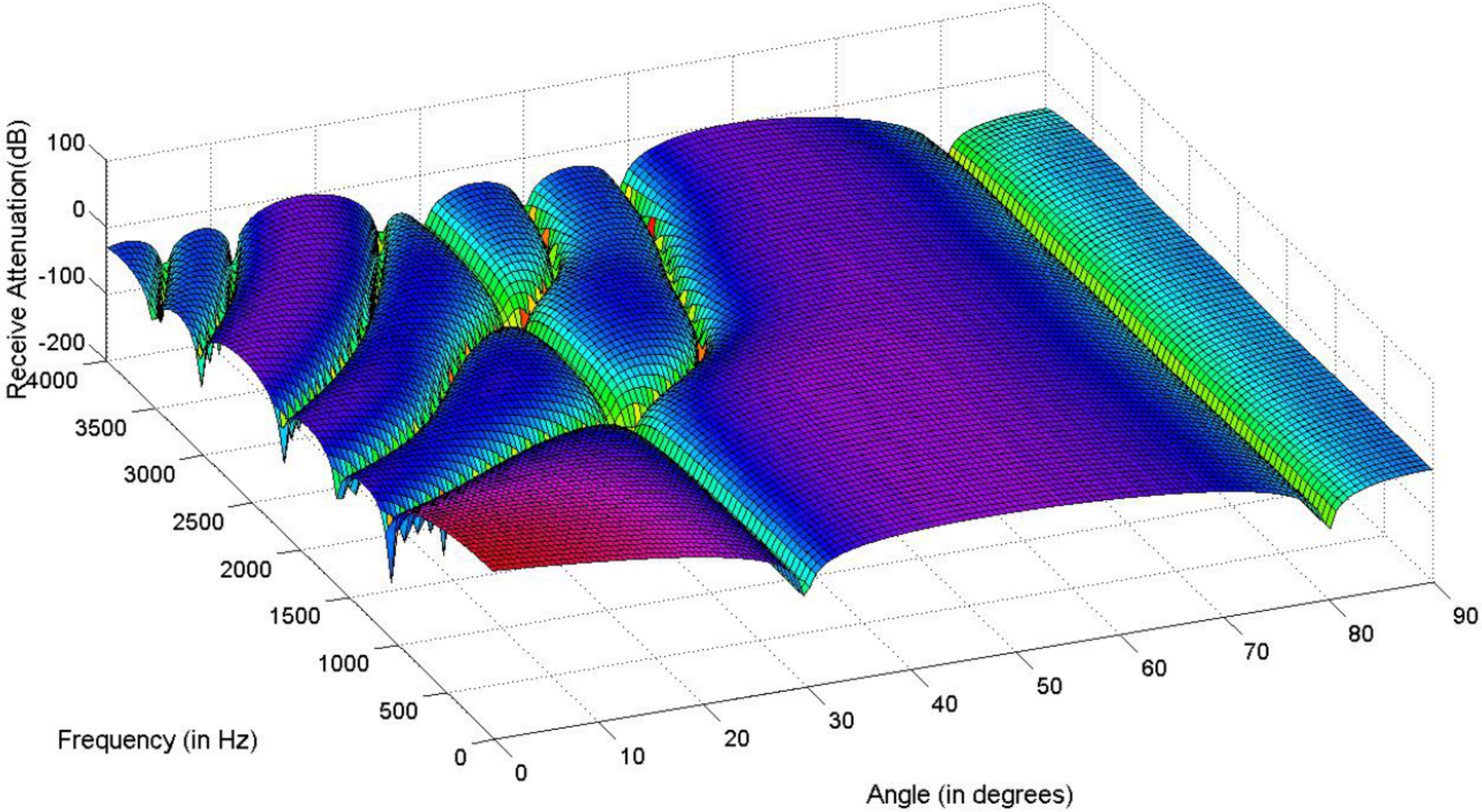
Beampattern for doubly constrained beamformer for signal of interest DOA as 60° and interfering signal DOA as 30° for a 5 microphone array.

LCMV Beamformer : Beampattern Analysis



Beampattern for doubly constrained beamformer for signal of interest DOA as 60° and interfering signal DOA as 80° for a 5 microphone array.

LCMV Beamformer : Beampattern Analysis



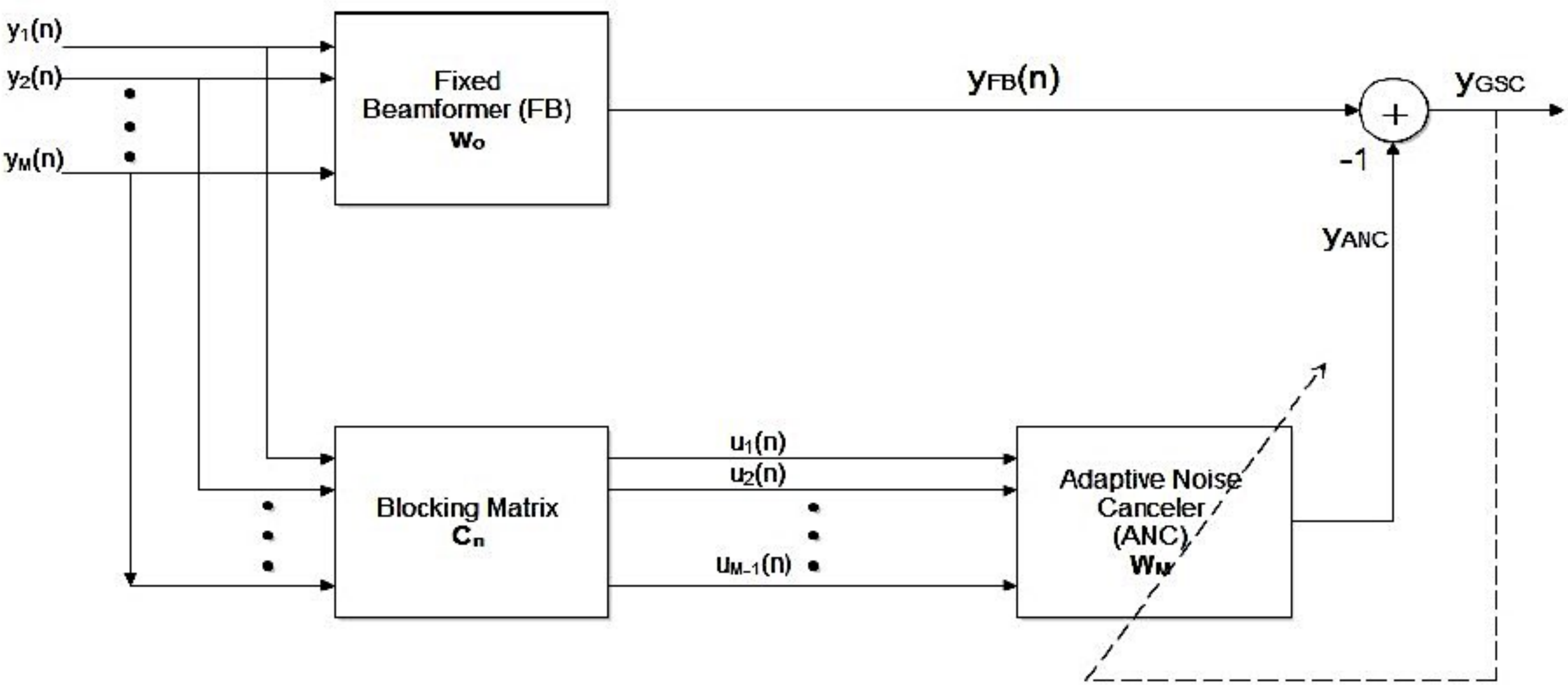
Beampattern for beamformer designed for signal of interest DOA as 60° and interfering signal DOA as 80° and at 30° for a 5 microphone array.

LCMV Beamformer : Beampattern Analysis

- Null has been created at multiple DOAs.
- Performance depends on number of sensors which decide the degree of freedom.
- Null at multiple DOAs pulls down the beampattern over a range of non-look direction.
- Performs well under correlated interference.

Generalized Sidelobe Canceler (GSC) Beamforming

- The GSC consists of three basic part: Beamformer (e.g. Delay and sum beamformer, MVDR), Blocking matrix and Adaptive filter.



Generalized Sidelobe Canceler (GSC) Beamforming

- The GSC splits the traditional beamformer into two orthogonal sub-spaces: the first subspace satisfies the constraints, and thus ideally contains undistorted desired signal.

$$d(k) = \mathbf{S}^T(k)\mathbf{w}_c(k)$$

- Second subspace (lower path in Fig) is orthogonal to $w_c(k)$. Orthogonality is assured by an matrix $B(k)$, which is orthogonal to each column of $S(k)$.

$$\mathbf{S}^T(k)\mathbf{B}(k) = 0$$

$\mathbf{B}(k)$ is called Blocking matrix.

Generalized Sidelobe Canceler (GSC) Beamforming

- Ideally, the output of blocking matrix fitler does not contain desired signal components, and thus is a reference for the noise.
- Then using Adaptive filter is used to minimize error using noise reference signal in output obtained in first path.
- Final output of GSC is given by

$$\hat{x} = (\mathbf{w}_c(k) - \mathbf{B}(k)\mathbf{w}_a(k))^T \mathbf{S}(k)$$

Where $\mathbf{w}_a$ is adaptive filter coefficients matrix

Digital Beamforming using Subspace based Methods
MUSIC, root-MUSIC, and MUSIC-Group Delay

Subspace based Methods : MUSIC

- Correlation-based and beamforming-based source localization methods are limited by resolution ability.
- The methods fail in multi-source environments when sources are closely spaced.
- The limitation arises because these methods do not exploit the sensor array data efficiently.
- Schmidt proposed MUltiple SIgnal Classification (MUSIC) algorithm [5], based on decomposition of array covariance matrix into noise and signal subspace.

$$\begin{bmatrix} p_1(t) \\ p_2(t) \\ \vdots \\ p_I(t) \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1(\phi_1, k) & \mathbf{a}_2(\phi_2, k) & \cdots & \mathbf{a}_L(\phi_L, k) \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_L(t) \end{bmatrix} + \mathbf{v}(t)$$

$$\mathbf{p}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t).$$

Subspace based Methods : MUSIC

$$\mathbf{p}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t).$$

- The array covariance matrix can be written as

$$\mathbf{R_p} = E[\mathbf{p}\mathbf{p}^H] = E[\mathbf{A}\mathbf{s}\mathbf{s}^H\mathbf{A}^H] + E[\mathbf{v}\mathbf{v}^H]$$
$$= \mathbf{A}\mathbf{R_s}\mathbf{A}^H + \sigma^2\mathbf{I} = \mathbf{R}_i + \sigma^2\mathbf{I}$$

- $\mathbf{R_s}$ is signal covariance matrix given by

$$\mathbf{R_s} = E[\mathbf{s}\mathbf{s}^H] = \begin{bmatrix} E[|s_1|^2] & \dots & \dots & 0 \\ 0 & E[|s_2|^2] & \dots & 0 \\ \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & E[|s_L|^2], \end{bmatrix}$$

and $\mathbf{I}$ is identity matrix.

- $\mathbf{R_s}$ is an $L \times L$ diagonal matrix that has all the eigenvalues positive, making $\mathbf{R_s}$ to be positive definite matrix.
- Steering matrix $\mathbf{A}$ comprises of steering vectors which are linearly independent. Hence, $\mathbf{A}$ has full column rank.

Subspace based Methods : MUSIC

- Full column rank of $\mathbf{A}$ and positive definiteness of $\mathbf{R}_s$ guarantees that, when number of sources L is less than number of sensors I, the $I \times I$ matrix $\mathbf{R}_i$ is positive semidefinite with rank L.
- It implies that $I - L$ eigenvalues of $\mathbf{R}_i$ will be zero.

- Assuming $\mathbf{q}_u$ to be u^{th} eigenvector corresponding to zero eigenvalue, we have

$$\mathbf{R}_i\mathbf{q}_u = \mathbf{A}\mathbf{R}_s\mathbf{A}^H\mathbf{q}_u = 0$$

$$\mathbf{q}_u^H\mathbf{A}\mathbf{R}_s\mathbf{A}^H\mathbf{q}_u = 0$$

- As $\mathbf{R}_s$ is positive definite matrix, the following condition holds.

$$\mathbf{A}^H\mathbf{q}_u = 0$$

$$\mathbf{a}_l^H(\phi_l)\mathbf{q}_u = 0 \forall l = 1, 2, \cdots, L \text{ and } \forall u = 1, 2, \cdots, I - L.$$

Subspace based Methods : MUSIC

$$\mathbf{a}_l^H(\phi_l)\mathbf{q}_u = 0 \forall l = 1, 2, \cdots, L \text{ and } \forall u = 1, 2, \cdots, I - L.$$

implies that all the $(I - L)$ noise eigenvectors $(\mathbf{q}_u)$ are orthogonal to the L steering vectors.

- All such noise eigenvectors are denoted by $\mathbf{Q}_n$, as a $I \times (I - L)$ matrix. $\mathbf{Q}_n$ is called the noise subspace.
- The MUSIC spectrum is formulated as

$$P_{MUSIC}(\phi) = \frac{1}{\sum_{u=1}^{I-L} |\mathbf{a}^H(\phi)\mathbf{q}_u|^2} = \frac{1}{||\mathbf{Q}_n^H\mathbf{a}(\phi)||^2} = \frac{1}{(\mathbf{a}^H(\phi)\mathbf{Q}_n\mathbf{Q}_n^H\mathbf{a}(\phi))}.$$

- As the noise eigenvector is orthogonal to steering vector, the denominator becomes zero for $\phi = DOA$.
- Hence, the DOA is estimated from the L largest peaks in MUSIC spectrum corresponding to L incident sources.

Subspace based Methods : Implementation of MUSIC in Practice

- In practice, array covariance matrix $\mathbf{R_p}$ is available for processing, not $\mathbf{R}_i$.
- $\mathbf{R_p}$ is to be estimated from sample covariance matrix for N_s snapshots as

$$\hat{\mathbf{R}}_{\mathbf{p}} = \frac{1}{N_s}\sum_{t=1}^{N_s}\mathbf{p}(t)\mathbf{p}^H(t)$$

- When the data is Gaussian, the sample covariance matrix converges to true covariance matrix.
- Now, let $\mathbf{q}_i$ be any eigenvector of $\hat{\mathbf{R}}_i$ with eigenvalue as λ_i, then

$$\begin{aligned}\hat{\mathbf{R}}_i\mathbf{q}_i &= \lambda_i\mathbf{q}_i\\ \hat{\mathbf{R}}_{\mathbf{p}}\mathbf{q}_i &= \hat{\mathbf{R}}_i\mathbf{q}_i + \sigma^2\mathbf{I}\mathbf{q}_i\\ &= (\lambda_i + \sigma^2)\mathbf{q}_i\end{aligned}$$

- It means that any eigenvector of $\hat{\mathbf{R}}_i$ is also an eigenvector of $\hat{\mathbf{R}}_{\mathbf{p}}$ with eigenvalue as $(\lambda_i + \sigma^2)$.

Subspace based Methods : Implementation of MUSIC in Practice

- So, if $\hat{\mathbf{R}}_i = \mathbf{Q}\Lambda\mathbf{Q}^H$ then

$$\hat{\mathbf{R}}_\mathbf{p} = \mathbf{Q}[\Lambda + \sigma^2\mathbf{I}]\mathbf{Q}^H$$

$$\hat{\mathbf{R}}_\mathbf{p} = \mathbf{Q}\begin{bmatrix} \lambda_1+\sigma^2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2+\sigma^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_L+\sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}\mathbf{Q}^H$$

- The eigenvector matrix $\mathbf{Q}$ is decomposed into signal subspace $\mathbf{Q}_s$ and noise subspace $\mathbf{Q}_n$.
- The eigenvectors corresponding to the highest L eigenvalues, form the signal subspace matrix of order $I \times L$.
- The other $I - L$ columns of $\mathbf{Q}$ (noise eigenvectors) form noise subspace, $\mathbf{Q}_n$ with eigenvalues σ^2.
- Now the MUSIC spatial spectrum can be computed as

$$P_{MUSIC}(\phi) = \frac{1}{(\mathbf{a}^H(\phi)\mathbf{Q}_n\mathbf{Q}_n^H\mathbf{a}(\phi))}.$$

Subspace based Methods : MUSIC for Broadband Signals

There are two variants of non cohorent wideband DOA estimation: NCO1 and NCO2 which use spectral MUSIC as a tool. The NCO1 estimate is given as

$$\hat{\theta} = \arg\min_{\theta} \quad \Sigma_{j=0}^{L-1} S^H(\omega_j)Q_n(\omega_j)Q_n^H(\omega_j)S(\omega_j)$$

where L is the no of frequency bins, $S(\omega_j)$ and $Q_n(\omega_j)$ are the noise subspace and array steering vector at frequency (ω_j). Similarly, the NCO2 estimate is

$$\hat{\theta} = \arg\max_{\theta} \quad \Sigma_{j=0}^{L-1} \frac{1}{S^H(\omega_j)Q_n(\omega_j)Q_n^H(\omega_j)S(\omega_j)}$$

It turns out that NCO1 and NCO2 perform similarly even though their formulations are quite different.

Subspace based Methods : MUSIC-Group Delay

- MUSIC-Group delay spectrum is defined as

$$P_{MGD}(\phi) = (\sum_{u=1}^{I-L} |\nabla arg(\mathbf{a}^H(\phi)\mathbf{q}_u)|^2) P_{MUSIC}(\phi)$$

  where ∇ arg indicates gradient of unwrapped phase spectrum of $(\mathbf{a}^H(\phi)\mathbf{q}_u)$.

- The gradient is with respect to the spatial variables ϕ.
- A sharp transition at the DOAs is observed in unwrapped phase spectrum of MUSIC.
- Gradient of the unwrapped phase spectrum (group delay of MUSIC) results in sharp peaks at the location of the DOAs.
- In practice, abrupt changes can occur in the phase due to small variations in the signal caused by microphone calibration errors. This leads to spurious peaks in group delay spectrum.
- However, the product of MUSIC and group delay spectra, called MUSIC-Group delay [6], removes such spurious peaks and gives high resolution estimation.

Subspace based Methods : MUSIC and MUSIC-Group Delay over a UCA

- The MUSIC-magnitude spectrum is given as

$$P_{MUSIC}(\theta,\phi) = \frac{1}{\sum_{i=1}^{N-M} |s^H(\theta,\phi).q_i|^2} = \frac{1}{||Q_n^H.s(\theta,\phi)||^2}$$

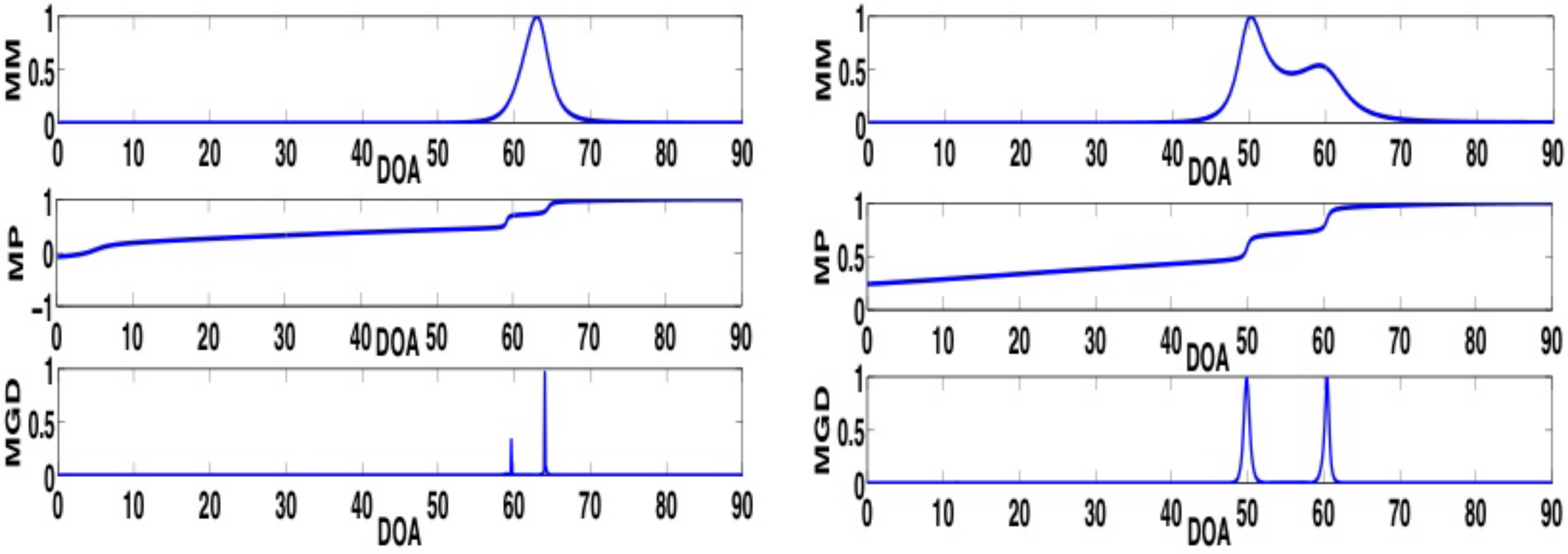
where $s(\theta,\phi)$ is a particular steering vector which forms S and Q_n is the $N \times (N-M)$ matrix of eigenvectors spanning the noise subspace. q_i is a particular eigenvector which forms Q_n

- The MUSIC-Group Delay for DOA estimation over UCA is defined as

$$P_{MGD}(\theta,\phi) = (\sum_{i=1}^{N-M} |\nabla \arg(s^H(\theta,\phi).q_i)|^2).P_{MUSIC}(\theta,\phi)$$

where $\nabla \arg$ indicates gradient of the argument of $s^H(\theta,\phi).q_i$

MUSIC and MUSIC-Group Delay: Some DOA Estimation Results

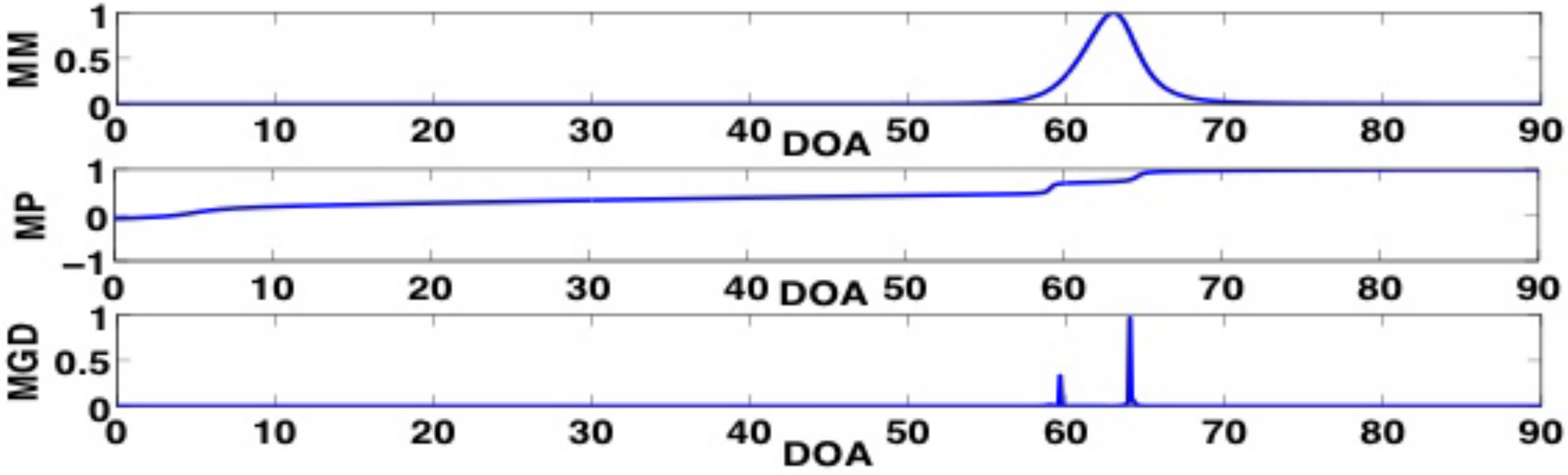


(a)

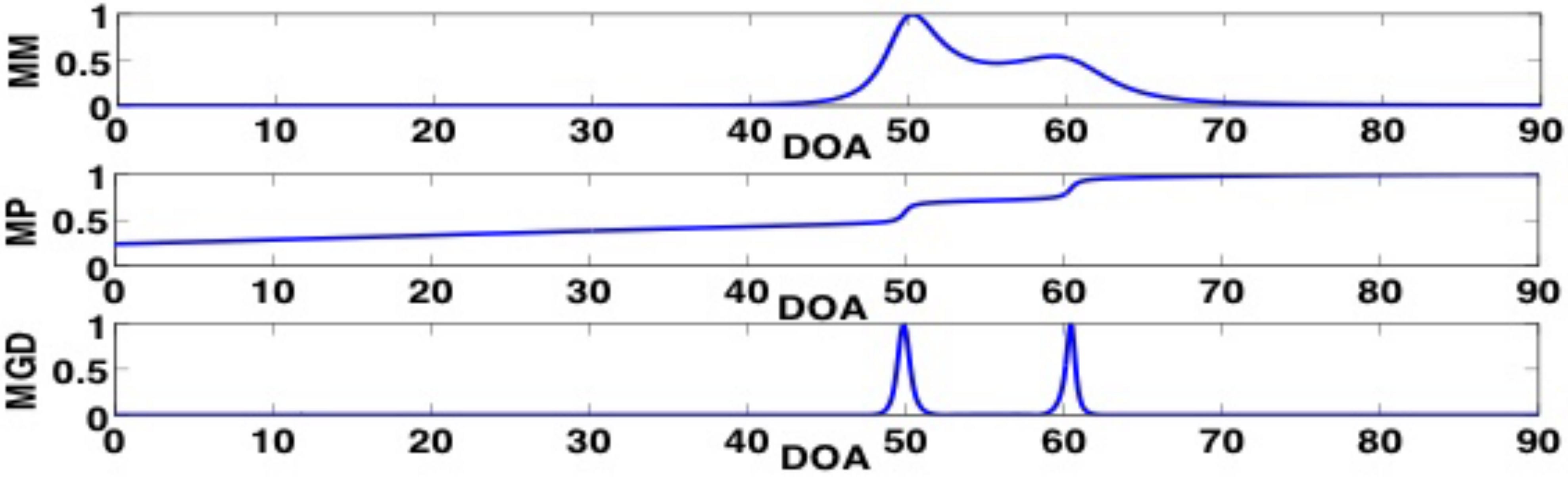
(b)

Figure: MUSIC, Unwrapped phase (of MUSIC) and MUSIC-Group delay spectra for two sources with azimuth (a) 60° and 65°, (b) 50° and 60°.

MM vs MGD for Sources at (a) 60° and 65° (b) 50° and 60°

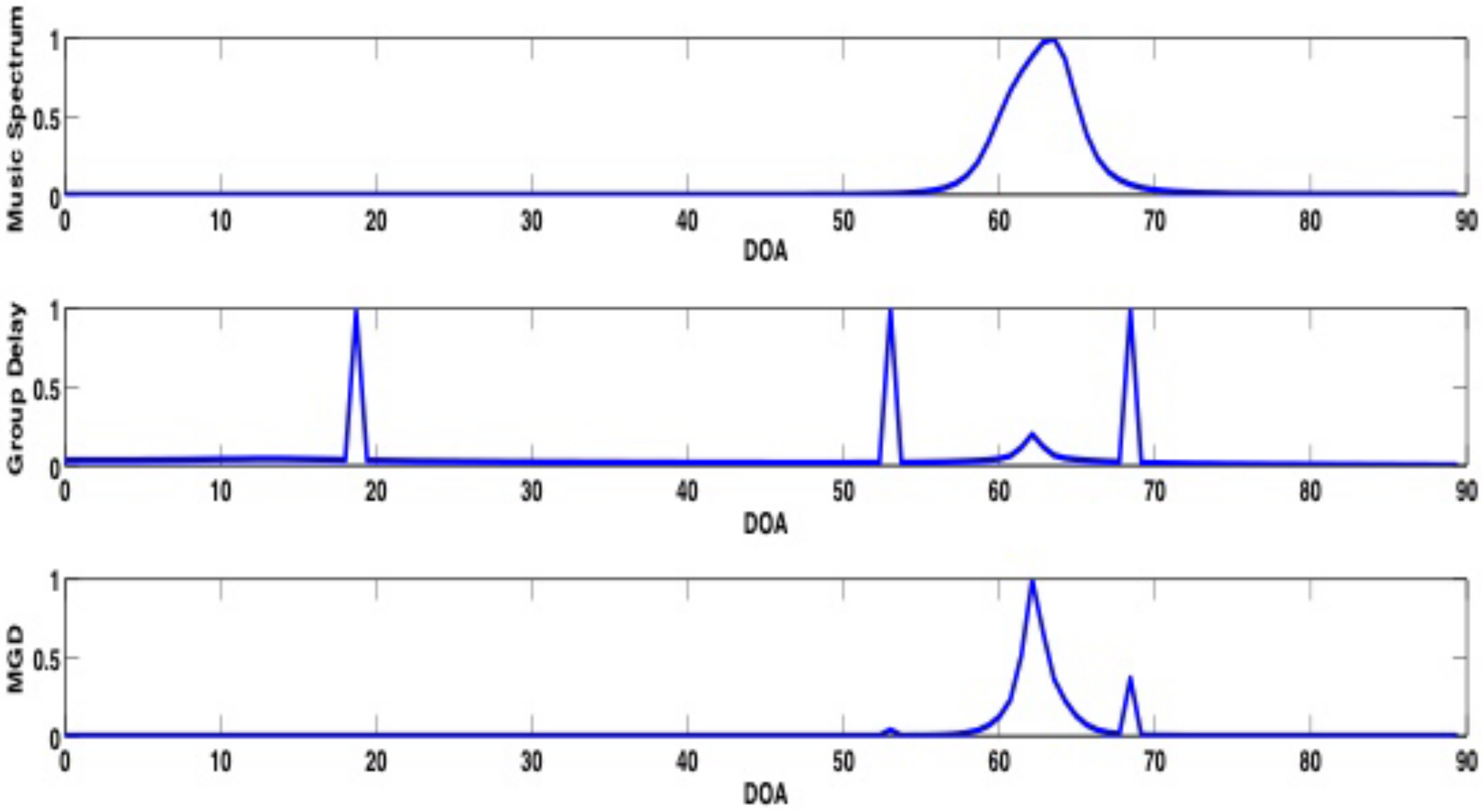


(a)

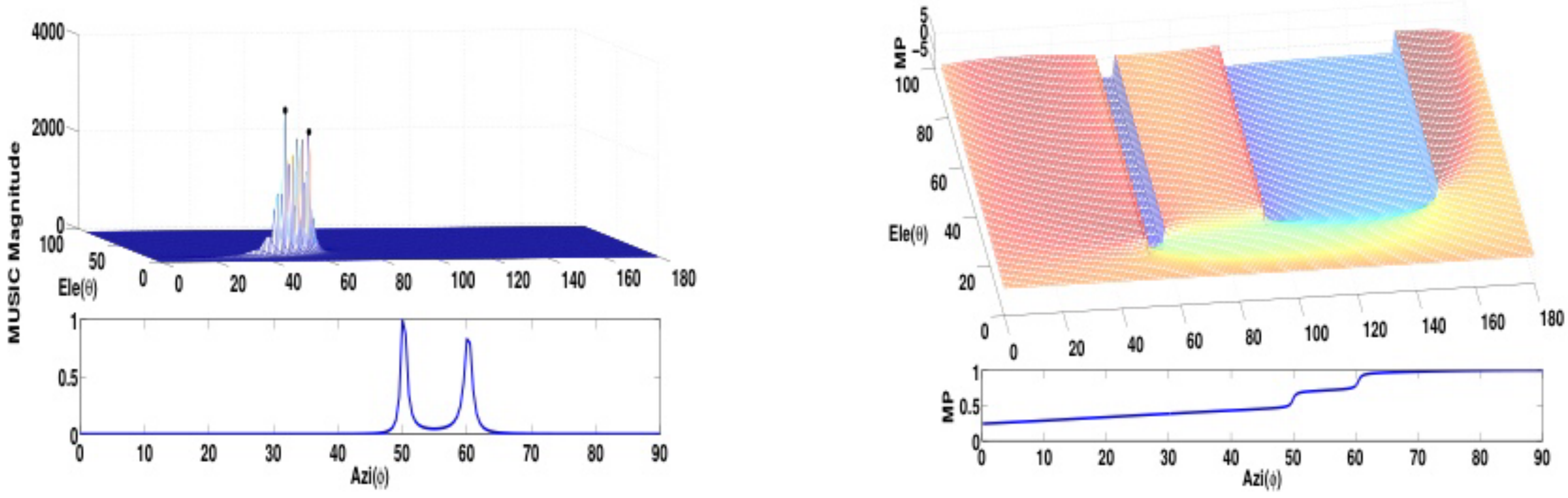


(b)

Spurious peak removal using product spectrum



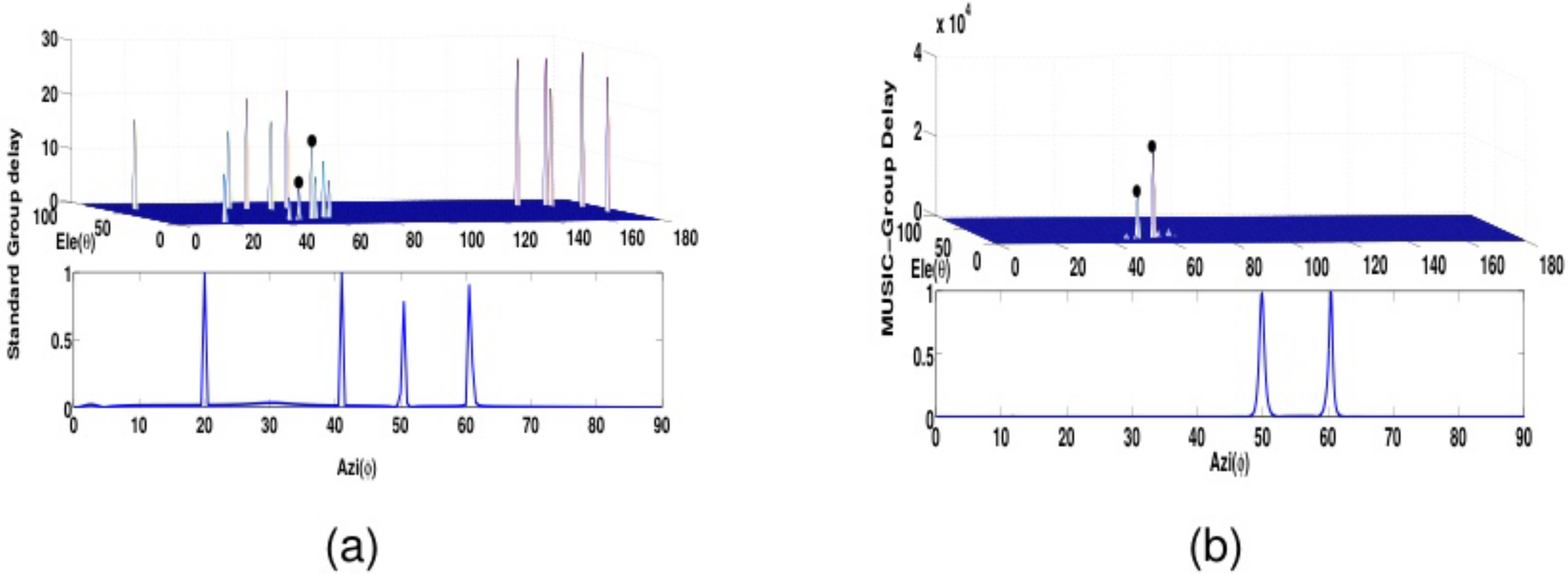
MUSIC Magnitude and MUSIC phase Spectrum



(a) (b)

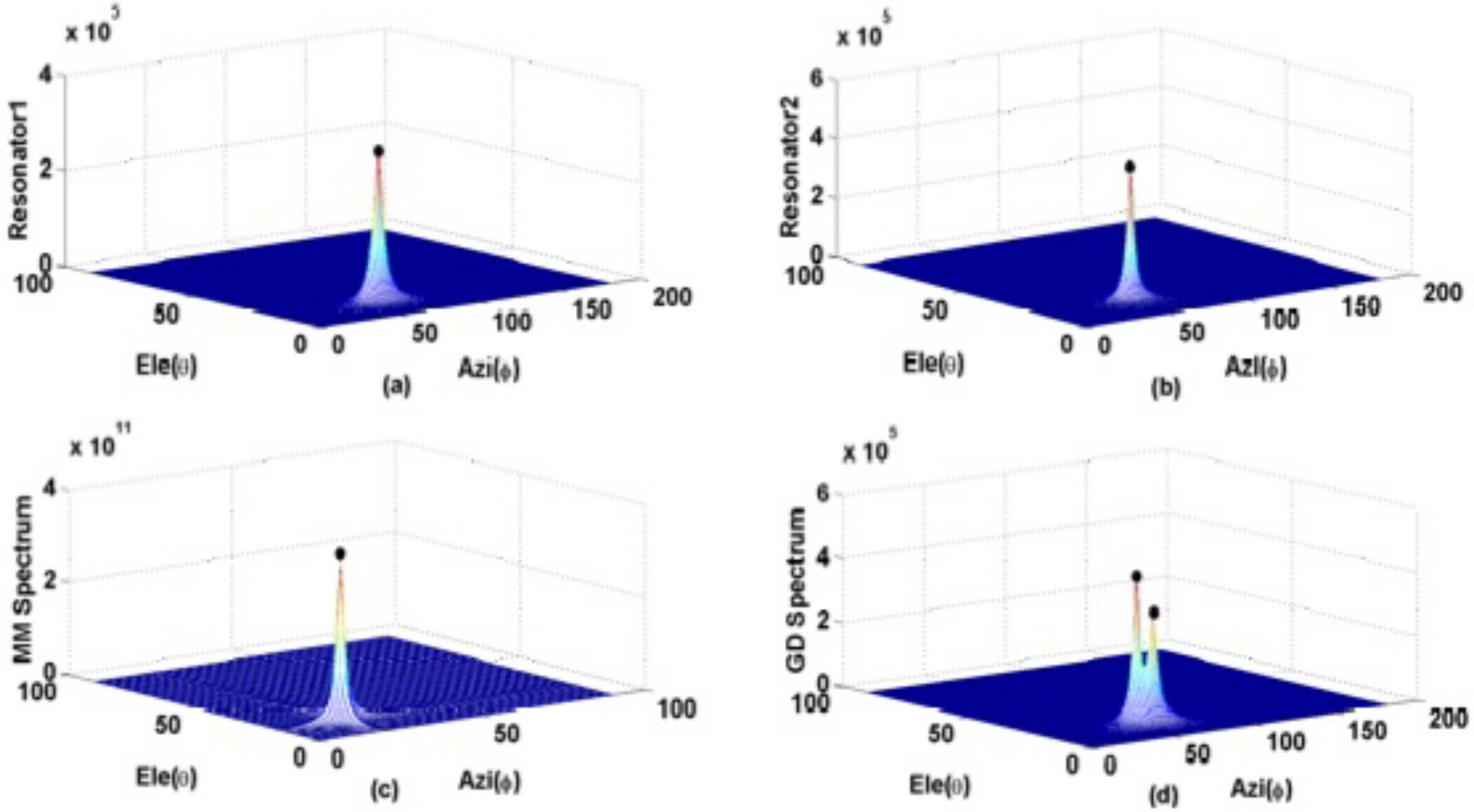
(a) Spectral magnitude of MUSIC for UCA (top) and ULA (bottom). (b)Spectral phase of MUSIC for UCA (top) and ULA (bottom). Sources at (15°,50°) and (20°,60°) for UCA. Sources at 50° and 60° for ULA.

Group Delay and MUSIC-Group Delay Spectrum



(a) Standard group delay spectrum of MUSIC for UCA (top) and ULA (bottom) (b) MUSIC-Group delay spectrum for UCA (top) and ULA (bottom).

Additive Property of MUSIC-Group Delay Spectrum [Expand]



2D spectral plots for the cascade of two individual DOAs (resonators), (a) Source with DOA (15°,60°) (b) Source with DOA (18°,55°) (c) MUSIC-Magnitude spectrum (d) MUSIC-GD spectrum.

Subspace based Methods : root-MUSIC

- The accuracy is limited by the discretization at which the spectrum ($P_{MVDR}(\phi)$, $P_{MUSIC}(\phi)$, or $P_{MGD}(\phi)$) is estimated.
- Moreover, it requires a comprehensive search algorithm for deciding candidate peak corresponding to DOA of a source.
- root-MUSIC proposed in [8], is a search free algorithm, and it estimates DOAs as roots of MUSIC polynomial. Hence, the solution is exact and not limited by the discretization.
- The MUSIC spectrum can also be written as

$$\begin{aligned} P_{MUSIC}^{-1}(\phi) &= \mathbf{a}^H(\phi)\mathbf{Q}_n\mathbf{Q}_n^H\mathbf{a}(\phi) \\ &= \mathbf{a}^H(\phi)C\mathbf{a}(\phi) \end{aligned}$$

$$\text{where, } C = \mathbf{Q}_n\mathbf{Q}_n^H$$

- Substituting $z = e^{jkd\cos(\phi)}$ in Equation 28, steering vector for ULA can be expressed as

$$\mathbf{a}(\phi) = \begin{bmatrix} 1 & z & z^2 & \cdots & z^{I-1} \end{bmatrix}^T$$

Subspace based Methods : root-MUSIC

written in a polynomial form (called root-MUSIC polynomial), as shown below.

$$P^{-1}_{MUSIC}(z) = \sum_{m=0}^{I-1}\sum_{n=0}^{I-1} z^n C_{mn} z^{-m}$$

$$P(z) = \sum_{m=0}^{I-1}\sum_{n=0}^{I-1} z^{n-m} C_{mn}$$

Substituting $n - m = r$ which suggests

$$P(z) = \sum_{r=-(I-1)}^{(I-1)} C_r z^r$$

$$\text{where, } C_r = \sum C_{mn}.$$

- It can be observed that the root-MUSIC polynomial is of degree $(2I - 2)$ with $(2I - 2)$ roots.

Subspace based Methods : root-MUSIC

- If z is the root of the polynomial, $\frac{1}{z^*}$ is also the root.Since, z and $\frac{1}{z^*}$ have the same phase and reciprocal magnitude, one root is within the unit circle while the other is outside.
- Hence, $(I-1)$ roots are within the unit circle and rest $(I-1)$ roots are outside. Out of $(I-1)$ roots within the unit circle, L roots close to unit circle can be used for DOA estimation.

- The azimuth is estimated using

$$\phi = \cos^{-1}\{\frac{\Im ln(z)}{kd}\}$$

where $\Im$ is imaginary part.

Subspace based Methods : root-MUSIC

- The azimuth is estimated using

$$\phi = \cos^{-1}\{\frac{\Im ln(z)}{kd}\}$$

where $\Im$ is imaginary part.

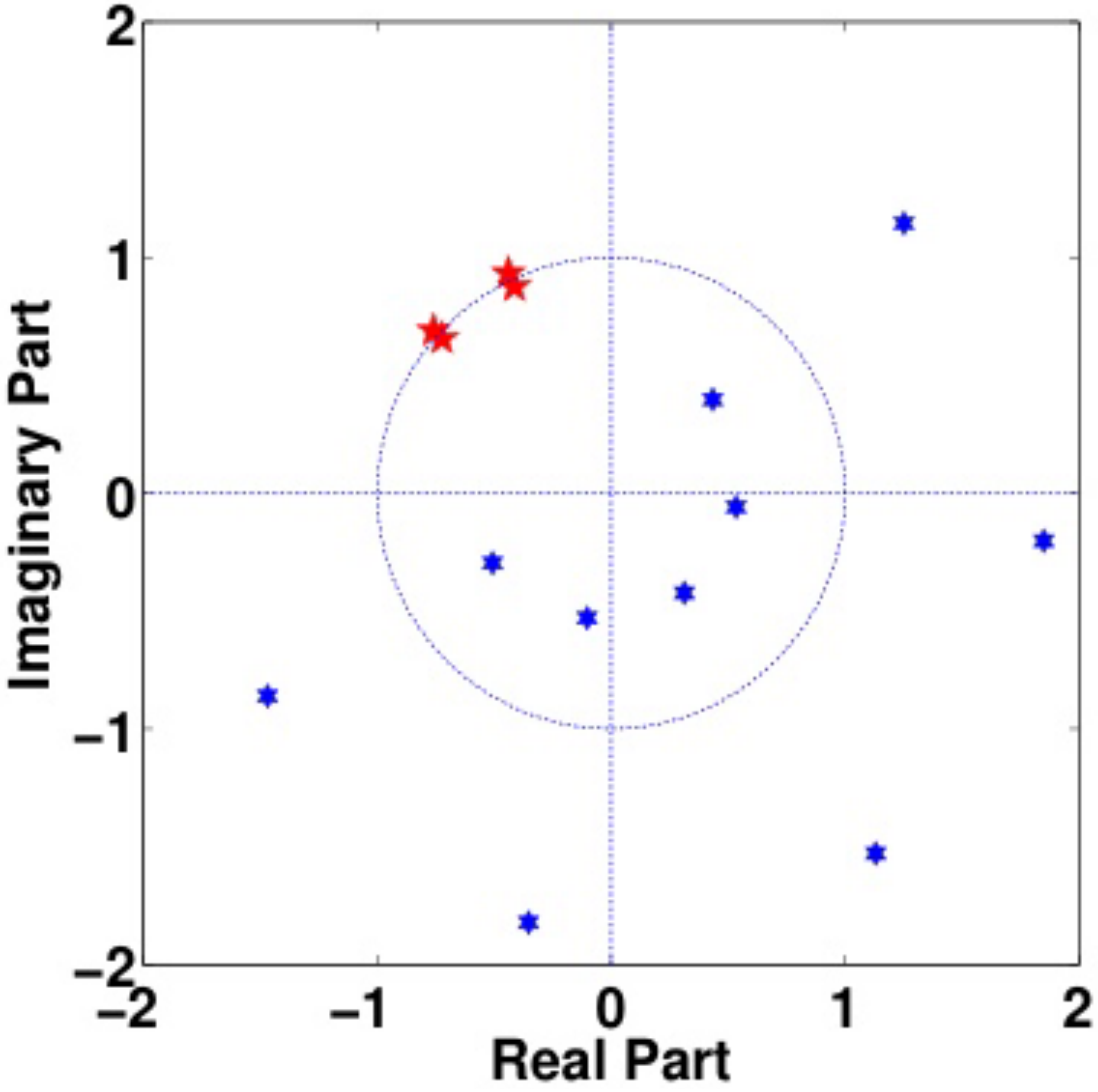
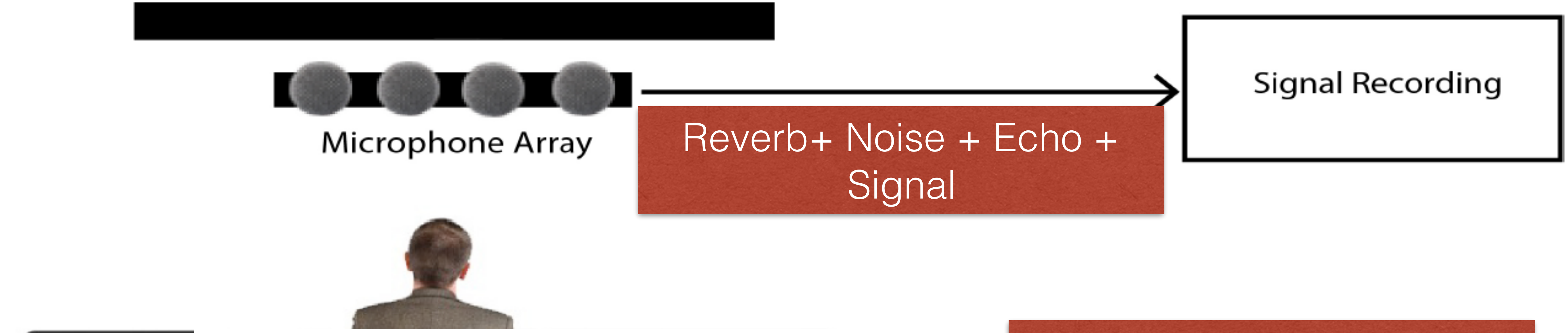
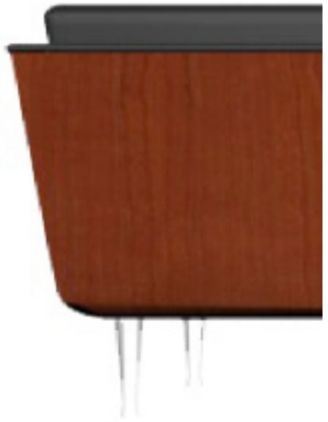


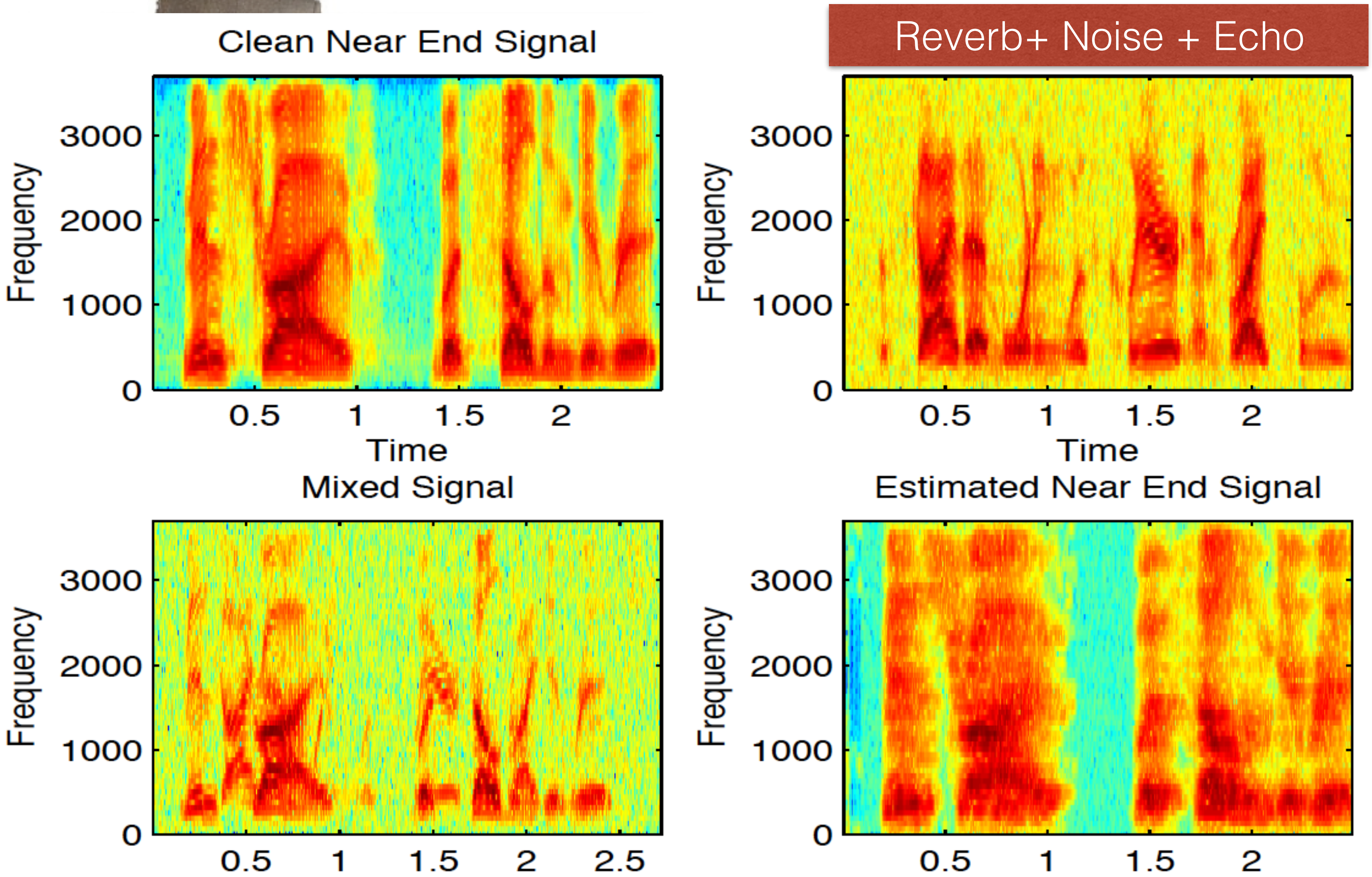
Figure: Z-Plane representation of all the roots of root-MUSIC polynomial using 8 sensors for 2 sources with locations 130° and 140°.

Applications to Speech Source Localization, Enhancement, and Distant Speech Recogntion

Speech Enhancement and Recognition Experiments : ULA





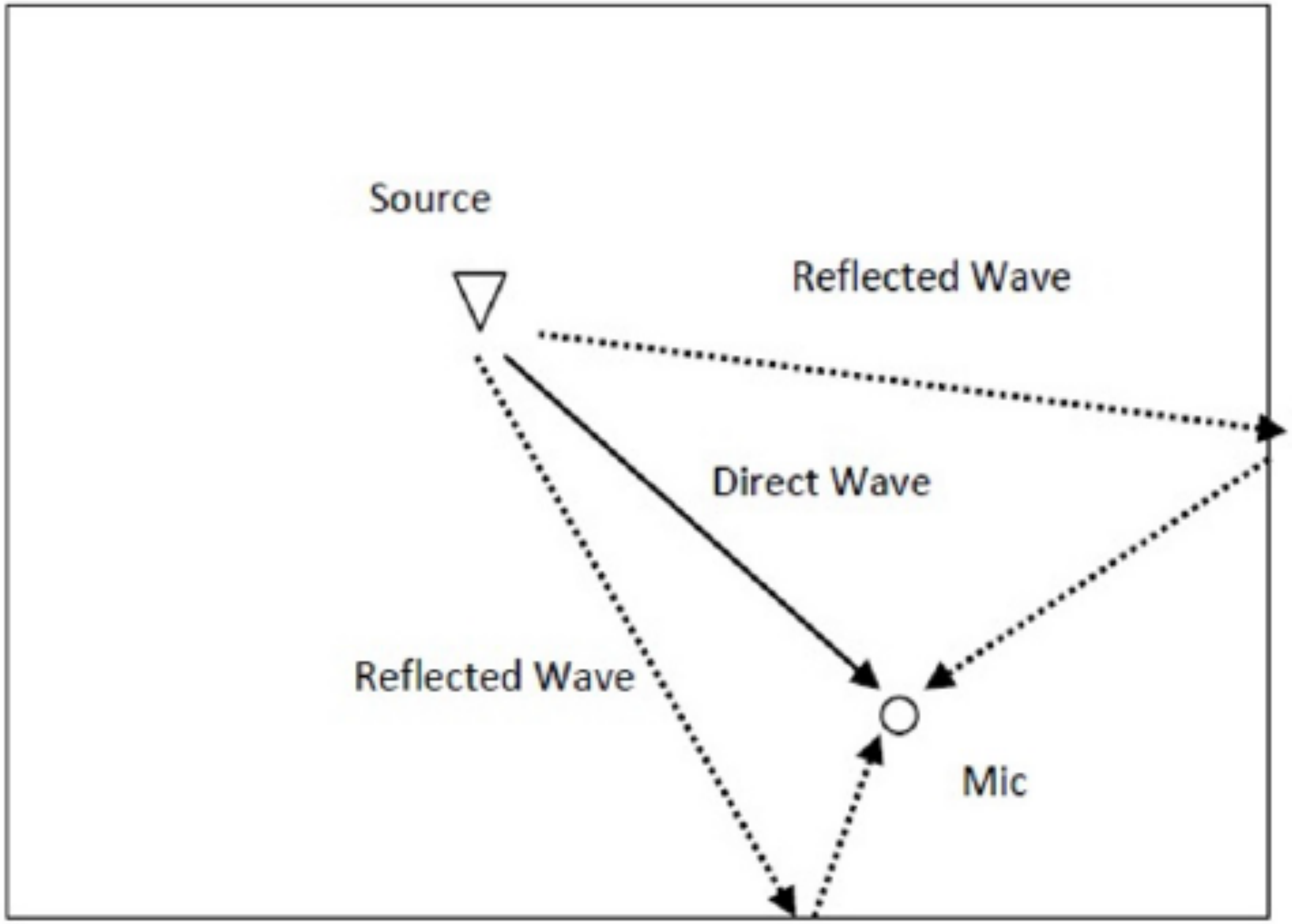

Reverb+ Noise + Echo

Clean Near End Signal

Mixed Signal

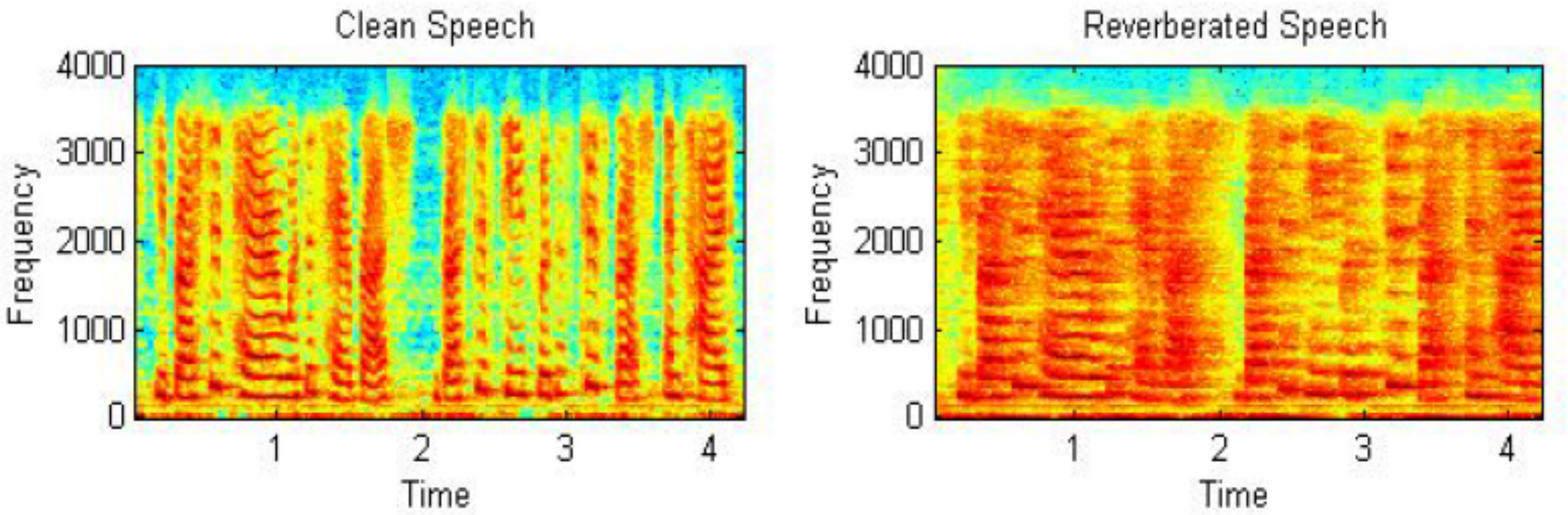
Estimated Near End Signal

Speech Enhancement and Recognition Experiments : ULA



- A speech signal captured using a distant microphone is smeared due to reverberation
- Reverberation is a phenomenon in which multiple delayed and attenuated versions of a signal are added to itself

Speech Enhancement and Recognition Experiments : ULA



- Perceptually we can interpret the reverberated speech signal as the same source signal coming from several different sources at different locations in the room
- Reverberation causes loss in intelligibility of speech.
- The characteristics of the speech signal are altered due to reverberation which causes significant reduction of performance in voice based applications

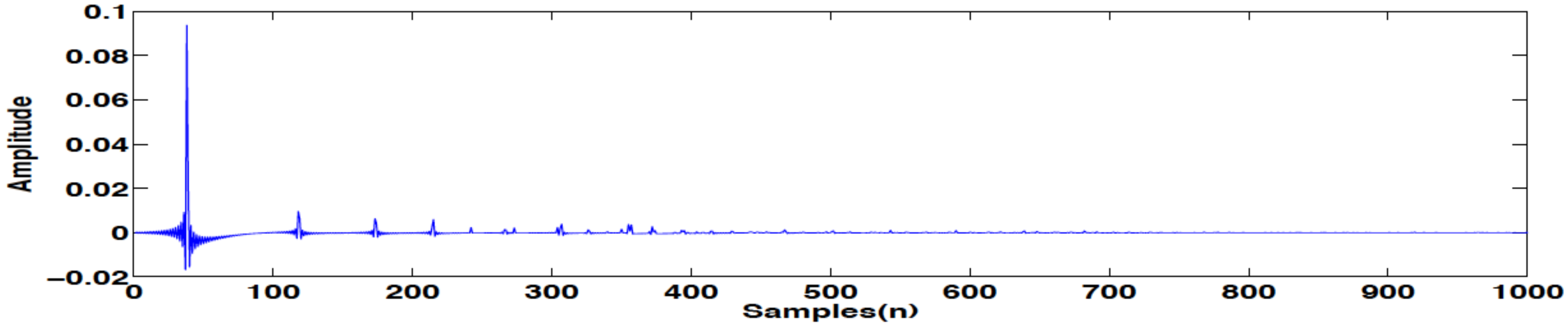
Speech Enhancement and Recognition Experiments : ULA

Analysis under reverberation

- The persistence of sound due to multipath is called reverberation
- Mathematically the multipath is modeled using original signal convolved with room impulse response (RIR)

$$x = p * h + n$$

where h is RIR.



Typical room impulse response, DRR 7.83dB

Speech Enhancement and Recognition Experiments : ULA

Analysis under reverberation

- The Direct to Reverberant energy ratio is calculated as the ratio of the energy in direct sound to the energy of the remaining impulse response

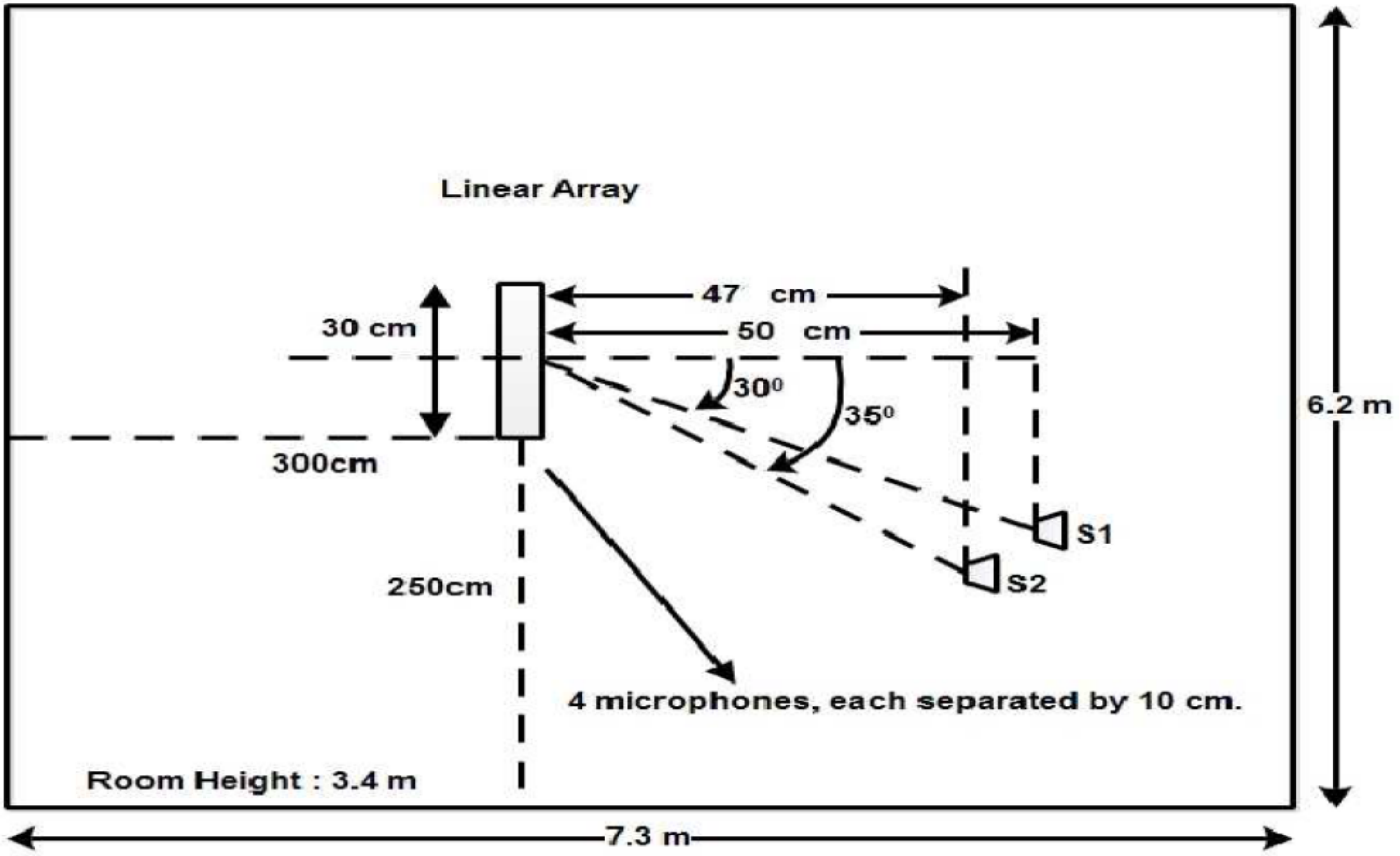
$$DRR = 10log_{10}\left(\frac{\sum_{n=0}^{n_d} h^2(n)}{\sum_{n=n_d+1}^{\infty} h^2(n)}\right) dB$$

  where n_d is direct path component.

- The reverberation time, T_{60}, is time taken for the reverberant energy to decay by 60 dB.
- Higher the DRR implies lower reverberation time.

Speech Enhancement and Recognition Experiments : ULA

Test bed for distant speech acquisition over a ULA



Speech Enhancement and Recognition Experiments : ULA

Experimental conditions for acquiring TIMIT Data over a ULA

- A room with dimensions $7.3m \times 6.2m \times 3.4m$ was used for experiments. The sources are placed at the following two locations with respect to one of the corners of the room.

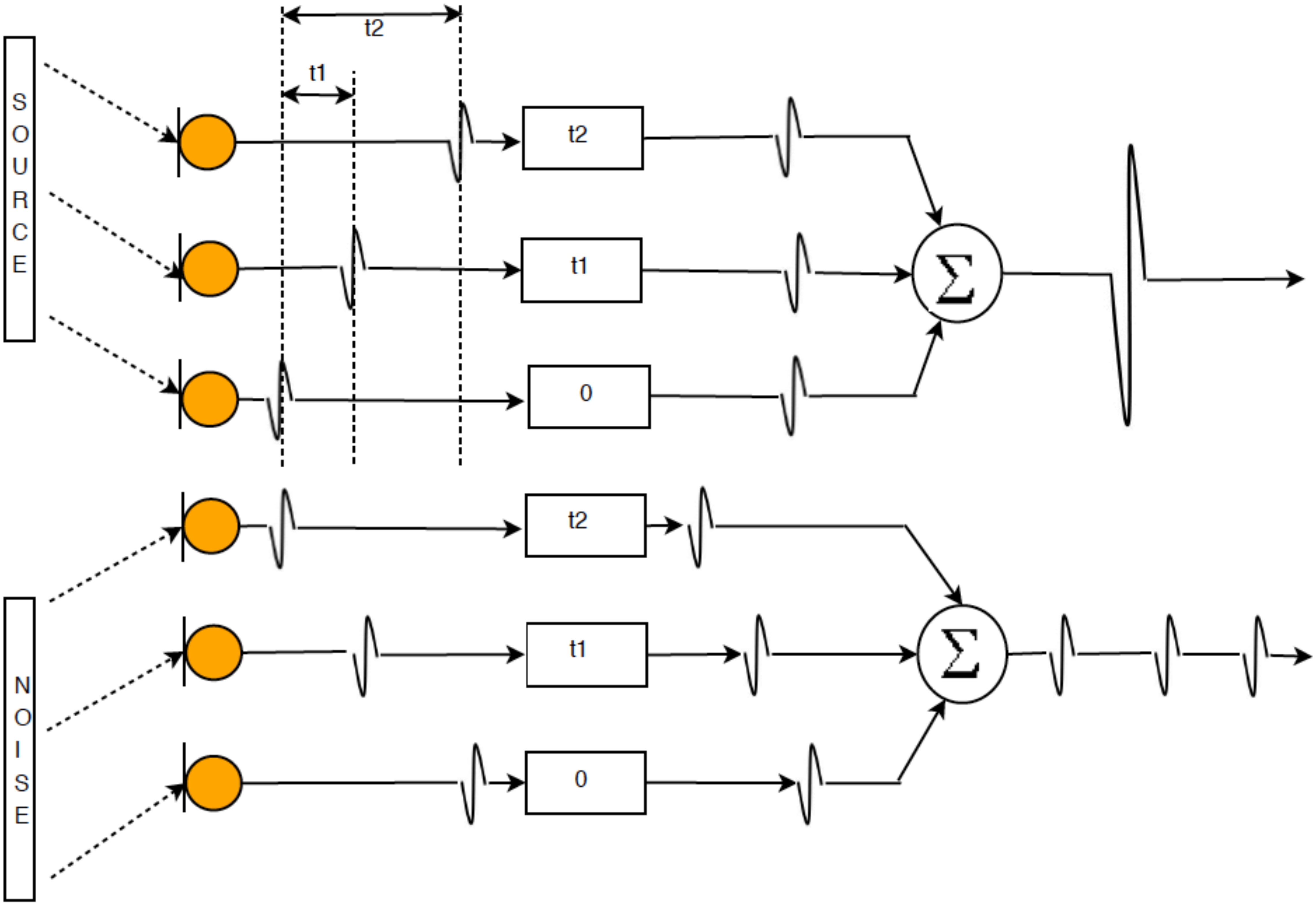
Source - 1 :	$[3.5\ 2.36\ 1.5]^T$
Source - 2 :	$[2.5\ 2.32\ 1.5]^T$

- A 4-element Uniform Microphone Array (ULA) with a spacing of 10 cm was used to perform the DOA estimation. The microphones were located at:

Mic - 1 :	$[3.0\ 2.5\ 1.0]^T$
Mic - 2 :	$[3.0\ 2.6\ 1.0]^T$
Mic - 3 :	$[3.0\ 2.7\ 1.0]^T$
Mic - 4 :	$[3.0\ 2.8\ 1.0]^T$

Speech Enhancement and Recognition Experiments : ULA

Reconstructing Speech using DSB beamformer for ASR Experiments



Speech Enhancement and Recognition Experiments : ULA

Speech recognition experiments

Word Error Rate (% WER)				
Method	**Recognition** Clean	**Recognition** DRR=20dB	**Recognition** DRR=13.77dB	**Recognition** DRR=7.96dB
CTM	7.48	11.55	14.45	21.95
MGD	25.56	39.44	42.77	46.65
RM	35.19	53.64	60.48	62.87
MM	50.61	54.69	63.92	67.44
GCCP	50.61	59.33	63.92	67.87
GCCR	53.77	61.45	70.75	72.17

Large Vocabulary Speech Recognition performance of various DOA Estimation methods on S-TIMIT Data

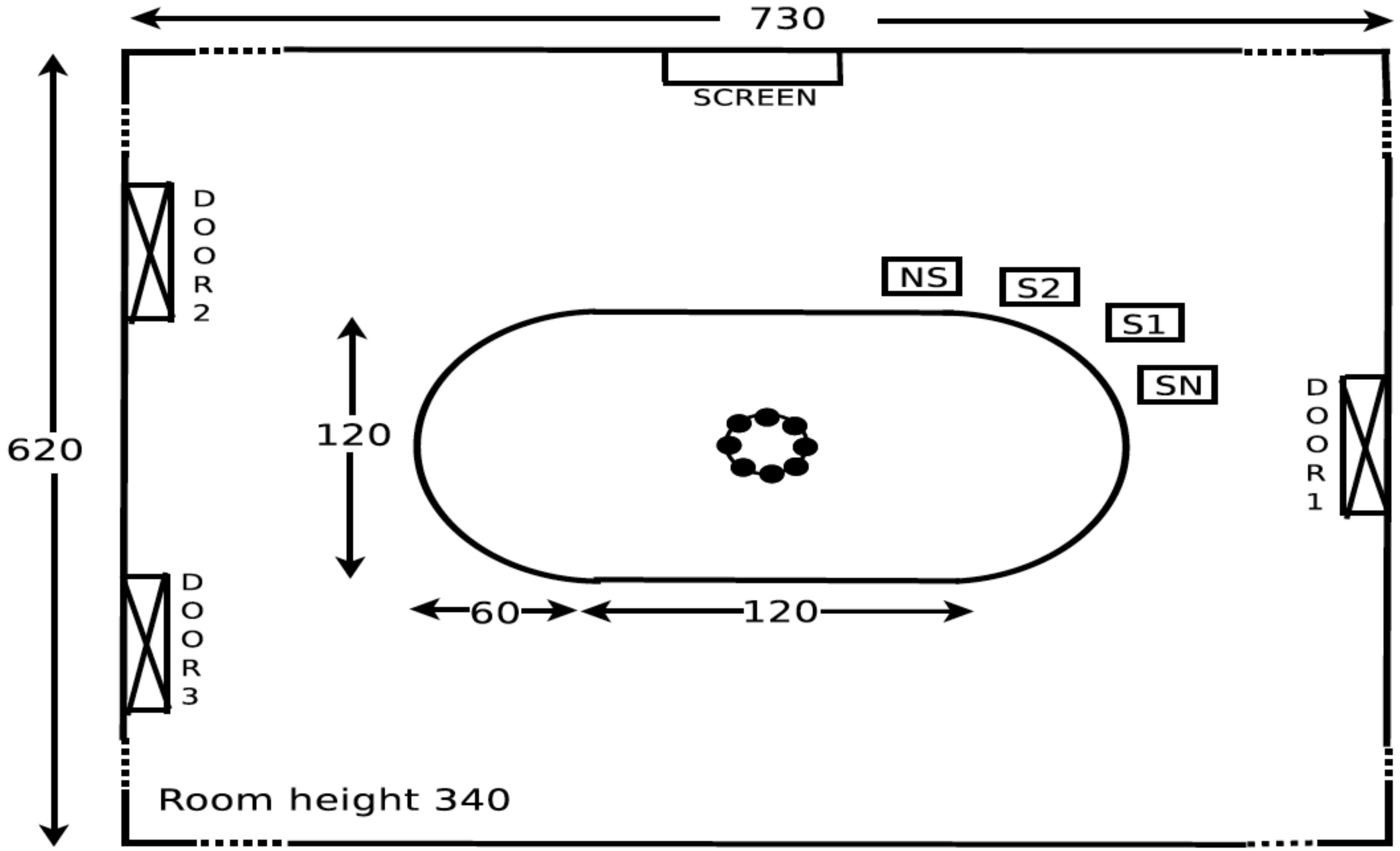
Word Error Rate (% WER)				
Method	**Recognition** Clean	**Recognition** DRR=20dB	**Recognition** DRR=13.77dB	**Recognition** DRR=7.96dB
CTM	11.78	21.66	27.88	29.76
MGD	15.98	25.66	33.11	35.44
RM	19.06	31.66	36.88	43.66
MM	20.55	35.78	40.55	44.35
GCCP	23.66	38.77	42.65	44.77
GCCR	24.63	36.89	44.19	46.53

Comparison of continuous digit speech recognition performance of two spatially contiguous speakers in a meeting room for various methods on the MONC database

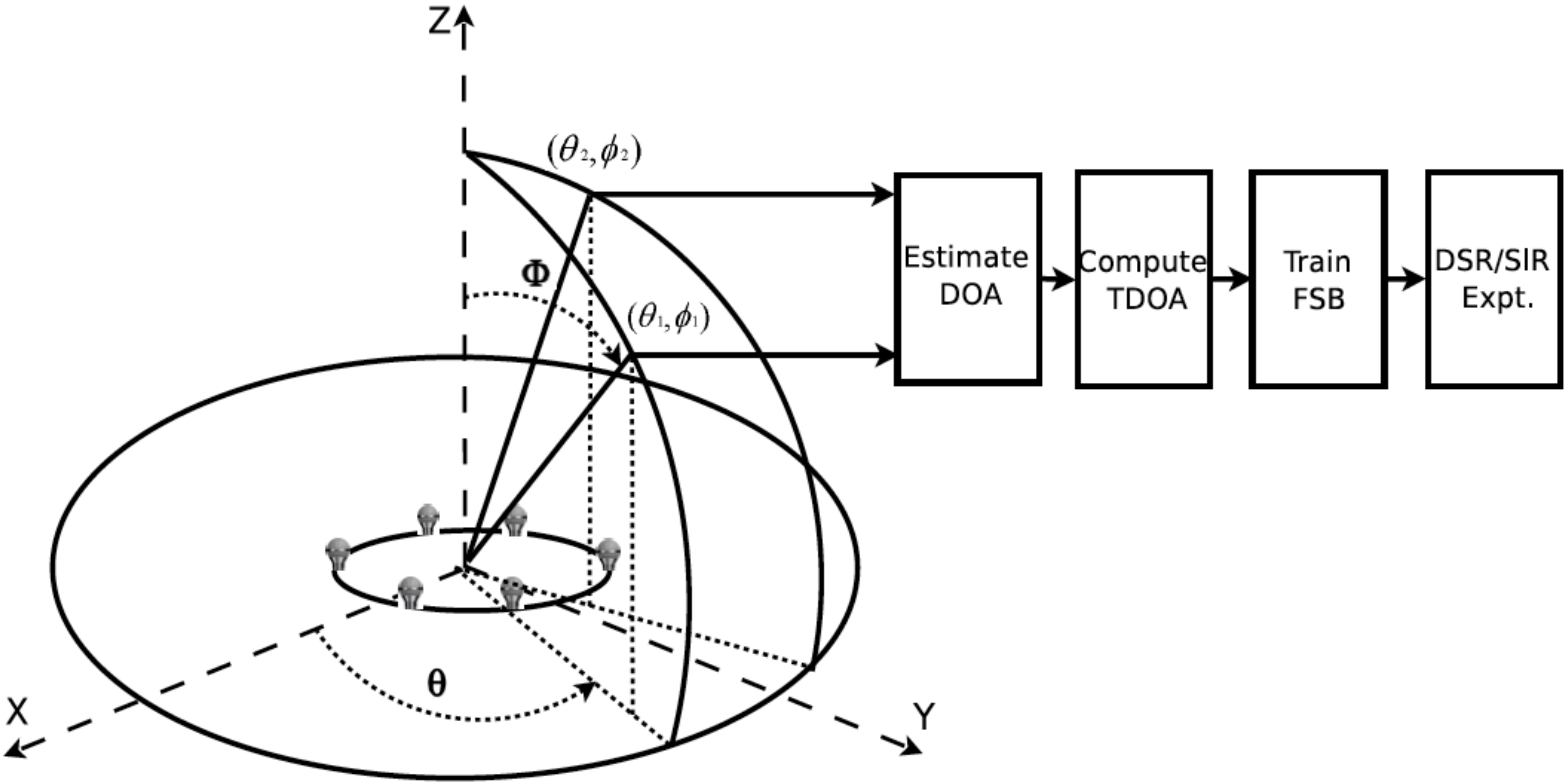
Speech Enhancement and Recognition Experiments : UCA

Experimental Setup

- Meeting room with two speakers (S1 and S2) and two interference (stationary noise source SN and non-stationary noise source NS), located at (17°,35°), (19°,40°), (15°,30°) and (21°,45°) respectively.
- Uniform circular, 15 channel microphone array with a radius of 10 cm was used.
- White noise and babble noise from NOISEX-92 database were used as stationary and non-stationary interfering sources respectively.



Speech Enhancement and Recognition Experiments : UCA



Estimate DOA → Compute TDOA → Train FSB → DSR/SIR Expt.

Methodology Followed in Performance Evaluation

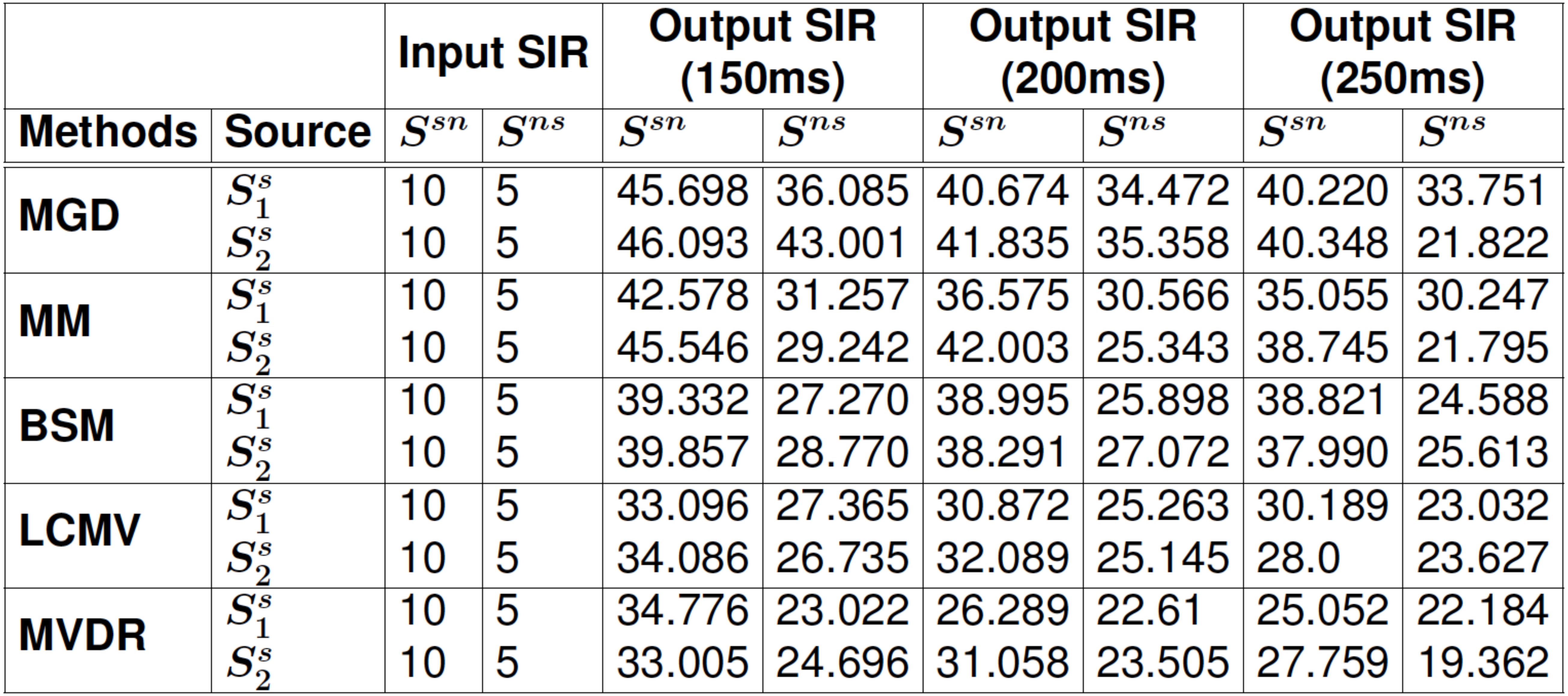
Speech Enhancement and Recognition Experiments : UCA

- Proposed method is evaluated by computing objective measures of perceptual evaluation on enhanced speech.
- The objective measures are, Log-Likelihood Ratio measure (LLR) [20], segmental SNR (segSNR) [20], Weighted-Slope Spectral (WSS) distance [21] and Perceptual Evaluation of Speech Quality, PESQ [22].
- PESQ and segSNR scores should be high while LLR and WSS scores should be low for better reconstruction.
- Desired speaker and stationary noise source pair is considered for evaluation.
- Six hundred sentences from TIMIT database were taken.

Method	T_{60}	LLR	SegSNR	WSS	PESQ
MGD	150	1.3879	-3.1843	35.5345	2.2819
	250	1.6193	-4.8298	36.7986	2.2229
MM	150	1.62	-3.1847	35.6	2.2815
	250	1.6487	-4.9995	37.73	2.2215
BSM	150	1.6657	-3.2	35.5639	2.28
	250	1.6878	-5.108	38.5765	2.2
LCMV	150	1.668	-3.22	36.2	2.2826
	250	1.6994	-5.095	40.0321	2.1746
MVDR	150	1.67	-3.4	36.4	2.2815
	250	1.7379	-5.0356	40.0647	2.1753

Speech Enhancement and Recognition Experiments : UCA

Experiments on Speech Enhancement [SIR]



		Input SIR		Output SIR (150ms)		Output SIR (200ms)		Output SIR (250ms)	
Methods	**Source**	S^{sn}	S^{ns}	S^{sn}	S^{ns}	S^{sn}	S^{ns}	S^{sn}	S^{ns}
MGD	S_1^s	10	5	45.698	36.085	40.674	34.472	40.220	33.751
	S_2^s	10	5	46.093	43.001	41.835	35.358	40.348	21.822
MM	S_1^s	10	5	42.578	31.257	36.575	30.566	35.055	30.247
	S_2^s	10	5	45.546	29.242	42.003	25.343	38.745	21.795
BSM	S_1^s	10	5	39.332	27.270	38.995	25.898	38.821	24.588
	S_2^s	10	5	39.857	28.770	38.291	27.072	37.990	25.613
LCMV	S_1^s	10	5	33.096	27.365	30.872	25.263	30.189	23.032
	S_2^s	10	5	34.086	26.735	32.089	25.145	28.0	23.627
MVDR	S_1^s	10	5	34.776	23.022	26.289	22.61	25.052	22.184
	S_2^s	10	5	33.005	24.696	31.058	23.505	27.759	19.362

Speech Enhancement and Recognition Experiments : UCA

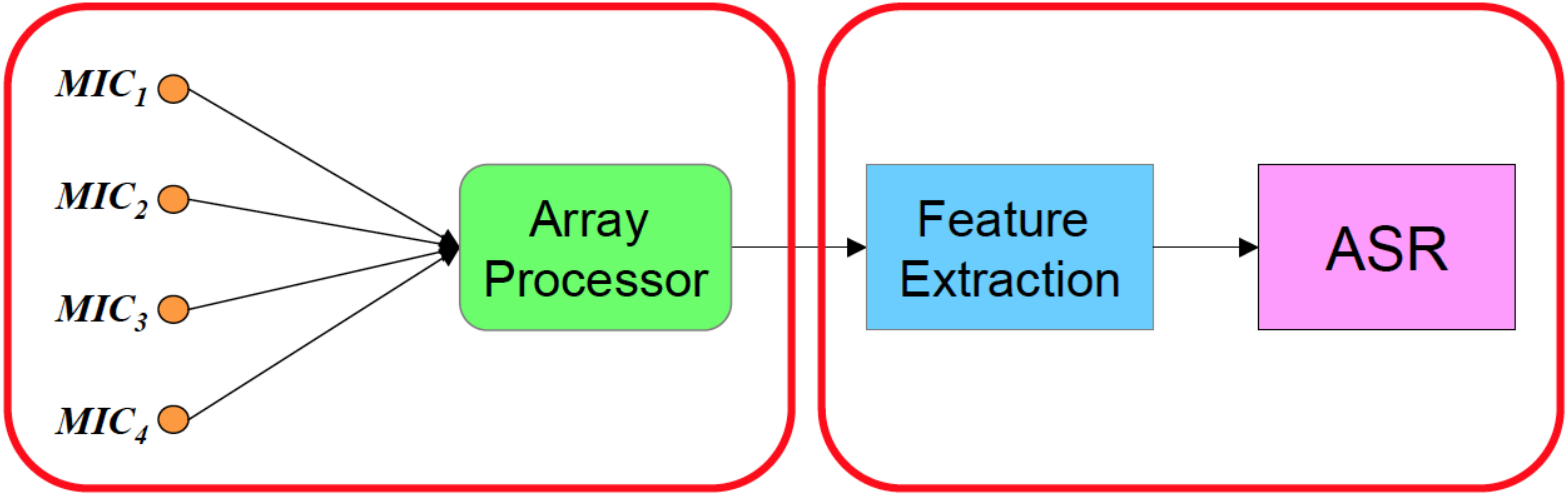
Experiments on distant speech recognition

- DSR result is presented as word error rate (WER).
- WER is given by $WER = 100 - \frac{(W_n - (W_s + W_d + W_i))}{W_n} \cdot 100$.
- TIMIT and MONC both databases were used.

reverberation time

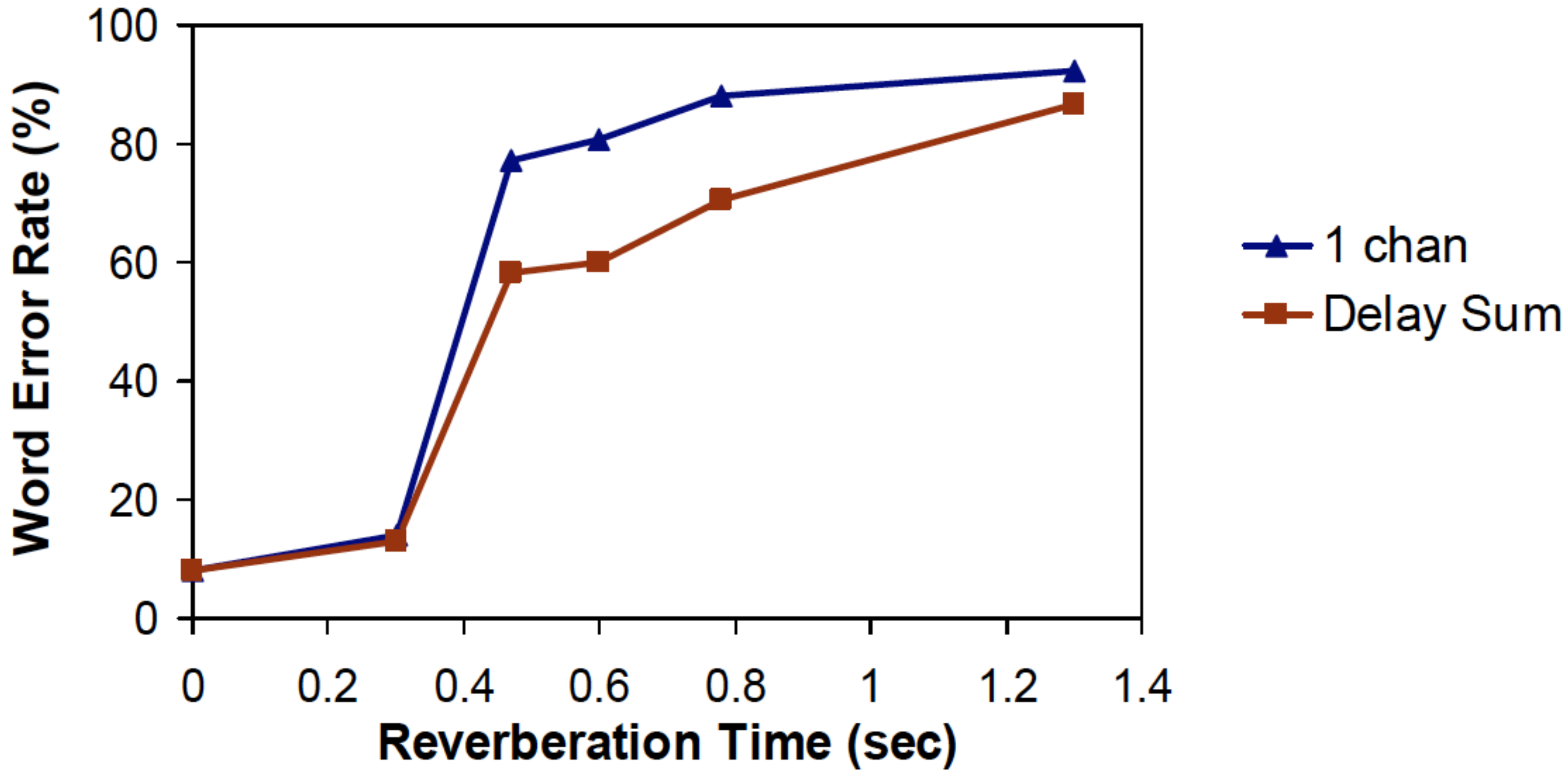
			s_1^s		s_2^s	
	Methods	**CTM**	T_{60} **(150ms)**	T_{60} **(250ms)**	T_{60} **(150ms)**	T_{60} **(250ms)**
MONC	**MGD**	**9.2**	12.98	23.96	11.99	23.58
	MM		14.21	26.01	13.78	25.56
	BSM		15.02	27.99	15.22	27.32
	LCMV		16.59	29.04	16.3	28.39
	MVDR		17.04	30.16	16.96	29.86
TIMIT	**MGD**	**6.73**	8.81	15.79	9.16	16.02
	MM		10.15	18.06	10.92	18.68
	BSM		10.98	19.16	12.1	20.12
	LCMV		12.18	20.44	15.25	21.67
	MVDR		14.08	22.47	17.41	24.37

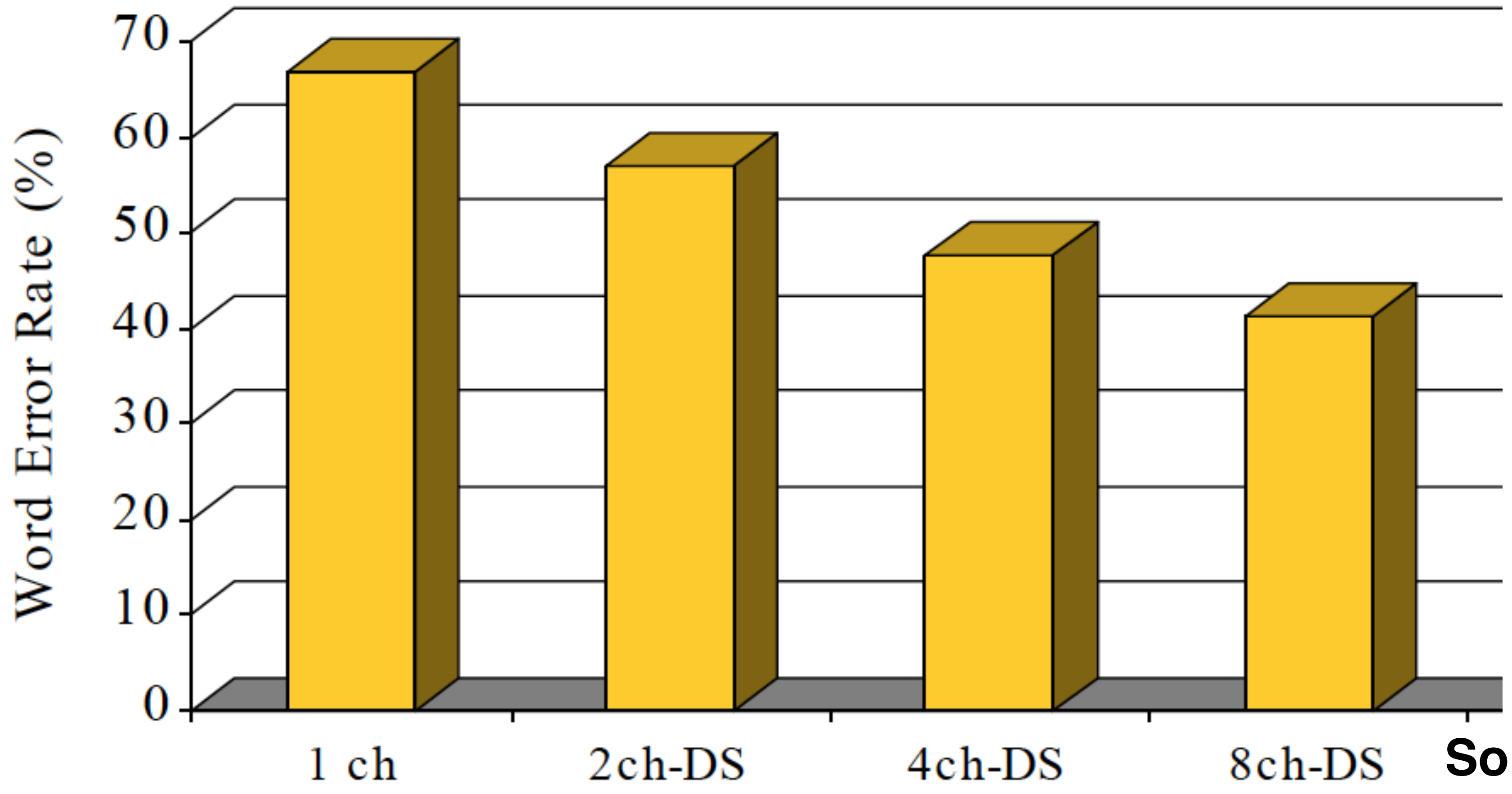
A General Framework for ASR using Microphone Arrays



Source : Bhiksha Raj CMU

WER versus Reverberation Time

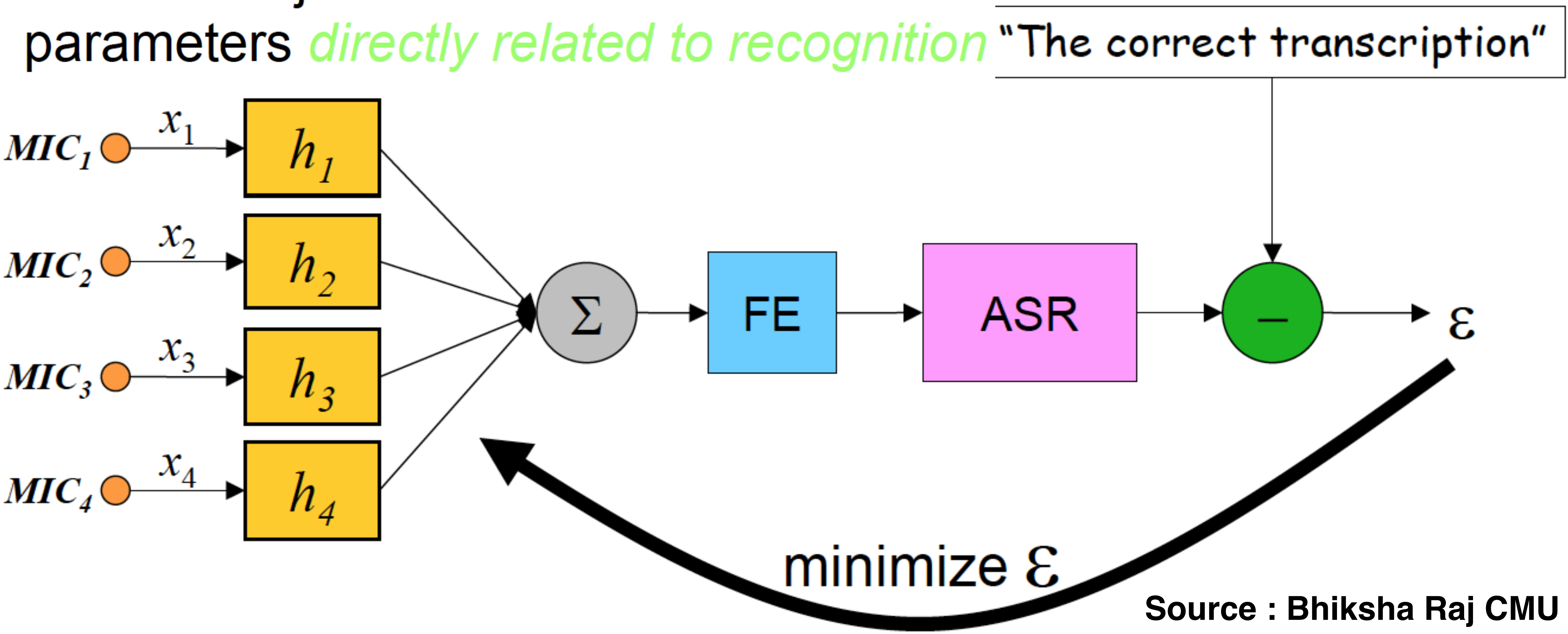




Source : Bhiksha Raj CMU

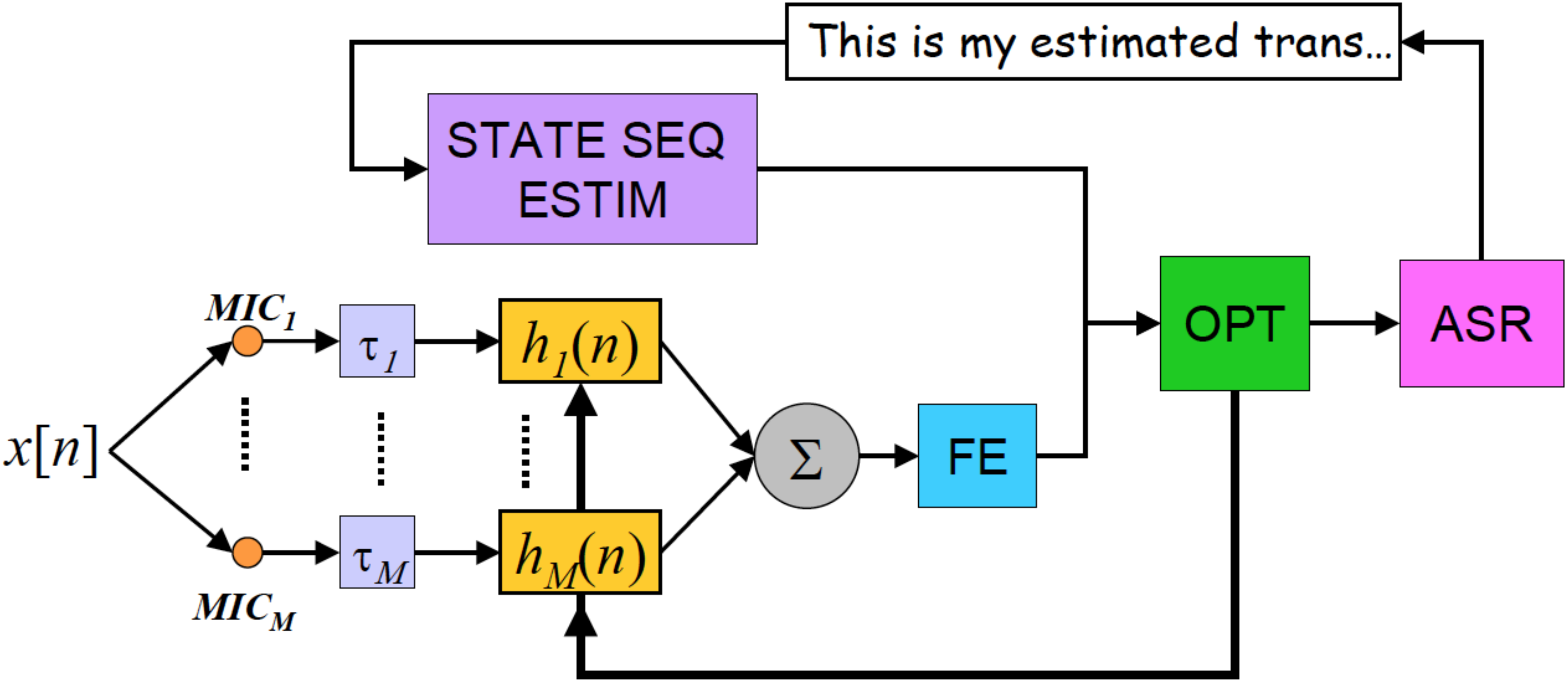
Microphone Array based Speech Recognition : LIMABEAM

- Likelihood Maximizing Beamforming [Seltzer 2004].
  - specifically targeted at improved speech recognition performance without regard to conventional signal-domain objective criteria.
- Want an objective function that uses parameters *directly related to recognition*



Source : Bhiksha Raj CMU

Unsupervised LIMABEAM



Source : Bhiksha Raj CMU

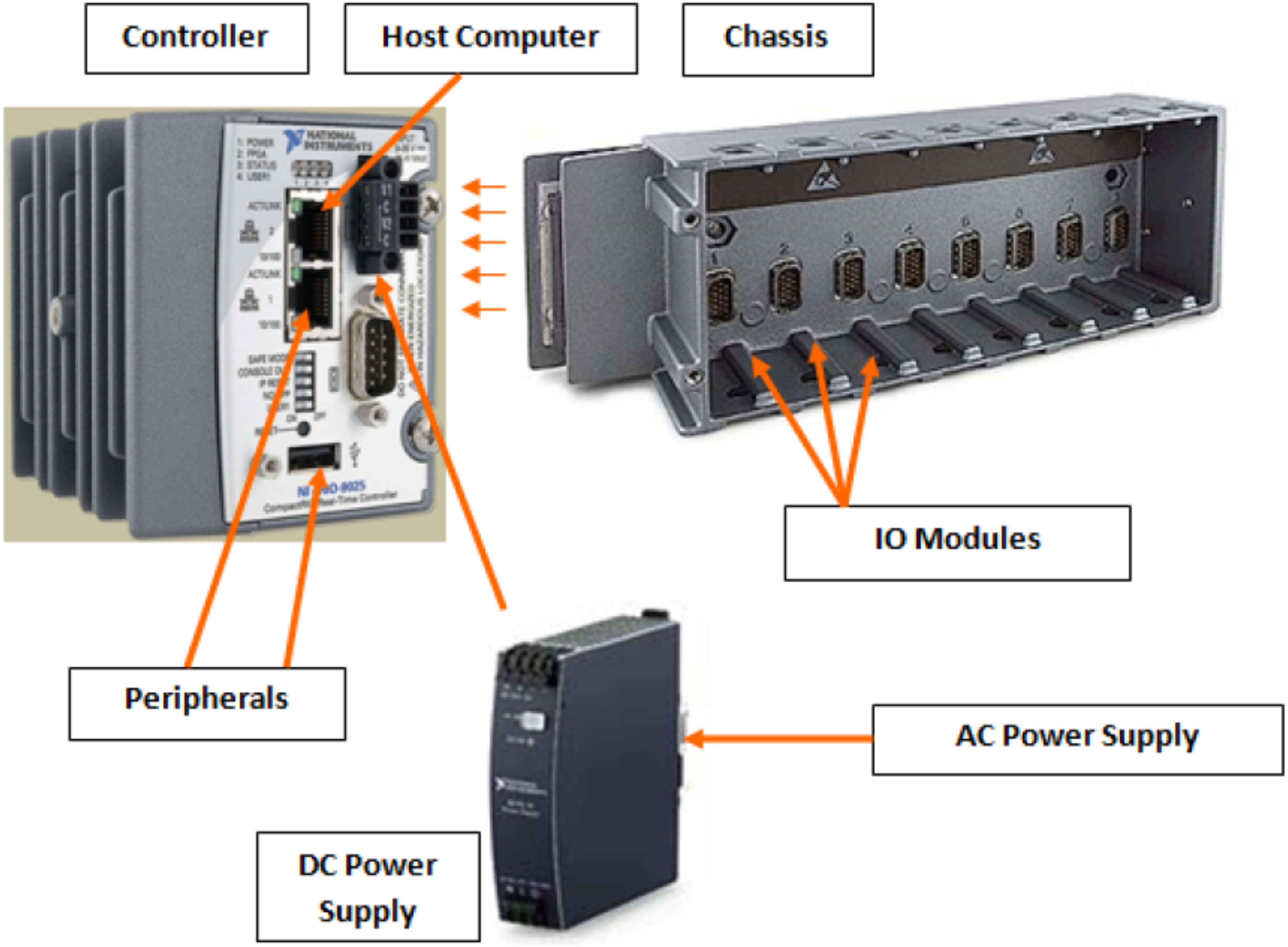
Databases for Microphone Array based Speech Recognition

- Small publicly available corpora (these are just a handful)
  - CMU Microphone Array
    - http://www.speech.cs.cmu.edu/databases/micarray
  - CMU PDA
    - http://www.speech.cs.cmu.edu/databases/pda
  - IDIAP Multi-channel Overlapping Numbers
    - http://www.cslu.ogi.edu/corpora/corpCurrent.html
  - ICSI Meeting Recorder Digits
    - http://www.icsi.berkeley.edu/Speech/mr/mrdigits.html

- Many sites have collected large corpora for research involving meeting transcription and annotation
  - ICSI Meeting Recorder project
    - http://www.icsi.berkeley.edu/Speech/mr
  - CHIL project
    - http://chil.server.de
  - AMI project
    - http://www.amiproject.org

Source : Bhiksha Raj CMU

Rapid Prototyping (on NI- cRIO) of an Intelligent Meeting Capture System using Microphone Array Processing



Controller

Host Computer

Chassis

IO Modules

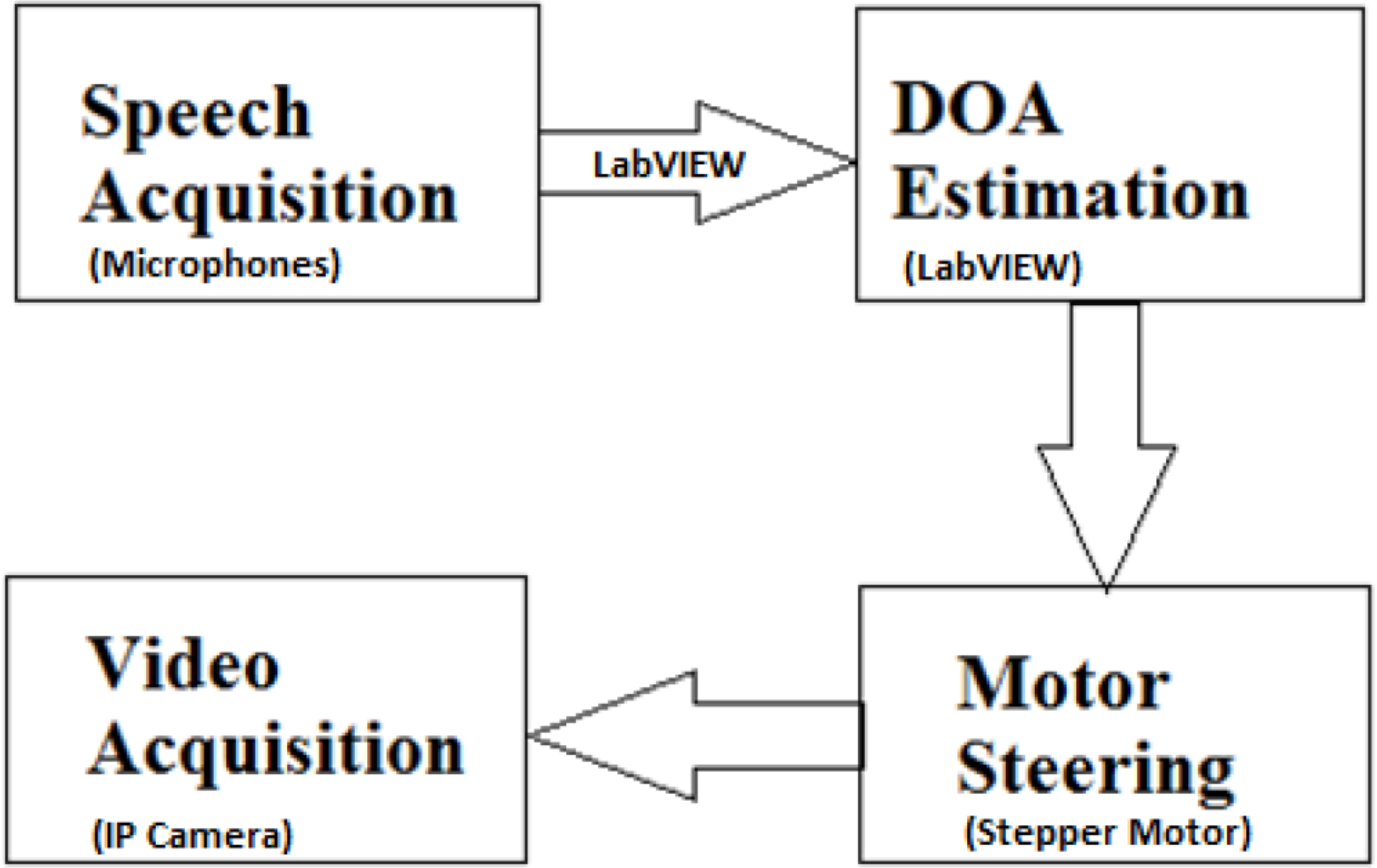
Peripherals

AC Power Supply

DC Power Supply

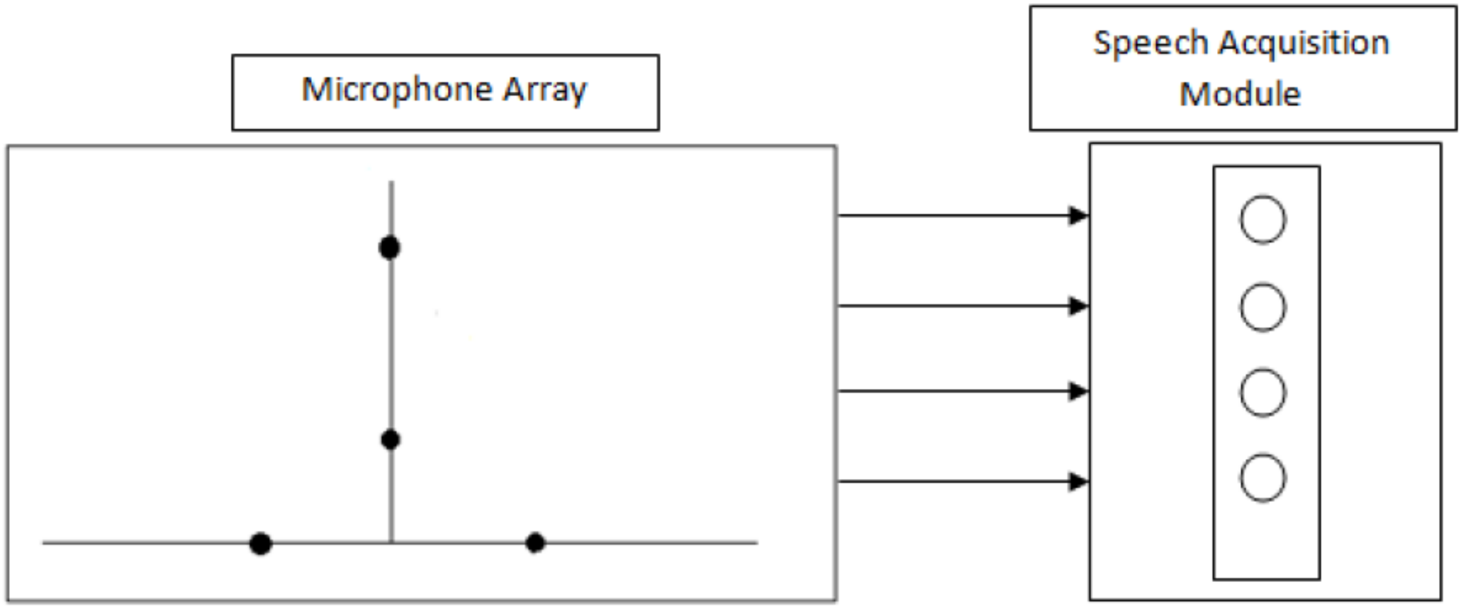
Modular Blocks Required in a Meeting Capture System

- Speech Acquisition Module
- DOA Estimation Module
- Motor Steering Module
- Video Acquisition Module



Speech Acquisition Module

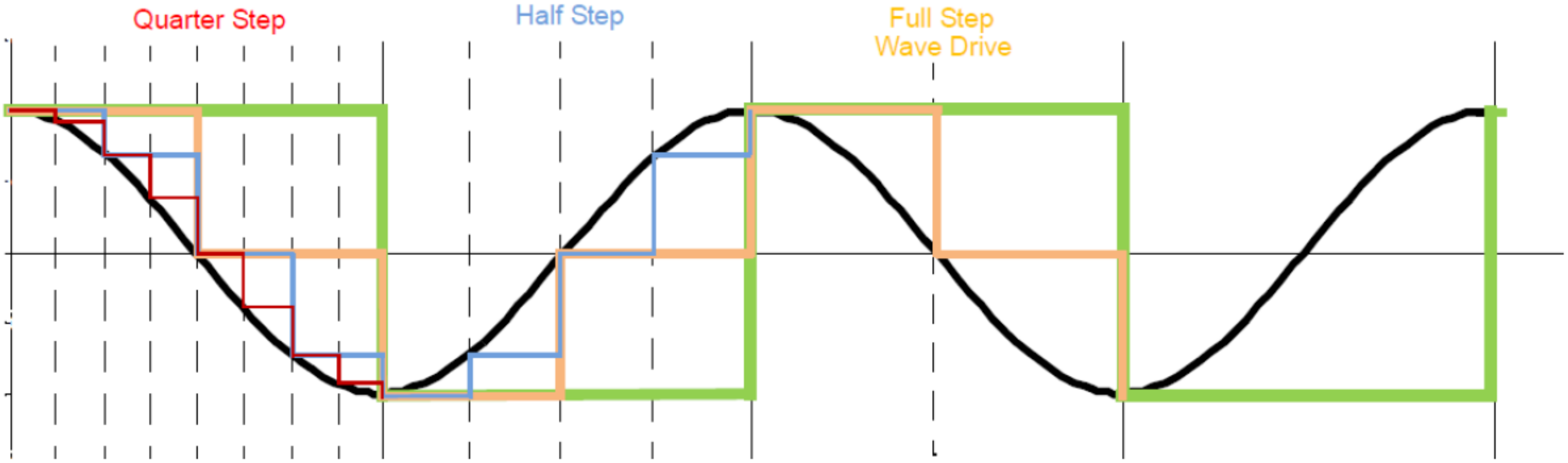
- Consists of four Channels
- Two channels are used to calculate the Azimuth, the other two for Elevation
- Simultaneous sampling at a rate of 51.2 KS/s
- Minimum excitation voltage of 18 VDC



Stepper Motor Microstepping

- ▶ **Ideal waveform to drive a stepper motor is a Sine wave**
- ▶ **Microstepping reduces the vibrations and improves accuracy**

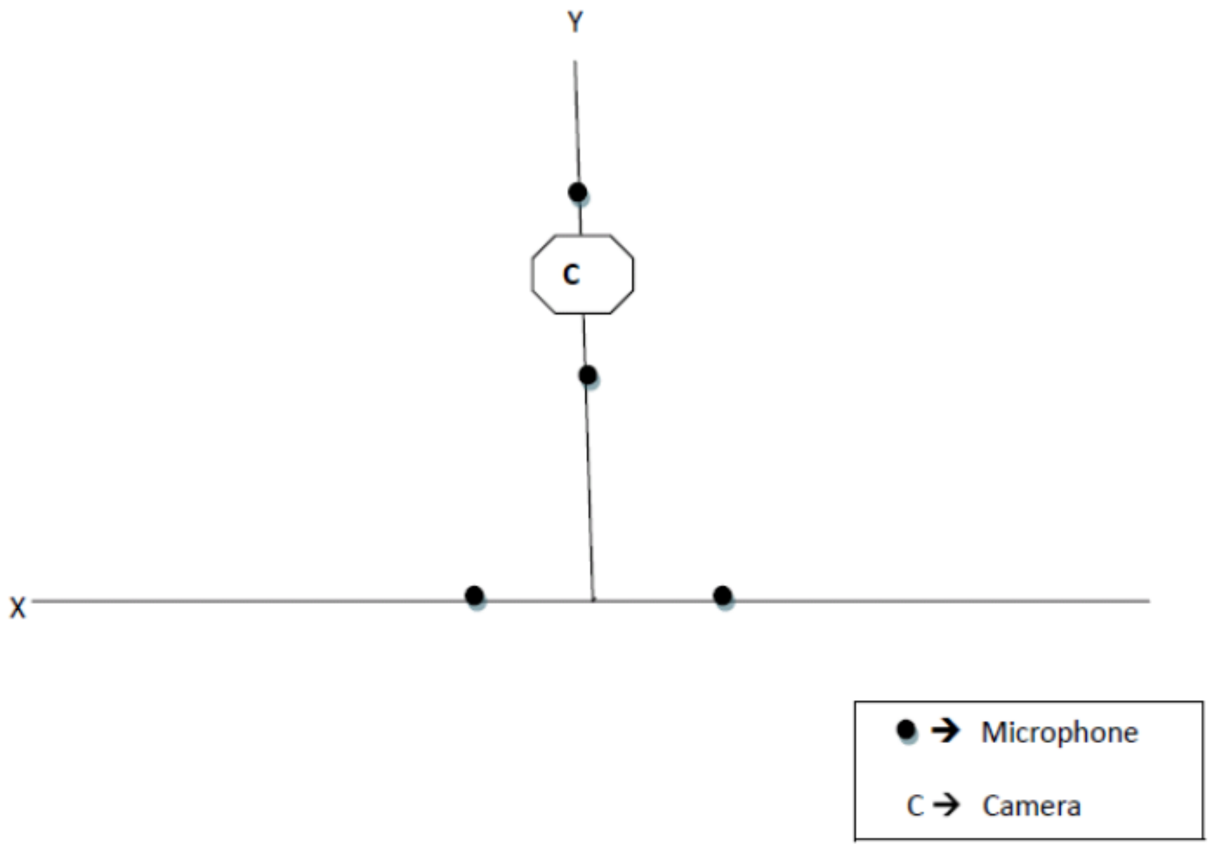
Microstepping Rate	Least count
1	1.8
2	0.9
4	0.45
8	0.225
16	0.1125
32	0.05625
64	0.028125
128	0.0140625
256	0.00703125



Microstepping in Stepper Motors

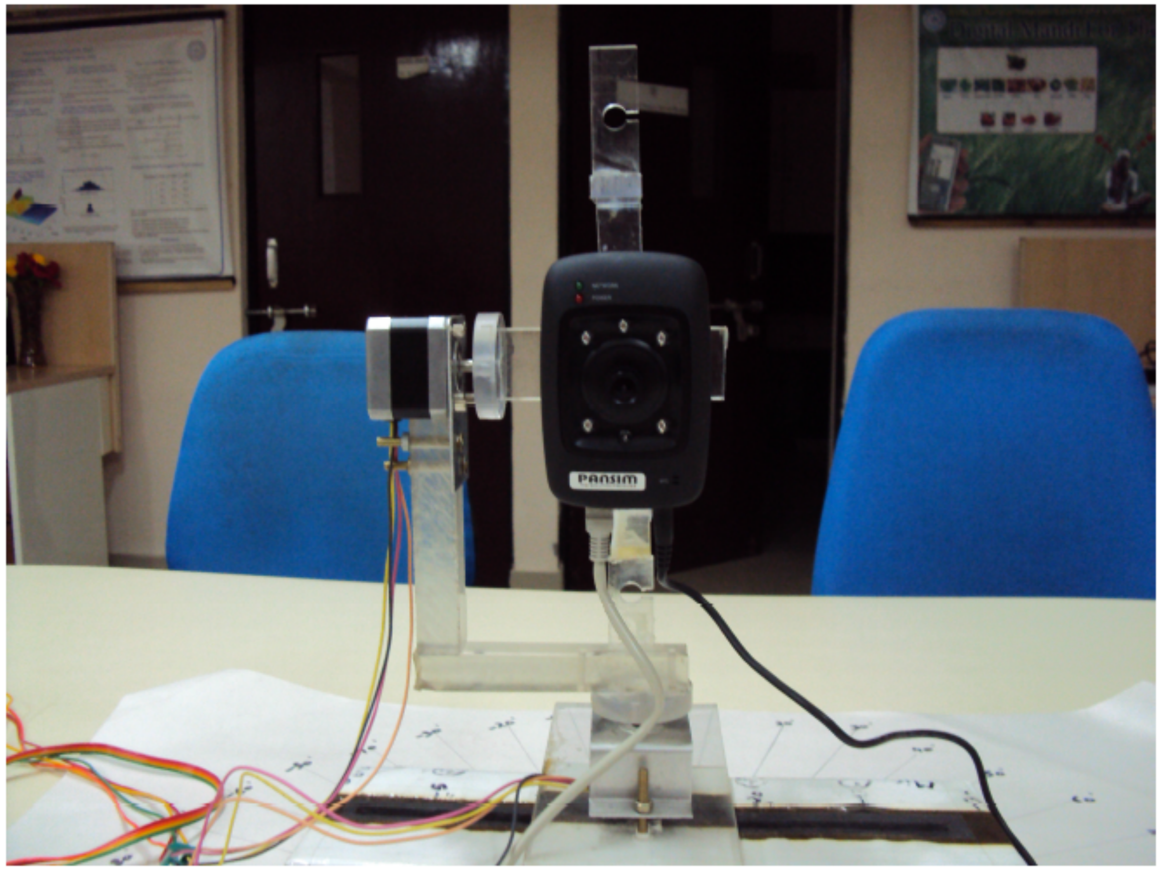
Video Acquisition Module

- Consists of an IP Camera connected to the two motors
- Camera is instructed to take a picture every five seconds
- The data is stored in a USB flash drive connected to the cRIO controller

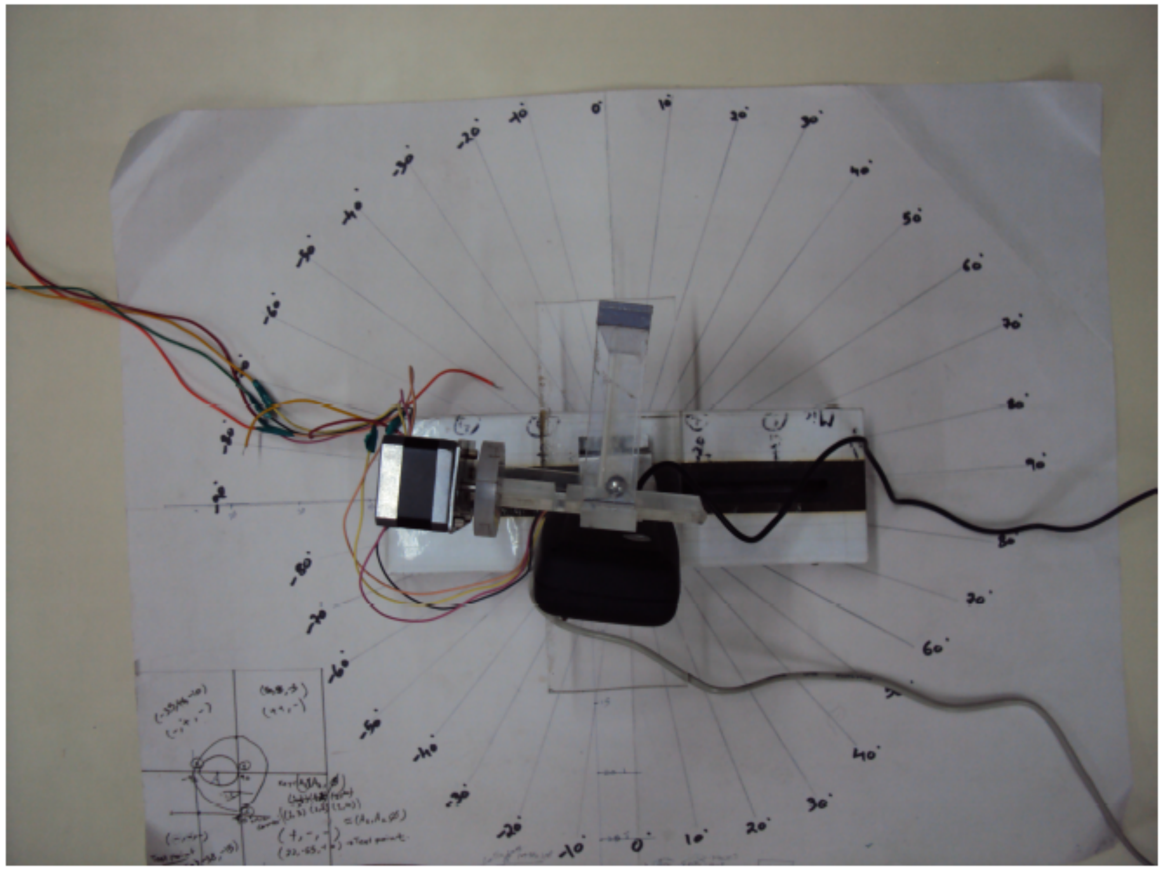


Position of Camera in the Steering Frame

Experimental Setup

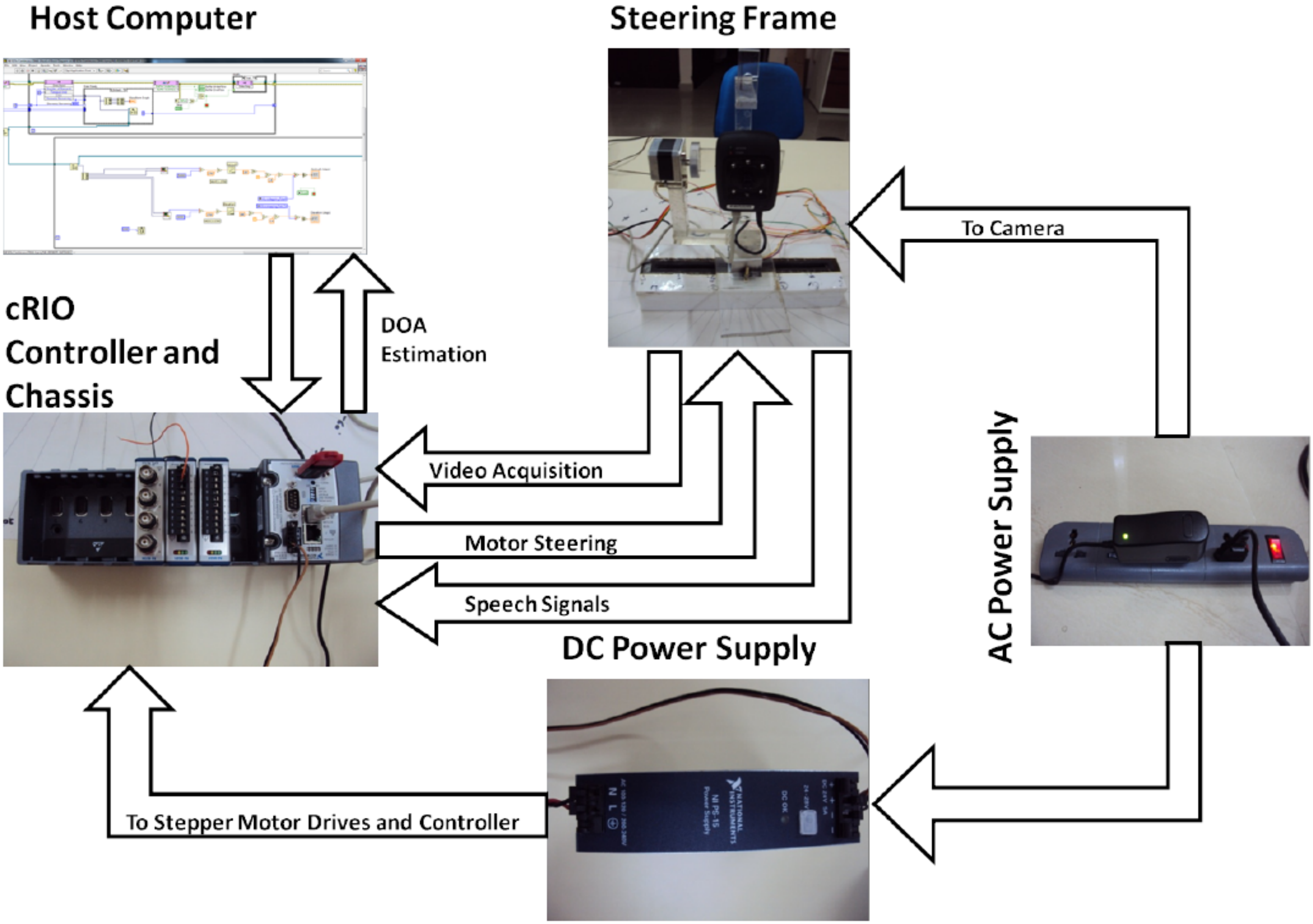


The Steering Frame used to rotate the Camera



Angular Chart for measuring actual DOA

Flow Diagram of the Intelligent Meeting Capture System



Host Computer

Steering Frame

cRIO Controller and Chassis

DOA Estimation

To Camera

Video Acquisition

Motor Steering

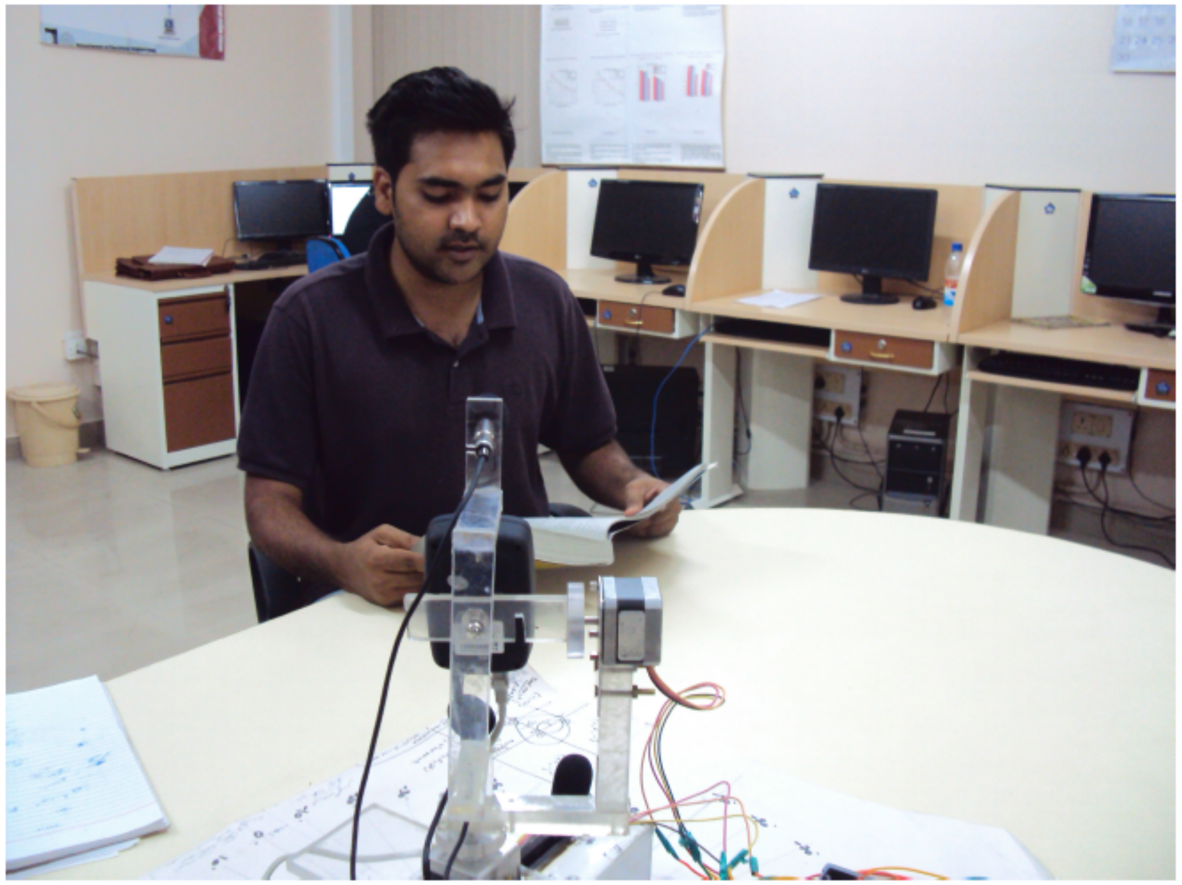
Speech Signals

AC Power Supply

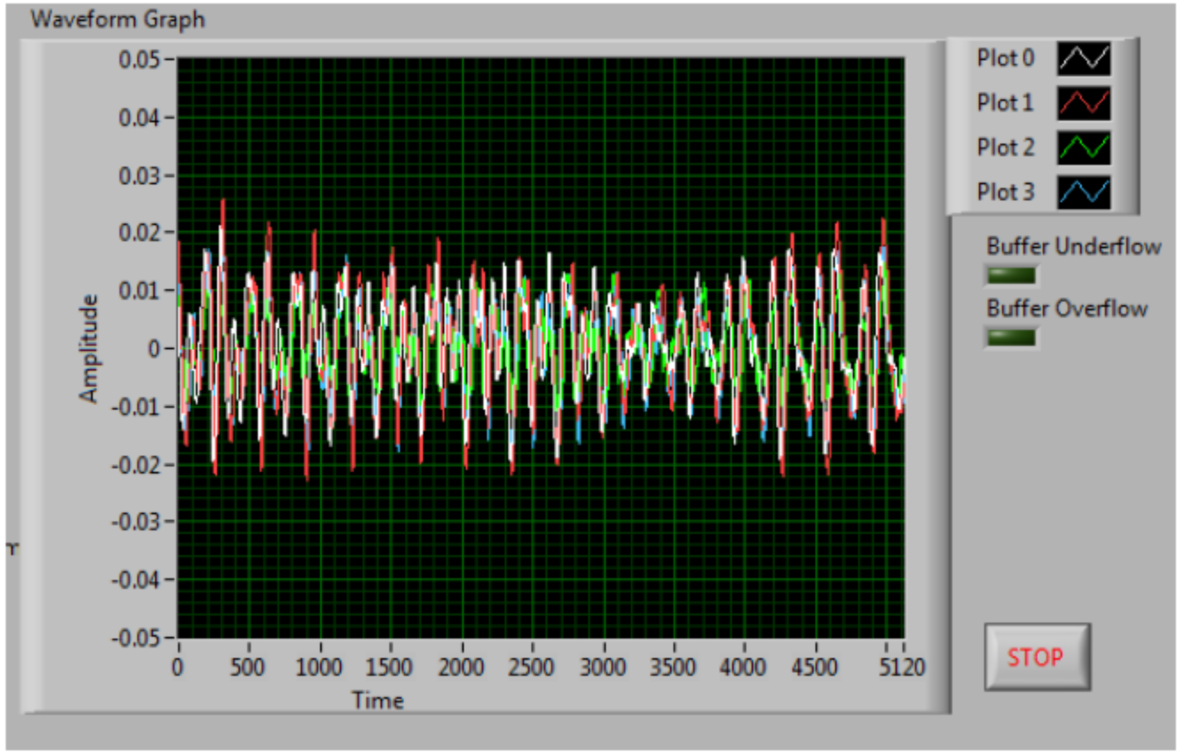
DC Power Supply

To Stepper Motor Drives and Controller

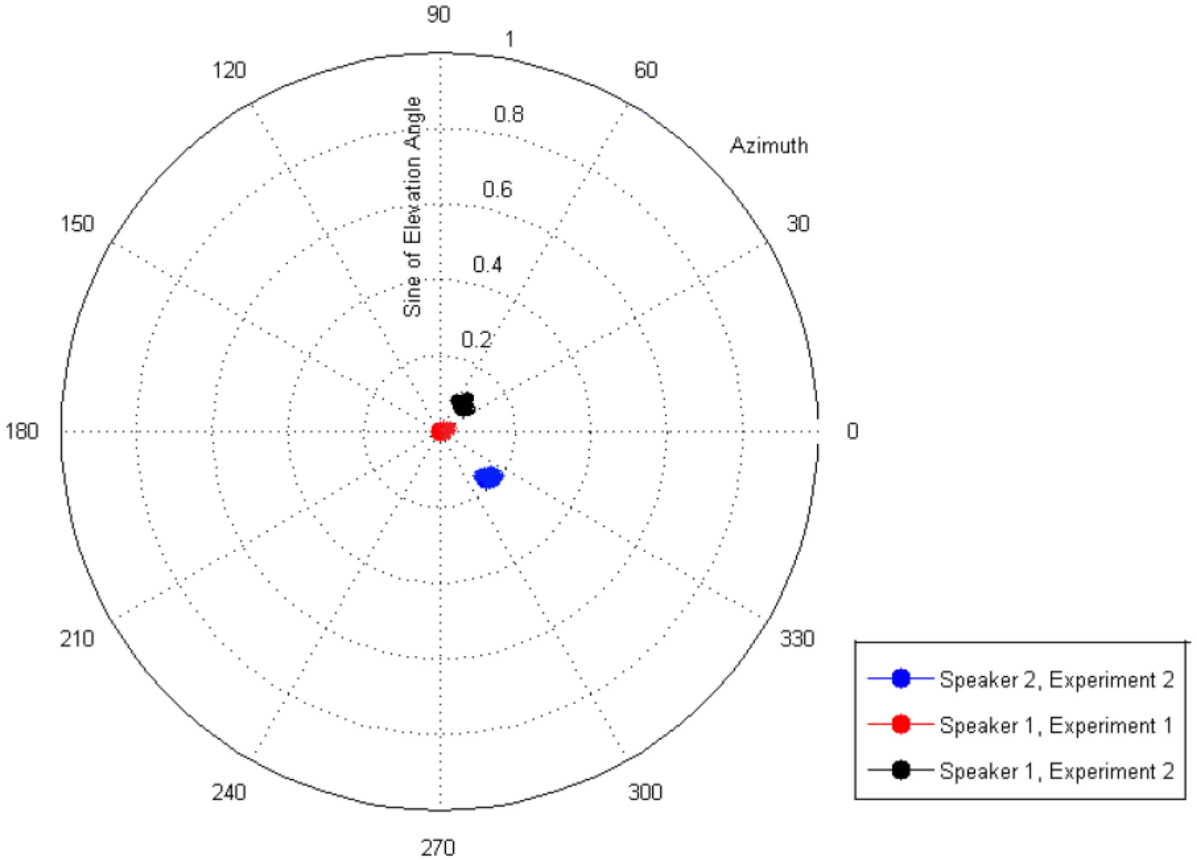
Speaker localization in Real Time and Speech Acquisition Waveform



Locating a Speaker in Real-time



Waveform graph showing Speech Signals being recorded in Real-time

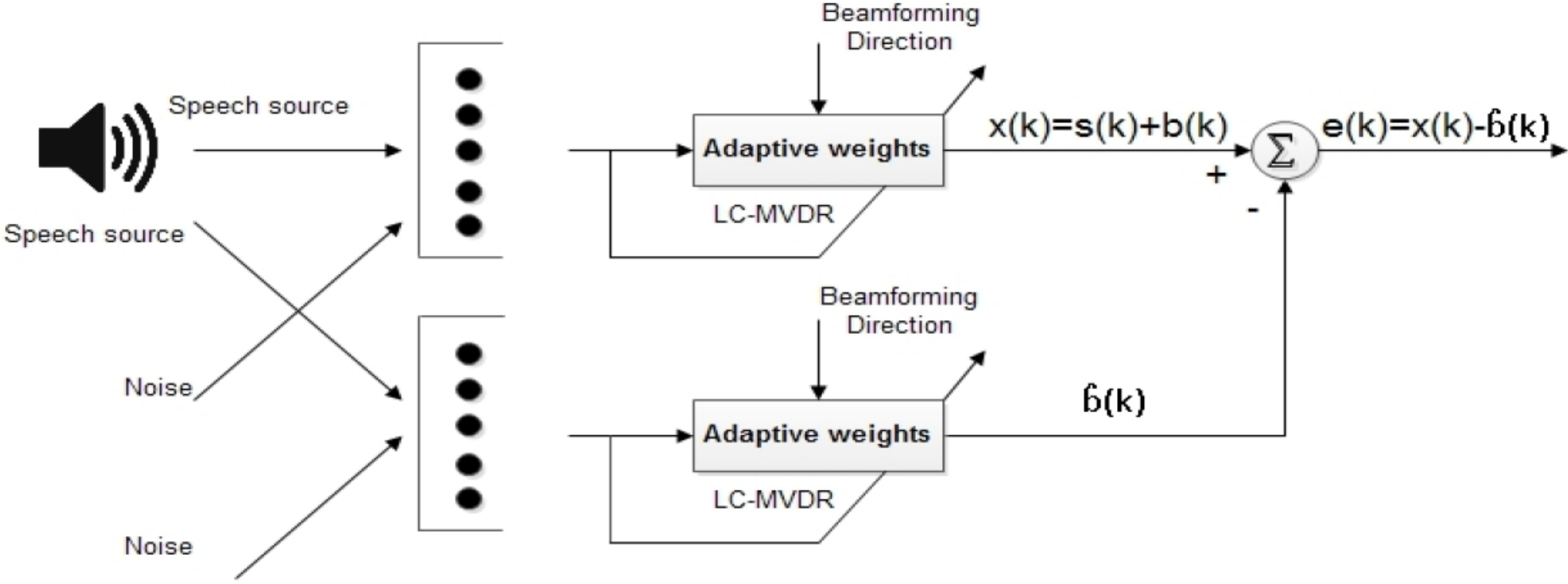


Demo Video : Intelligent Meeting Capture System

The Non Reference Anchor Array Framework for Speech Enhancement and Recognition

- Speech recognition over microphone arrays is challenging under multi source environments
- Generally a single primary microphone array (PMA) is used along with a LCMV beamformer
- A non reference anchor array framework uses an additional Auxiliary microphone array (AMA)
- Auxiliary microphone array is anchored such that it cancels noise in the same direction as the SOI
- Both PMA and AMA use an Adaptive LCMV beamformer

The Non Reference Anchor Array Framework for Speech Enhancement and Recognition



- Difference of the output at two arrays is given by,

$$e(k) = s(k) + b(k) - \hat{b}(k)$$

- When $\hat{b}$(k) is a good approximation for b(k),

$$\text{Desired signal, d(k)} = s(k) + b(k) - \hat{b}(k)$$

Determining Optimal Location of Non-Reference Anchor Array

- Auxiliary Microphone Array (AMA) placement plays crucial role in performance.
- It should not be arbitrary.
- AMA should be placed such that it receives the same noise signal as the Primary Microphone Array (PMA).
- AMA should be placed such that it receives least of the signal of interest.
- To obtain a good location for AMA correlation evaluation can be used.
- Let only noise source be present.
- Signal at PMA is denoted by O1 and signal at AMA is denoted by O2.
- For delay ‘n’ between the signals, correlation ‘C’ is evaluated as,

$$C = \sum_{m=0}^{N1} O1^{*}[m] O2[n+m]$$

  where N1 is the number of samples.
- Location at which C is maximum (DOA of Noise at PMA and AMA is same) is the optimal location.

Determining Optimal Location of Non-Reference Anchor Array

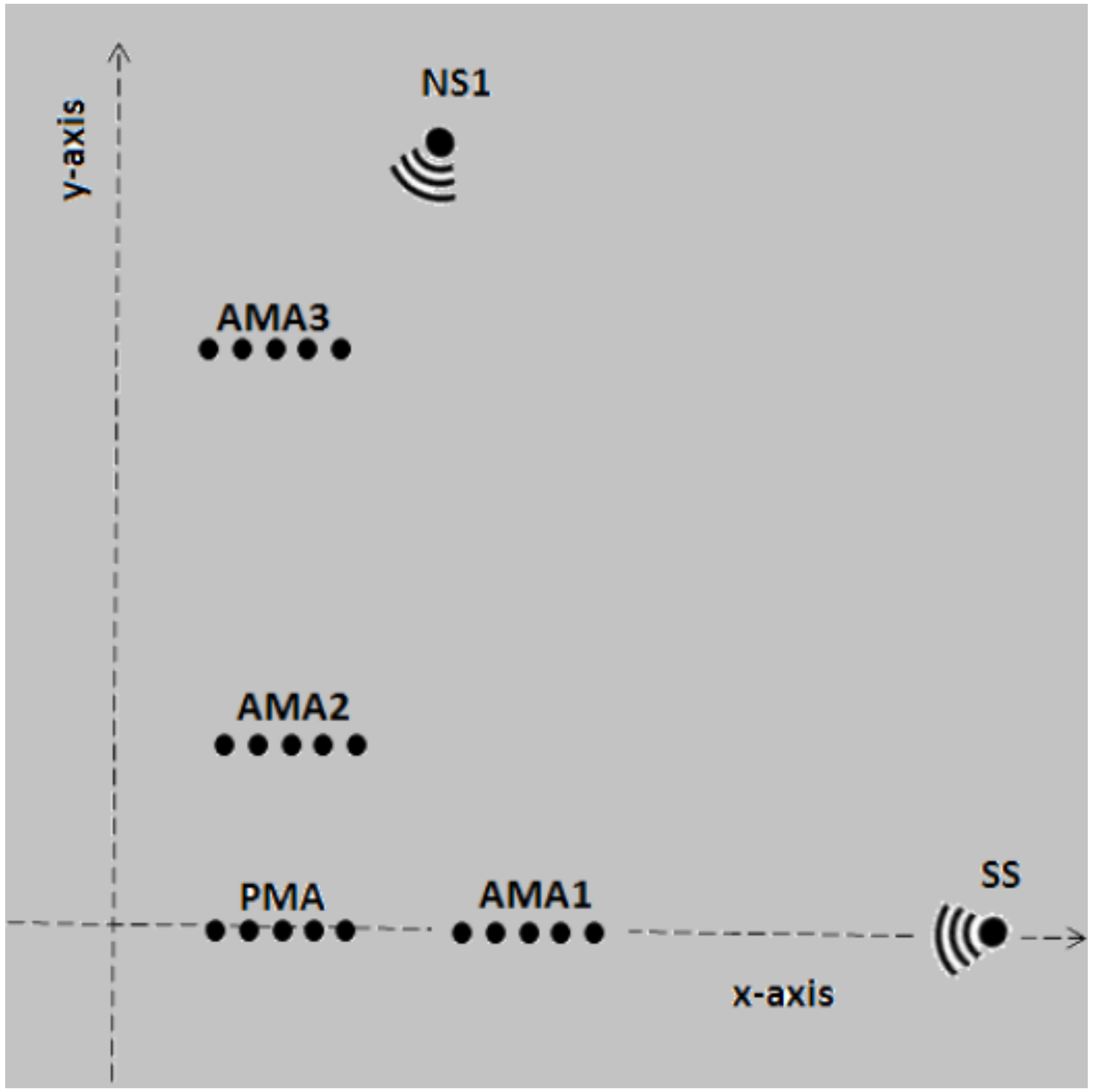
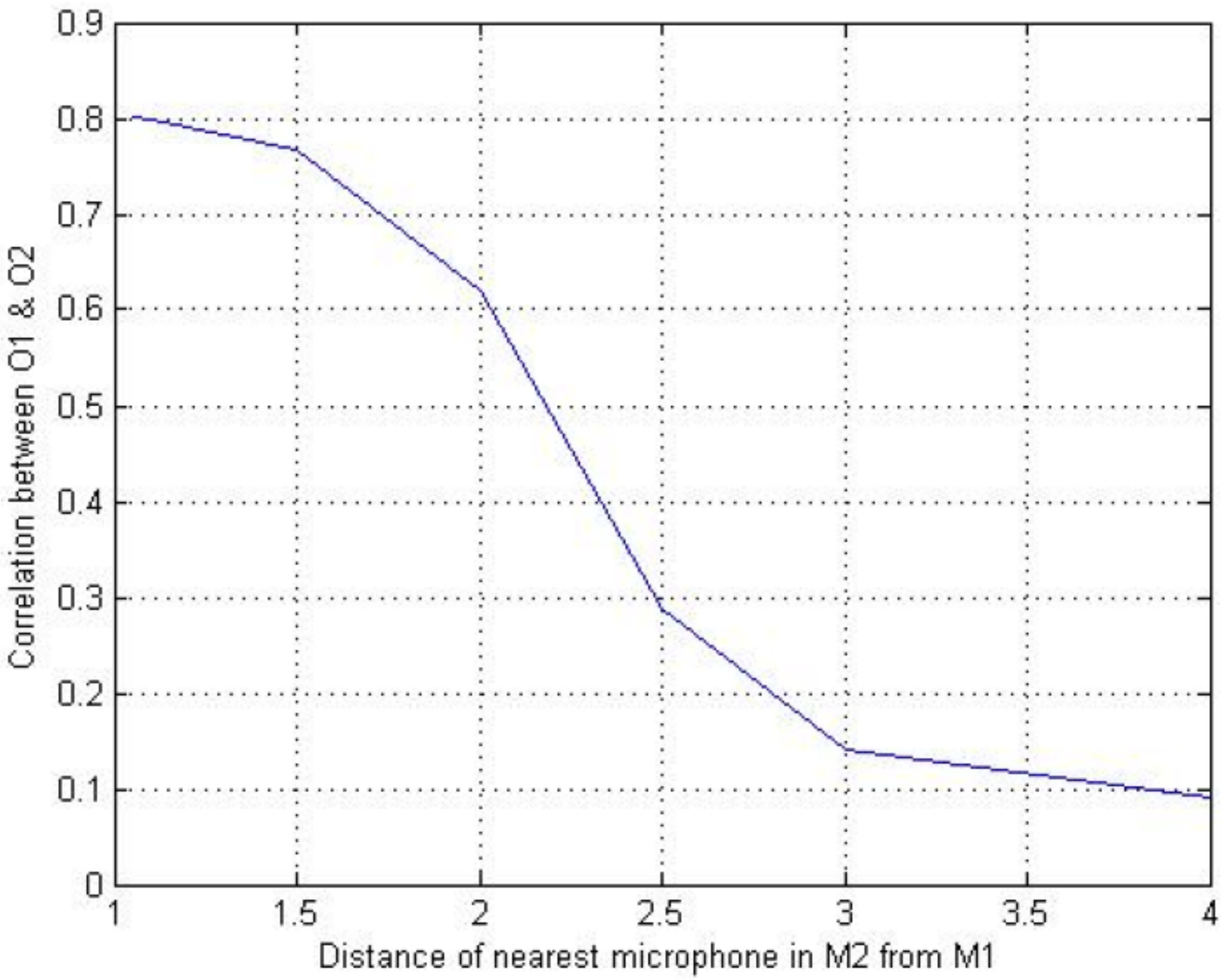


Illustration for the optimal placement of AMA. While AMA1 and AMA3 are a bad choice for the placement of AMA, AMA2 represents a better location for the placement of AMA. NS1 is the noise source and SS is the signal source



Normalized Correlation between output O1 (output of array recording signal) & O2 (output of array to model noise). Position of M1 (PMA) is kept fixed and M2 (AMA) position is varied. Leftmost mic in PMA was placed at (1,0,0). Noise source was placed at (6.06,3.5,0). AMA position is varied along the x-axis

Determining Optimal Location of Non-Reference Anchor Array

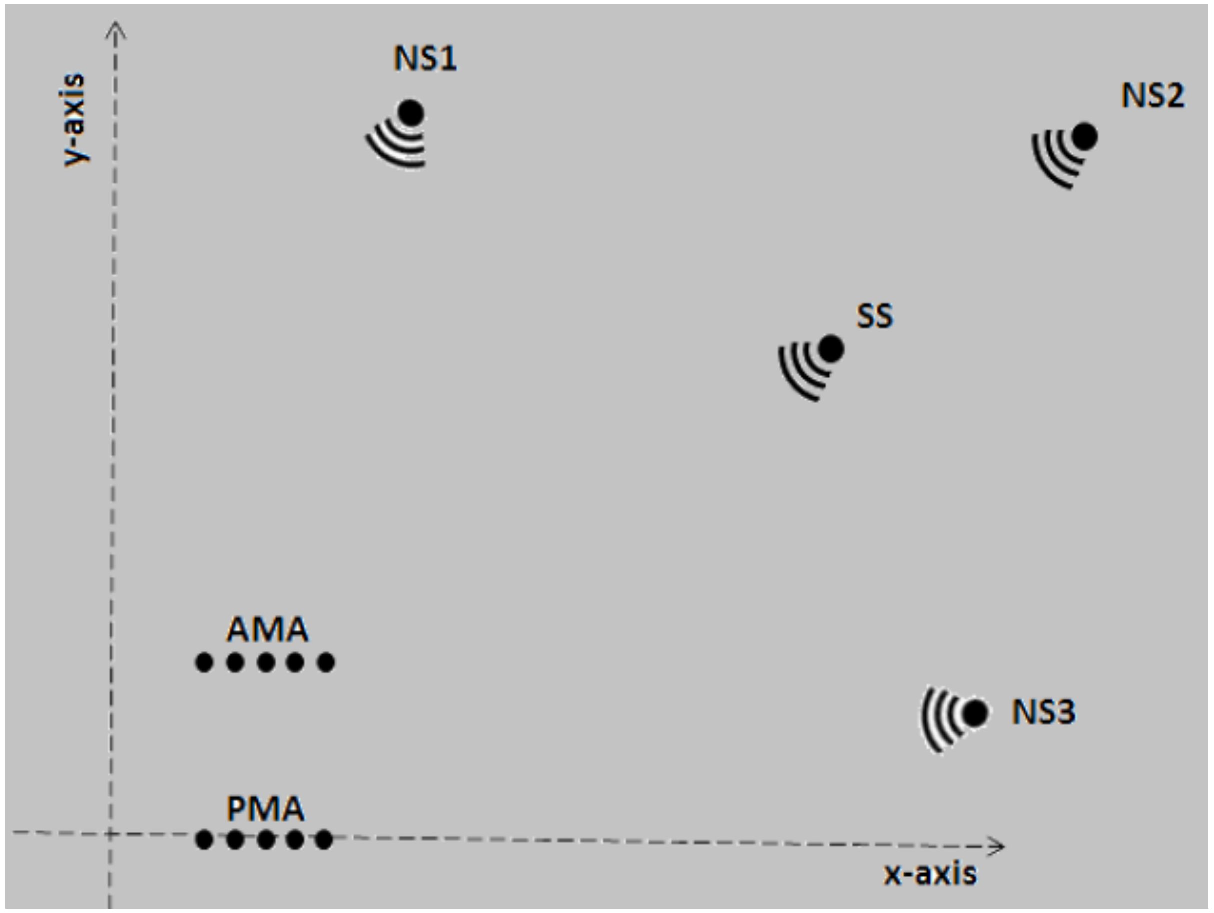
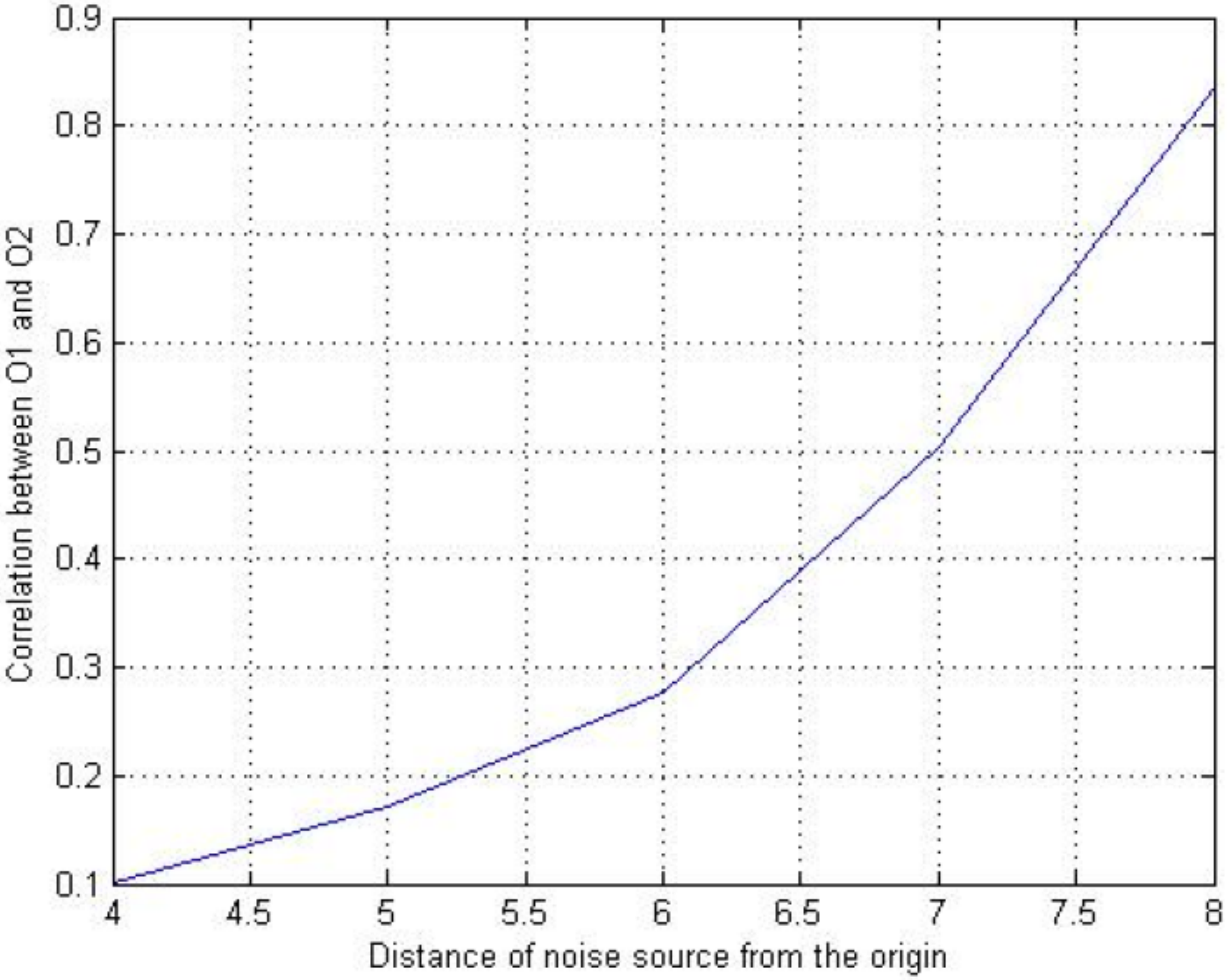


Figure showing the projection of experimental setup on x-y plane (not to scale). NS1, NS2, NS3 represent three noise sources and SS represent SOI source. Note that there is a misalignment in thepositions of PMA (Primary Microphone Array) and AMA (Auxiliary Microphone Array)



Normalized Correlation between output O1 (output of array recording signal) & O2 (output of array to model noise). Location of, array recording signal, is kept fixed. Leftmost mic in PMA was placed at (1,0,0) and leftmost mic in AMA was placed at (1,0.01,0.12). Noise source is moved farther from PMA and AMA

Optimal Location of NRA for Cell Phone

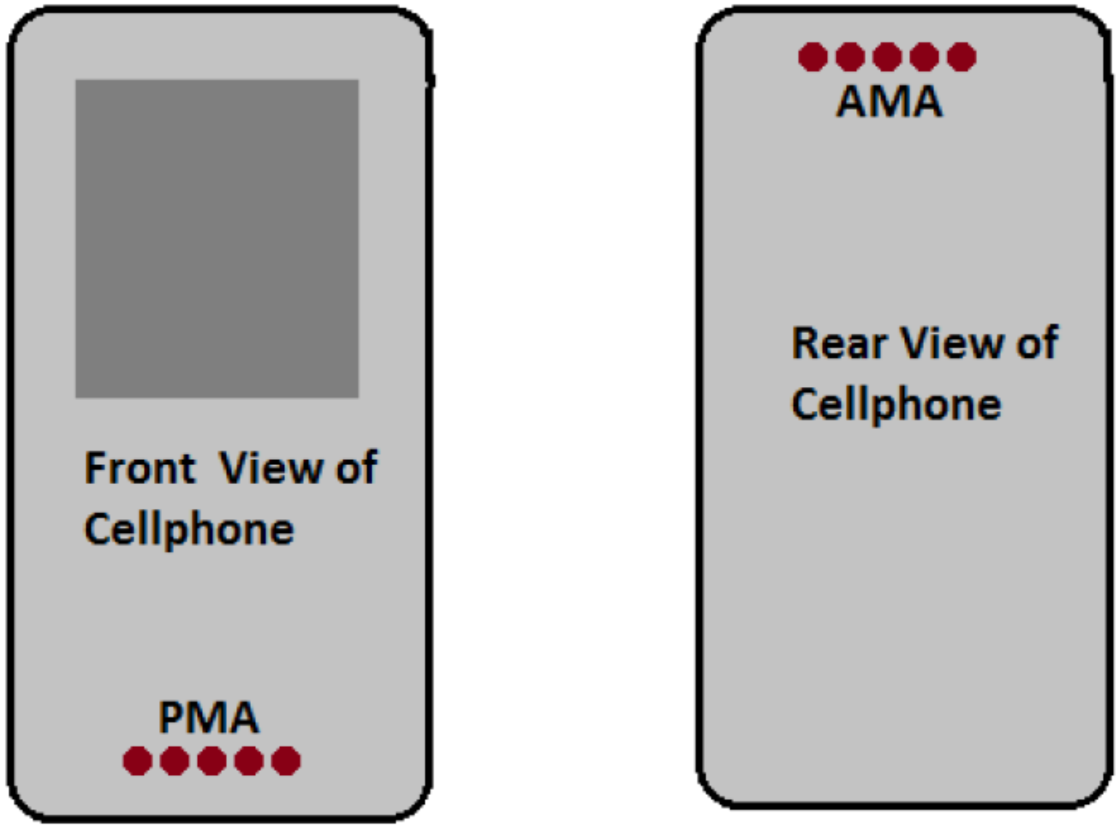


Illustration of reference and non-reference anchor array using front & rear view. Both PMA (Primary Microphone Array) & AMA (Auxiliary Microphone Array) are shown.

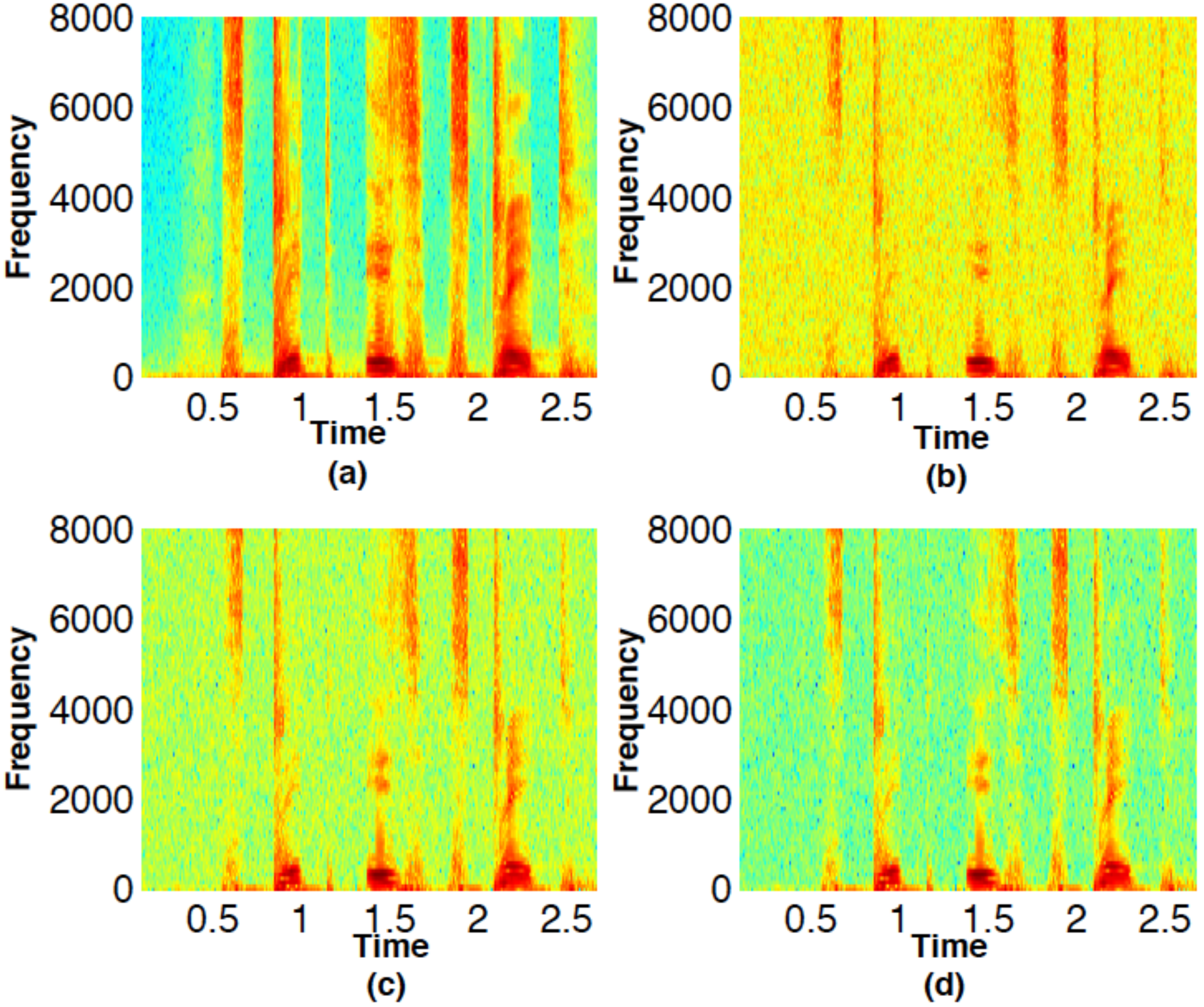
This method is able to minimize the effects of correlated interferences coming from the direction of the signal of interest.

Performance Evaluation : Speech Enhancement and Recognition

- The performance of the proposed adaptive LCMV beamforming method in a non-reference anchor (NRA) array framework is evaluated by conducting large vocabulary speech recognition (speaker dependent) experiments on the spatialized version of the TIMIT database.
- TIMIT database has been used for experiments.
- The TIMIT corpus of read speech is designed to provide speech data for acoustic-phonetic studies and for the development and evaluation of automatic speech recognition systems.
- MONC database has also been used for the experiments.
- Digit recognition experiments are also conducted on the MONC database
- The MONC is derived from the Numbers Corpus release 1.0, prepared by the Center for Spoken Language Understanding at the Oregon Graduate Institute.

Performance Evaluation : Speech Enhancement

- Speech signal is taken for signal of interest and interfering signal.
- DOA of SOI at primary array is $40°$.
- Interfering signal direction is taken as $40°$,$10°$ and $80°$.
- SNR is 25 dB.
- 60 % of noise come from same direction of SOI and 20 % of noise comes each from $10°$ and $80°$.



Spectrograms for (a) clean speech signal, (b) Noisy speech signal, (c) Standard MV beam former with a NRA array & (d) Adaptive LC-MV beam former with a NRA array

Performance Evaluation : Speech Enhancement and Recognition

- Signal source was located at radial distance of 0.6m from central mic in PMA and subtends on it an angle of 40°.
- Noise source 1,2 and 3 were located at distance of 7m from central mic in PMA and subtends on it an angle of 80°, 40° and 10° respectively.
- Performance is observed in terms % Word Error Rate, evaluated as,

$$\textbf{WR} = 100 - \frac{(N - S - D - I)}{N} 100 \ .$$

where N is the total number of words, S is the total number of substitutions, D the total number of deletions and I is the total number of insertions in the recognized word list.

Performance Evaluation : Speech Enhancement and Recognition

- **Alignment Factor, A_f is defined as,**

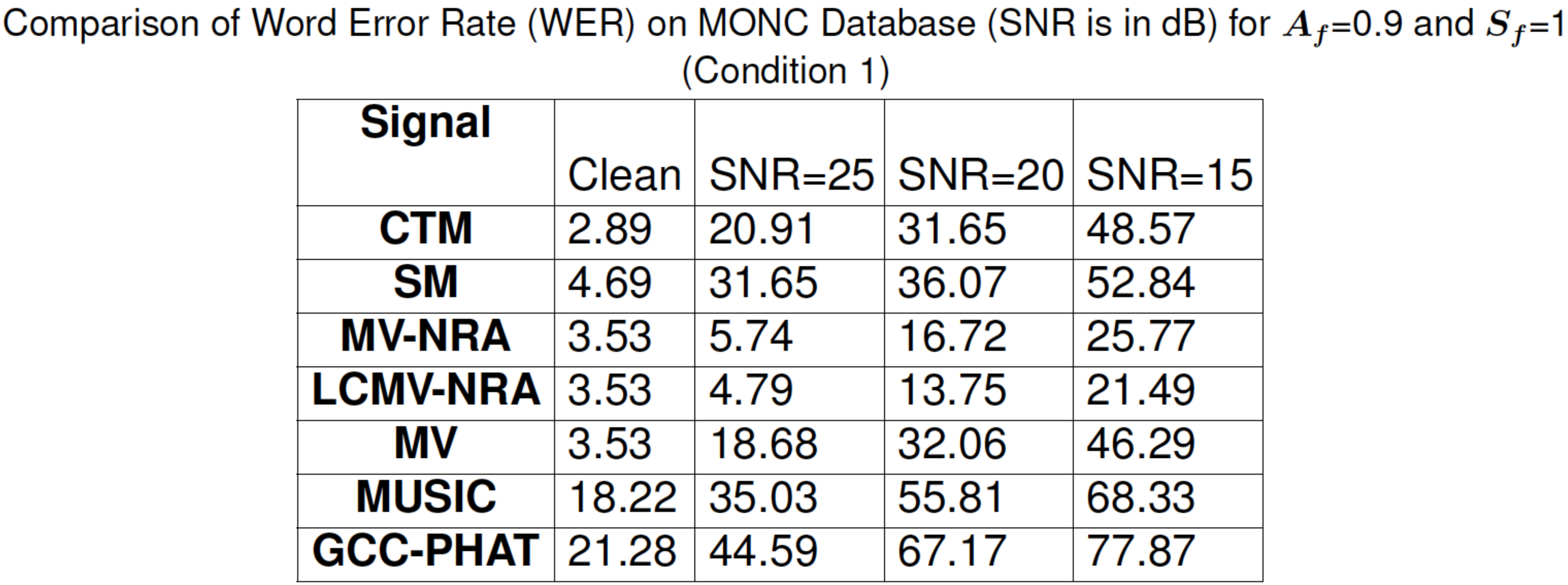
$$A_f = \frac{\text{Noise energy from DOA of SOI}}{\text{Total noise energy}}$$

- **Split factor, S_f is defined as,**

$$S_f = \text{Number of noise sources with DOA different from DOA of SOI.}$$

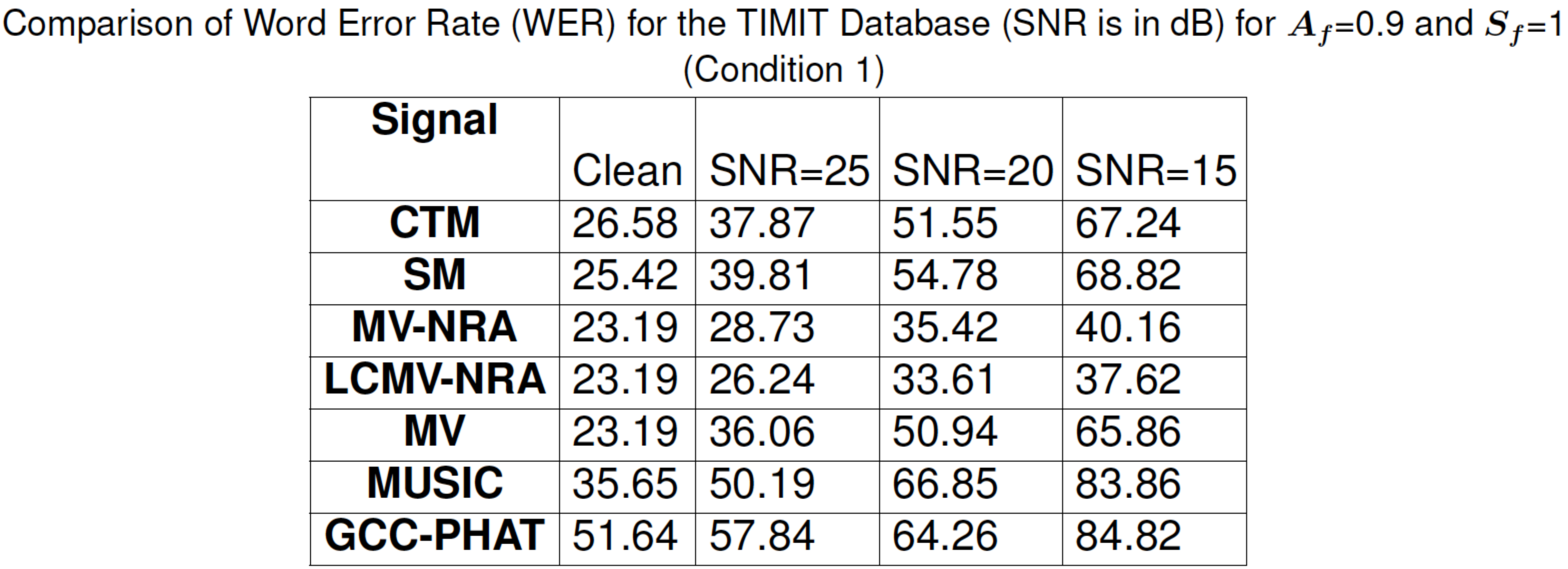
- **Condition 1 : DOA of SOI is taken as 40 degrees. 90 % of the noise energy is incident from direction of SOI whereas 10 % of the noise energy is incident from DOA of $80°$. Therefore, Condition 1 has A_f = 0.9 and S_f = 1.**

Performance Evaluation : Speech Enhancement and Recognition



Comparison of Word Error Rate (WER) on MONC Database (SNR is in dB) for A_f=0.9 and S_f=1 (Condition 1)

Signal	Clean	SNR=25	SNR=20	SNR=15
CTM	2.89	20.91	31.65	48.57
SM	4.69	31.65	36.07	52.84
MV-NRA	3.53	5.74	16.72	25.77
LCMV-NRA	3.53	4.79	13.75	21.49
MV	3.53	18.68	32.06	46.29
MUSIC	18.22	35.03	55.81	68.33
GCC-PHAT	21.28	44.59	67.17	77.87



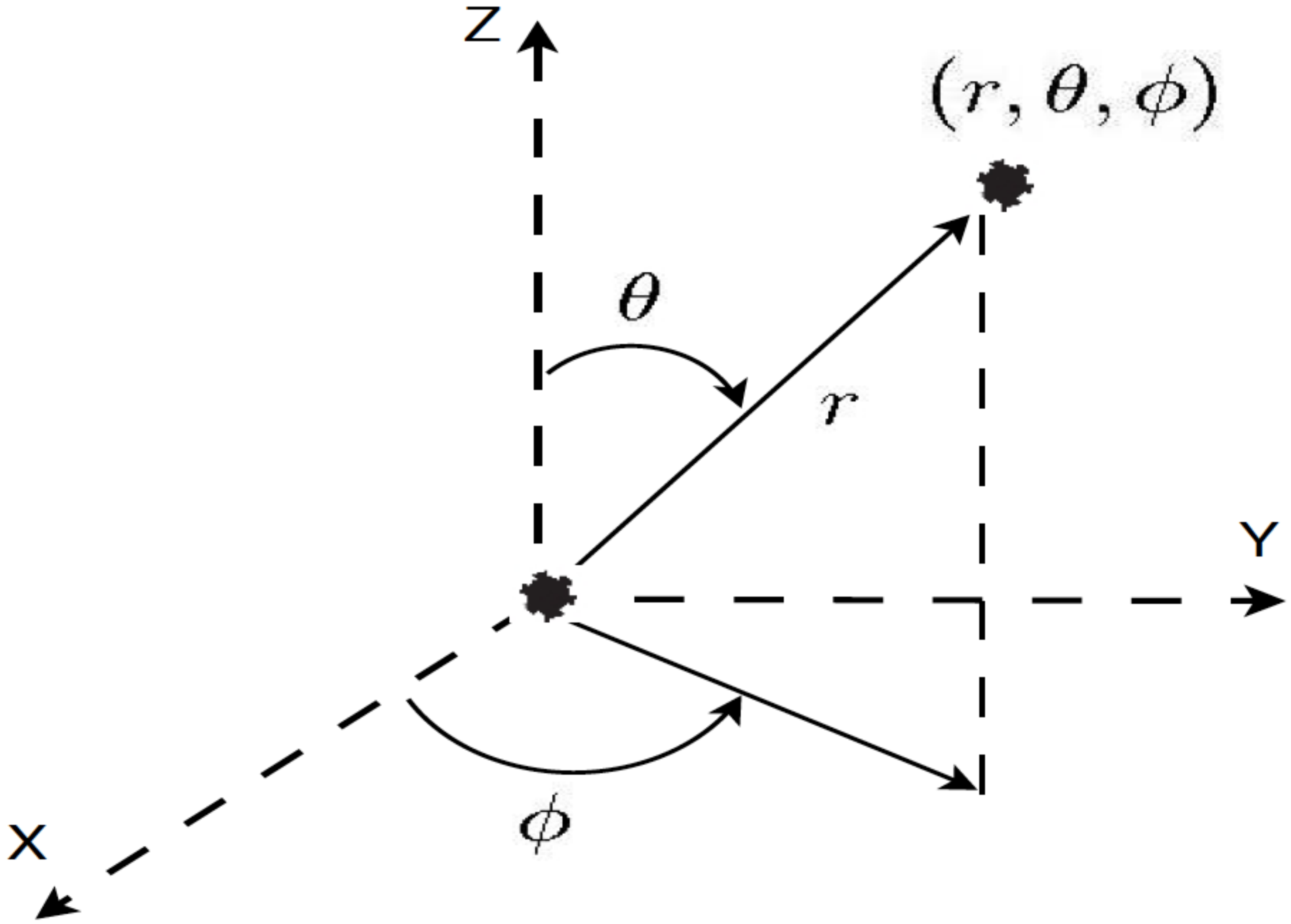
Comparison of Word Error Rate (WER) for the TIMIT Database (SNR is in dB) for A_f=0.9 and S_f=1 (Condition 1)

Signal	Clean	SNR=25	SNR=20	SNR=15
CTM	26.58	37.87	51.55	67.24
SM	25.42	39.81	54.78	68.82
MV-NRA	23.19	28.73	35.42	40.16
LCMV-NRA	23.19	26.24	33.61	37.62
MV	23.19	36.06	50.94	65.86
MUSIC	35.65	50.19	66.85	83.86
GCC-PHAT	51.64	57.84	64.26	84.82

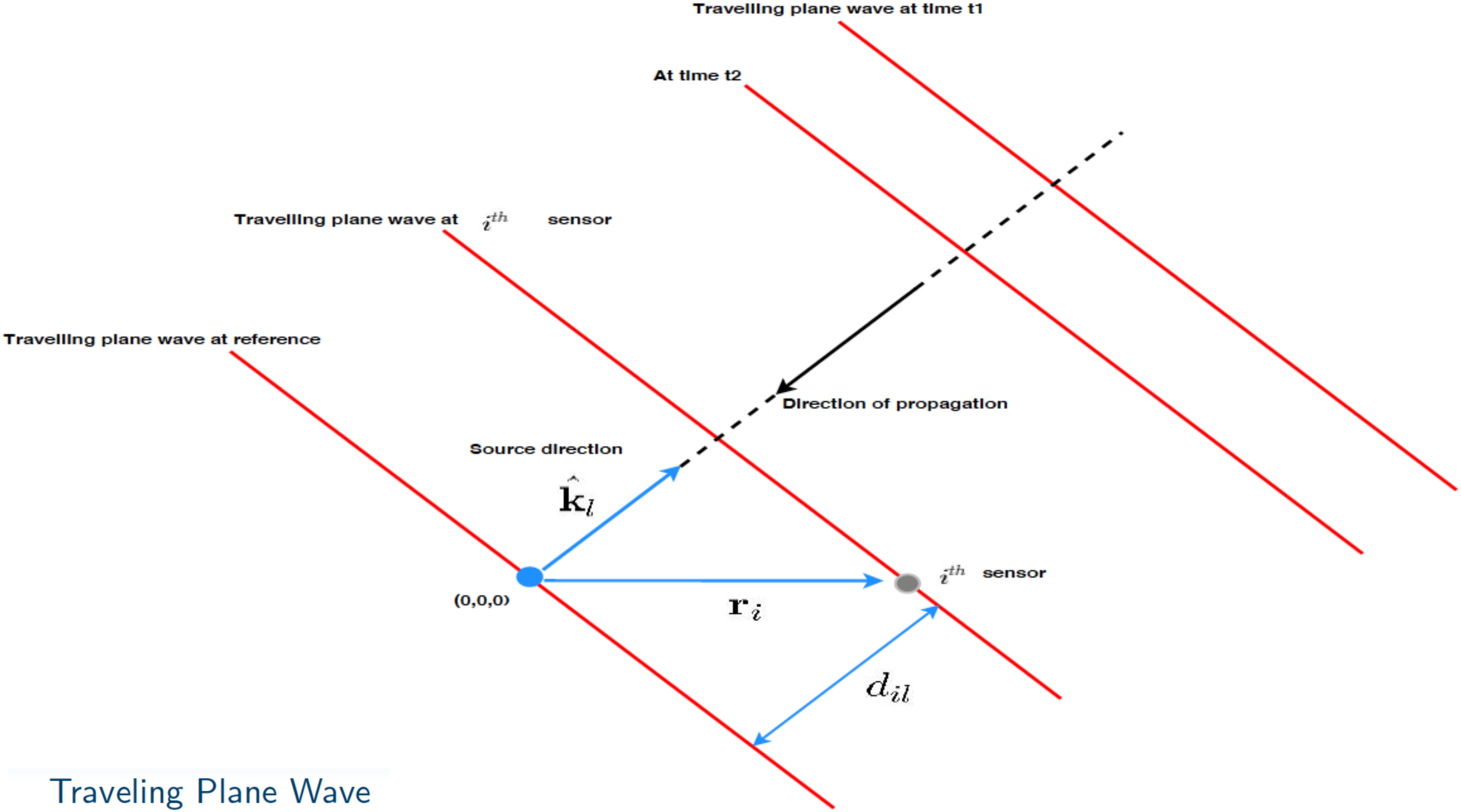
Spherical Microphone Array Processing

Spherical Coordinate System

- Location of a source is given by $\mathbf{r}_l = (r_l, \Psi_l)$, with $\Psi_l = (\theta_l, \phi_l)$.
- Location of a receiver is denoted as $\mathbf{r}_i = (r_i, \mathbf{\Phi}_i)$, where $\mathbf{\Phi}_i = (\theta_i, \phi_i)$.
- The range (r), elevation (θ) and azimuth (ϕ) takes values as $r \in (0, \infty)$, $\theta \in [0, \pi]$, $\phi \in [0, 2\pi)$



Principles of Acoustic Wave Propagation



Traveling Plane Wave

$$\mathbf{r}_i = (r_i \sin\theta_i \cos\phi_i, r_i \sin\theta_i \sin\phi_i, r_i \cos\theta_i)^T$$

$$\mathbf{k}_l = -(k \sin\theta_l \cos\phi_l, k \sin\theta_l \sin\phi_l, k \cos\theta_l)^T$$

$$|\tau_{il}| = \hat{\mathbf{k}}_l.\mathbf{r}_i/c = \mathbf{k}_l.\mathbf{r}_i/\omega = \mathbf{k}_l^T \mathbf{r}_i/\omega$$

Principles of Acoustic Wave Propagation

Wave Equation in Cartesian Coordinate

- The infinitesimal variation of acoustic pressure from its equilibrium, $p(\mathbf{r}, t)$ satisfies the following wave equation,

$$\nabla^2 p = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}$$

where ∇^2 represents the Laplacian operator and c is the speed of sound wave propagation in a particular medium.

- The Equation 4 can be written in Cartesian coordinate as

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}.$$

- Planewave solution of form $p(t-\tau)$ for far-field sources and spherical wave solution of the form $p(t-\tau)/r$ for near-field will satisfy above Eqn.

Principles of Acoustic Wave Propagation

Solution to Wave Equation in Cartesian Coordinate

- Writing the acoustic pressure at a point $\mathbf{r}_i$ due to a source at $\mathbf{r}_l$, we have

$$p_{il}(\mathbf{r}, t) = p_l(t - \mathbf{k}_l^T \mathbf{r}_i / \omega).$$

- Monochromatic plane wave solution is (after taking FT)

$$P_{il}(\mathbf{r}, k) = e^{-\mathbf{k}_l^T \mathbf{r}_i} P_l(k)$$

- Spherical wave[1] has the form $\frac{p(t-\tau_i(\Psi_l))}{|\mathbf{r}_i - \mathbf{r}_l|}$ with $\tau_i(\Psi_l) = \frac{|\mathbf{r}_i - \mathbf{r}_l|}{c}$.
- Monochromatic spherical wave solution, can be written as

$$P_{il}(\mathbf{r}, k) = \frac{e^{-k|\mathbf{r}_l - \mathbf{r}_i|}}{|\mathbf{r}_l - \mathbf{r}_i|} P_l(k).$$

[1] R. P. Feynman, R. B. Leighton, and M. Sands, The Feynman Lectures on Physics, 2013, vol. 1

Principles of Acoustic Wave Propagation

Solution to Wave Equation in Spherical Coordinate

- Wave equation in (4) can be written in spherical coordinates as

$$\frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial p}{\partial r}) + \frac{1}{r^2 \sin{(\theta)}}\frac{\partial}{\partial \theta}(\sin{(\theta)}\frac{\partial p}{\partial \theta}) + \frac{1}{r^2 \sin^2{(\theta)}}\frac{\partial^2 p}{\partial \phi^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}$$

- General solution to above Equation for standing wave type is

$$P(r,\theta,\phi,\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(A_{mn}j_n(kr) + B_{mn}y_n(kr)\right)Y_n^m(\theta,\phi)$$

and for traveling wave type is

$$P(r,\theta,\phi,\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(C_{mn}h_n^1(kr) + D_{mn}h_n^2(kr)\right)Y_n^m(\theta,\phi)$$

- The general solution for exterior problems (sources inside the spherical surface at $r = a$) is

$$P(r,\theta,\phi,\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} C_{mn}h_n^1(kr)Y_n^m(\theta,\phi)$$

Solution to Wave Equation in Spherical Coordinate

- The radiated pressure field is completely defined when the coefficients C_{mn} are determined.
- The general solution for interior problems (sources are located outside a sphere of radius $r = b$) is

$$P(r, \theta, \phi, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_{mn} j_n(kr) Y_n^m(\theta, \phi)$$

- The temporal dependency is implicit in frequency dependence of the co-efficients A_{mn}, B_{mn}, C_{mn} and D_{mn}.
- $j_n(kr)$ and $y_n(kr)$ are spherical Bessel functions, $h_n^1(kr)$ and $h_n^2(kr)$ are spherical Hankel function of first and second kind respectively.

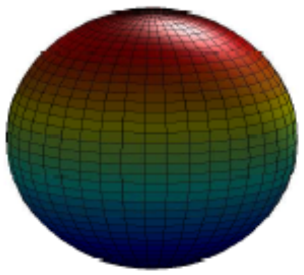
Spherical Harmonics

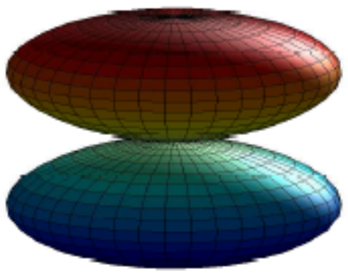
- Angular dependence of the solution is described by spherical harmonics.
- Y_n^m represents spherical harmonic of order n and degree m given by

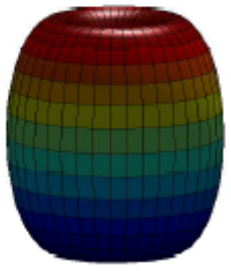
$$Y_n^m(\theta,\phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(cos\theta)e^{jm\phi}.$$

$$\forall 0 \leq n \leq N, -n \leq m \leq n$$

  where P_n^m are the associated Legendre function.
- Spherical harmonics form a orthonormal basis set and any arbitrary function on a sphere can be expanded in terms of them.
- Spherical harmonics plot : Y_0^0, Y_1^0, Y_1^1


Y_0^0


Y_1^0


Y_1^1

Principles of Acoustic Wave Propagation

Scattering from Rigid Sphere : Plane Wave

- The pressure field due to unit amplitude plane wave (interior problem) on an open sphere (without scattering) can be written as

$$e^{-j\mathbf{k}_l^T\mathbf{r}_i} = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(4\pi j^n j_n(kr)\right)[Y_n^m(\theta_l,\phi_l)]^* Y_n^m(\theta_i,\phi_i).$$

- In presence of scattering (exterior problem) from a spherical surface of radius r_a, the resultant pressure field due to unit amplitude plane wave on a rigid sphere can be written as

$$e^{-j\mathbf{k}_l^T\mathbf{r}_i} = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} 4\pi j^n\left(j_n(kr) - \frac{j_n'(kr_a)}{h_n'(kr_a)}h_n(kr)\right)[Y_n^m(\theta_l,\phi_l)]^* Y_n^m(\theta_i,\phi_i).$$

- Combining above two equations, the plane wave solution over sphere is

$$e^{-j\mathbf{k}_l^T\mathbf{r}_i} = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} b_n(k,r)[Y_n^m(\theta_l,\phi_l)]^* Y_n^m(\theta_i,\phi_i)$$

- $b_n(k,r)$ is called far-field mode strength.

Principles of Acoustic Wave Propagation

Scattering from Rigid Sphere : Spherical Wave

- Unit amplitude spherical wave can be written in spherical coordinate using Jacobi-Anger expansion as

$$\frac{e^{-jk|\mathbf{r}_i-\mathbf{r}_l|}}{|\mathbf{r}_i-\mathbf{r}_l|} = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} b_n(k,r,r_l)Y_n^m(\theta_l,\phi_l)^* Y_n^m(\theta_i,\phi_i).$$

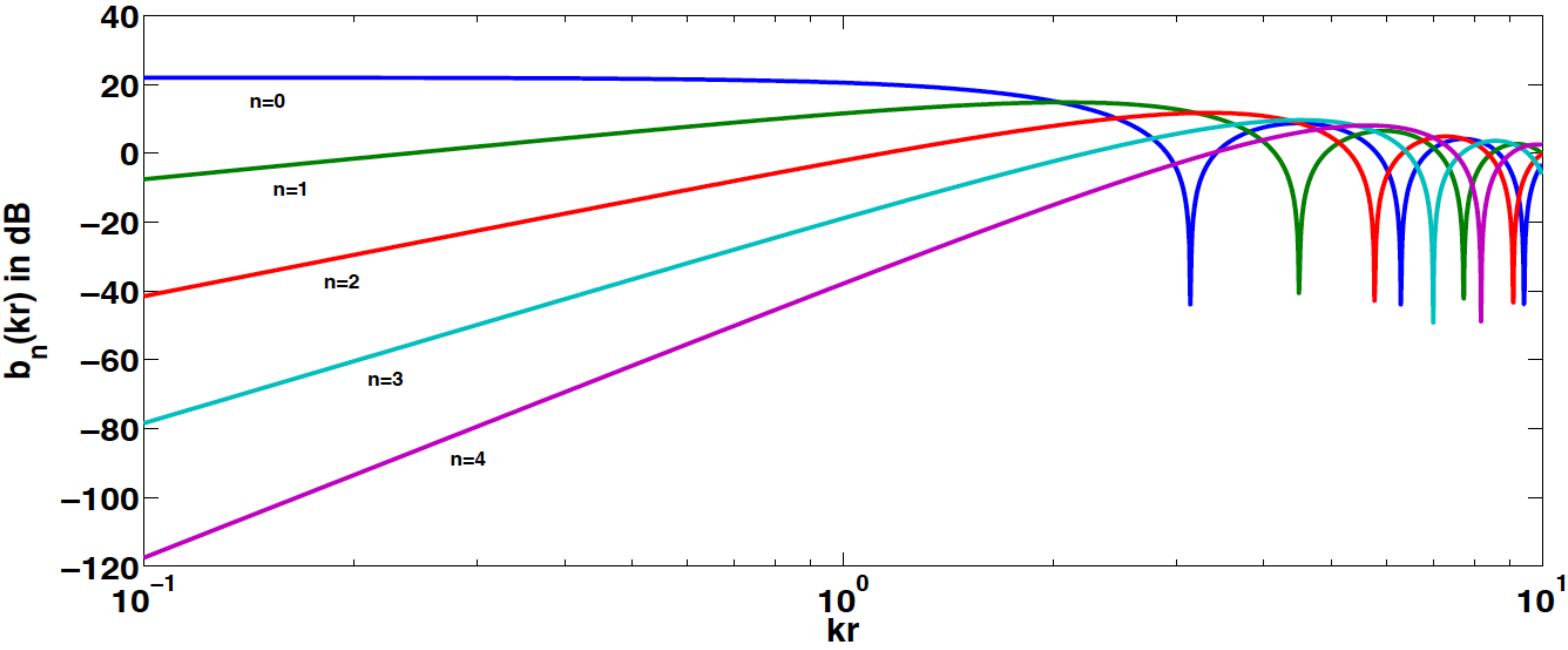
- $b_n(k,r,r_l)$ is near-field mode strength. It is related to far-field mode strength $b_n(k,r)$ as

$$b_n(k,r,r_l) = j^{-(n-1)}kb_n(k,r)h_n(kr_l)$$

- Far-field mode strength $b_n(k,r)$ is given by

$$\begin{aligned} b_n(k,r) &= 4\pi j^n j_n(kr), \text{ open sphere} \\ &= 4\pi j^n\left(j_n(kr) - \frac{j_n'(kr_a)}{h_n'(kr_a)}h_n(kr)\right), \text{ rigid sphere, radius } r_a \leq r \end{aligned}$$

Spherical Mode Strength



- b_n decreases significantly for $n > kr$. The summation can be truncated to some finite $N \geq kr$, called array order.

Near-field Criteria

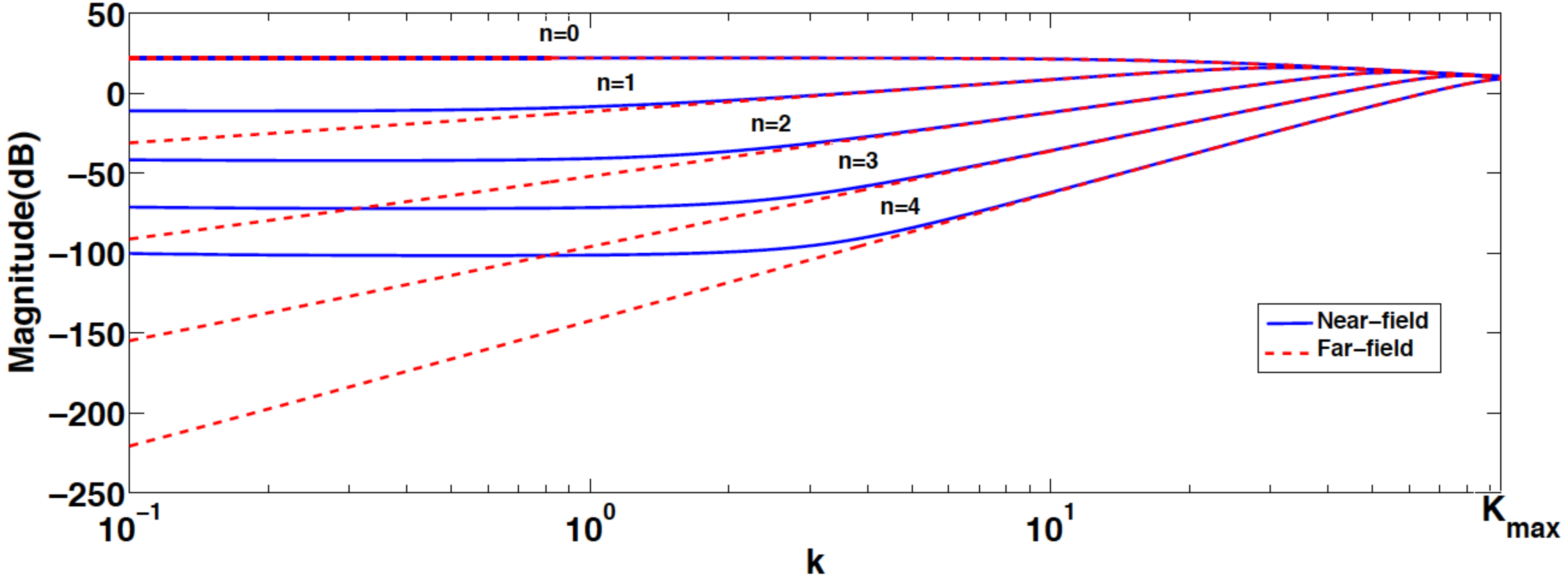


Figure: Far-field and near-field mode strength for source at $r_l = 1m$.

- Traditionally, transition between the near-field and far-field of sensor arrays is determined by the Fraunhofer or Fresnel distances.
- The near-field criteria for spherical array is based on similarity of near-field mode strength ($|b_n(k, r_a, r_l)|$) and far-field mode strength ($|b_n(k, r_a)|$).
- The two functions start behaving in similar way at $kr_l \approx N$, for array of order N. Hence, near-field condition for spherical array turns out to be $r_{NF} \approx \frac{N}{k}$ and $r_a \leq r_l \leq \frac{N}{k}$ [5].

The Spherical Fourier Transform

- Linear array : front-back ambiguity, planar array : up-down ambiguity. Spherical arrays has no spatial ambiguity and can localize sources anywhere in 3D space .
- Let the signal received at $(r, \Phi) = (r, \theta, \phi)$ be denoted by $p(t, r, \Phi) \leftrightarrow P(k, r, \Phi)$ with $r \geq r_a$ and k is wavenumber.
- The spherical Fourier transform (SFT) or spherical harmonics decomposition of the received signal is

$$P_{nm}(k, r) = \int_{\Omega \in S^2} P(k, r, \Phi)[Y_n^m(\Phi)]^* d\Omega$$

where Y_n^m is spherical harmonics of order n and degree m, $d\Omega = \sin\theta d\theta d\phi$ is elemental area over sphere of unit radius, and $(.)^*$ denotes complex conjugate of $(.)$.

- Substituting for $d\Omega$, the SFT can be expressed as

$$P_{nm}(k, r) = \int_0^{2\pi} \int_0^{\pi} P(k, r, \Phi)[Y_n^m(\Phi)]^* \sin(\theta) d\theta d\phi.$$

Far Field Source Localisation and Beamforming in SH Domain

The Spherical Fourier Transform

- In practice, the signal is sampled at the sensor locations. Hence, SFT of pressure is approximated by a summation as

$$P_{nm}(k,r) \cong \sum_{i=1}^{I} a_i P_i(k,r,\Phi_i)[Y_n^m(\Phi_i)]^*.$$

- For validity of above Eqn., orthogonality of spherical harmonics must be maintained

$$\sum_{i=1}^{I} a_i Y_{n'}^{m'}(\Phi_i)[Y_n^m(\Phi_i)]^* = \delta_{nn'}\delta_{mm'}$$

- One of the widely utilized sampling scheme is Gaussian Sampling where the azimuth angle is sampled at $2(N+1)$ equal-angle samples, and the elevation angle is sampled at $(N+1)$ samples with nearly equally spaced.

Far Field Source Localisation and Beamforming in SH Domain

The Spherical Fourier Transform

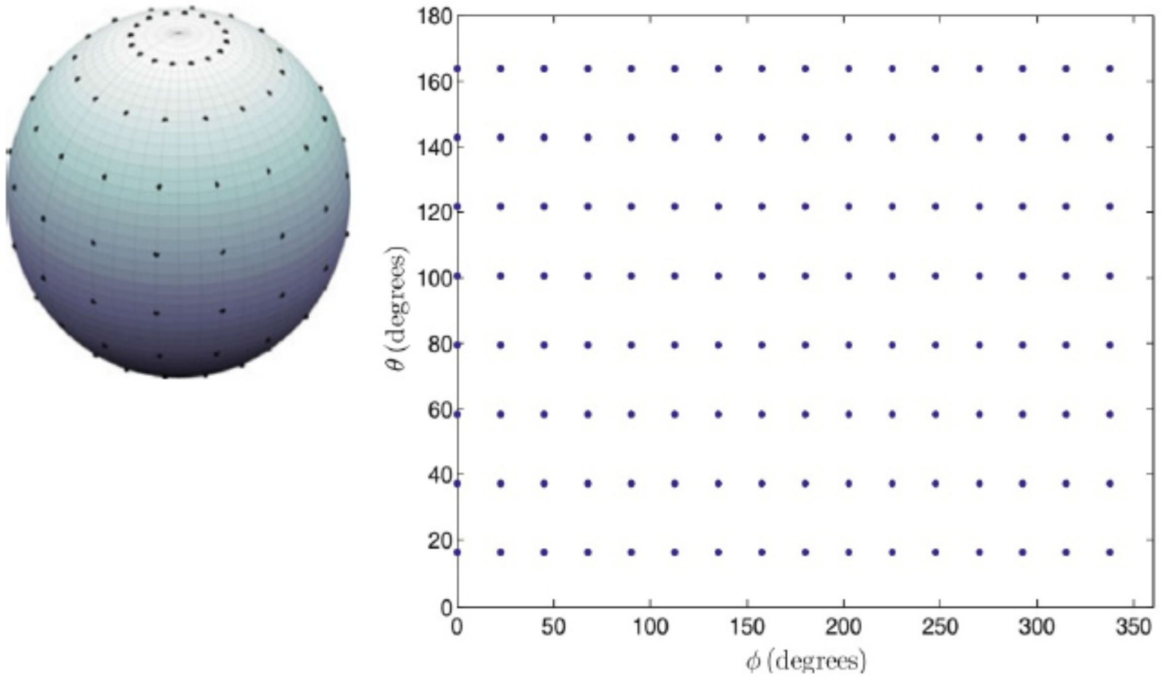


Figure: Gaussian sampling distribution for N = 7 and a total of 128 samples.

- In matrix form for all $n \in [0, N]$, $m \in [-n, n]$ and I, the SFT becomes

$$\mathbf{P_{nm}}(k, r) \cong \mathbf{Y}^H(\Phi)\mathbf{\Gamma}\mathbf{P}(k, r, \Phi)$$

- $\mathbf{P_{nm}} = \begin{bmatrix} P_{00} & P_{1(-1)} & P_{10} & P_{11} & \cdots & P_{NN} \end{bmatrix}^T$ is $(N+1)^2 \times 1$ matrix
- $\mathbf{Y}(\Phi)$ is $I \times (N+1)^2$ matrix whose i^{th} row is given as

$$\mathbf{y}(\Phi_i) = \begin{bmatrix} Y_0^0(\Phi_i) & Y_1^{-1}(\Phi_i) & Y_1^0(\Phi_i) & Y_1^1(\Phi_i) & \dots & Y_N^N(\Phi_i) \end{bmatrix}$$

- $\mathbf{\Gamma} = \mathrm{diag}(a_1, a_2, \cdots, a_I)$ is $I \times I$ matrix of sampling weights.

Far Field Source Localisation and Beamforming in SH Domain

Spatio-Temporal to Spherical Harmonics Data Model

- The total signal at received at the i^{th} sensor is

$$p_i(\Psi; t) = \sum_{l=1}^{L} s_l\big(t - \tau_i(\Psi_l)\big) + v_i(t)$$

- Computing the discrete Fourier transform (DFT) of above Equation, the spatio-frequency data model can be written as

$$P_i(\Psi; f_\nu) = \sum_{l=1}^{L} e^{-j2\pi f_\nu \tau_i(\Psi_l)} S_l(f_\nu) + V_i(f_\nu),\ \nu = 1, \cdots, N_s.$$

where the frequency f_ν is related to FFT index ξ_ν of DFT as

$$f_\nu = \frac{\xi_\nu}{T_s N_s}$$

- Matrix form of the data model in spatio-frequency domain becomes

$$\mathbf{P}(\Psi; k) = \mathbf{A}(\Psi; k)\mathbf{S}(k) + \mathbf{V}(k),$$

Far Field Source Localisation and Beamforming in SH Domain

Spatio-Temporal to Spherical Harmonics Data Model

- The spatio-frequency data model is

$$\mathbf{P}(\Psi;k) = \mathbf{A}(\Psi;k)\mathbf{S}(k) + \mathbf{V}(k)$$

$$\mathbf{A}(\Psi;k) = [\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a_L}], \text{ where}$$

$$\mathbf{a}_l = [e^{-j\mathbf{k}_l^T\mathbf{r}_1}, e^{-j\mathbf{k}_l^T\mathbf{r}_2}, \ldots, e^{-j\mathbf{k}_l^T\mathbf{r}_I}]^T,$$

- $e^{-j\mathbf{k}_l^T\mathbf{r}_i}$ is plane wave solution to the wave equation in Cartesian co-ordinates which can be written for a rigid sphere in spherical co-ordinate as [4]

$$e^{-j\mathbf{k}_l^T\mathbf{r}_i} = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} b_n(k,r)[Y_n^m(\Psi_l)]^* Y_n^m(\mathbf{\Phi}_i)$$

Far Field Source Localisation and Beamforming in SH Domain

Spatio-Temporal to Spherical Harmonics Data Model

- Using expression for plane wave solution obtained in last Eqn., the steering matrix can be finally simplified as

$$\mathbf{A}(\Psi; k) = \mathbf{Y}(\Phi)\mathbf{B}(k, r)\mathbf{Y}^H(\Psi)$$

- $\mathbf{Y}(\boldsymbol{\Phi})$ is $I \times (N+1)^2$ matrix whose i^{th} row is given as

$$\mathbf{y}(\boldsymbol{\Phi}_i) = [Y_0^0(\boldsymbol{\Phi}_i), Y_1^{-1}(\boldsymbol{\Phi}_i), Y_1^0(\boldsymbol{\Phi}_i), Y_1^1(\boldsymbol{\Phi}_i), \ldots, Y_N^N(\boldsymbol{\Phi}_i)].$$

- The $L \times (N+1)^2$ matrix $\mathbf{Y}(\Psi)$ can be expanded on similar lines.
- The $(N+1)^2 \times (N+1)^2$ matrix $\mathbf{B}(kr)$ is given by

$$\mathbf{B}(k, r) = diag\big(b_0(k, r), b_1(k, r), b_1(k, r), b_1(k, r), \ldots, b_N(k, r)\big)$$

Far Field Source Localisation and Beamforming in SH Domain

Advantage of Data Model Formulation in SH Domain

- The steering matrix in spherical harmonics domain turns out to be $\mathbf{A_{nm}}(\Psi) = \mathbf{Y}^H(\Psi)$. A particular steering vector can be written as

$$\mathbf{a_{nm}}(\Psi_l) = \mathbf{y}^H(\Psi_l) = [Y_0^{0^*}(\Psi_l), Y_1^{-1^*}(\Psi_l), Y_1^{0^*}(\Psi_l), Y_1^{1^*}(\Psi_l), \ldots, Y_N^{N^*}(\Psi_l)]^T$$

- Data dimensionality reduced from $I \times N_s$ to $(N+1)^2 \times N_s$ with $I \geq (N+1)^2$.
- Frequency independent nature of steering matrix allows frequency smoothing to restore the rank of signal covariance matrix.

$$\mathbf{R_{D_{nm}}}(k) = E[\mathbf{D_{nm}}(k)\mathbf{D_{nm}}(k)^H] = \mathbf{Y}^H(\Psi)\mathbf{R_S}(k)\mathbf{Y}(\Psi) + \mathbf{R_{Z_{nm}}}(k)$$

- Ease of beamforming, due to reduced dimensionality of array covariance matrix and simple structure of steering vector component.

Far Field Source Localisation and Beamforming in SH Domain

SH-MUSIC and SH-MGD

- Spherical harmonics MUSIC spectrum can now be written as

$$P_{SH-MUSIC}(\Psi) = \frac{1}{\mathbf{a_{nm}}^H(\Psi)\mathbf{Q_{nm}}\mathbf{Q_{nm}}^H\mathbf{a_{nm}}(\Psi)}$$

  where, $\mathbf{a_{nm}} = \mathbf{y}^H(\Psi_l)$.

- The Spherical Harmonics MUSIC-Group delay (SH-MGD) spectrum is computed as

$$P_{SH-MGD}(\Psi) = \Big(\sum_{u=1}^{U} |\nabla arg(\mathbf{a_{nm}}^H(\Psi)\mathbf{q}_u)|^2\Big) P_{SH-MUSIC}(\Psi).$$

Far Field Source Localisation and Beamforming in SH Domain

SH-MUSIC and SH-MGD Spectrum [3]

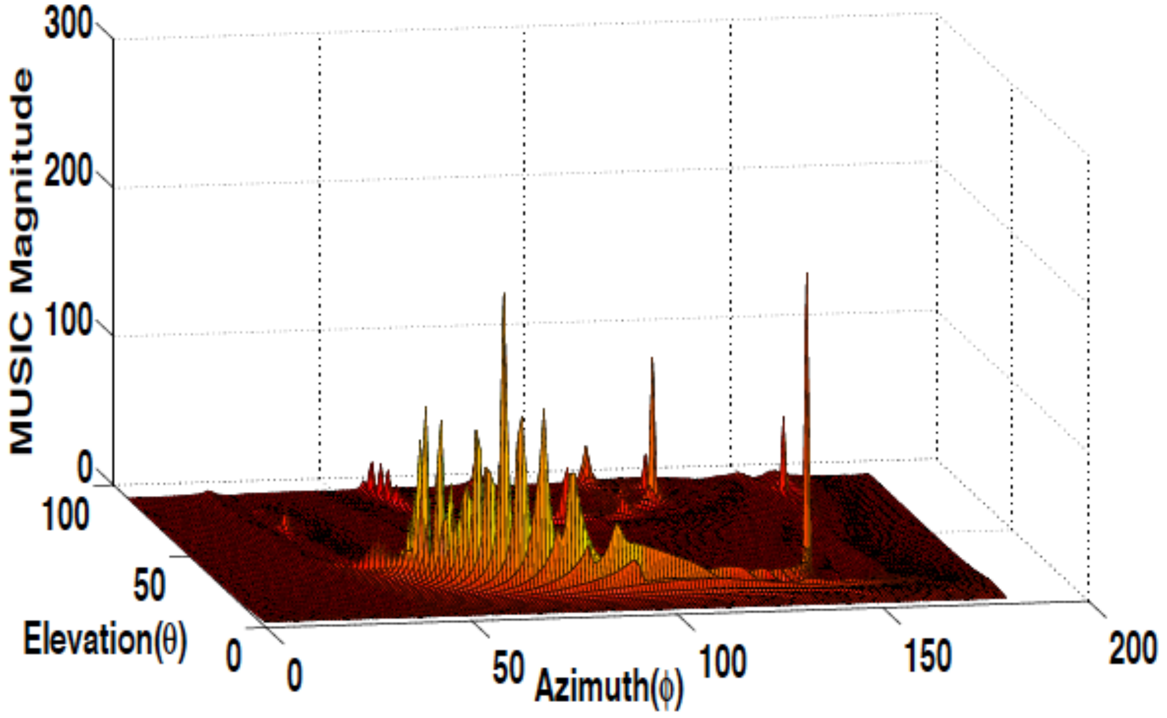


Figure: SH-MUSIC spectrum

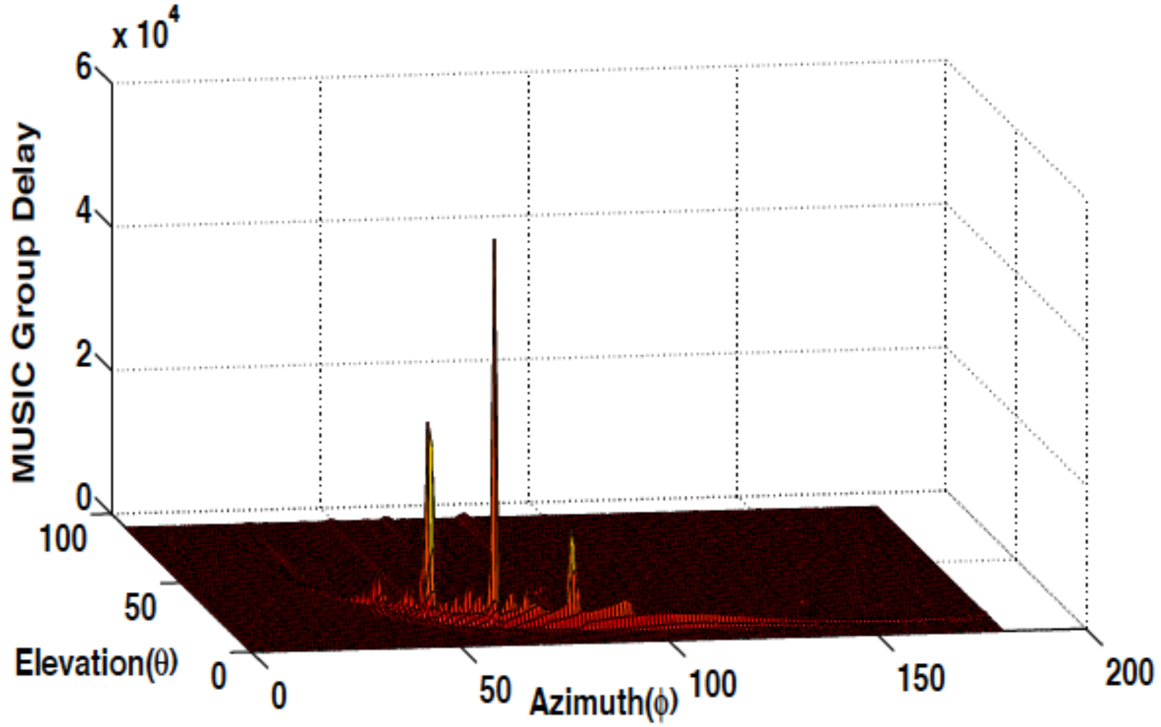
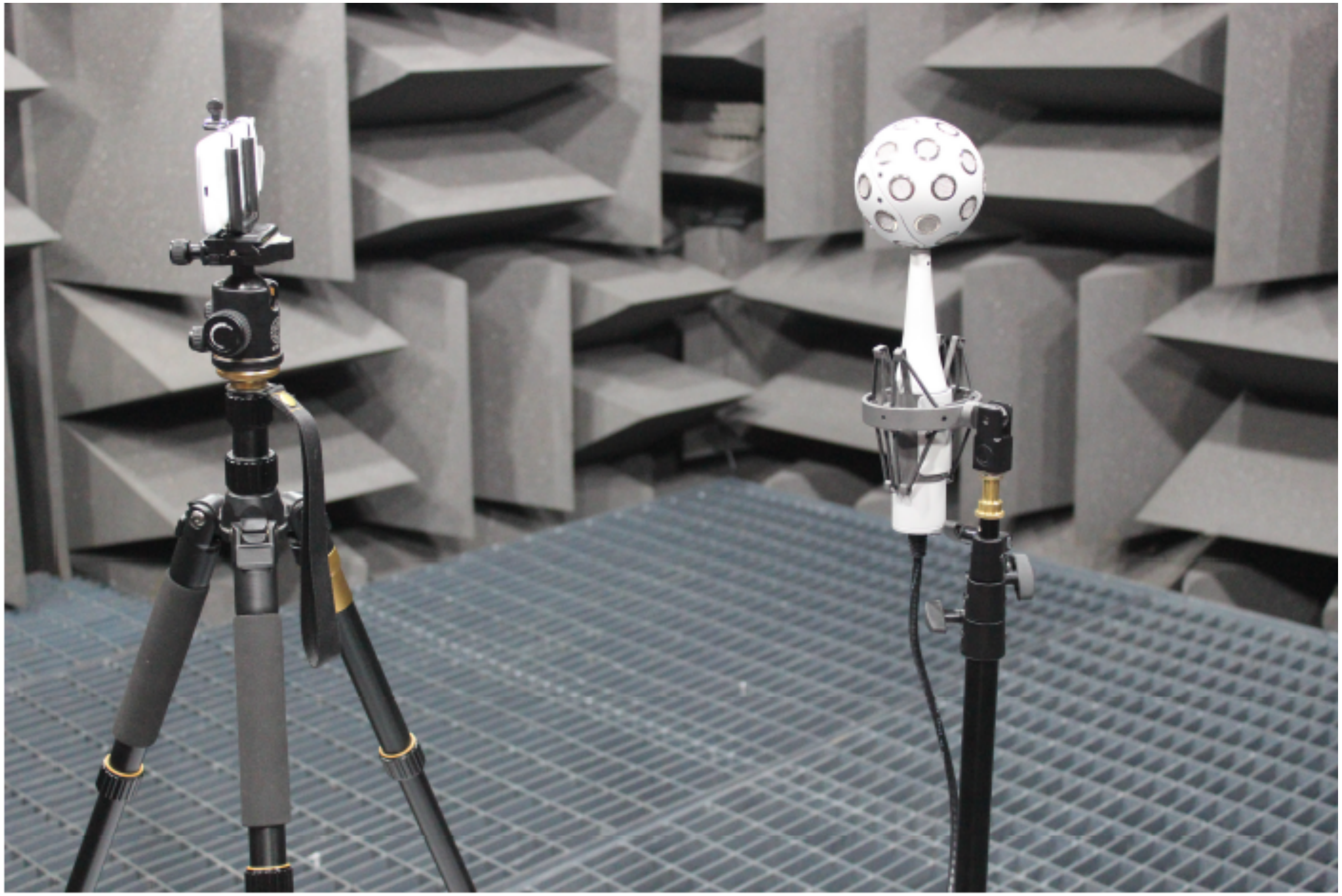


Figure: SH-MGD spectrum

- Two sources are taken at $(20^\circ, 50^\circ)$ and $(15^\circ, 60^\circ)$. Simulation is done assuming open sphere at SNR $10dB$.

Far Field Source Localisation and Beamforming in SH Domain



The Eigenmike setup in an anechoic chamber at IIT Kanpur for acquiring a far-field source

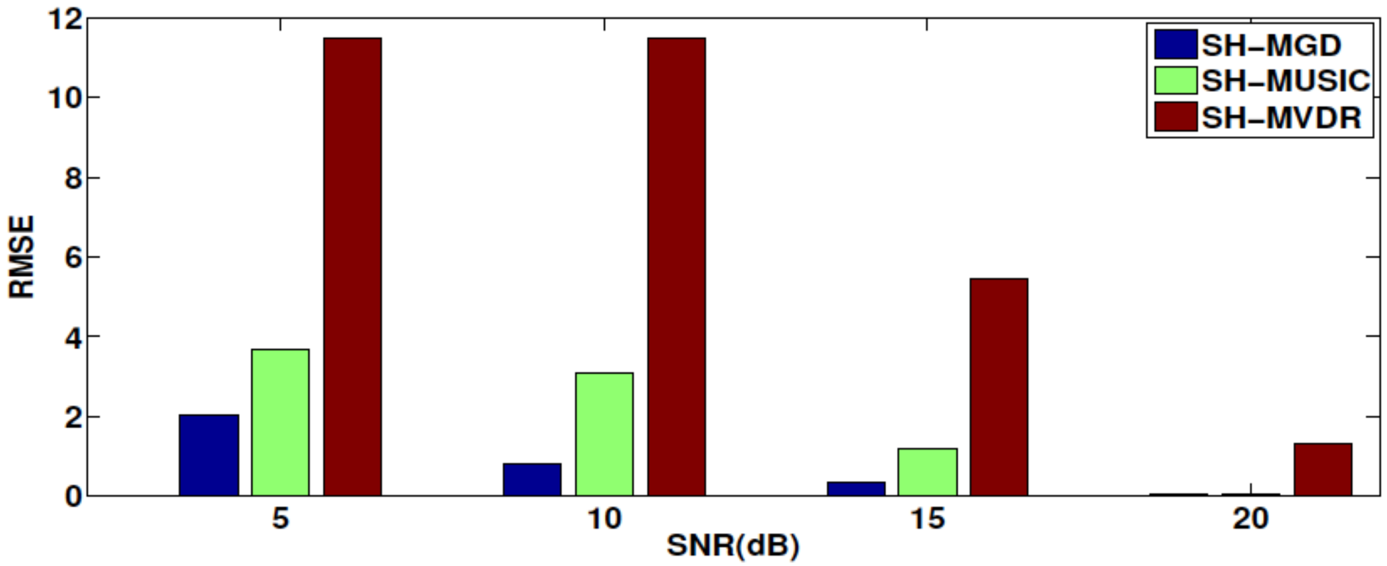
Far Field Source Localisation and Beamforming in SH Domain

Experiments on Far-field Source Localization in Noisy Environments

- Two sources at $(30°, 35°)$ and $(50°, 60°)$ are incident over fourth order Eigenmike system.
- The cumulative RMSE is defined as

$$RMSE = \frac{1}{4T}\sum_{t=1}^{T}\sum_{l=1}^{2}[(\theta_l - \hat{\theta}_l^{(t)})^2 + (\phi_l - \hat{\phi}_l^{(t)})^2],$$

where, t is trial number, T is the total trials and l denotes the source number.



Far Field Source Localisation and Beamforming in SH Domain

Experiments on Far-field Source Localization in Noisy Environments

- **The probability of resolution is given by**

$$P_r = \frac{1}{4T}\sum_{t=1}^{T}\sum_{l=1}^{2}\left[Pr(|\theta_l - \hat{\theta}_l^{(t)}| \le \zeta) + Pr(|\phi_l - \hat{\phi}_l^{(t)}| \le \zeta)\right]$$

where ζ is confidence interval.

Probability of resolution at various SNRs for 200 iterations. Sources are taken at $(30°, 35°)$ and $(50°, 60°)$.

Methods	SNR (5dB)	SNR (10dB)	SNR (15dB)	SNR (20dB)
SH-MGD	0.9167	0.9971	1	1
SH-MUSIC	0.9444	0.9829	0.9987	1
SH-MVDR	0	0	0.4179	1

Far Field Source Localisation and Beamforming in SH Domain

Experiments on Far-field Source Localization in Reverberant Environments

- **Proposed algorithm was tested for a room with dimensions, $7.3m \times 6.2m \times 3.4m$.**
- **Two sources at $\Psi_1 = (30°, 60°)$ and $\Psi_2 = (35°, 50°)$ were considered.**
- **Localization experiments were conducted for 300 iterations at three different reverberation level.**

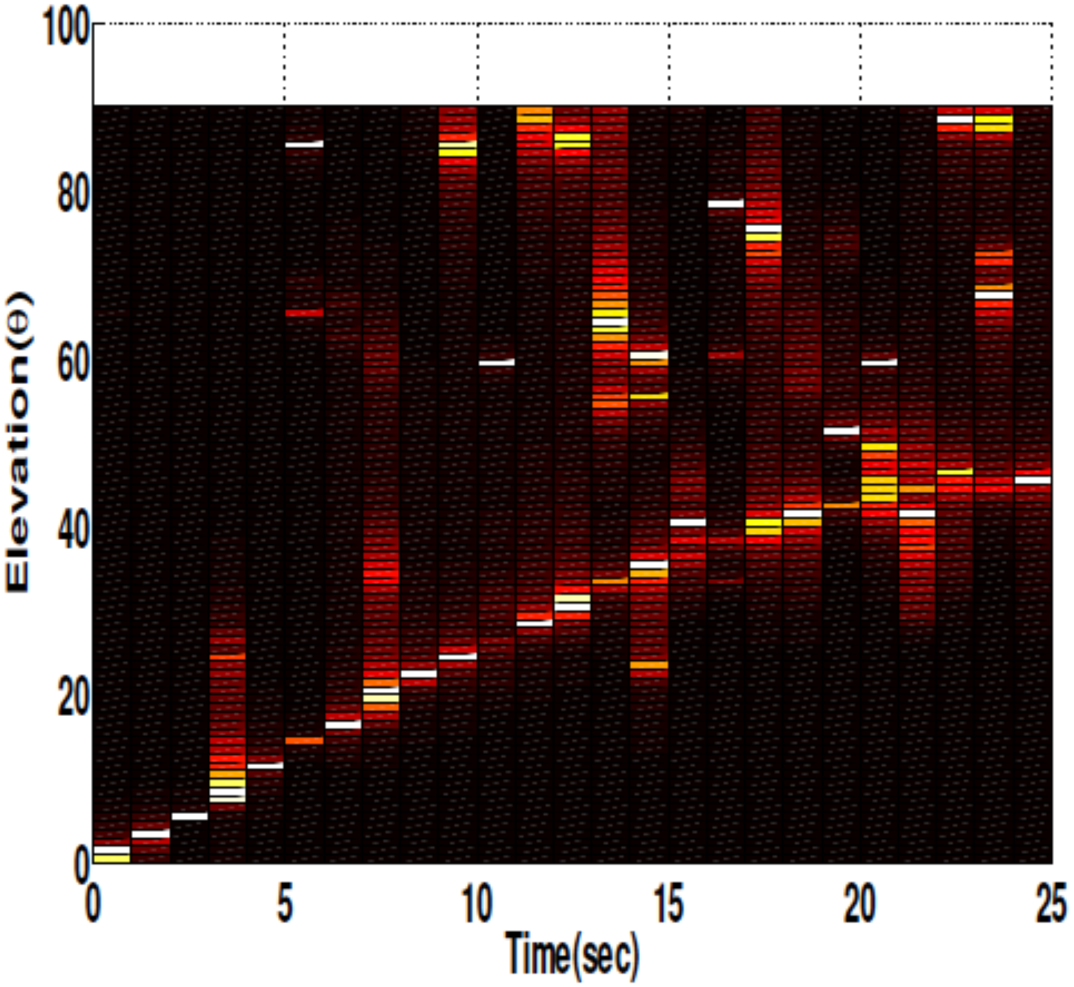
Comparison of RMSE of various methods at different reverberation time, T_{60}.

Angle	**Method**	T_{60} **(150ms)**	T_{60} **(200ms)**	T_{60} **(250ms)**
θ	**SH-MGD**	0.6403	0.6419	0.6475
	SH-MUSIC	0.6688	0.8144	0.7989
	SH-MVDR	1.1034	1.1579	1.1738
ϕ	**SH-MGD**	1.4387	1.4665	1.4866
	SH-MUSIC	1.7866	1.9127	1.6484
	SH-MVDR	2.276	2.3481	2.4927

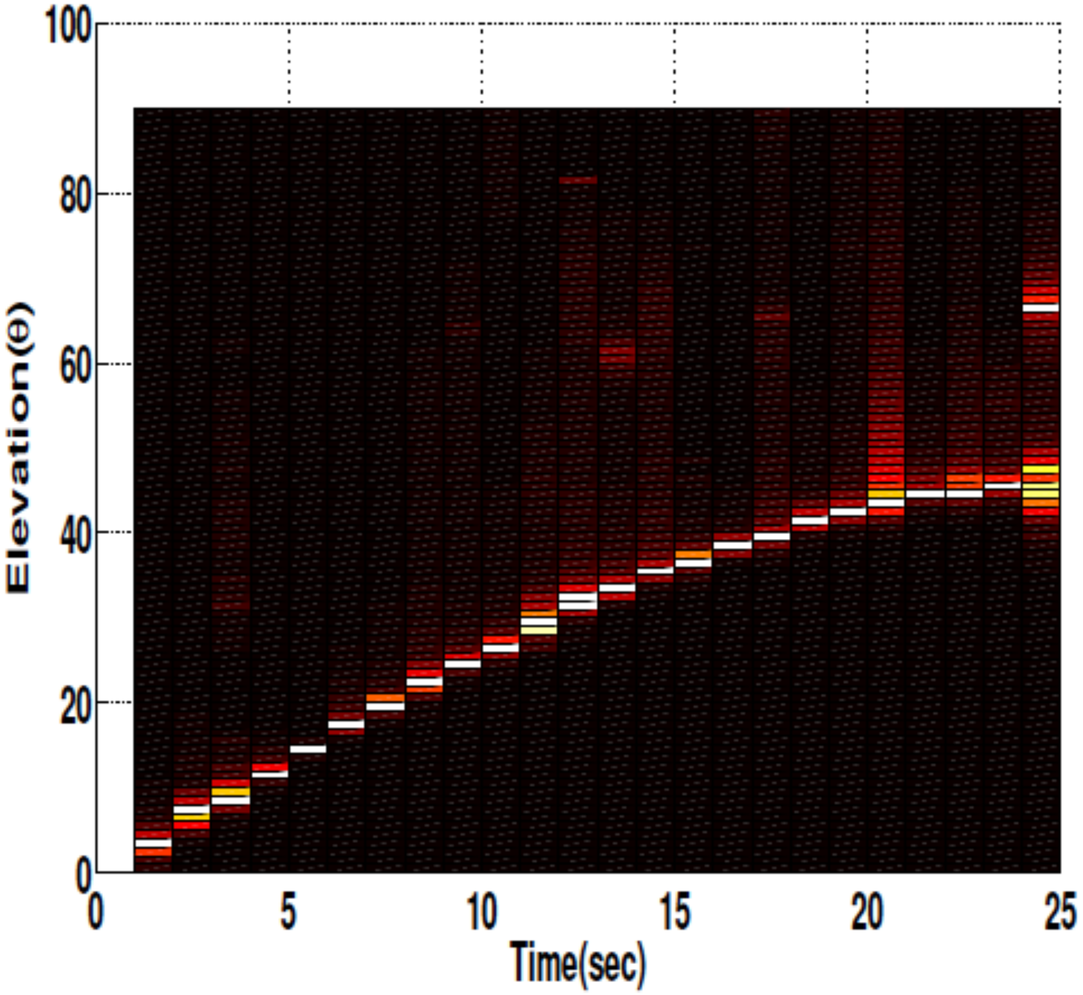
Far Field Source Localisation and Beamforming in SH Domain

Experiment on Narrowband Source Tracking

- Elevation of a narrowband source is tracked at fixed azimuth of 45°. The elevation is varied as in the below Figure.



Elevation tracking using SH-MUSIC

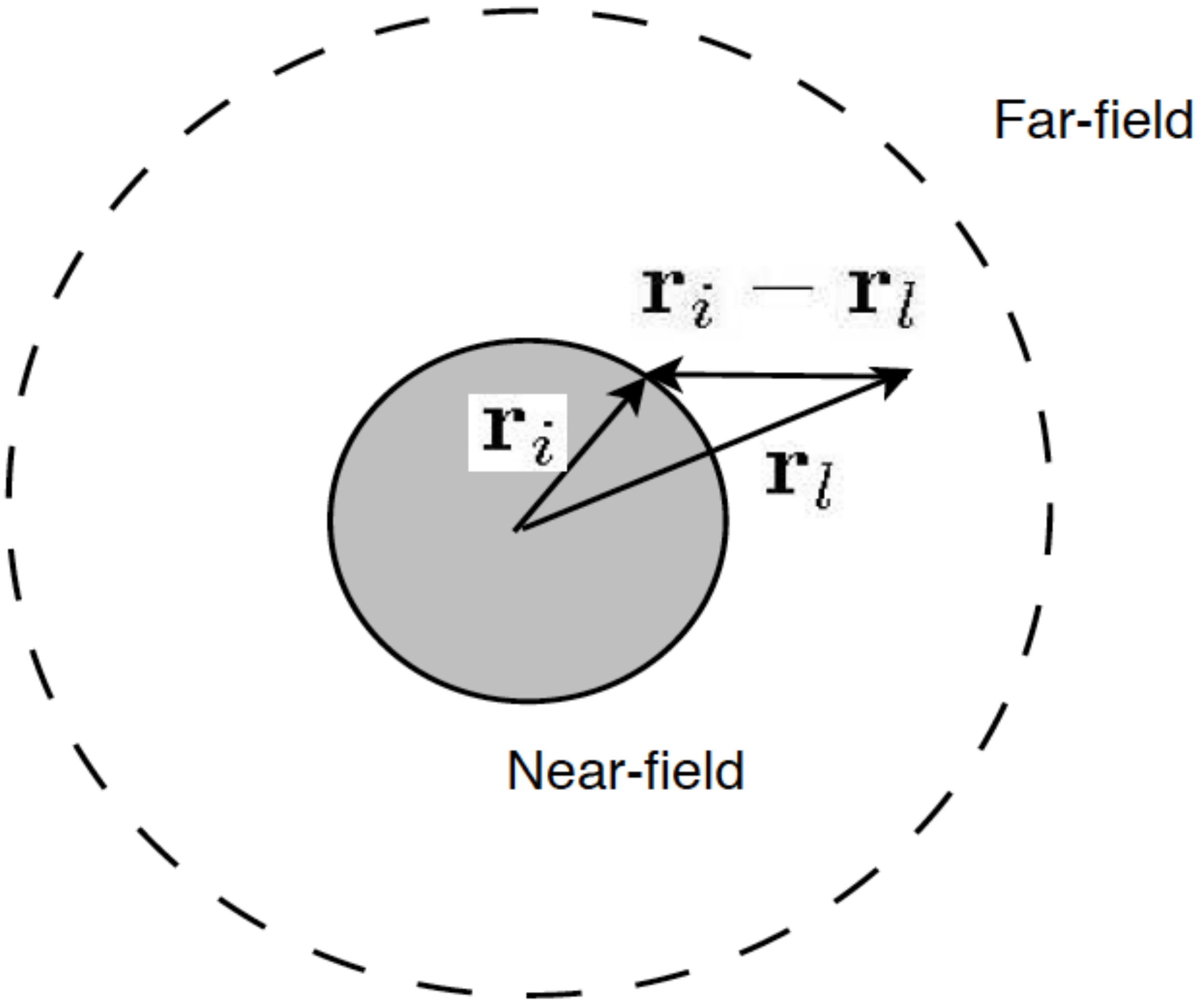


Elevation tracking using SH-MGD

Near Field Source Localisation and Beamforming in Spherical Harmonic Domain

Near Field Source Localisation and Beamforming in SH Domain

- The position vector of i^{th} microphone is given as $\mathbf{r}_i = (r_a, \mathbf{\Phi}_i)$ where r_a is radius of the spherical array and $\mathbf{\Phi}_i = (\theta_i, \phi_i)$.
- The position vector of l^{th} source at $\mathbf{r}_l = (r_l, \Psi_l)$



Near Field Source Localisation and Beamforming in SH Domain

Data Model Formulation

- Utilizing $\frac{e^{-jk|\mathbf{r}_i-\mathbf{r}_l|}}{|\mathbf{r}_i-\mathbf{r}_l|} = \sum_{n=0}^{N}\sum_{m=-n}^{n} b_n(k, r_a, r_l) Y_n^m(\Psi_l)^* Y_n^m(\boldsymbol{\Phi}_i)$ and SFT, the final near-field data model in SH domain can be written as

$$\mathbf{P_{nm}}(k) = \begin{bmatrix}\mathbf{B}(r_1)\mathbf{y}^H(\boldsymbol{\Psi}_1) & \cdots & \mathbf{B}(r_L)\mathbf{y}^H(\Psi_L)\end{bmatrix} \mathbf{S}(k) + \mathbf{V_{nm}}(k).$$

- Re-writing the data model in more compact way, we have

$$\mathbf{P_{nm}}(k) = \mathbf{A_{nm}}(r, \Psi)\mathbf{S}(k) + \mathbf{V_{nm}}(k)$$
$$\text{where, } \mathbf{A_{nm}}(r, \Psi) = \begin{bmatrix}\mathbf{B}(r_1)\mathbf{y}^H(\Psi_1), & \cdots, & \mathbf{B}(r_L)\mathbf{y}^H(\Psi_L)\end{bmatrix}$$

- A steering vector can be written as $\mathbf{a_{nm}}(r, \Psi) = \mathbf{B}(r)\mathbf{y}^H(\Psi)$.

Near Field Source Localisation and Beamforming in SH Domain

Near-field MUSIC, MUSIC-GD, MVDR in SH Domain

- The near-field spherical harmonics MUSIC spectrum can now be written as

$$P_{SH-MUSIC}(r,\Psi) = \frac{1}{\mathbf{a}_{\mathbf{nm}}^H(r,\Psi)\mathbf{Q}_{\mathbf{nm}}\mathbf{Q}_{\mathbf{nm}}^H\mathbf{a}_{\mathbf{nm}}(r,\Psi)}$$

  where $\mathbf{a}_{\mathbf{nm}}(r,\Psi) = \mathbf{B}(r)\mathbf{y}^H(\Psi)$.

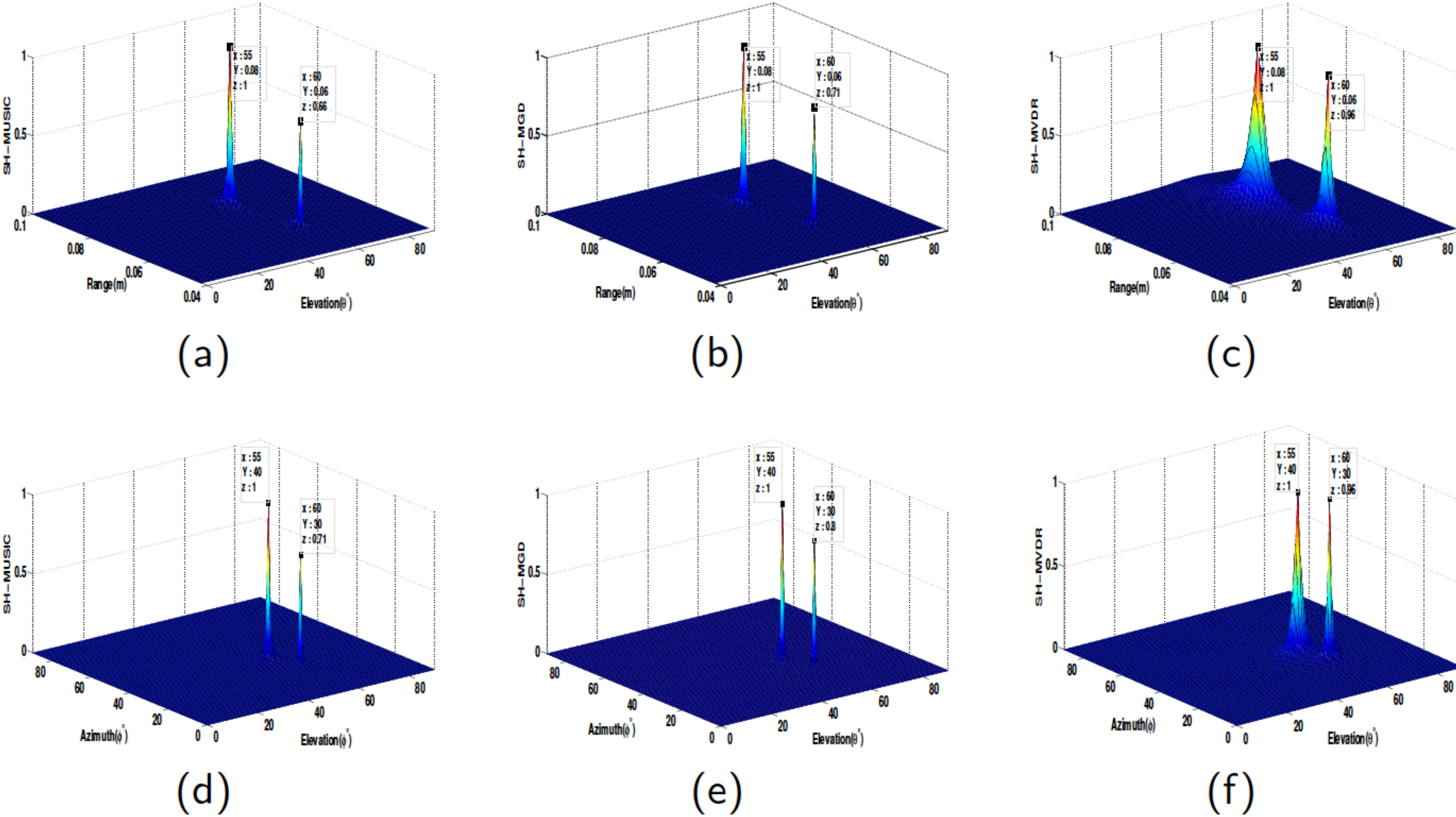
- The Spherical Harmonics MUSIC-Group delay spectrum is computed as

$$P_{SH-MGD}(r,\Psi) = \left(\sum_{u=1}^{U} |\nabla arg(\mathbf{a}_{\mathbf{nm}}^H(r,\Psi)\mathbf{q}_u)|^2\right) P_{SH-MUSIC}$$

- The SH-MVDR spectrum for near-field source localization, is written as

$$P_{SH-MVDR}(r,\Psi) = \frac{1}{\mathbf{a}_{\mathbf{nm}}^H(r,\Psi)\mathbf{R}_{\mathbf{P_{nm}}}^{-1}\mathbf{a}_{\mathbf{nm}}(r,\Psi)}.$$

Near-field Source Localization Result: SH-MUSIC, SH-MGD, SH-MVDR



(a) (b) (c)

(d) (e) (f)

Figure: The sources are at (0.06m,60°,30°) and (0.08m,55°,40°) with SNR 10dB.

Near Field Source Localisation and Beamforming in SH Domain

Experiments on Near-field Source Localization

- Two narrowband sources with location $\mathrm{r}_1 = (0.1, 30^\circ, 45^\circ)$ and $\mathrm{r}_2 = (0.8, 30^\circ, 45^\circ)$ with respective frequency as 220Hz and 250Hz, are taken.
- The DOA of the sources is assumed to be known, and range is estimated at various SNRs.
- The ability of the proposed methods to radially discriminate aligned sources is analyzed herein.
- The cumulative RMSE is computed as
- The cumulative RMSE is computed as

$$RMSE = \frac{1}{2T}\sum_{t=1}^{T}\sum_{l=1}^{2}[(r_l - \hat{r}_l^{(t)})^2],$$

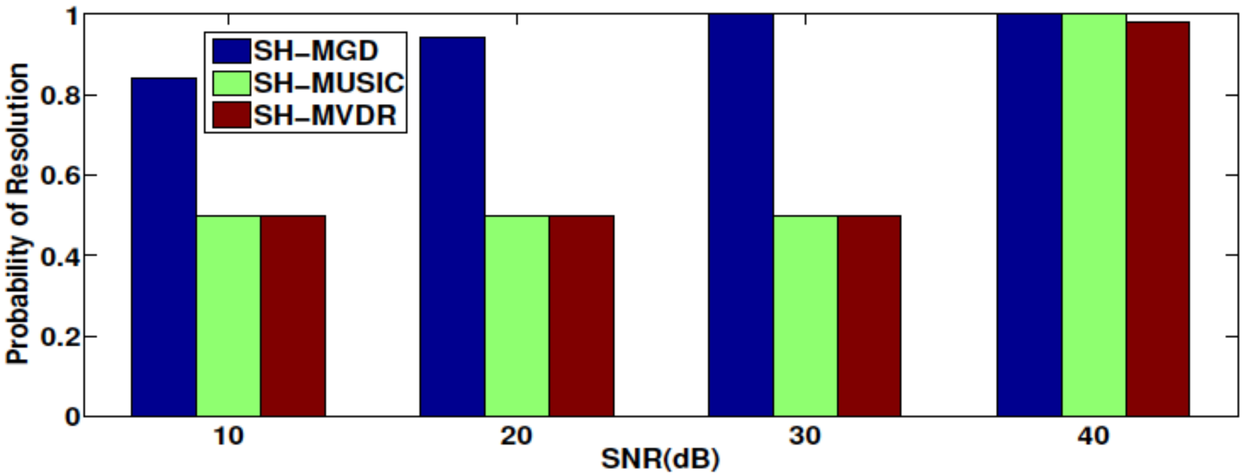
- The probability of resolution for range is defined as

$$P_r = \frac{1}{2T}\sum_{t=1}^{T}\sum_{l=1}^{2}[Pr(|r_l - \hat{r}_l^{(t)}| \leq \zeta)]$$

Experiments on Near-field Source Localization

Cumulative RMSE in range r, at various SNRs for 100 iterations. Sources are at $(0.1m, 30°, 45°)$ and $(0.8m, 30°, 45°)$.

Methods	SNR (10dB)	SNR (20dB)	SNR (30dB)	SNR (40dB)
SH-MGD	0.0847	0.0785	0.0389	0.0217
SH-MUSIC	0.495	0.495	0.2891	0.0049
SH-MVDR	0.495	0.495	0.495	0.0562



Range estimation performance of SH-MGD, SH-MUSIC and SH-MVDR in terms of probability of resolution.

Joint Range and Bearing Estimation

- The experiments are performed for both simulated and actual signals acquired from a spherical microphone array.
- A narrowband source with frequency 600Hz, is fixed at location $(0.3m, 90^\circ, 90^\circ)$.

Joint Range and Bearing Estimation : Simulation

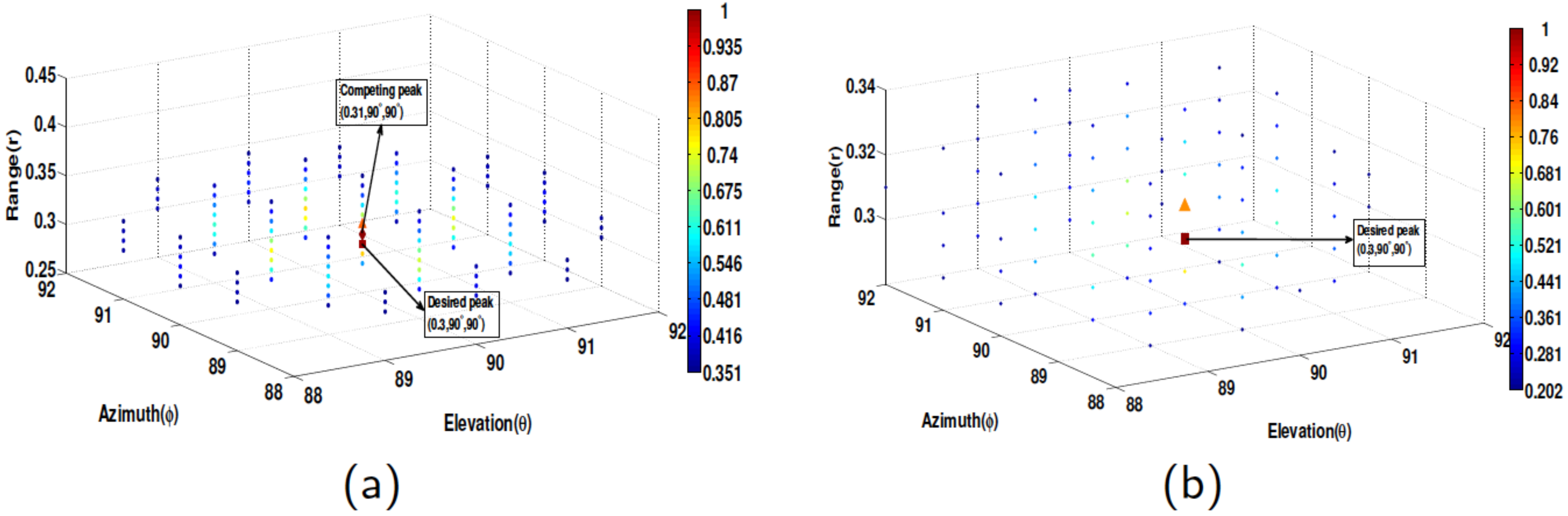
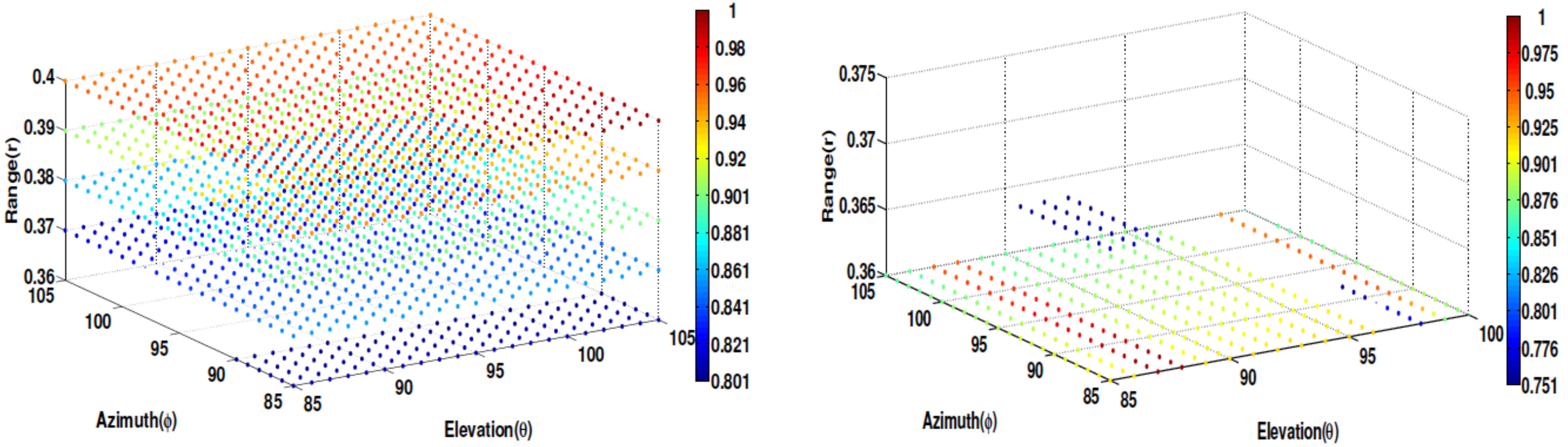


Figure: 4D scatter plots using, (a) SH-MUSIC for simulated signal, (b) SH-MGD for simulated signal.

Joint Range and Bearing Estimation : Real



(c) (d)

Figure: 4D scatter plots using (c) SH-MUSIC for signal acquired over SMA (d) SH-MGD for acquired over SMA.

References and Appendix