13. Microphone Arrays
Localization
Beamforming

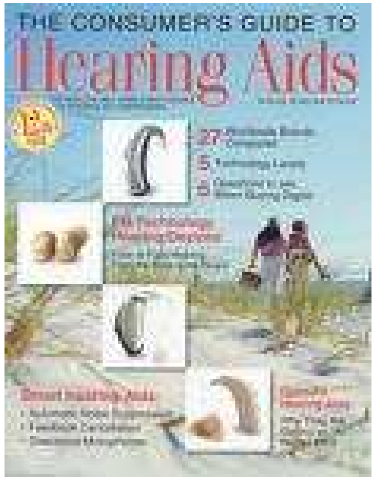
Rahil Mahdian

5.04.2016

Overview

- Microphone Array
- Speaker position localization
- Spatial filtering (Beamforming)

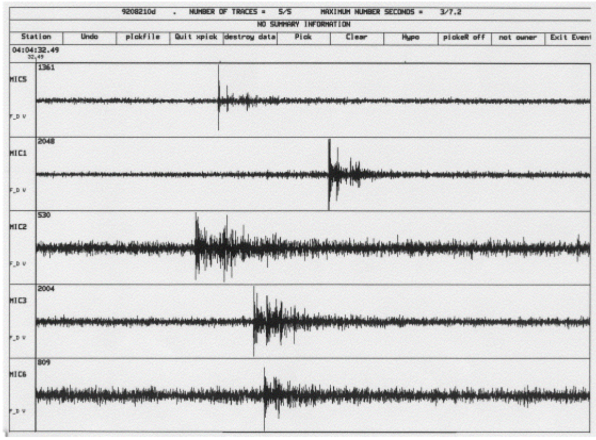
Applications Microphone Arrays



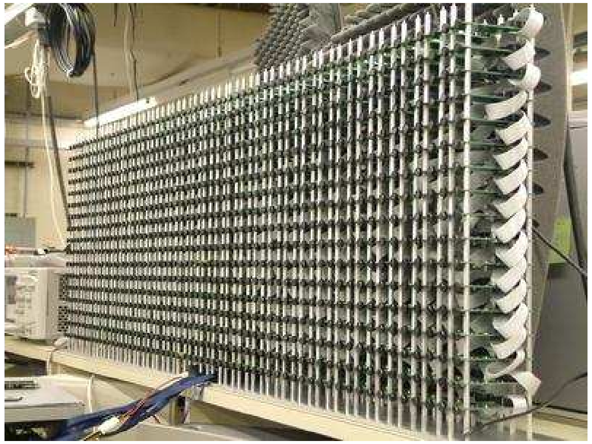
Speech Control

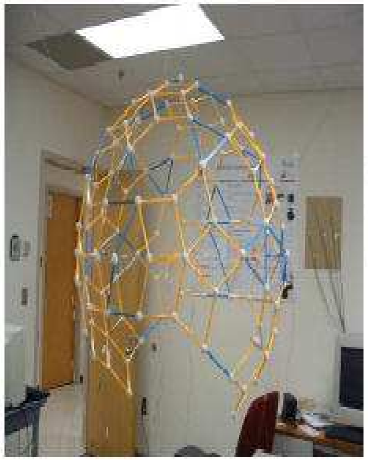
Robotics

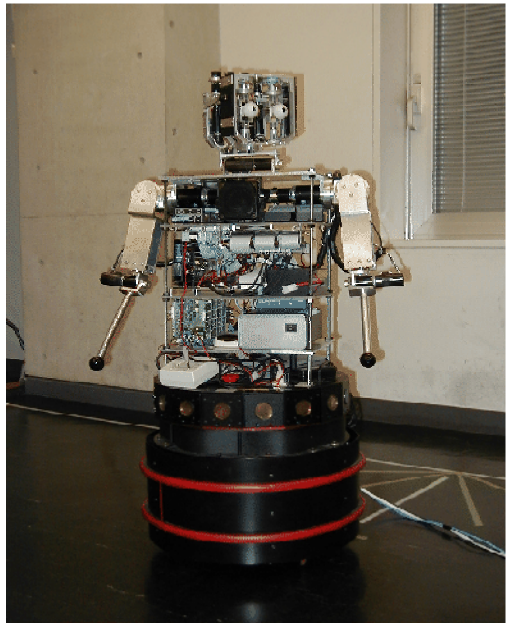
Seismology



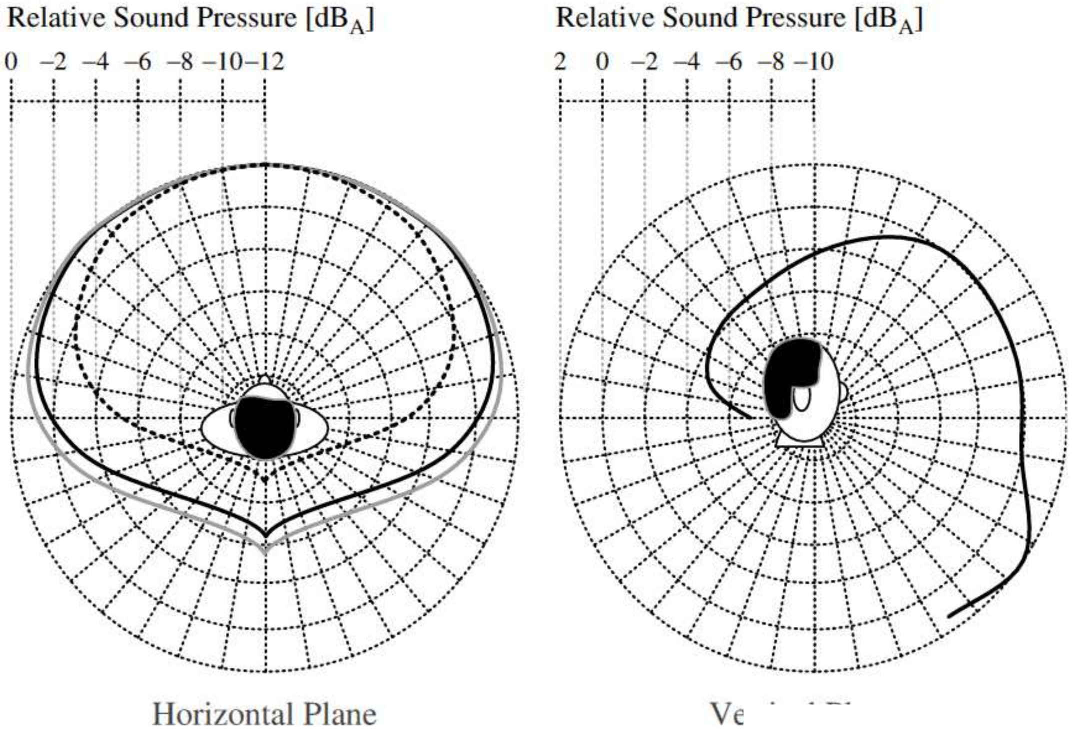
Types of Arrays

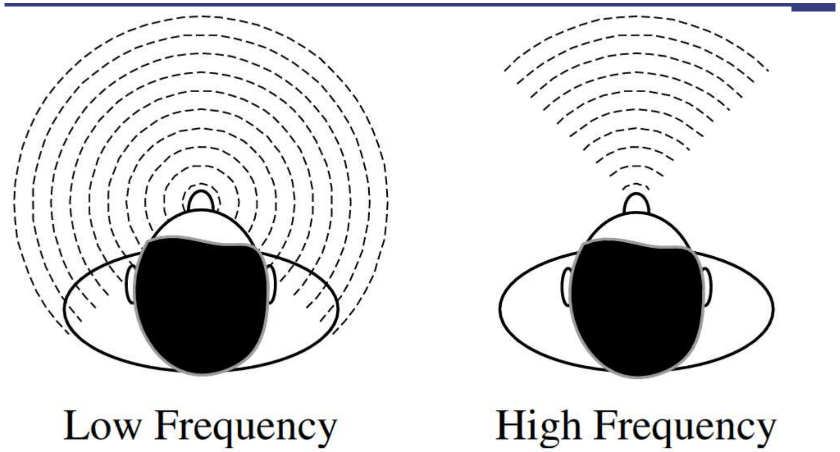


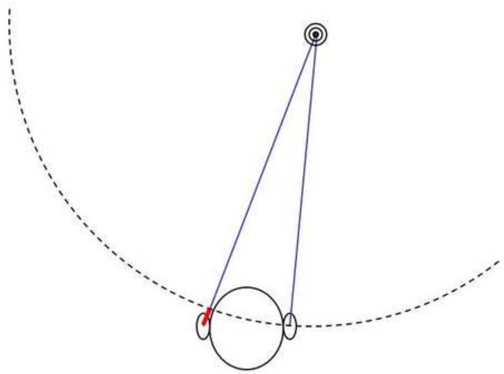


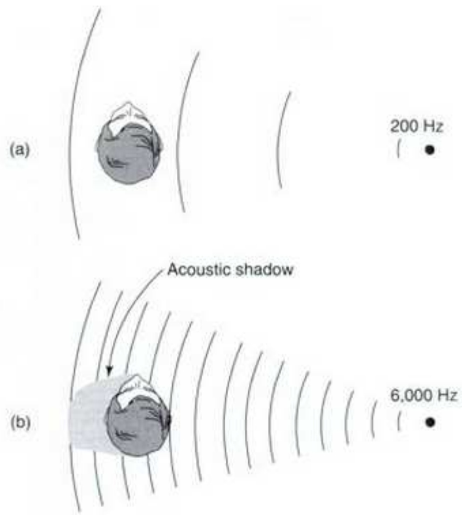


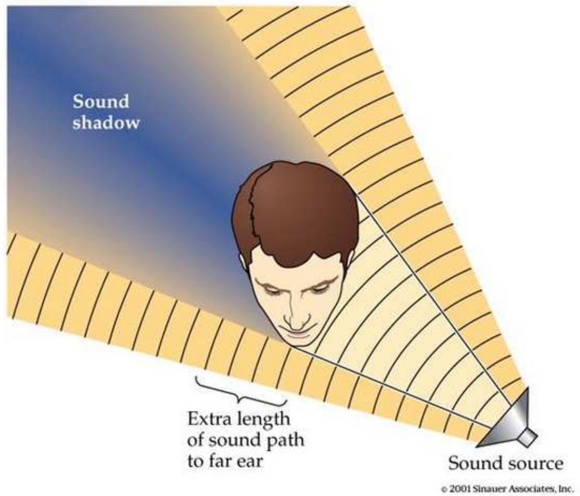
ITD and IID – Human listening



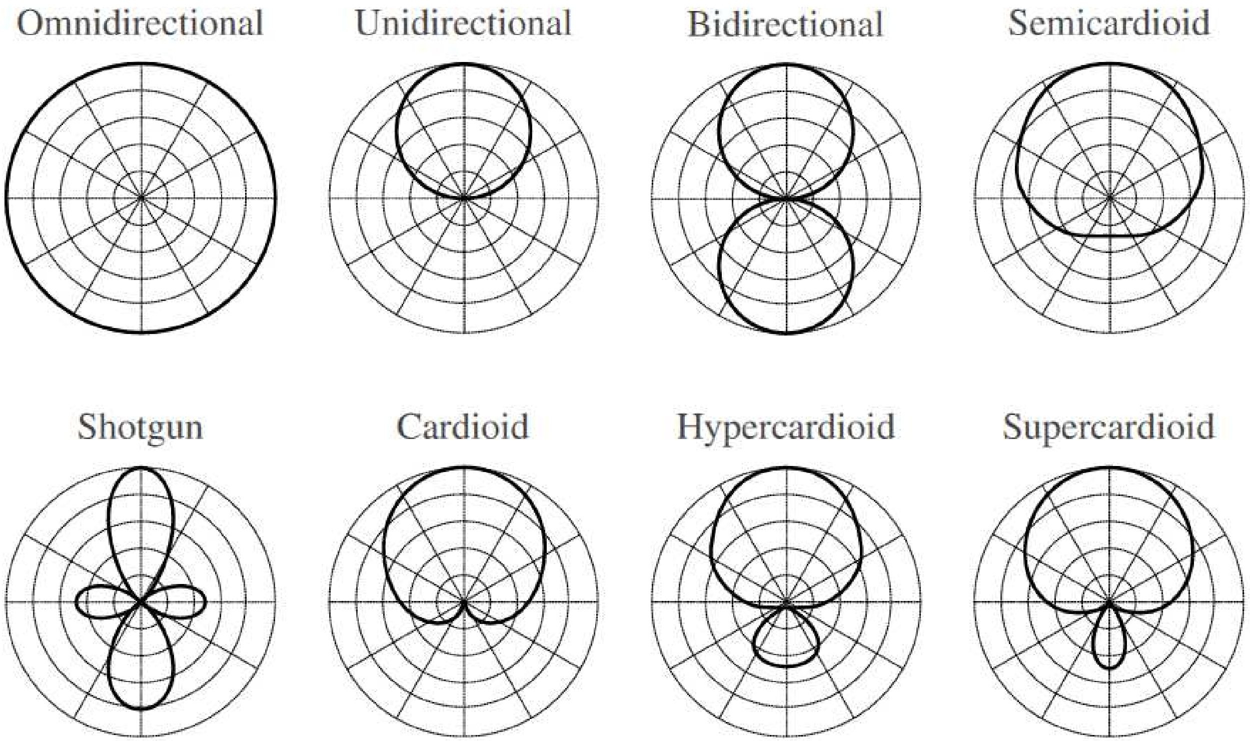








Microphone directivity



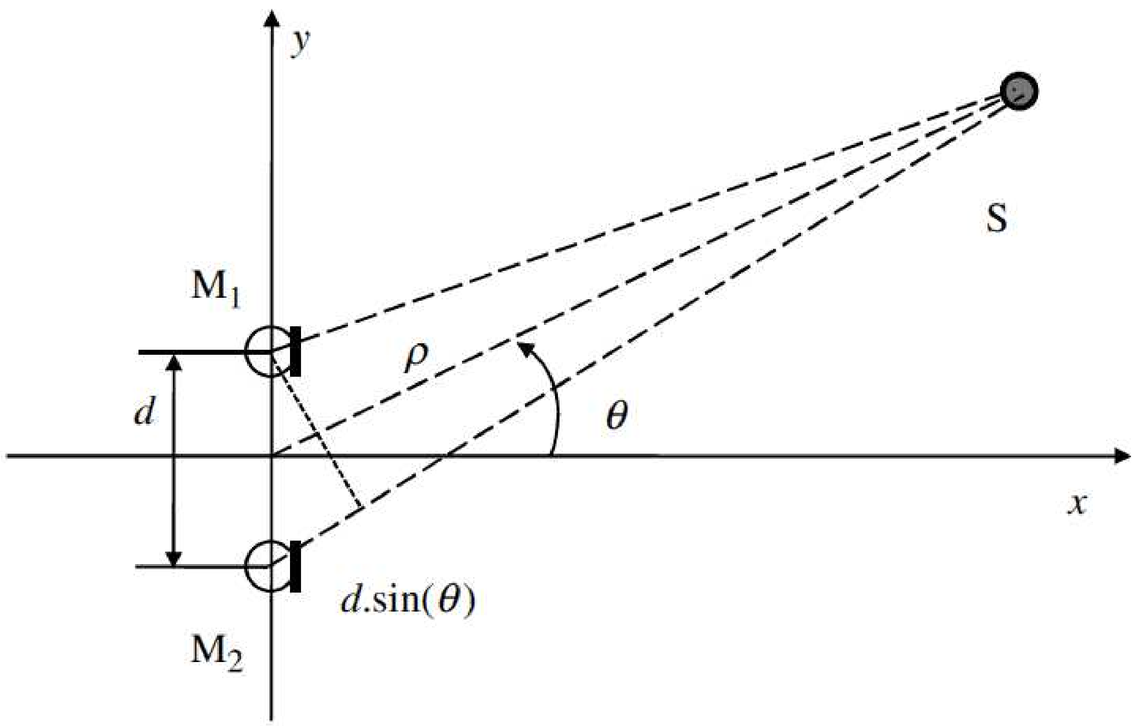
Other Example Application

From "HEAD VISOR - System for Real-Time Identification of Sound Sources"

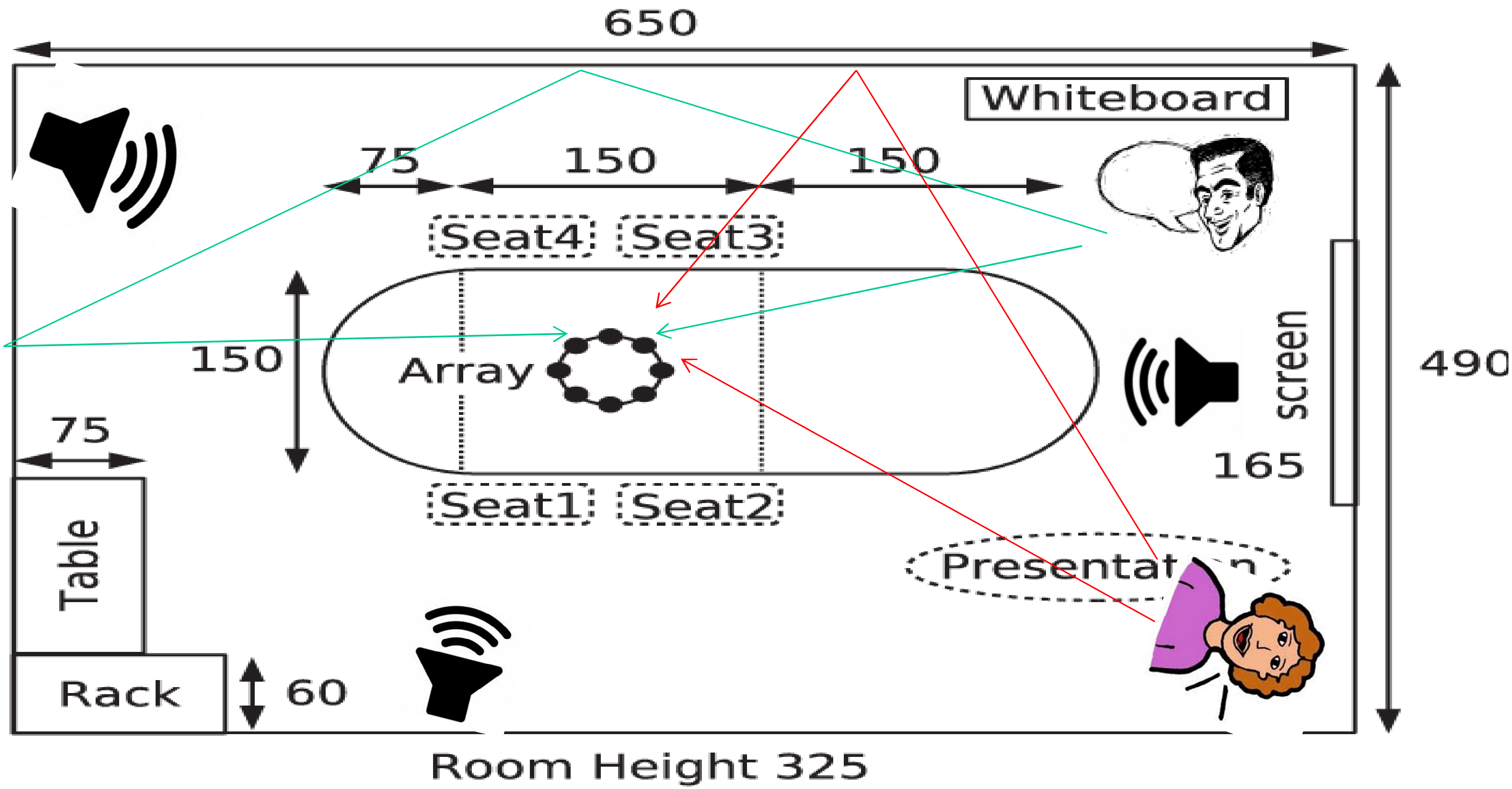
Array

Application to sound emission from car

Multi-sensor and localization



Imagination of the problem



Existing Sound Source Localization Strategies

1) Based on Maximizing Steered Response Power (SRP) of a beamformer.

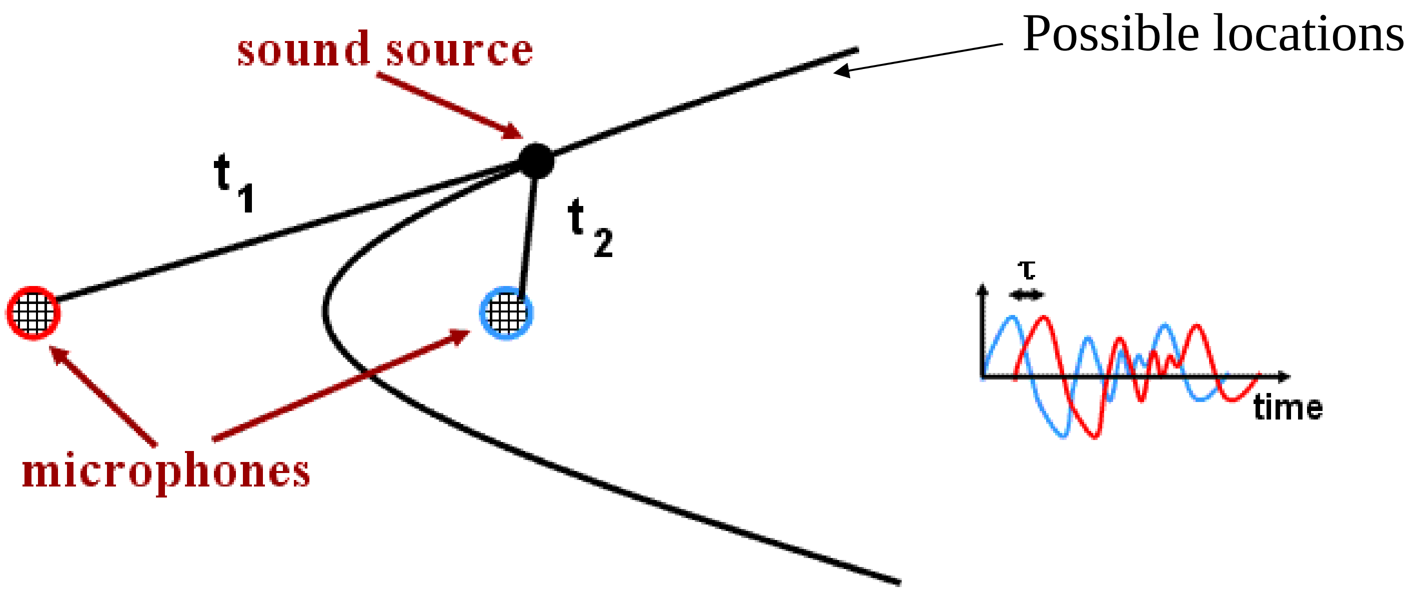
2) Techniques adopting high-resolution spectral estimation concepts.

3) Approaches employing Time Difference of Arrival (TDOA) information.

TDOA Based Locators

- Two-step method
  - TDOA estimation of sound signals between two spatially separated microphones (TDE).
  - Given array geometry and calculated TDOA estimate the 3D location of the source.
- High Quality of TDE is crucial.

Cross Correlation function



Generalized Cross-Correlation -Phase Transform: **GCC-PHAT**
Steered Response power: **SRP-PHAT**

$$R_{12}(\tau) = \frac{1}{2\pi}\int_0^{2\pi} \psi_{12}(\omega) X_1(\omega) X_2^*(\omega) e^{j\omega\tau}\, d\omega$$

$$\psi_{12}(\omega) = \frac{1}{|X_1(\omega) X_2^*(\omega)|}$$

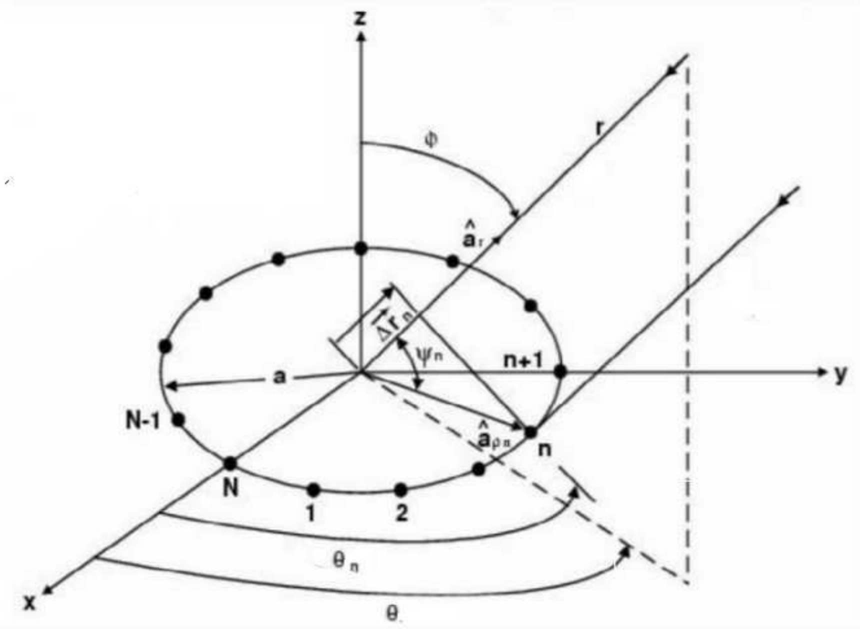
$$\hat{\tau}_{12} = \underset{\tau}{\operatorname{argmax}}\, R_{12}(\tau)$$

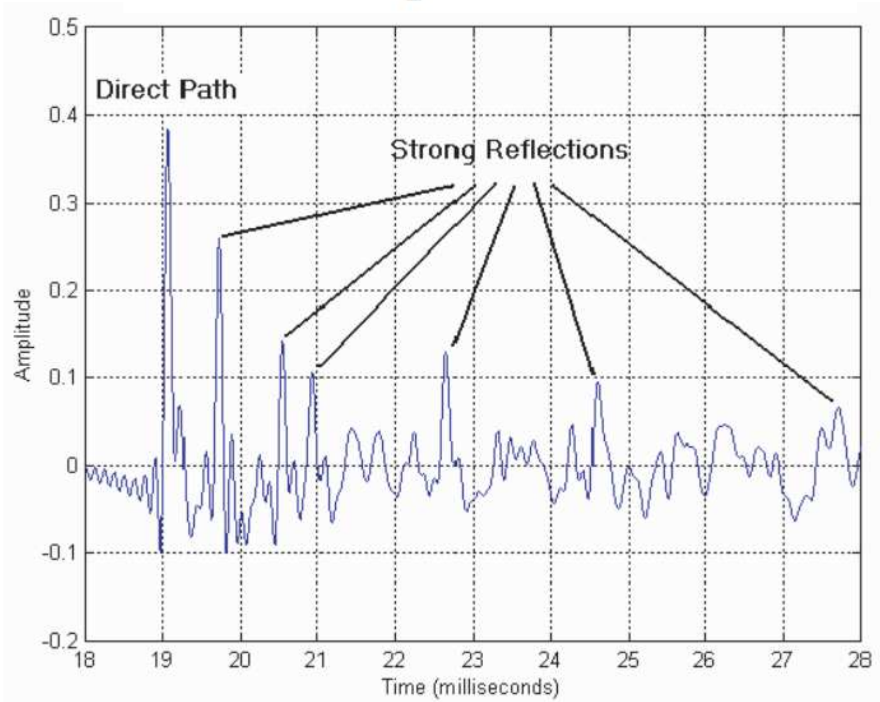
$$\tau_q(p) = \frac{\|p - m_i\| - \|p - m_j\|}{c}$$

$$SRP(p) = 4\pi \sum_{q=1}^{Q} R_q(\tau_q(p))$$

$$p(r, \theta, \varphi)$$

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmax}}\, SRP(\mathbf{p})$$

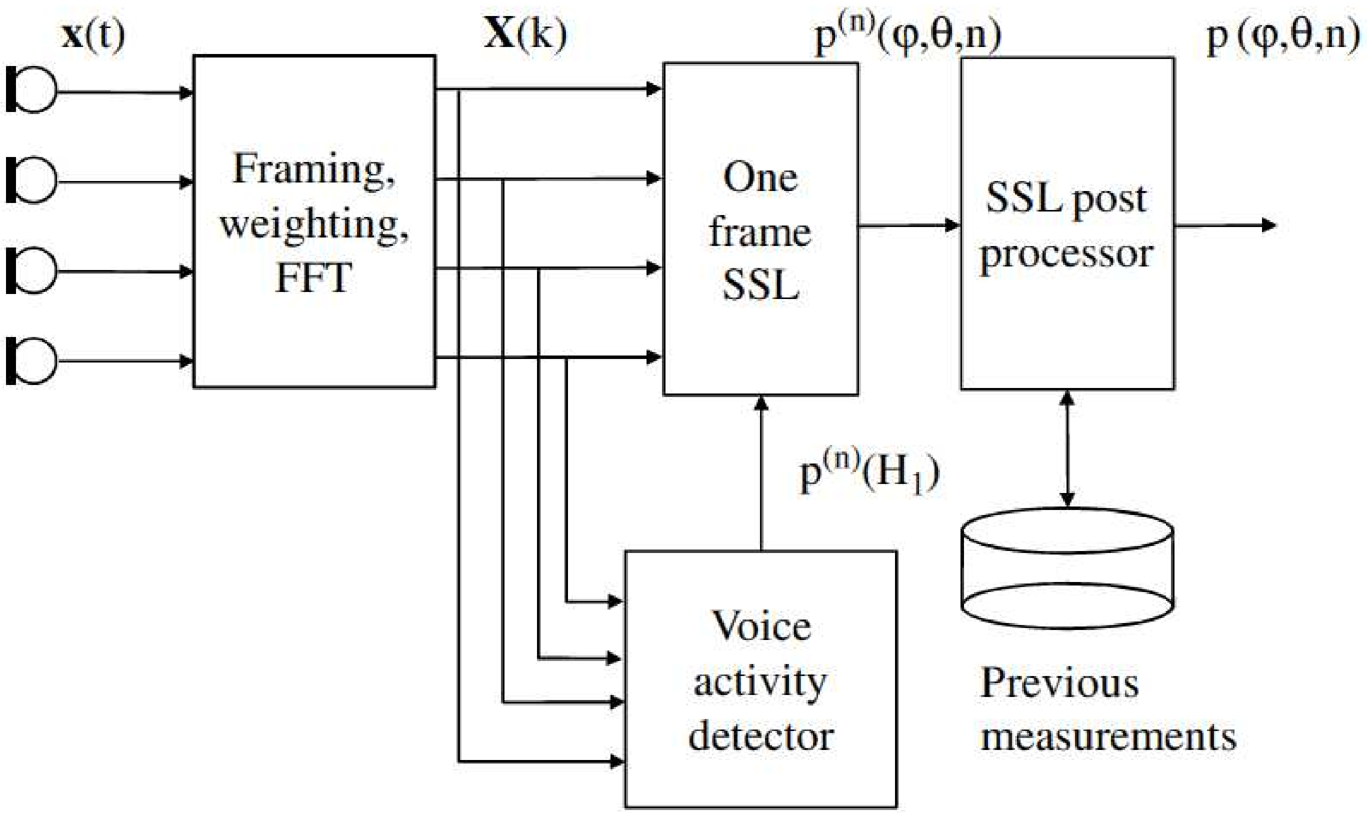




Remarks

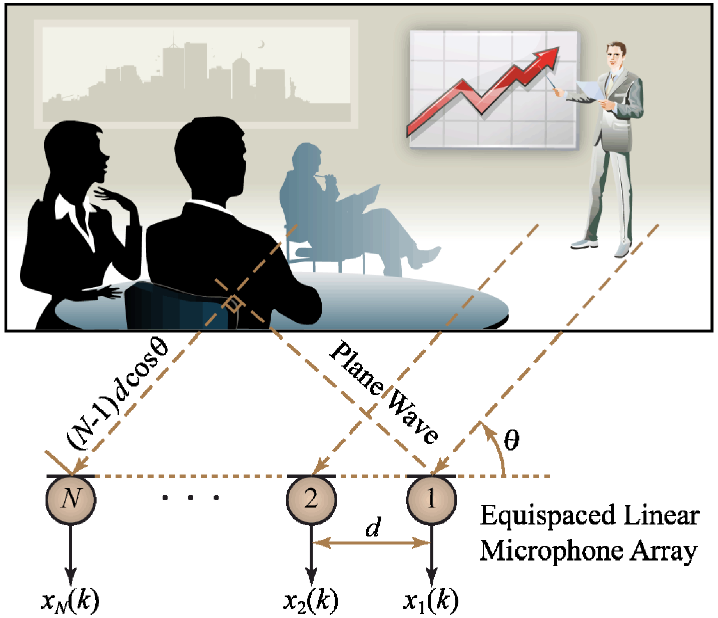
- Use TDOA techniques for real time applications.
- Use Steered-Beamformer strategies in critical applications where robustness is important.

VAD – localize for active source (Tracking)

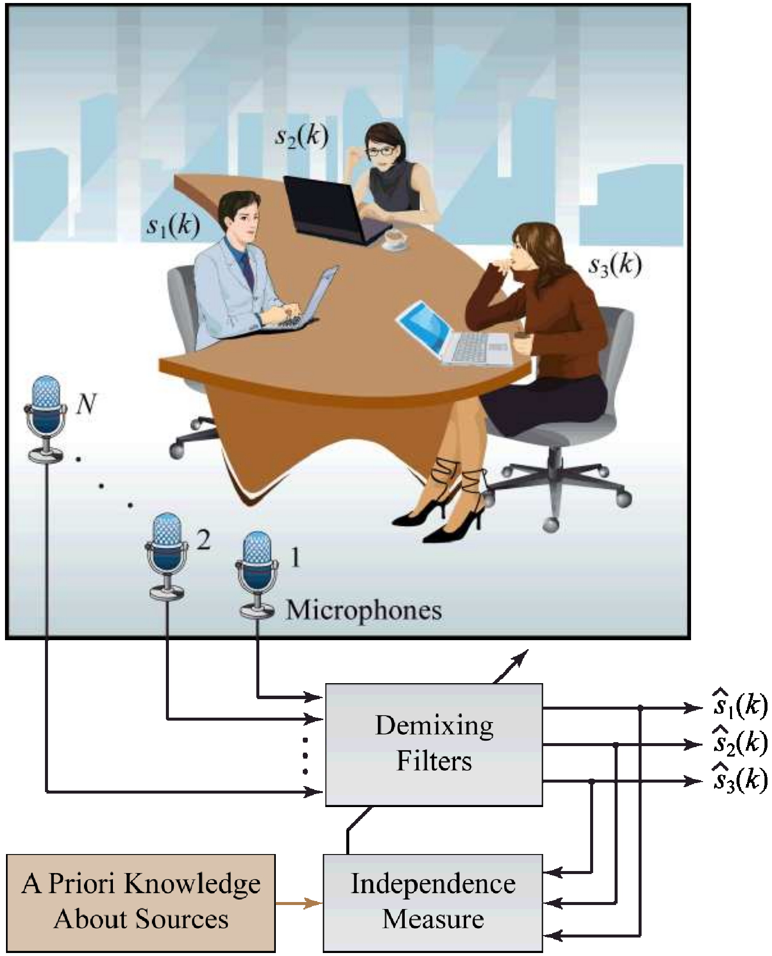


Beamforming vs. Blind Source Separation

Beamforming

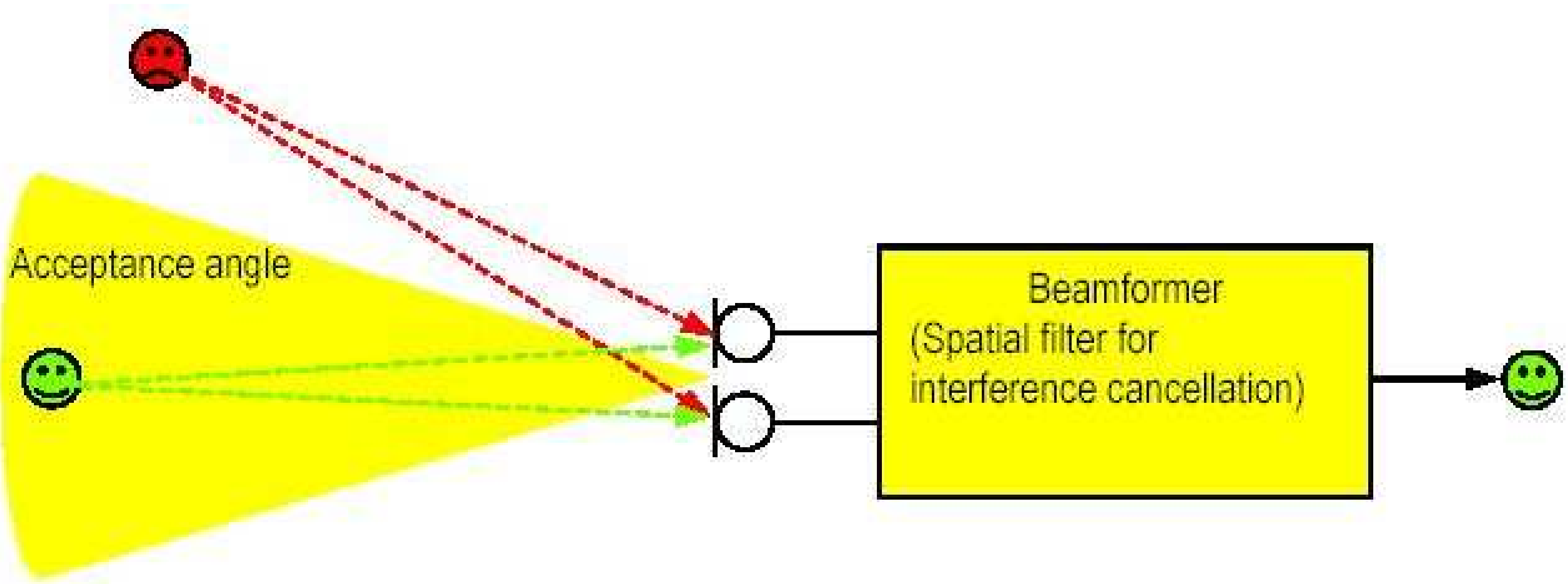


Blind Source Separation

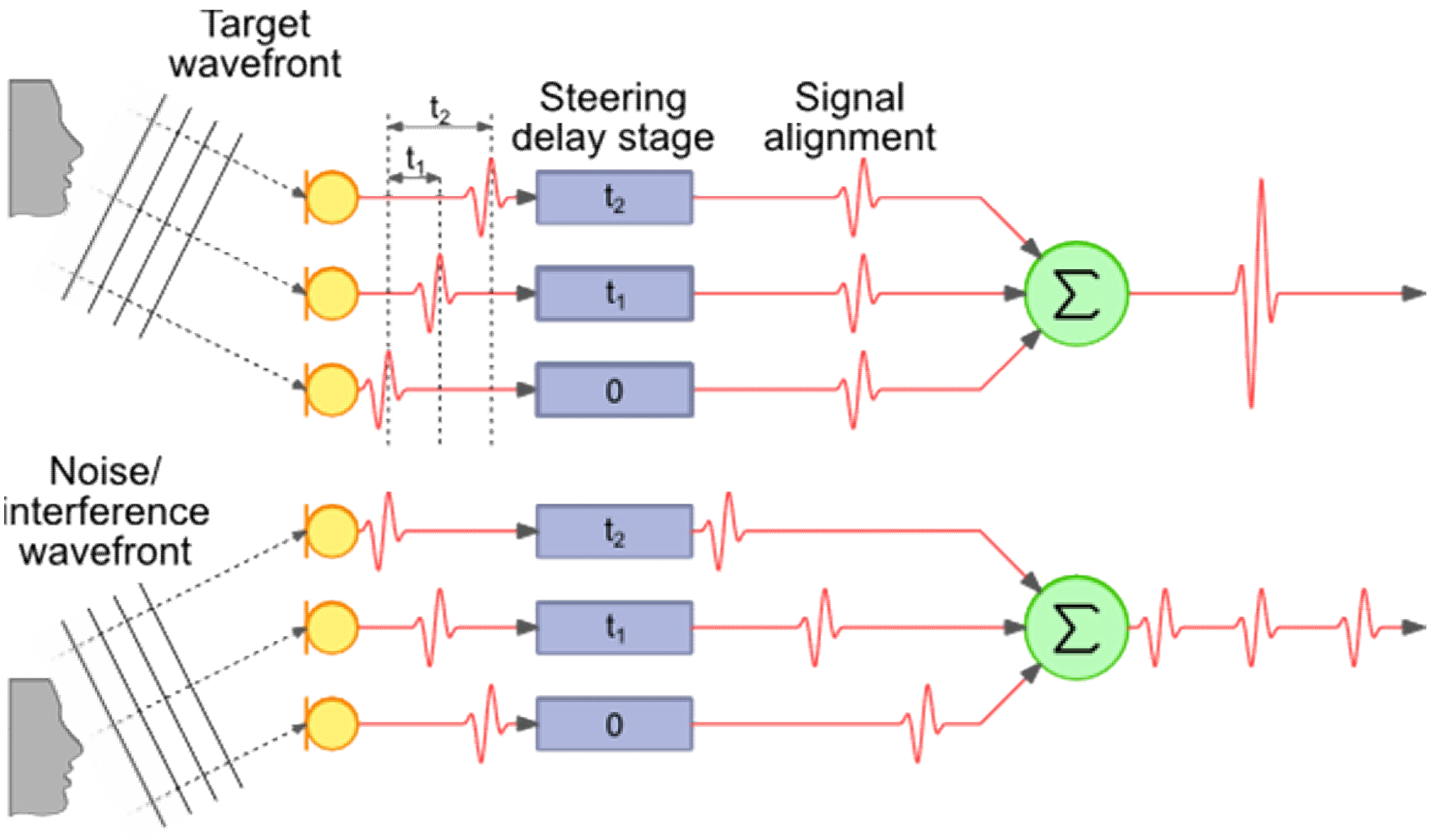


Beamforming

Is one of the simplest and most robust means of Spatial Filtering, i.e., discriminating between signals based on the physical locations of the signal source

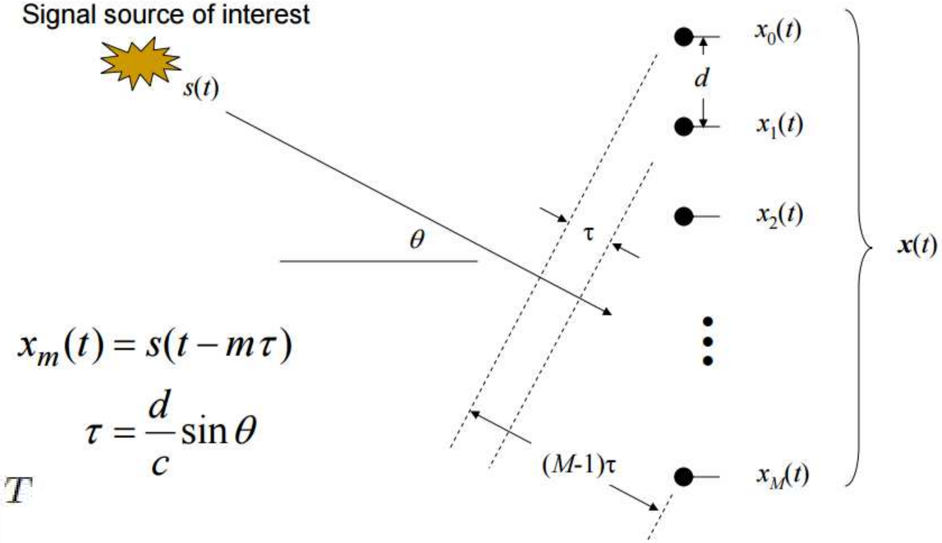


Beam Steering of Beamformer



Courtesy of Andrew Greensted

Delay & Sum beamforming



$$x_m(t) = s(t - m\tau)$$

$$\tau = \frac{d}{c}\sin\theta$$

$$\tau_i = -\mathbf{a}^T \mathbf{m}_i / c$$

$$\mathbf{a} = \begin{bmatrix} \cos\theta \sin\phi & \sin\theta\sin\phi & \cos\phi \end{bmatrix}^T$$

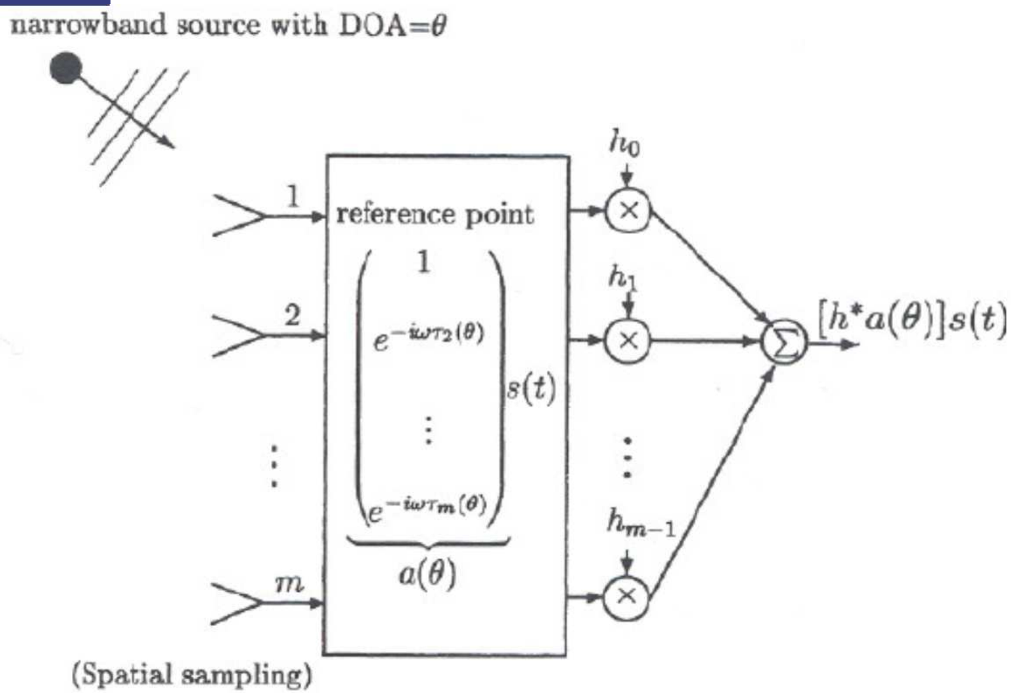
$$\mathbf{v}(\omega) = \begin{bmatrix} e^{-j\omega\tau_1} & \cdots & e^{-j\omega\tau_L} \end{bmatrix}^T$$

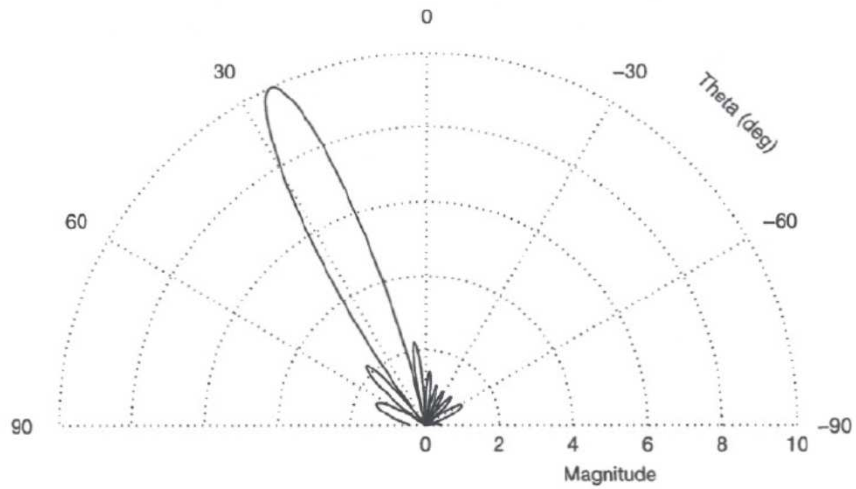
$$\mathbf{X}(\omega) = \begin{bmatrix} X_1(\omega) & \cdots & X_L(\omega) \end{bmatrix}^T \qquad \mathbf{w}(\omega) \triangleq \frac{1}{N}\mathbf{v}(\omega)$$

$$\begin{aligned} Y(\omega) &= \mathbf{w}^H(\omega) \cdot \mathbf{X}(\omega) \\ &= \underbrace{\mathbf{w}^H(\omega) \cdot S(\omega)\mathbf{v}}_{S(\omega)} + \mathbf{w}^H(\omega) \cdot \mathbf{N}(\omega) \end{aligned}$$

Optimization based BF – Goal

- The spatially filtered signal: $x(t) = \left[\mathbf{h}^{*}\mathbf{a}(\theta)\right]s(t)$

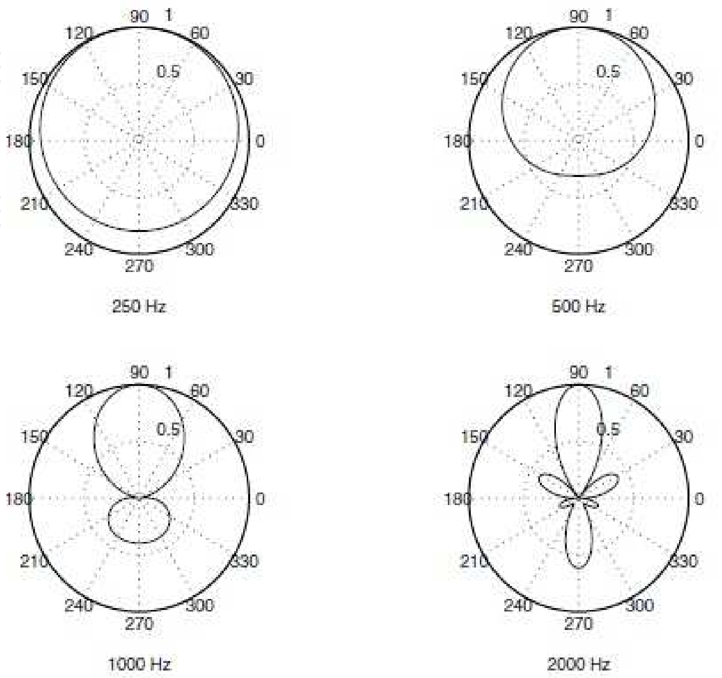




20

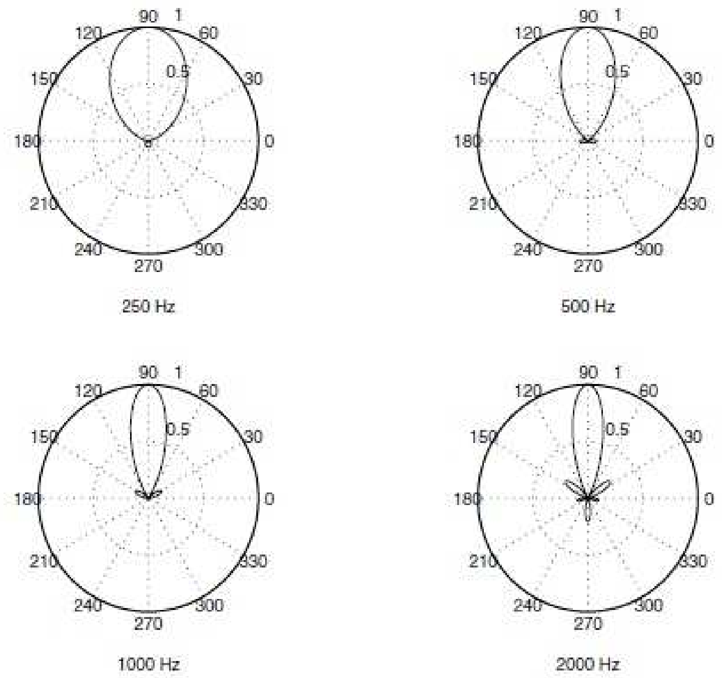
Beampatterns (Gainpattern)

Delay-sum Beamformer



Frequency varying directivity pattern

Superdirective Beamformer



Robust directivity over frequency

Best Linear Unbias Estimate – MVDR BF

$$\underline{x}(t) = \underline{v}(t) * \underline{s}(t) + \underset{\downarrow}{\underline{n}}(t) \qquad \underline{x} \in \mathbb{R}^M$$

$$\mathcal{N}(0, \sigma_n^2) \qquad \underline{v} \in \mathbb{R}^M$$

FFT

$$\forall n : \quad \underline{X}(\omega) = \underline{V}(\omega)\,\underline{S}(\omega) + N(\omega) \qquad \underline{s} \in \mathbb{R}^N$$

$$\underline{n} \in \mathbb{R}^M$$

⇒ The goal is to estimate sources based on observed $\underline{X}$ values at each frame. let's assume that we can have a Linear estimator, such that : $\hat{\underline{S}} = \underline{W}^H \underline{X}$. we try to find the Best Linear Unbiased Estimation $\hat{\underline{S}}$.

$$\text{Unbiased :} \quad \mathbb{E}\{\hat{\underline{S}}\} = \mathbb{E}\{s\} \Rightarrow \underline{W}^H \mathbb{E}\{\underline{X}\} = \mathbb{E}\{\underline{S}\}$$

$$\Rightarrow \underline{w}^H \underline{v}\, \mathbb{E}\{\underline{S}\} = \mathbb{E}\{\underline{S}\} \Longrightarrow \boxed{\underline{w}^H \underline{v} = 1}$$

$$\text{Min\_Variance :} \quad \mathrm{Var}\{\hat{\underline{S}}\} = \mathbb{E}\{(\underline{w}^H \underline{x} - \underline{w}^H \mathbb{E}\{\underline{x}\})(\underline{w}^H \underline{x} - \underline{w}^H \mathbb{E}\{\underline{x}\})^H\}$$

$$= \underline{W}^H \mathbb{E}\{(\underline{x} - \underline{\mu}_x)(\underline{x} - \underline{\mu}_x)^H\}\,\underline{W} = \boxed{\underline{W}^H R_{xx} \underline{W}}$$

$$\Rightarrow \text{Problem: } \underline{W}_{opt} = \underset{\underline{w}}{\operatorname{argmin}}\ \underline{w}^H R_{xx} \underline{w} \quad \text{subject to} \quad \underline{w}^H \underline{v} = 1$$

Minimum Variance Distortionless Response BeamFormer

$$\underline{X} = \underline{V}\,\underline{S} + \underline{N} \rightarrow R_{XX} = \mathbb{E}\{\underline{X}\,\underline{X}^H\}$$

$$\Rightarrow R_{XX} = \mathbb{E}\{(\underline{V}\,\underline{S} + \underline{N})\underbrace{(\underline{V}S + \underline{N})^H}_{(\underline{S}^H\underline{V}^H + \underline{N}^H)}\} = \underline{V}\,R_{SS}\,\underline{V}^H + R_{NN}$$

Signal & Noise are independent $\mathbb{E}\{\underline{S}\,\underline{N}^H\} = \mathbb{E}\{\underline{N}\,\underline{S}^H\} = 0$

$$\Rightarrow \min W^H R_{XX}\underline{W} = \min \underbrace{\underline{W}^H\underline{V}}_{=1} R_{SS} \underbrace{\underline{V}^H\underline{W}}_{=1} + \underline{W}^H R_{NN}\,\underline{W}$$

$$\Rightarrow \quad // = \underbrace{R_{SS}}_{\text{doesn't depend on } \underline{W}} + W^H R_{NN}\,\underline{W}$$

doesn't depend on $\underline{W}$

$$\Rightarrow \text{Problem: } \boxed{\underline{W}_{MVDR} = \arg\min_{\underline{W}} \underline{W}^H R_{NN}\underline{W} \quad \text{subject to} \quad \underline{W}^H\underline{V} = 1}$$

Noise Coherence Matrix-isotrpoic field

When noise field is homogeneous (isotropic)

$$R_{nn} = \Phi_{nn} \Gamma_{nn} \quad ; \quad \Phi_{nn} = \text{noise power}$$

$$\underset{\sim}{\Gamma}_{nn} = \text{noise coherence matrix}$$

$$\overset{m_i}{S_n}(\omega) = \overset{m_j}{S_n}(\omega) \rightarrow \quad S_{ij}(\omega) = S_i(\omega)\, e^{-j\frac{\omega}{c} d\cos\phi}$$

$$\begin{cases} x_j = x_i\left(t - \frac{\Delta}{c}\right) \\ \Delta = d\cos\phi \\ \tau = \frac{\Delta}{c} \quad c \rightarrow \text{speed of wave} \end{cases}$$

Spatial coherence:

$$\gamma_{ij}(\omega) = \frac{\oint_A S_{ij}(\omega)\, dA}{\oint_A \sqrt{S_i(\omega) S_j(\omega)}\, dA}$$

$$\Rightarrow \gamma_{ij}(\omega) = \frac{1}{A} \oint_A e^{-j\omega/c\, d\cos\phi}\, dA \xrightarrow[\text{isotropic}]{\text{Spherically}} dA = r^2 \sin\phi\, d\phi\, d\theta$$

$$A = 4\pi r^2, \ \phi \in [0, \pi], \theta \in [0, 2\pi)$$

$$\Rightarrow \gamma_{ij}(\omega) = \frac{1}{4\pi r^2} \int_0^{2\pi} \int_0^{\pi} e^{-j\frac{\omega}{c} d\cos\phi}\, r^2 \sin\phi\, d\phi\, d\theta = \boxed{\frac{\sin(\omega d/c)}{\omega d/c}} = \left(\Gamma_{nn}\right)_{i,j}(\omega)$$

Also, $d = \frac{\| m_i - m_j \|}{c}$, m_i : microphone i

Super Directive BF

Therefore, in a spherically isotropic noise field:

$$\underset{\sim}{R}_{nn} = \Phi_{nn} \underset{\sim}{\Gamma}_{nn} \quad , \text{where} \quad \underset{\sim}{\Gamma}_{nn} = \begin{bmatrix} 1 & \Gamma_{0,1} & \Gamma_{0,2} & \cdots & \Gamma_{0,M-1} \\ \Gamma_{1,0} & 1 & \Gamma_{12} & \cdots & \Gamma_{1,M-1} \\ \vdots & & \ddots & & \vdots \\ \Gamma_{M-1,0} & \cdots & & & 1 \end{bmatrix}$$

$$\left(\underset{\sim}{\Gamma}_{nn}\right)_{ij}(\omega) = \mathrm{Sinc}\left(\omega \frac{\|m_i - m_j\|}{c}\right)$$

Super-Directive BF:

$$\underline{W}_{SDBF} = \underset{\underline{W}}{\arg\min}\ \underline{W}^H \underset{\sim}{\Gamma}_{nn} \underline{W} \quad \text{subject to } \underline{W}^H \underline{V} = 1$$

Lagrange multiplier:

$$\boxed{\underline{W}_{SDBF} = \frac{\Gamma_{nn}^{-1}\, \underline{V}}{\underline{V}^H \Gamma_{nn}^{-1}\, \underline{V}}}$$

Super-Directive BF is the best BLUE BF in isotropic noise field.

Calculation --optional

$$L = \underline{W}^H \Gamma_{nn} \underline{W} - \lambda(\underline{w}^H \underline{v} - 1)$$

$$\nabla_{w} L = 2\,\Gamma_{nn} \underline{W} - 2\lambda \underline{v} = 0 \Rightarrow \underline{W} = \lambda \Gamma_{nn}^{-1} \underline{v}$$

$$\nabla_{\lambda} L = 0 \rightarrow \underline{W}^H \underline{v} = 1 \rightarrow \lambda \underline{v}^H \Gamma_{nn}^{-1} \underline{v} = 1$$

Γ_{nn} → Hermition matrix

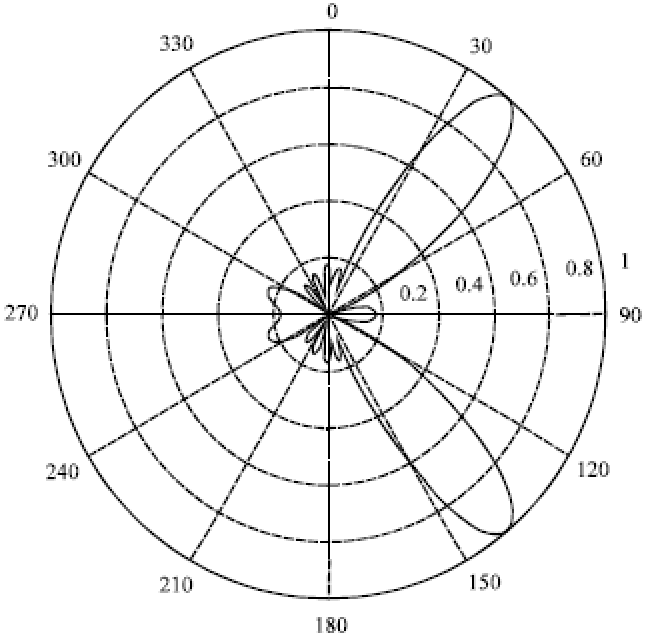
$$A^H = A$$

$$\Rightarrow \lambda = \frac{1}{v^H \Gamma_{nn}^{-1} v}$$

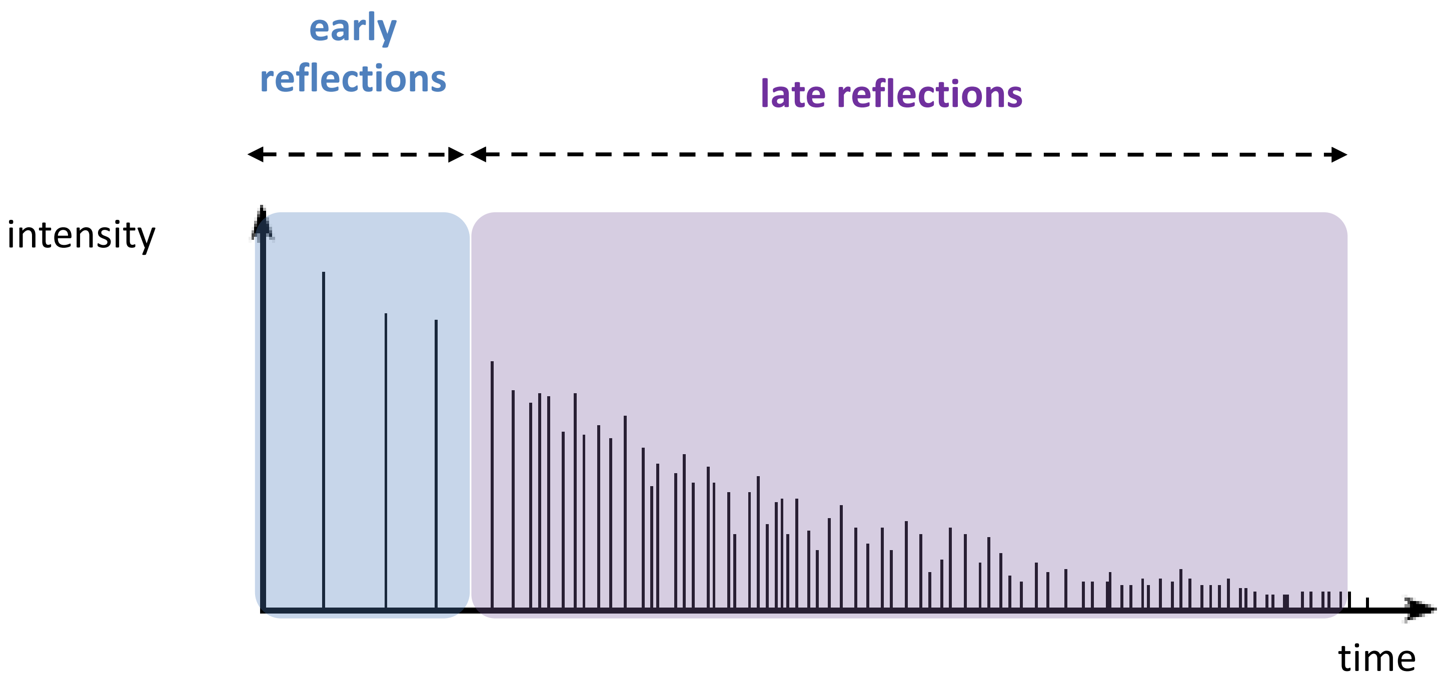
$$\underline{W} = \frac{\Gamma_{nn}^{-1} \underline{v}}{\underline{v}^H \Gamma_{nn}^{-1} \underline{v}}$$

Linear Constrained Minimum Variance BF

$$\mathbf{W}^H = \Gamma_{nn}^{-1}\left[V_1\ V_2\right]\left(\left[V_1\ V_2\right]^H \Gamma_{nn}^{-1}\left[V_1\ V_2\right]\right)^{-1}\left[\begin{array}{c} 1 \\ 0 \end{array}\right]$$



Room Impulse Response



Beamforming problems

- Very sensitive to array geometry, need good calibration
- Has only directivity, no selectivity in range or other location parameters
- Frequency response is not flat
- Ambient noises are assumed to be spatially white
- Beam width (or selectivity) depends on the size of the array
- Spatial aliasing problem (For high frequencies- sampling ambiguity)

Thank you

Directivity of microphone arrays
(omni directional microphones)