# Precedence Effect

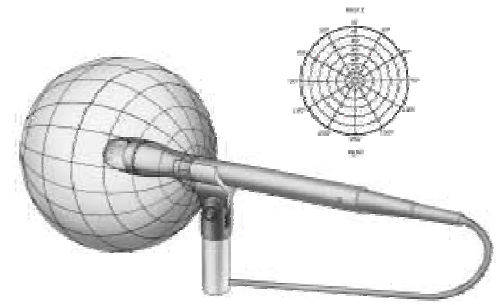
# Beamforming

# Demo of the Franssen effect

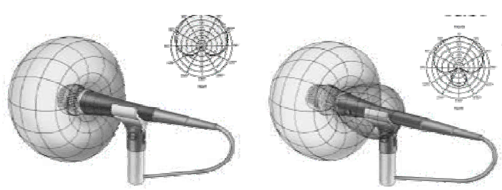
- Demonstrates precedence

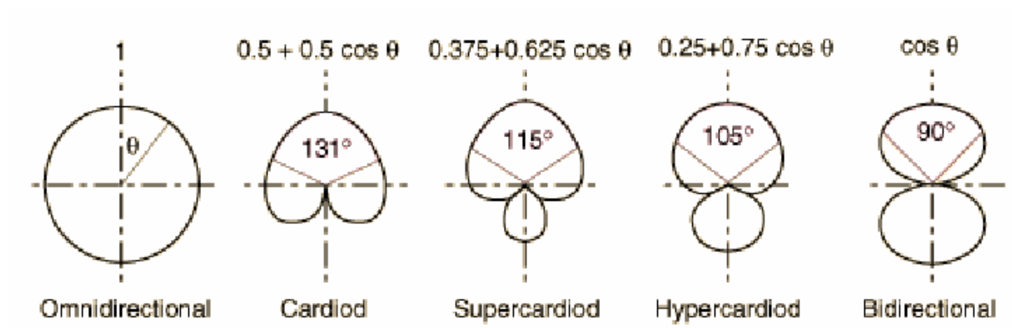
# Introduction to 3D Audio (capture)

- Directivity of microphone.
  - Omni-directional
  - Advantages are that microphones capture all sound including that of interest



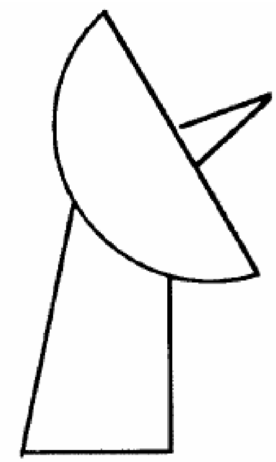
  - Directional
  - Capture sound from a preferred direction

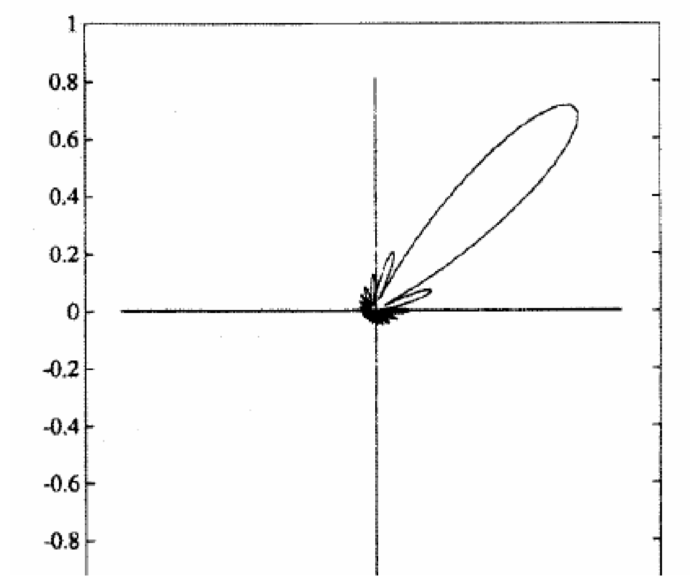


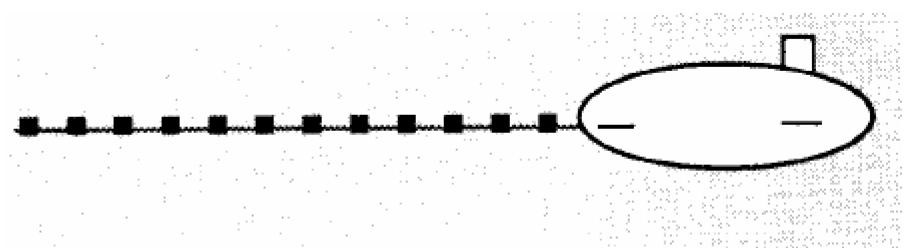


# Beamforming

- Given *N* microphones combine their signals in a way that some desired result occurs
- Word arises from the use of parabolic reflectors to form pencil “beams” for broadcast and reception
- Alternate word: “spatial filtering”
- Towed array and fixed array sonars





# Delay and Sum Beamforming

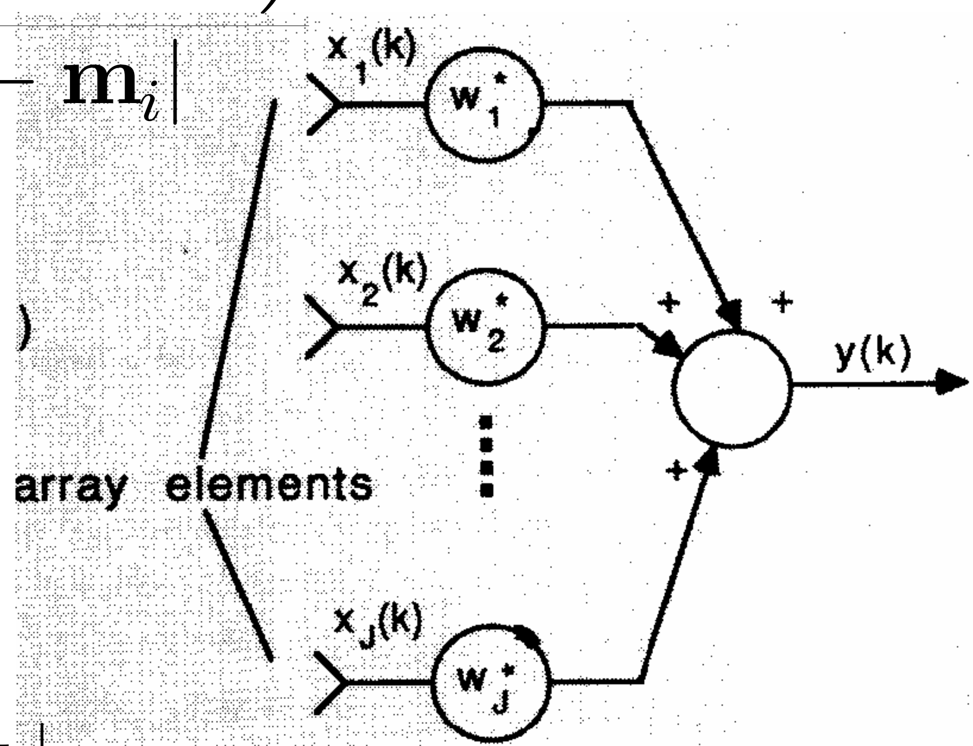
- If the source location is known, delays relative to the microphone can be obtained
- Signal $x$ at location $\mathbf{s}$ arrives at microphone $\mathbf{m}_i$ as

$$\frac{x\left(t - \frac{|\mathbf{s}-\mathbf{m}_i|}{c}\right)}{|\mathbf{s}-\mathbf{m}_i|}$$

- Signals at microphones can be appropriately delayed and weighted.
- Output signal is

$$y(k) = \frac{1}{N}\sum_{l=1}^{N} w_l^* x_l(k-\Delta_l)$$

$$\Delta_l = |\mathbf{s}-\mathbf{x}_l|/c \quad w_l = 1/|\mathbf{s}-\mathbf{x}_l|$$



# Behavior of simple beamformer

- Usually source is assumed to be far away.
  - Weights are approximately the same in this case
- Signal from source direction adds in phase
  - So the signal is amplified $N$ times
- Signals from other directions will add up with random phase and the power will decrease by a factor of 1/N
- Directivity index is a measure of the gain of the array in the look direction (location of the delays) in decibels
  - For $N$ microphones $10 \log_{10} (N)$
- Requires an ability to store the signal (at least for max $\{\Delta_l\}$
- Jargon: "taps" number of samples in time that are stored

- Data independent beamforming:
  - Weights are fixed
- Data dependent (adaptive)
  - Weights change according to the data
- Simple example:
  - Fixed: Delay and sum looking at a particular point (direction)
  - Adaptive: Delay and sum looking at a particular moving source

# More general beamforming

- Suppose we want to take advantage of the stored data
- Write the beamformer output as

$$y(k) = \sum_{l=1}^{N} \sum_{m=k-M}^{k} w_{lm}^* x_l(k-m)$$

- Can be written as $\mathbf{y}=\mathbf{w}^H \mathbf{x}$
- Take Fourier transform of the weights and the signal

# Speech and Audio Processing

# Microphone Array Processing

Slides adapted from those of
Marc Moonen/Simon Doclo

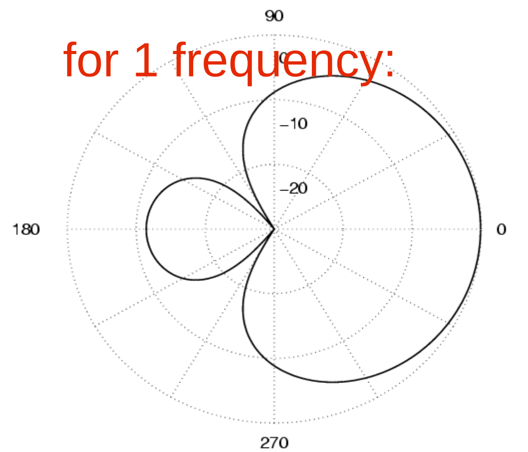
Dept. E.E./ESAT, K.U.Leuven

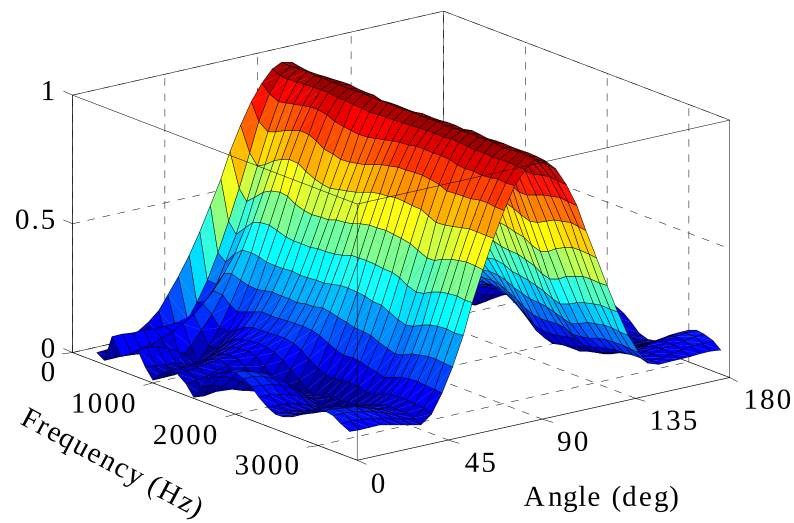
www.esat.kuleuven.ac.be/~moonen/

# Introduction

- **Each microphone is characterized by a `directivity pattern' which specifies the gain (& phase shift) that the microphone gives to a signal coming from a certain direction (`angle-of-arrival').**
- **Directivity pattern is a function of angle-of-arrival and frequency**
- **Directivity pattern is a (physical) microphone design issue.**

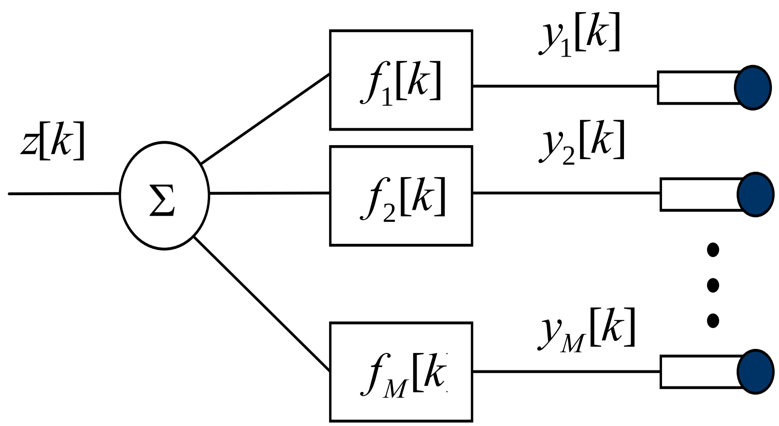
for 1 frequency:



# Introduction

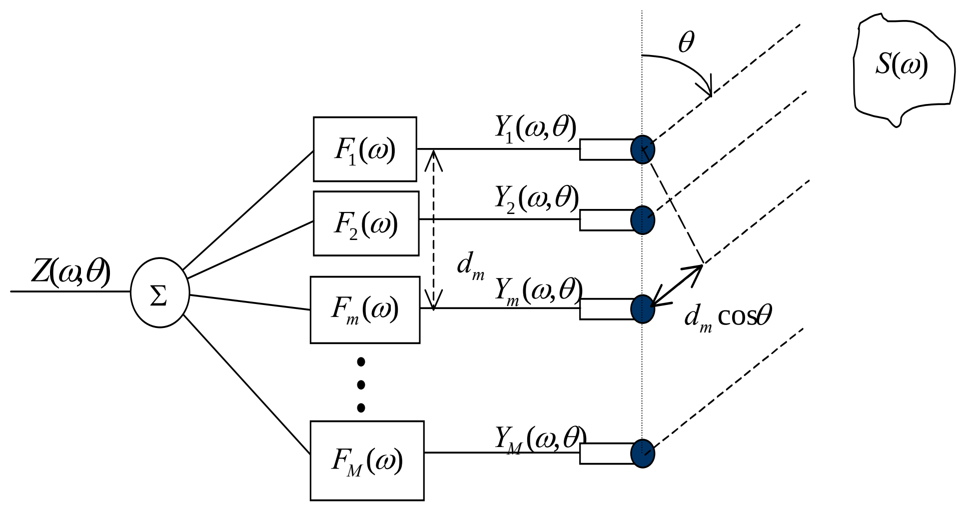
- **By weighting/filtering and summing signals from different microphones, a `virtual' directivity pattern may be produced**




- **This is `spatial filtering' and `spatial filter design', based on given microphone characteristics (with correspondences to traditional (spectral) filter design)**
- **Applications: teleconferencing, hands-free telephony, hearing aids, voice-controlled systems, …**

# Introduction

- **An important aspect is that different microphones in a microphone array are in different positions/locations, hence receive different signals**
- **Example : linear array, with uniform inter-microphone distances, under far-field (plane waveforms) conditions. Each microphone receives the same signal, but with different delays.**




- **Hence `spatial filter design' based on microphone characteristics + microphone array configuration.**

  **Often simple assumptions are made, e.g. microphone gain = 1 for all frequencies and all angles.**

# Introduction

- **Background/history: ideas borrowed from antenna array design/processing for RADAR & (later) wireless comms.**

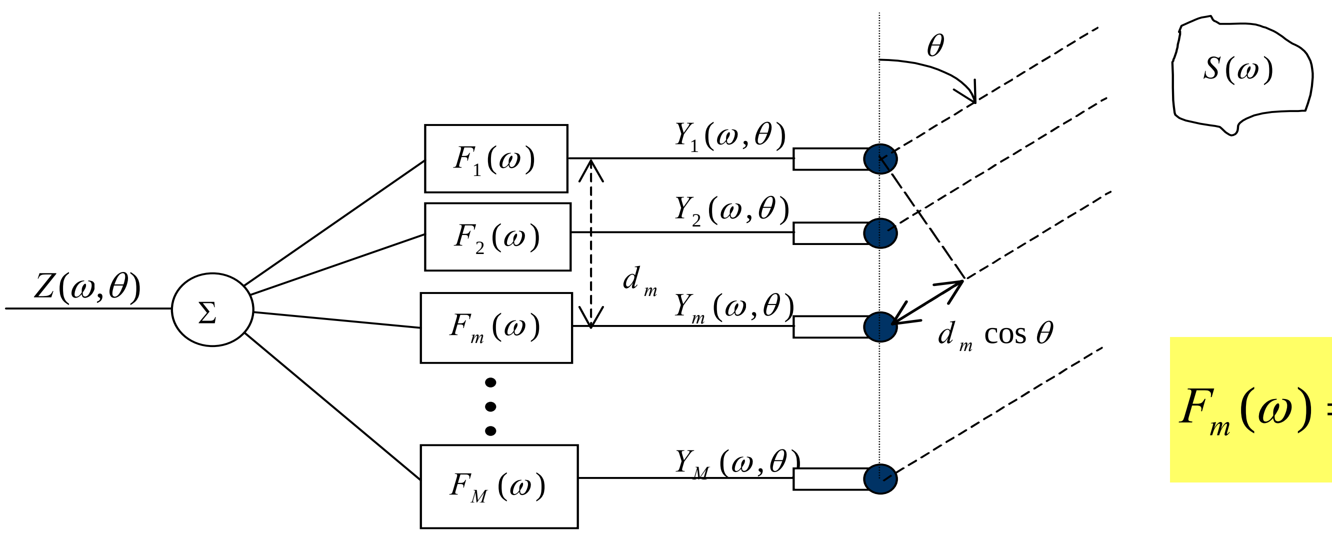
- **Microphone array processing considerably more difficult than antenna array processing:**
  - narrowband radio signals versus broadband audio signals
  - far-field (plane wavefronts) versus near-field (spherical wavefronts)
  - pure-delay environment versus multi-path reverberant environment

- **Classification:**
  - <u>fixed beamforming</u>: data-independent, fixed filters $f_m[k]$
    *e.g.* delay-and-sum, weighted-sum, filter-and-sum
  - <u>adaptive beamforming</u>: data-dependent, adaptive filters $f_m[k]$
    *e.g.* LCMV-beamformer,

# Beamforming basics

## General form: **filter-and-sum beamformer**

- linear microphone array with $M$ microphones and inter-micr. distance $d_m$
- Microphone gains are assumed to be equal to 1 for all freqs./angles
  (otherwise, this characteristic is to be included in the steering vector, see next page)
- source $S(\omega)$ at angle $\theta$ *(far-field, no multipath)*
- filters $f_m[k]$ with filter length $L$




$$F_m(\omega) = \sum_{k=0}^{L-1} f_m[k] e^{-jk\omega}$$

**Terminology: `Broadside' direction: $\theta = 90^o$ `End-fire' direction: $\theta = 0^o$**

# Near-field beamforming

- **Far-field assumptions not valid for sources close to microphone array**
  - spherical wavefronts instead of planar waveforms
  - include attenuation of signals
  - 3 spherical coordinates $\theta,\phi,r$ (=position **q**) instead of 1 coordinate $\theta$
- **Different steering vector:**

$$\mathbf{d}(\omega,\theta) \longrightarrow \mathbf{d}(\omega,\mathbf{q}) = \begin{bmatrix} a_1 e^{-j\omega\tau_1(\mathbf{q})} & a_2 e^{-j\omega\tau_2(\mathbf{q})} & \dots & a_M e^{-j\omega\tau_M(\mathbf{q})} \end{bmatrix}^T$$

$$a_m = \frac{\left\|\mathbf{q}-\mathbf{p}_{ref}\right\|}{\left\|\mathbf{q}-\mathbf{p}_m\right\|} \qquad \tau_m(\mathbf{q}) = \frac{\left\|\mathbf{q}-\mathbf{p}_{ref}\right\| - \left\|\mathbf{q}-\mathbf{p}_m\right\|}{c} f_s$$

with **q** position of source

$\mathbf{p}_{ref}$ position of reference microphone

$\mathbf{p}_m$ position of $m^{th}$ microphone

# Beamforming basics

Data model:

- **<u>Microphone signals</u> are delayed versions of $S(\omega)$**

$$Y_m(\omega,\theta) = e^{-j\omega\tau_m(\theta)}.S(\omega)$$

$$y_m[k] = s[k-\tau_m(\theta)] \qquad \tau_m(\theta) = \frac{d_m \cos\theta}{c} f_s$$

- **<u>Stack</u> all microphone signals in a vector**

$$\mathbf{Y}(\omega,\theta) = \mathbf{d}(\omega,\theta).S(\omega) \qquad \mathbf{d}(\omega,\theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_2(\theta)} & \ldots & e^{-j\omega\tau_M(\theta)} \end{bmatrix}^T$$

**d is `steering vector'**

- **<u>Output signal</u> $Z(\omega,\theta)$ is**

$$Z(\omega,\theta) = \sum_{m=1}^{M} F_m^*(\omega) Y_m(\omega,\theta) = \mathbf{F}^H(\omega)\cdot\mathbf{Y}(\omega,\theta)$$

# Beamforming basics

## Data model:

- **<u>Microphone signals</u> are corrupted by additive noise**

$$y_m[k] = s[k - \tau_m(\theta)] + n_m[k]$$

- **<u>Stack</u> all noise signals in a vector**

$$\mathbf{N}(\omega) = \begin{bmatrix} N_1(\omega) & N_2(\omega) & \ldots & N_M(\omega) \end{bmatrix}^T$$

- **Define <u>noise correlation matrix</u> as**

$$\mathbf{\Phi}_{NN}(\omega) = E\{\mathbf{N}(\omega).\mathbf{N}(\omega)^H\}$$

- **We assume noise field is <u>homogeneous</u>, i.e. diagonal elements of** $\mathbf{\Phi}_{NN}(\omega)$

$$\Phi_{ii}(\omega) = \Phi_{noise}(\omega) \quad , \quad \forall i$$

- **Then** <u>noise coherence matrix</u> **is**

$$\mathbf{\Gamma}_{NN}(\omega) = \frac{1}{\phi_{noise}(\omega)}.\mathbf{\Phi}_{NN}(\omega)$$

# Beamforming basics

Definitions:

- **<u>Spatial directivity pattern</u>: `transfer function' for source at angle $\theta$**

$$H(\omega,\theta)=\frac{Z(\omega,\theta)}{S(\omega)}=\sum_{m=1}^{M}F_m^*(\omega)e^{-j\omega\tau_m(\theta)}=\mathbf{F}^H(\omega)\cdot\mathbf{d}(\omega,\theta)$$

- **<u>Steering direction</u> $\theta_{max}$ = angle $\theta$ with maximum amplification (for 1 freq.)**

- **<u>Beamwidth</u> = region around $\theta_{max}$ with amplification > -3dB (for 1 freq.)**

- **<u>Array Gain</u> = improvement in SNR**

$$G(\omega,\theta)=\frac{SNR_{Output}}{SNR_{Input}}=\frac{\left|\mathbf{F}^H(\omega)\cdot\mathbf{d}(\omega,\theta)\right|^2}{\mathbf{F}^H(\omega)\cdot\mathbf{\Gamma}_{NN}(\omega)\cdot\mathbf{F}(\omega)}$$

# Beamforming basics

## Definitions:

- **Array Gain = improvement in SNR**

$$G(\omega,\theta)=\frac{SNR_{Output}}{SNR_{Input}}=\frac{\left|\mathbf{F}^H(\omega)\cdot\mathbf{d}(\omega,\theta)\right|^2}{\mathbf{F}^H(\omega)\cdot\mathbf{\Gamma}_{NN}(\omega)\cdot\mathbf{F}(\omega)}$$

- **Directivity = array gain for $\theta_{max}$ and diffuse noise (=coming from all directions)**

$$DI(\omega)=\frac{\left|\mathbf{F}^H(\omega)\cdot\mathbf{d}(\omega,\theta_{\max})\right|^2}{\mathbf{F}^H(\omega)\cdot\mathbf{\Gamma}_{NN}^{diffuse}(\omega)\cdot\mathbf{F}(\omega)}$$

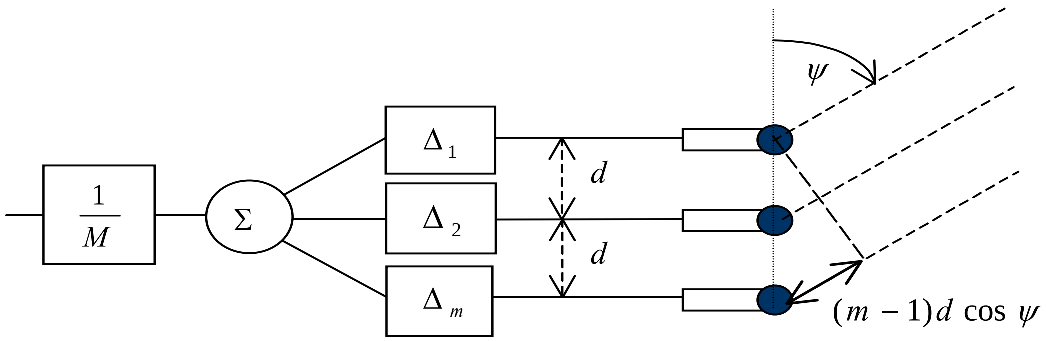
- **White Noise Gain = array gain for $\theta_{max}$ and spatially uncorrelated noise ($\Gamma_{NN}=\mathrm{I}$) (e.g. sensor noise)**

$$WNG(\omega)=\frac{\left|\mathbf{F}^H(\omega)\cdot\mathbf{d}(\omega,\theta_{\max})\right|^2}{\mathbf{F}^H(\omega)\cdot\mathbf{F}(\omega)}$$

**ps: often used as a measure for robustness**

# Delay-and-sum beamforming

- **Microphone signals are delayed and summed together**
  **Array can be virtually steered to angle $\psi$**




$$z[k] = \frac{1}{M} \cdot \sum_{m=1}^{M} y_m[k + \Delta_m]$$

$$F_m(\omega) = \frac{e^{-j\omega\Delta_m}}{M}$$

$$\Delta_m = \frac{d_m \cos\psi}{c} f_s$$

- **Angular selectivity is obtained, based on constructive (for $\theta = \psi$) and destructive (for $\theta \neq \psi$) interference**

**For $\theta = \psi$, this is referred to as a `matched filter' :**

$$\mathbf{F}(\omega) = \frac{\mathbf{d}(\omega,\psi)}{M}$$

$$H(\omega, \theta = \psi) = 1$$

- **For <u>uniform</u> linear array :**

$$d_m = (m-1)d \qquad \Delta_m = (m-1)\Delta$$

- **PS:** $H(\omega,\theta) = H(\omega,-\theta)$ **plain!) (if microphone characteristics are ignored)**

# Delay-and-sum beamforming

- **Spatial directivity pattern $H(\omega,\theta)$ for <u>uniform</u> DS-beamformer**

$$H(\omega,\theta) = \sum_{m=1}^{M} e^{-j(m-1)\omega \frac{d(\cos\theta - \cos\psi)}{c} f_s} = \frac{e^{-jM\gamma/2}\sin(M\gamma/2)}{e^{-j\gamma/2}\sin(\gamma/2)}$$

$\gamma$

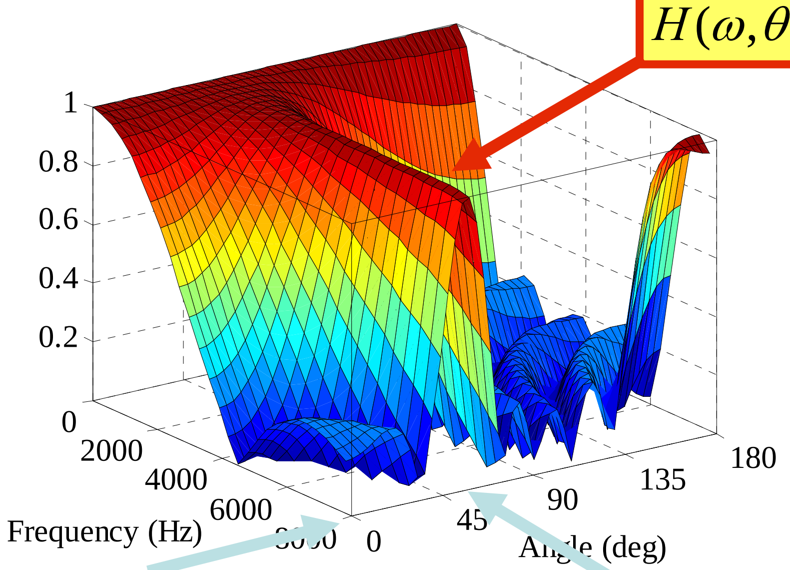
*M*=5 microphones

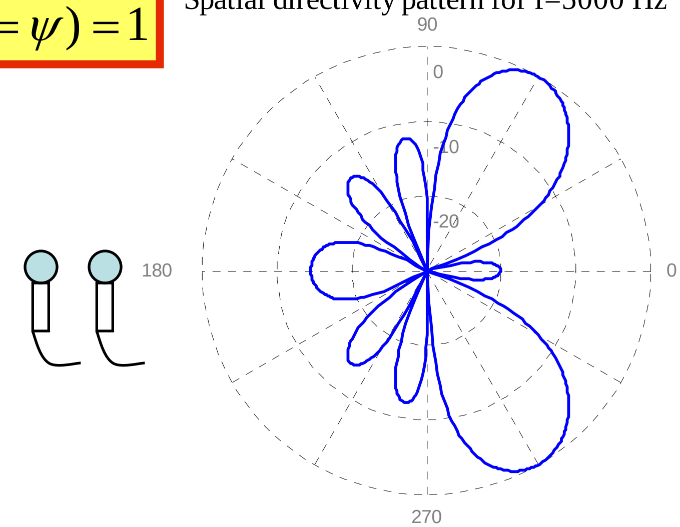
*d*=3 cm inter-microphone distance

$\psi$=60° steering angle

$f_s$=16 kHz sampling frequency

- **$H(\omega,\theta)$ has sinc-like shape and is frequency-dependent**



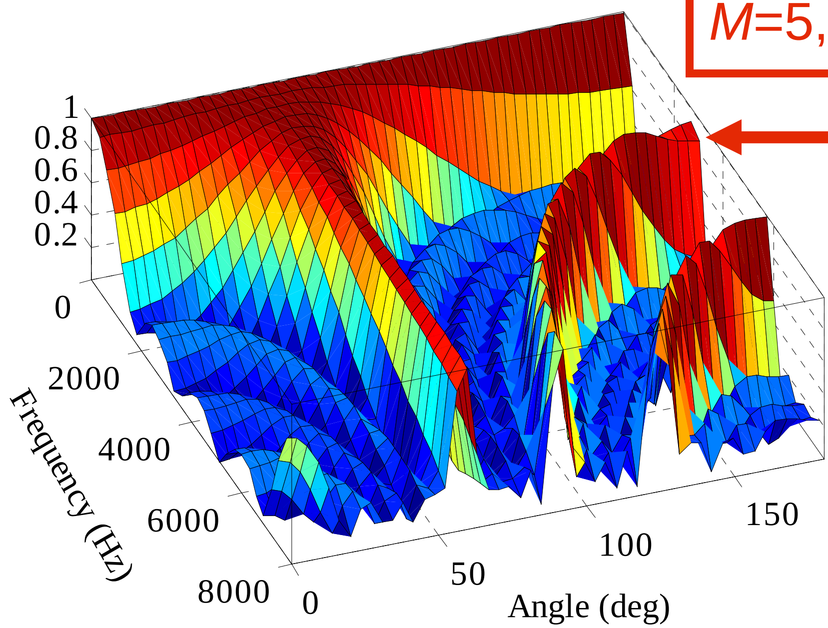
# Delay-and-sum beamforming

- **For** $f \geq \frac{c}{d(1+|\cos\psi|)}$ **an ambiguity, called *spatial aliasing*, occurs.**

**This is analogous to time-domain aliasing where now the spatial sampling (=*d*) is too large.**
**Aliasing does not occur (for any** $\psi$**) if**

$$d \leq \frac{c}{f_s} = \frac{c}{2.f_{max}} = \frac{\lambda_{min}}{2}$$

$M$=5, $\psi$=60°, $f_s$=16 kHz, $d$=8 cm




$$f = \frac{c}{d.(1+\cos\psi)}$$

Details...

$H(\omega,\theta)=1$ *iff* $\gamma = 2\pi.p$ for integer $p$

1) $\gamma = 0$ for $\theta = \psi$ (for all $\omega$)

2) if $\psi \leq \frac{\pi}{2}$ then $\gamma = 2\pi$ occurs for $\theta = \pi$ and $f = .. = \frac{c}{d.(1+\cos\psi)}$

3) if $\psi \geq \frac{\pi}{2}$ then $\gamma = 2\pi$ occurs for $\theta = 0$ and $f = .. = \frac{c}{d.(1-\cos\psi)}$

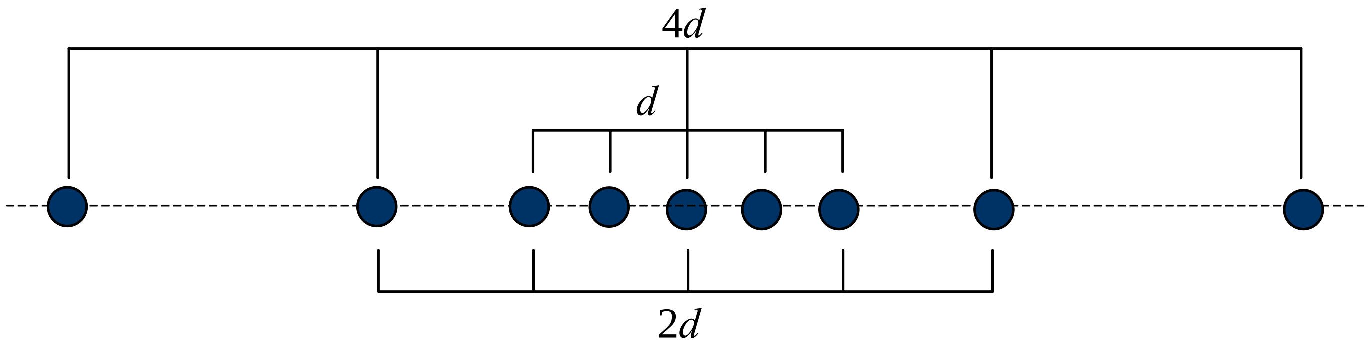
# Delay-and-sum beamforming

- **Beamwidth: for a uniform delay-and-sum beamformer**

$$BW \approx c\frac{\sqrt{96(1-\nu)}}{\omega d M}\sec\psi \qquad \text{with e.g. } \nu = \sqrt{\frac{1}{2}} \quad \text{(-3 dB)}$$

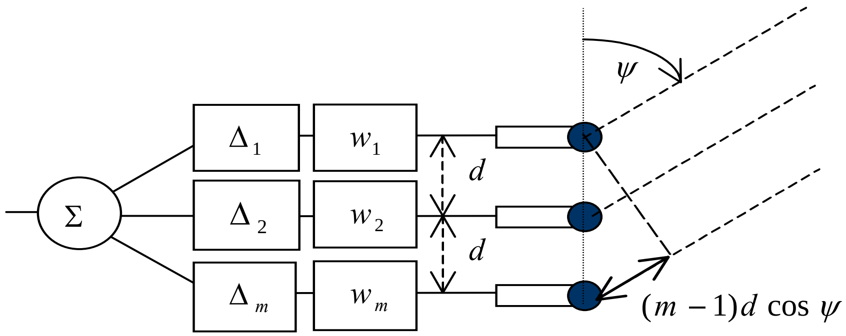
**hence large dependence on # microphones, distance (compare p14 & 15) and frequency (e.g. BW infinitely large at DC)**

- **Array topologies:**
  - Uniformly spaced arrays
  - Nested (logarithmic) arrays (small *d* for high *ω*, large *d* for small *ω*)
  - Planar / 3D-arrays



# Weighted-sum beamforming `delay-and-weight/sum’

- **Sensor-dependent complex weight + delay (compare to p. 13)**

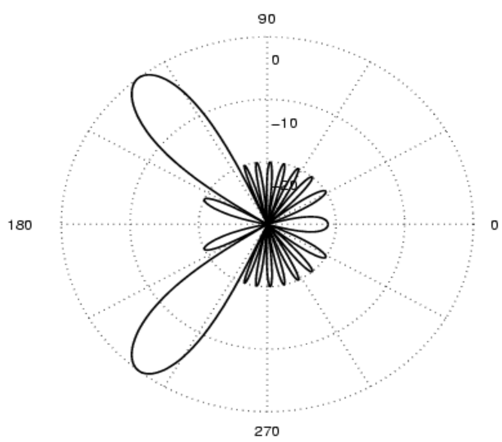



$$z[k] = \sum_{m=1}^{M} w_m . y_m[k + \Delta_m]$$

$$H(\omega,\theta) = \sum_{m=1}^{M} w_m \cdot e^{-j(m-1)\omega\frac{d(\cos\theta-\cos\psi)}{c} f_s}$$

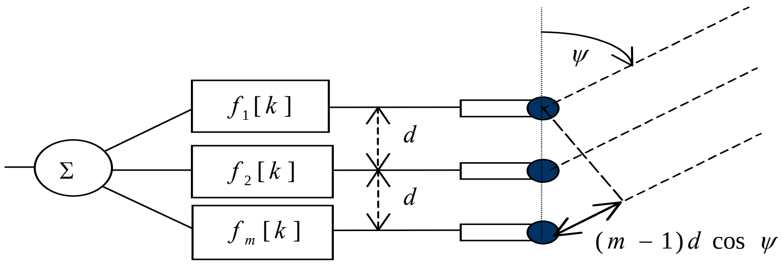
- **Weights added to allow for better beam shaping**
- **Design similar to traditional (spectral) filter design**

Ex: Dolph-Chebyshev design: beampattern with uniform sidelobe level (`equiripple’)



# Filter-and-sum beamforming

- **Sensor-dependent filters implement frequency-dependent complex weights to obtain a desired response over the whole frequency/angle range of interest**




$$z[k]=\sum_{m=1}^{M} f_m[-k]\otimes y_m[k]$$

$$H(\omega,\theta)=\sum_{m=1}^{M} F_m^*(\omega)\cdot e^{-j(m-1)\omega\frac{d\cos\theta}{c}f_s}$$

- **Design strategies : desired beampattern is *P*(*ω*,*θ*)**
  - Non-linear:

$$\min_{f_m[k],m=1\ldots M}\int_{\theta_1}^{\theta_2}\int_{\omega_1}^{\omega_2}\left(\left|H(\omega,\theta)\right|-\left|P(\omega,\theta)\right|\right)^2 d\omega\, d\theta$$

  - Quadratic:

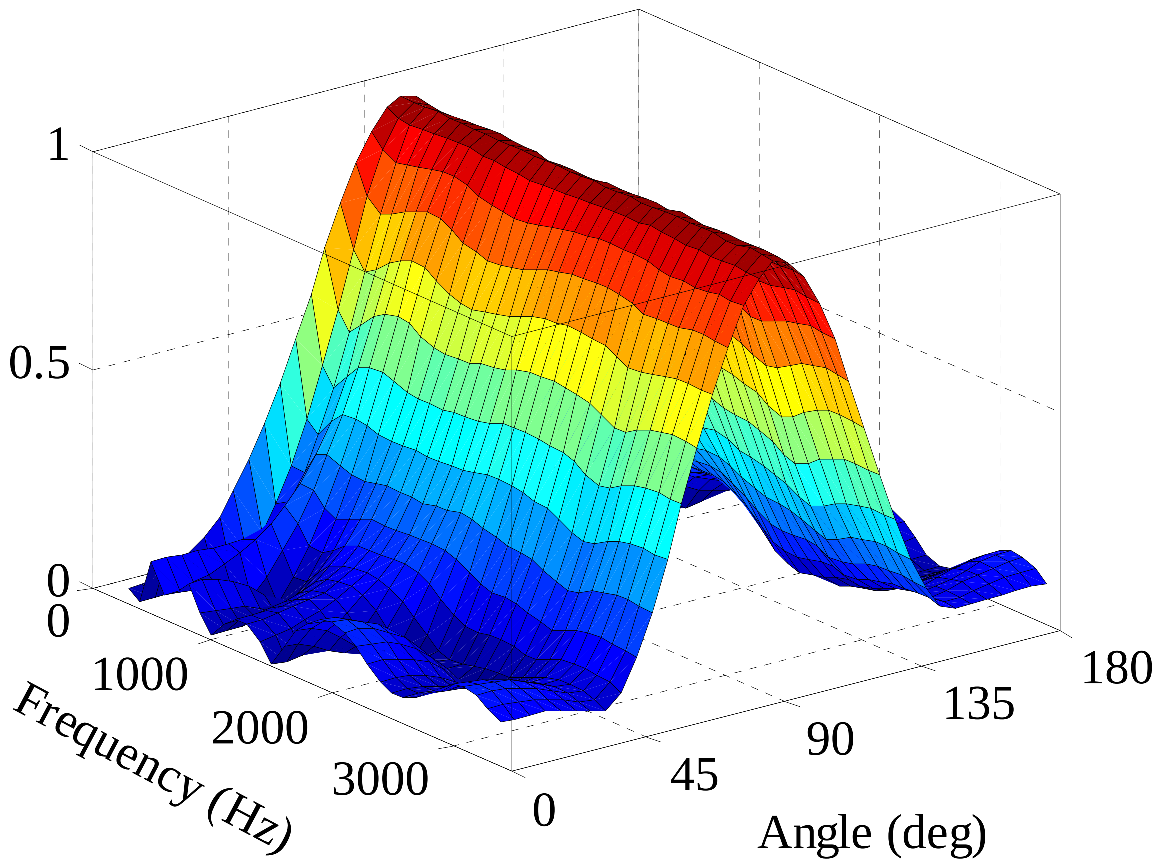
$$\min_{f_m[k],m=1\ldots M}\int_{\theta_1}^{\theta_2}\int_{\omega_1}^{\omega_2}\left|H(\omega,\theta)-P(\omega,\theta)\right|^2 d\omega\, d\theta$$

  - Frequency sampling, i.e. design weights for sampling frequencies $\omega_I$ and then interpolate :

$$\min_{F_m(\omega_i),m=1\ldots M}\int_{\theta_1}^{\theta_2}\left|H(\omega_i,\theta)-P(\omega_i,\theta)\right|^2 d\theta$$

# Filter-and-sum beamforming

- **Example-1: frequency-independent beamforming** (continued)




$M$=8
Logarithmic array
$L$=50
$\psi$=90°
$f_s$=8 kHz

# Filter-and-sum beamforming

- **Example-2: `superdirective' beamforming**
  - Maximize directivity for known (diffuse) noise fields
  - Maximum directivity $=M^2$ obtained for diffuse noise & endfire steering ($\theta$=0$^o$)

**<u>Design</u>: find** F($\omega$) **that <u>maximizes</u>**

$$DI(\omega) = \frac{\left|\mathbf{F}^H(\omega)\cdot\mathbf{d}(\omega,\theta_{\max})\right|^2}{\mathbf{F}^H(\omega)\cdot\mathbf{\Gamma}_{NN}^{diffuse}(\omega)\cdot\mathbf{F}(\omega)}$$

for given steering angle theta_max

- Optimal solution is

$$\mathbf{F}(\omega) = \alpha\cdot\mathbf{\Gamma}_{NN}^{-1}(\omega)\cdot\mathbf{d}(\omega,\theta_{\max})$$

- This is equivalent to minimization of noise output power, subject to unit response for steering angle (**)

$$\min_{\mathbf{F}(\omega)} \mathbf{F}^H(\omega)\cdot\mathbf{\Gamma}_{NN}(\omega)\cdot\mathbf{F}(\omega), \text{ s.t. } \mathbf{F}^H(\omega)\cdot\mathbf{d}(\omega,\theta_{\max}) = 1$$

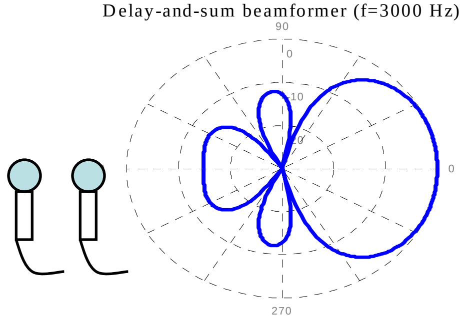
**<u>PS</u>: Delay-and-sum beamformer similarly <u>maximizes</u> WNG** ($\Gamma_{NN} = I$)

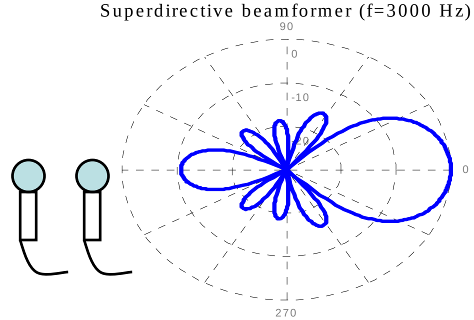
$$\mathbf{F}(\omega) = \alpha\cdot\mathbf{d}(\omega,\theta_{\max})$$

# Filter-and-sum beamforming

- **Example-2: `superdirective' beamforming (continued)**

  **<u>Directivity patterns</u> for endfire steering:**

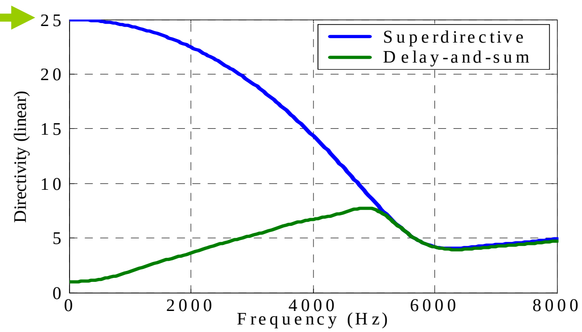


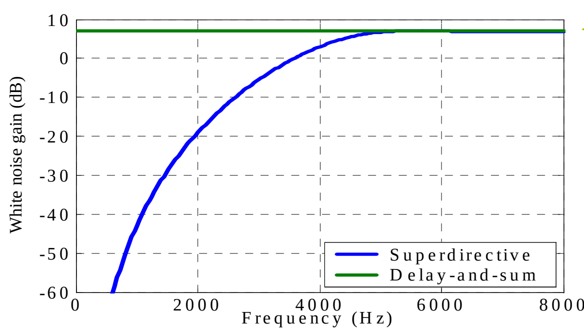




*M*=5
*d*=3 cm
*theta_max*=0°
$f_s$=16 kHz

**Superdirective beamformer has <u>highest</u> <u>DI</u>, but very <u>poor</u> <u>WNG</u>
hence problems with *robustness* (e.g. sensor noise) !**

$M^2$



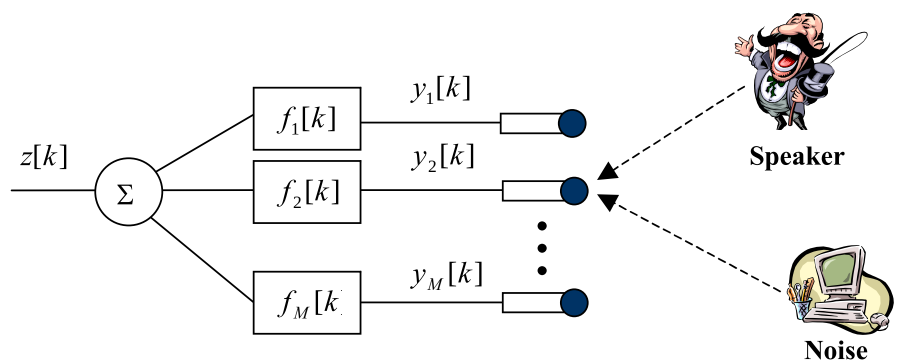




6.99=10.Log(5)

PS:
diffuse noise =
white noise for
high frequencies

# LCMV-beamforming

- **Adaptive filter-and-sum structure:**
  - Aim is to minimize noise output power, while maintaining a chosen frequency response in a given look direction (and/or other linear constraints, see below)
  - This corresponds to operation of a superdirective array (see (**) p25), but now noise field is <u>unknown</u>
  - Implemented as adaptive filter (e.g. constrained LMS algorithm)
  - Notation:




$$z[k] = \mathbf{f}^T \mathbf{y}[k] = \sum_{m=1}^{M} \mathbf{f}_m^T \mathbf{y}_m[k]$$

$$\mathbf{y}[k] = \left[\mathbf{y}_1^T[k] \quad \mathbf{y}_2^T[k] \quad \ldots \quad \mathbf{y}_M^T[k]\right]^T$$

$$\mathbf{y}_m[k] = \left[y_m[k] \quad y_m[k-1] \quad \ldots \quad y_m[k-L+1]\right]^T$$

$$\mathbf{f} = \left[\mathbf{f}_1^T \quad \mathbf{f}_2^T \quad \ldots \quad \mathbf{f}_M^T\right]^T$$

$$\mathbf{f}_m = \left[f_m[0] \quad f_m[1] \quad \ldots \quad f_m[L-1]\right]^T$$

# LCMV-beamforming

LCMV = **Linearly Constrained Minimum Variance**

- f designed to minimize variance of output $z[k]$ :

$$\min_{\mathbf{f}} E\{z^2[k]\} = \min_{\mathbf{f}} \mathbf{f}^T \cdot \mathbf{R}_{yy}[k] \cdot \mathbf{f}$$

- to avoid desired signal distortion/cancellation, add linear constraints:

$$\mathbf{C}^T \cdot \mathbf{f} = \mathbf{b}, \text{ with } \mathbf{C} \in \Re^{ML \times J}, \mathbf{b} \in \Re^J$$

- if noise and speech are uncorrelated, constrained output power minimization corresponds to <u>constrained noise power minimization</u>
- Type of constraints:
  - Frequency response in look-direction. Ex: $\sum_{m=1}^{M} F_m(z) = 1$ (for broadside)
  - Point, line and derivative constraints (=L constraints)
- Solution is (obtained using Lagrange-multipliers, etc..):

$$\mathbf{f}_{opt} = \mathbf{R}_{yy}^{-1}[k] \cdot \mathbf{C} \cdot \left(\mathbf{C}^T \cdot \mathbf{R}_{yy}^{-1}[k] \cdot \mathbf{C}\right)^{-1} \mathbf{b}$$