Lecture

Speech and Audio Signal Processing

Lecture 6: Beamforming, Part I

Content

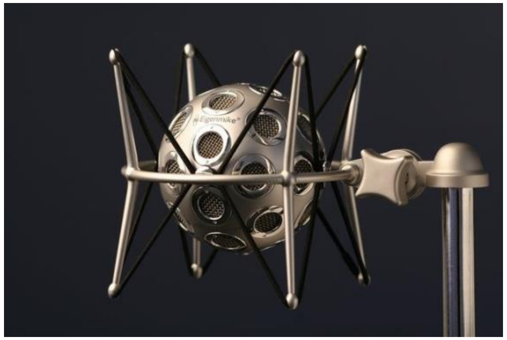
Part I:

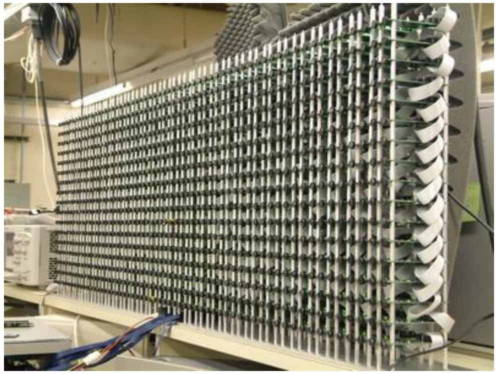
- Introduction
- Characteristics of multi-microphone systems
- *Differential* beamformer
- *Delay-and-sum* beamformer

Part II:

- *Filter-and-sum* beamformer
  - Minimum Variance Distortionless Response (MVDR) beamformer
- Multi-channel Wiener Filter
- Linear Constrained Min. Variance (LCMV) beamformer
- Audio examples and results

Introduction





References

Beamforming

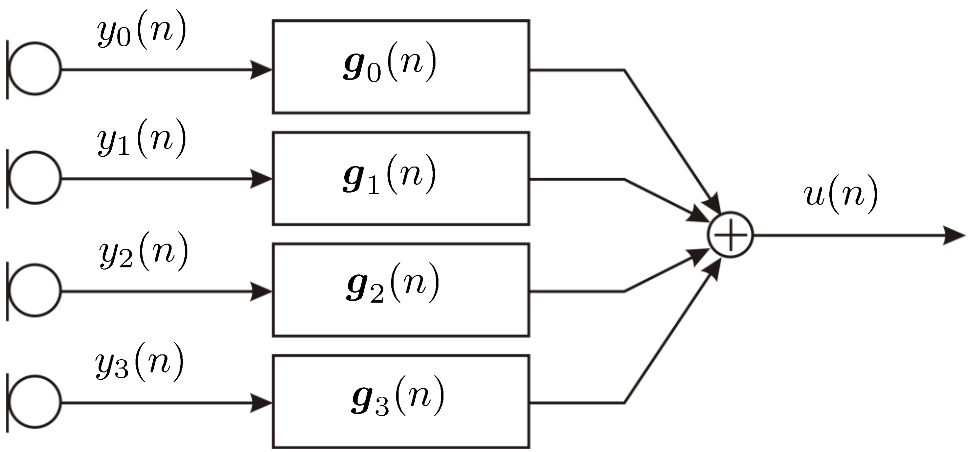
- E. Hänsler / G. Schmidt: *Acoustic Echo and Noise Control – Chapter 11 (Beamforming),* Wiley, 2004
- H. L. Van Trees: *Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory*, Wiley, 2002
- W. Herbordt: *Sound Capture for Human/Maschine Interfaces: Practical Aspects of Microphone Array Signal Processing*, Springer, 2005

Post-filtering

- K. U. Simmer, J. Bitzer, C. Marro: *Post-Filtering Techniques*, in M. Brandstein, D. Ward (Editors), Microphone Arrays, Springer, 2001
- S. Gannot, I. Cohen: *Adaptive Beamforming and Postfiltering,* in J. Benesty, M. M. Sondhi, Y. Huang (Editors), Springer Handbook of Speech Processing, Springer, 2007

Introduction

Basic structure:



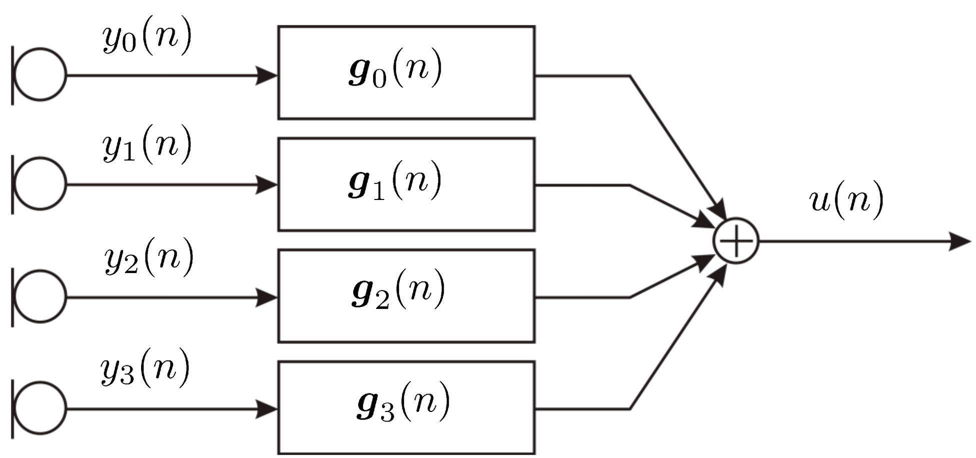
Output signal: additive convolution:

$$u(n) \quad = \quad \sum_{m=0}^{M-1} \sum_{i=0}^{N-1} y_m(n-i)\, g_{m,i}(n)$$

Different target signal directions

Target signal **parallel**
to the microphone axis:
=> **Endfire array**

Target signal **orthogonal**
to the microphone axis:
=> **Broadside array**



TECHNISCHE UNIVERSITÄT DARMSTADT

Additive convolution in vector notation:

$$
\begin{aligned}
u(n) &= \sum_{m=0}^{M-1} \sum_{i=0}^{N-1} y_m(n-i)\, g_{m,i}(n) \\
&= \sum_{m=0}^{M-1} \boldsymbol{y}_m^{\mathrm{T}}(n)\, \boldsymbol{g}_m(n)
\end{aligned}
$$

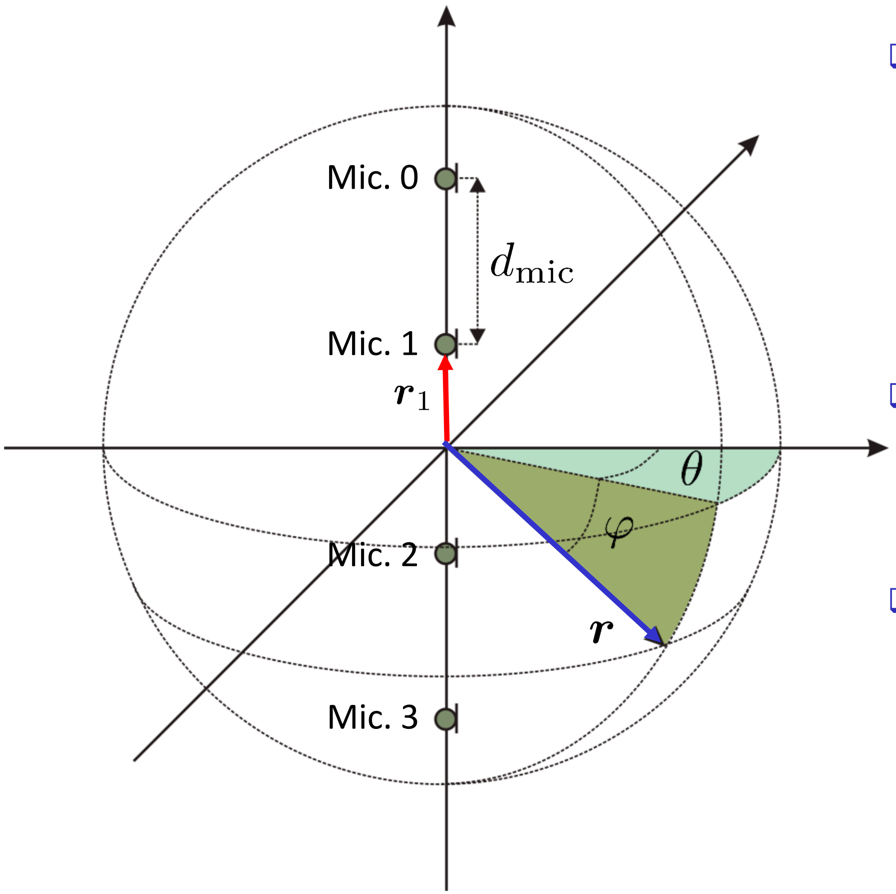
with:

$$
\begin{aligned}
\boldsymbol{y}_m(n) &= \big[y_m(n),\, y_m(n-1),\, \ldots,\, y_m(n-N+1)\big]^{\mathrm{T}} \\
\boldsymbol{g}_m(n) &= \big[g_{m,0}(n),\, g_{m,1}(n),\, \ldots,\, g_{m,N-1}(n)\big]^{\mathrm{T}}
\end{aligned}
$$

Fixed beamformer (no time dependent filtering):

$$
u(n) = \sum_{m=0}^{M-1} \boldsymbol{y}_m^{\mathrm{T}}(n)\, \boldsymbol{g}_m \quad \Longleftrightarrow \quad U(e^{j\Omega}) = \sum_{m=0}^{M-1} Y_m(e^{j\Omega})\, G_m(e^{j\Omega})
$$

Microphone positions and coordinate system:



- The center is chosen such that the sum of the position vectors showing to the microphones is equal to zero:

$$\sum_{m=0}^{M-1} \boldsymbol{r}_m = 0.$$

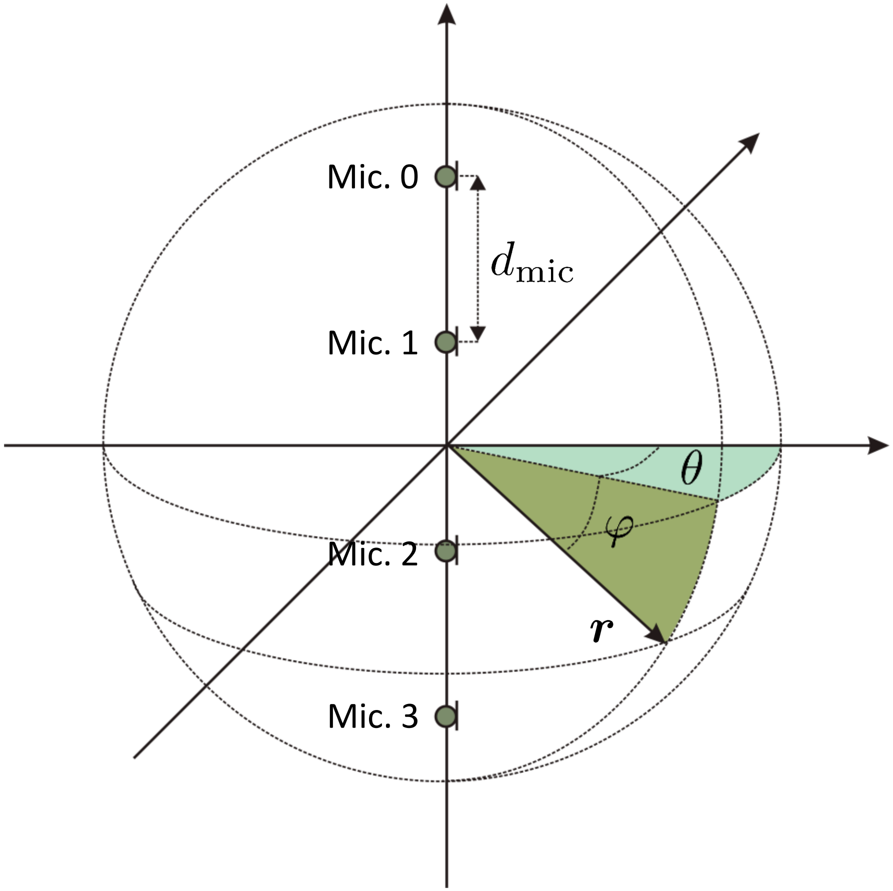
- The vector $\boldsymbol{r}$ is oriented in the direction of the incoming sound and has the length 1:

$$\|\boldsymbol{r}\| = 1.$$

- Assuming a plane wave (far-field assumption) the sound arrives with the following normalized (with the sampling time) propagation delays at the microphones:

$$\tau_m = \frac{f_{\text{s}}}{c} \boldsymbol{r}_m^{\text{T}} \boldsymbol{r} = \frac{f_{\text{s}}}{c} \|\boldsymbol{r}_m\| \sin(\varphi)$$

Directional processing obtained by filtering or by direction dependent sensors (microphones):



- A direction dependent processing can be obtained by filtering and summation of the microphone signals

$$u(n) = \sum_{m=0}^{M-1} \boldsymbol{y}_m^{\mathrm{T}}(n)\, \boldsymbol{g}_m$$

  or by sensors which already show a direction dependent sensitivity.

- Additionally, the beamformer can be used to calculate a noise reference signal (target signal cancellation) which can than be cancelled by adaptive (Wiener) filtering.

Notations

TECHNISCHE UNIVERSITÄT DARMSTADT

Assumptions of a spatial frequency response:

- The sound propagation of point sources is modelled as a plane wave:

$$S_m(e^{j\Omega}) = S(e^{j\Omega})\, e^{-j\Omega\tau_m}.$$

- Each microphone has a direction dependent sensitivity which can be described as follows:

$$M_m(e^{j\Omega}, \boldsymbol{r}) = M_m(e^{j\Omega}, \varphi, \theta)$$

Microphones with an omni-directional sensitivity:

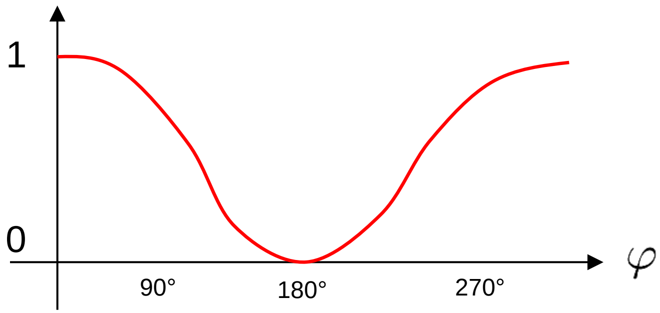
$$M_{m,\mathrm{omni}}(e^{j\Omega}, \varphi, \theta) = 1.$$

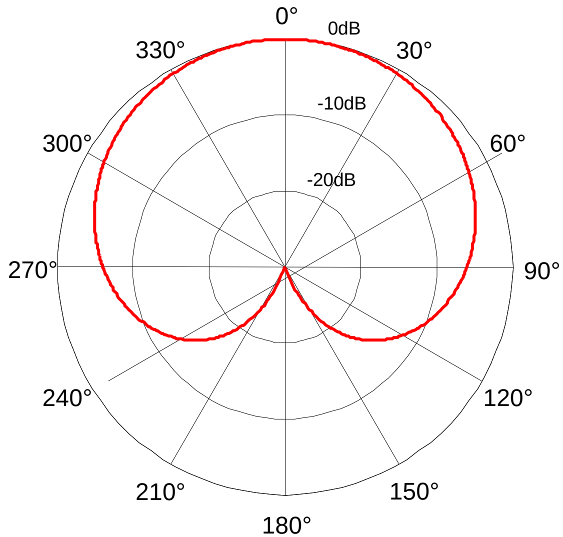
Other microphone characteristics possible (by design!), such as a cardioid sensitivity:

$$M_{m,\mathrm{cardioid}}(e^{j\Omega}, \varphi, \theta) = \frac{1}{2}\big[1 + \cos(\varphi)\big].$$

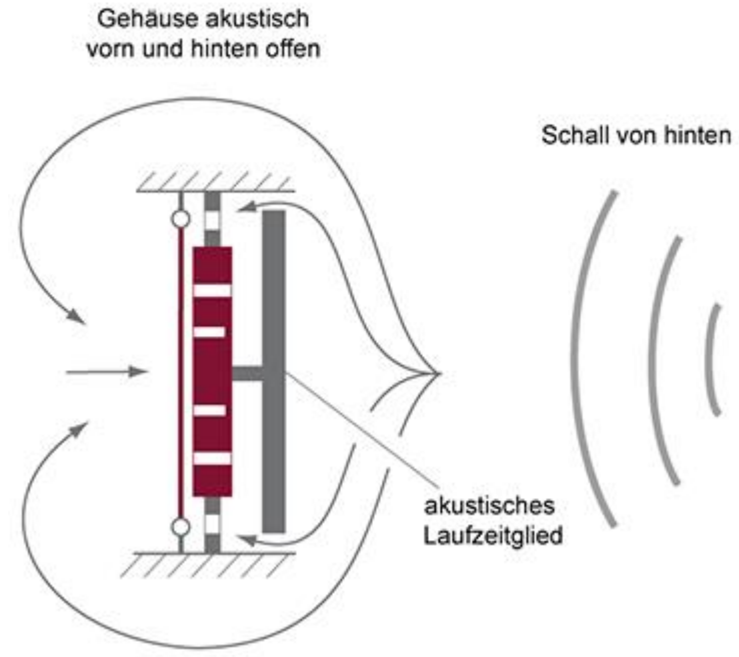
Cardioid microphones – per design

$$M_{m,\text{cardioid}}(e^{j\Omega}, \varphi, \theta) = \frac{1}{2}\left[1 + \cos(\varphi)\right].$$





- Mechanical design of a cardioid microphone



Source: https://www.masterclass-sounddesign.de/content_bonusmaterial/richtcharakteristiken.htm

TECHNISCHE UNIVERSITÄT DARMSTADT

Notations

„Spatial“ frequency response:

- The microphone signals for a point target source can be noted as follows:

$$\begin{aligned} Y_m(e^{j\Omega}, \boldsymbol{r}) &= S_m(e^{j\Omega})\, M_m(e^{j\Omega}, \boldsymbol{r}) \\ &= S(e^{j\Omega})\; M_m(e^{j\Omega}, \boldsymbol{r})\, e^{-j\Omega\tau_m}. \end{aligned}$$

- Resulting in the following output spectrum for the beamformer:

$$\begin{aligned} U(e^{j\Omega}, \boldsymbol{r}) &= \sum_{m=0}^{M-1} Y_m(e^{j\Omega}, \boldsymbol{r})\, G_m(e^{j\Omega}) \\ &= S(e^{j\Omega}) \sum_{m=0}^{M-1} M_m(e^{j\Omega}, \boldsymbol{r})\, G_m(e^{j\Omega})\, e^{-j\Omega\tau_m}. \end{aligned}$$

- With the following spatial frequency response:

$$G_{\mathrm{BF}}(e^{j\Omega}, \boldsymbol{r}) = \frac{U(e^{j\Omega}, \boldsymbol{r})}{S(e^{j\Omega})} = \sum_{m=0}^{M-1} M_m(e^{j\Omega}, \boldsymbol{r})\, G_m(e^{j\Omega})\, e^{-j\Omega\tau_m}.$$

Notations

TECHNISCHE UNIVERSITÄT DARMSTADT

Beampattern:

- The squared magnitude of the „spatial frequency response“ is noted as „beampattern“

$$\Phi(\Omega, \boldsymbol{r}) \quad = \quad \left| G_{\mathrm{BF}}(e^{j\Omega}, \boldsymbol{r}) \right|^2.$$

- In case all microphones have the same beampattern, the impact of the microphones and the directional filtering can be separated:

$$\begin{aligned} \Phi(\Omega, \boldsymbol{r}) &= \left| \sum_{m=0}^{M-1} M_m(e^{j\Omega}, \boldsymbol{r})\, G_m(e^{j\Omega})\, e^{-j\Omega\tau_m} \right|^2 \\ &= \left| M(e^{j\Omega}, \boldsymbol{r}) \right|^2 \left| \sum_{m=0}^{M-1} G_m(e^{j\Omega})\, e^{-j\Omega\tau_m} \right|^2 \\ &= \Phi_{\mathrm{Mic}}(\Omega, \boldsymbol{r})\, \Phi_{\mathrm{Sig}}(\Omega, \boldsymbol{r}). \end{aligned}$$

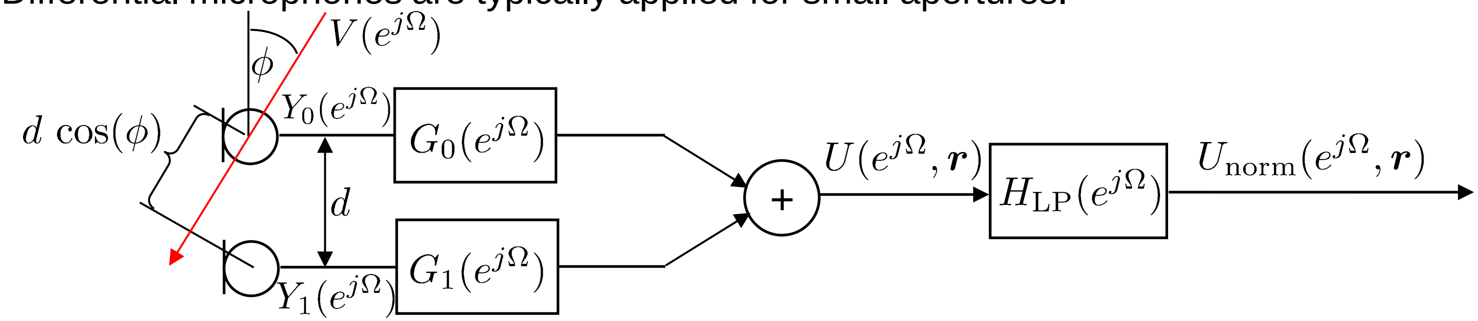
- In the following, we will only apply omni-directional microphones, if not explicitly mentioned:

$$\Phi(\Omega, \boldsymbol{r}) \quad = \quad \Phi_{\mathrm{Sig}}(\Omega, \boldsymbol{r}).$$

Differential beamformer

Differential beamformer: Setup

- Concept:
  - Delay second microphone signal and subtract it from the first.
  - Differential microphones are endfire arrays: Angle ϕ oriented with respect to the microphone axis.
  - Center of the coordinate system is chosen at the frontal microphone.
  - Differential microphones are typically applied for small apertures.



Differential setup:

$$G_0(e^{j\Omega}) = 1$$

$$G_1(e^{j\Omega}) = -e^{-j\,\Omega\,T f_s}$$

Microphone signal components:

$$Y_0(e^{j\Omega}, \boldsymbol{r}) = V(e^{j\Omega})$$

$$Y_1(e^{j\Omega}, \boldsymbol{r}) = V(e^{j\Omega}) e^{-j\Omega\, f_s\, \frac{d}{c}\cos(\phi)}$$

c : sound propagation speed

$$T_d = \frac{d\ \cos(\phi)}{c}$$: sound propagation delay

- Output signal spectrum:

$$U(e^{j\Omega}, \boldsymbol{r}) = \sum_{m=0}^{M-1} Y_m(e^{j\Omega}, \boldsymbol{r})\, G_m(e^{j\Omega}) = Y_0(e^{j\Omega}, \boldsymbol{r}) - Y_1(e^{j\Omega}, \boldsymbol{r})\, e^{-j\Omega\, T f_s}$$

$$U(e^{j\Omega}, \boldsymbol{r}) = V(e^{j\Omega}) \left[1 - e^{-j\Omega\, f_s\, (\frac{d}{c}\cos(\phi) + T)}\right]$$

Differential beamformer: Spatial frequency response

- **Normalization** of the frequency response **with respect to desired signal direction**
  I.e.: Signal $S(e^{j\Omega})$ arriving from the frontal direction $\phi = 0$

  - Output spectrum for target signal:

$$U(e^{j\Omega}, \boldsymbol{r}) \quad = \quad S(e^{j\Omega}) \left[1 - e^{-j\Omega\, fs\, (\frac{d}{c}\cos(\phi=0)+T)}\right]$$

  - Target signal should pass the beamformer unmodified
    => chose normalization low-pass $H_{\mathrm{LP}}(e^{j\Omega})$ appropriately :

$$U_{\mathrm{norm}}(e^{j\Omega}, \boldsymbol{r}(\phi=0)) = S(e^{j\Omega})\, H_{\mathrm{LP}}(e^{j\Omega}) \left[1 - e^{-j\Omega\, fs\, (\frac{d}{c}\cos(\phi=0)+T)}\right] \overset{!}{=} S(e^{j\Omega})$$

=> $H_{\mathrm{LP}}(e^{j\Omega}) = \dfrac{1}{1 - e^{-j\Omega\, f_s\, (\frac{d}{c}+T)}}$: Low pass to normalized high-pass frequency characteristic

  - Normalized output spectrum:

$$U_{\mathrm{norm}}(e^{j\Omega}, \boldsymbol{r}) = V(e^{j\Omega}) \frac{1 - e^{-j\Omega\, fs\, (\frac{d}{c}\cos(\phi)+T)}}{1 - e^{-j\Omega\, f_s\, (\frac{d}{c}+T)}}$$

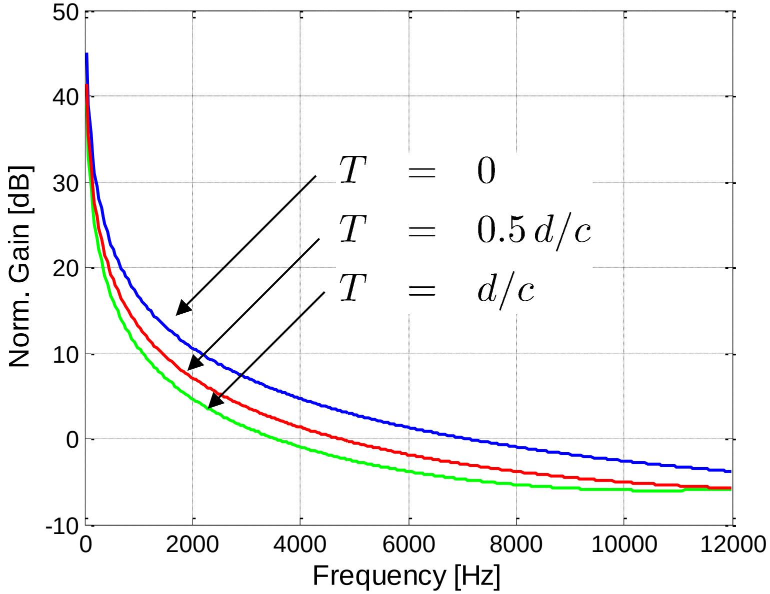
  - Spatial frequency response of the beamformer:

$$G_{\mathrm{BF}}(e^{j\Omega}, \boldsymbol{r}) = \frac{1 - e^{-j\Omega\, fs\, (\frac{d}{c}\cos(\phi)+T)}}{1 - e^{-j\Omega\, f_s\, (\frac{d}{c}+T)}}$$

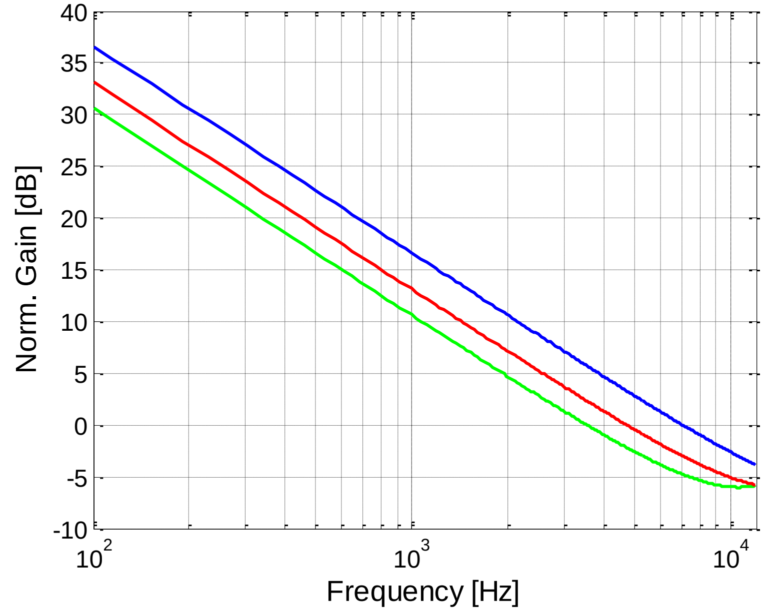
Differential beamformer: Normalization

- **Setup:** $$\begin{aligned} f_s &= 24\ \mathrm{kHz} \\ d &= 8\ \mathrm{mm} \\ c &= 340\ \mathrm{m/s} \end{aligned}$$ $T_0 = \dfrac{d}{c}$: Sound propagation delay between the microphones.
- **Normalization low-pass:** $$H_{\mathrm{LP}}(e^{j\Omega}) = \frac{1}{1 - e^{-j\Omega\, f_s\, (\frac{d}{c} + T)}}$$

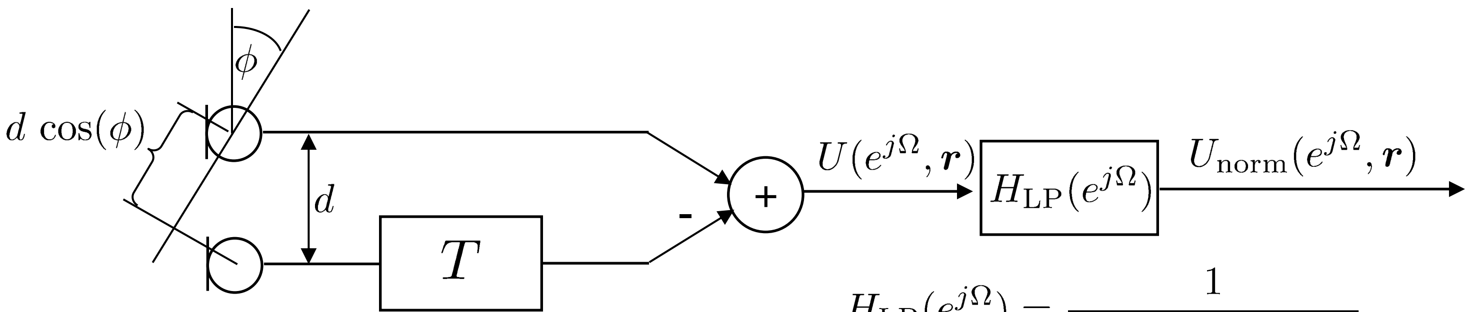
Linear frequency axis



Logarithmic frequency axis



Differential beamformer: Beampattern



$$H_{\mathrm{LP}}(e^{j\Omega}) = \frac{1}{1 - e^{-j\Omega\, f_s\, (\frac{d}{c} + T)}}$$

$$G_{\mathrm{BF}}(e^{j\Omega}, \boldsymbol{r}) = \frac{1 - e^{-j\Omega\, fs\, (\frac{d}{c}\cos(\phi) + T)}}{1 - e^{-j\Omega\, f_s\, (\frac{d}{c} + T)}}$$

- Beampattern:

$$\begin{aligned}
\Phi(\Omega, \boldsymbol{r}) &= \left| G_{\mathrm{BF}}(e^{j\Omega}, \boldsymbol{r}) \right|^2. \\
\Phi_{\mathrm{dB}}(\Omega, \boldsymbol{r}) &= 20\,\log_{10} \left| \frac{1 - e^{-j\Omega\, fs\, (\frac{d}{c}\cos(\phi) + T)}}{1 - e^{-j\Omega\, f_s\, (\frac{d}{c} + T)}} \right| \\
&= 20\,\log_{10} \left| \frac{\sin\left(\frac{1}{2}\Omega\, fs\, (\frac{d}{c}\cos(\phi) + T)\right)}{\sin\left(\frac{1}{2}\Omega\, f_s\, (\frac{d}{c} + T)\right)} \right| \\
&\approx 20\,\log_{10} \left| \frac{\frac{d}{c}\cos(\phi) + T}{\frac{d}{c} + T} \right|
\end{aligned}$$

for:

- Approximation valid for :

$$\begin{aligned}
\frac{1}{2}\Omega_{\max}\, f_s\, (\frac{d}{c} + T_{\max}) &\ll \pi \\
\frac{1}{2}\pi f_s\, 2\frac{d}{c} &\ll \pi \\
f_s \frac{d}{c} &\ll 1
\end{aligned}$$

f_s : Sampling frequency

f_N : Nyquist frequency

$$\boxed{d \ll \frac{c}{f_s} = \frac{c}{2 f_N} = \frac{\lambda_{\min}}{2}}$$

Differential beamformer: Beampattern

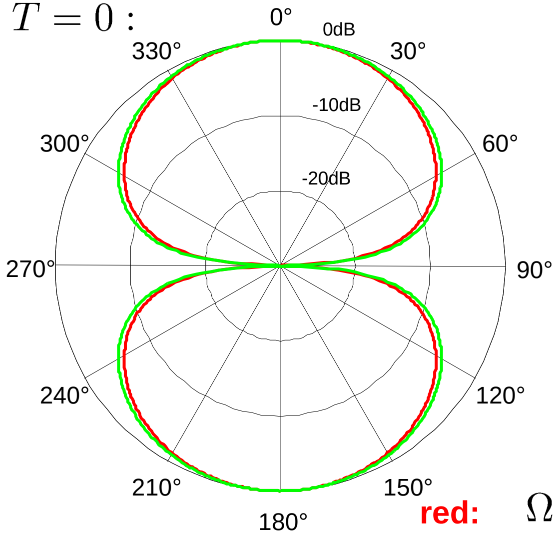
- **Setup:** $$f_s = 24\ \mathrm{kHz}, \quad d = 8\ \mathrm{mm}, \quad c = 340\ \mathrm{m/s}$$

  $T_0 = \frac{d}{c}$: Sound propagation delay between the microphones.

- **Beampattern:** $$\Phi_{\mathrm{dB}}(\Omega, \boldsymbol{r}) = 20\,\log_{10}\left|\frac{\sin\left(\frac{1}{2}\Omega\, f s\,(\frac{d}{c}\cos(\phi) + T)\right)}{\sin\left(\frac{1}{2}\Omega\, f_s\,(\frac{d}{c} + T)\right)}\right| \approx 20\,\log_{10}\left|\frac{\frac{d}{c}\cos(\phi) + T}{\frac{d}{c} + T}\right|$$

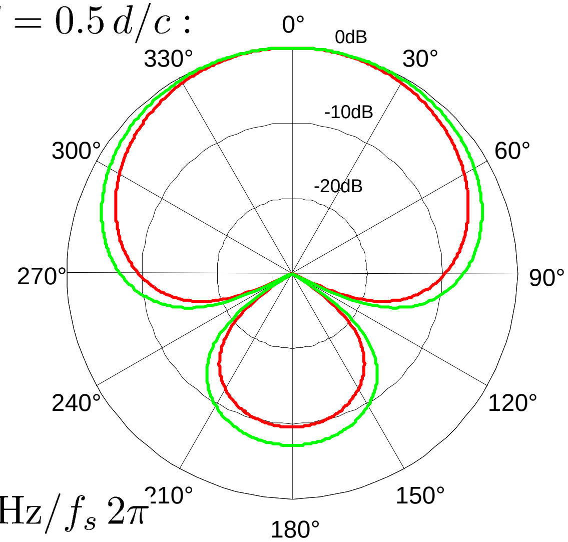
Figure “8”:

$T = 0:$



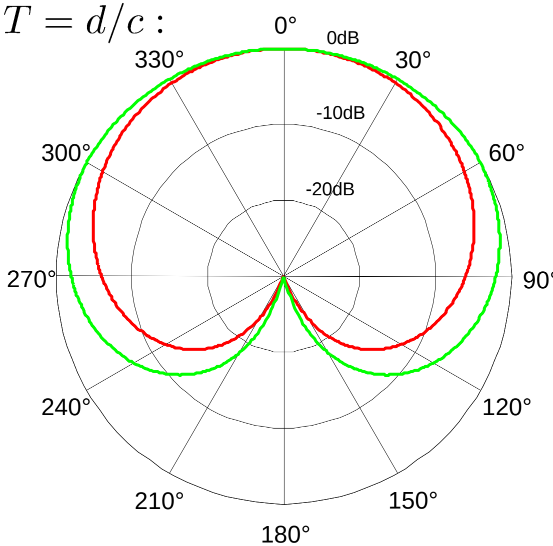
Hyper-Cardioid:

$T = 0.5\, d/c:$



Cardioid:

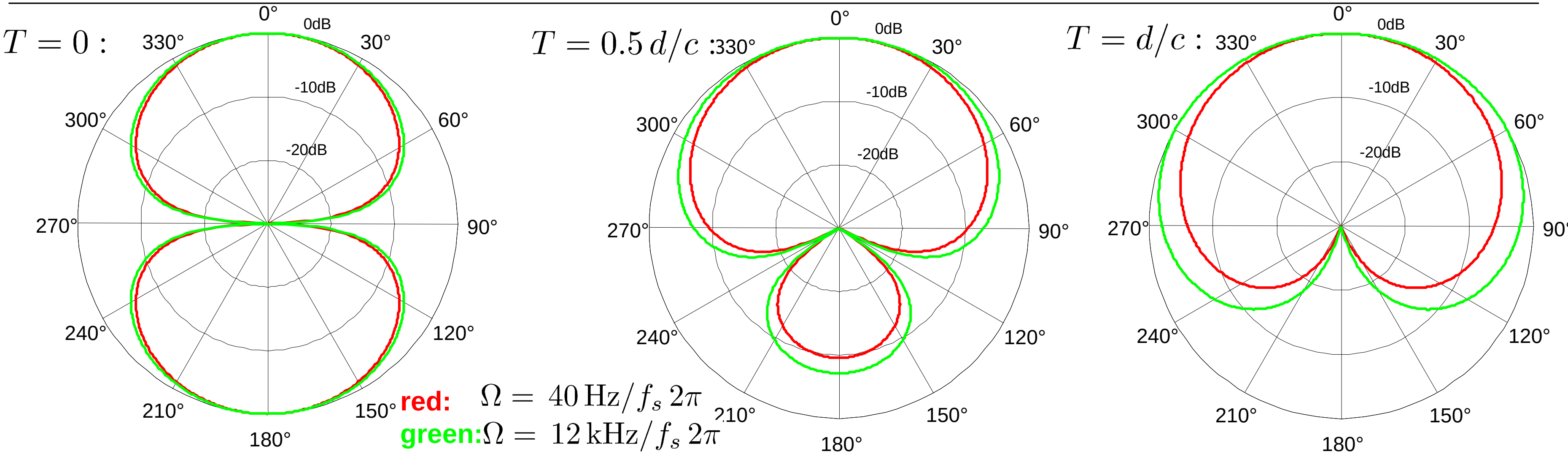
$T = d/c:$



red: $\Omega = 40\,\mathrm{Hz}/f_s\,2\pi$

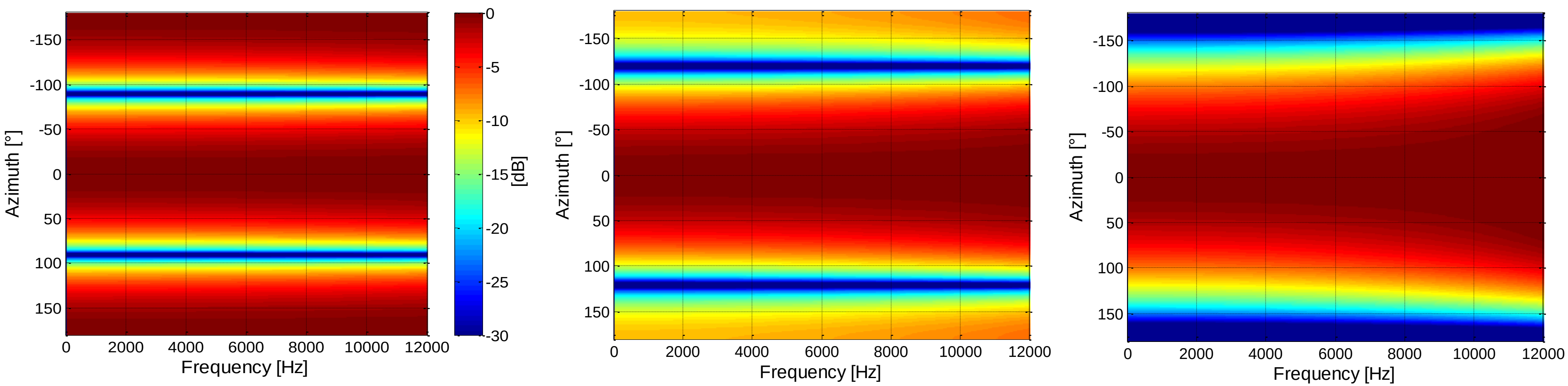
green: $\Omega = 12\,\mathrm{kHz}/f_s\,2\pi$

Differential beamformer: Beampattern



red: $\Omega = 40\,\mathrm{Hz}/f_s\,2\pi$
green: $\Omega = 12\,\mathrm{kHz}/f_s\,2\pi$

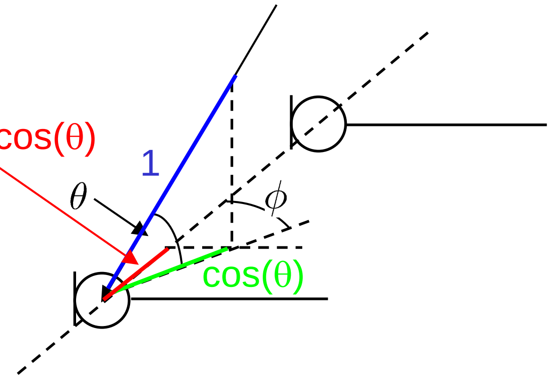
- **Beampattern with respect to angle and frequency:**



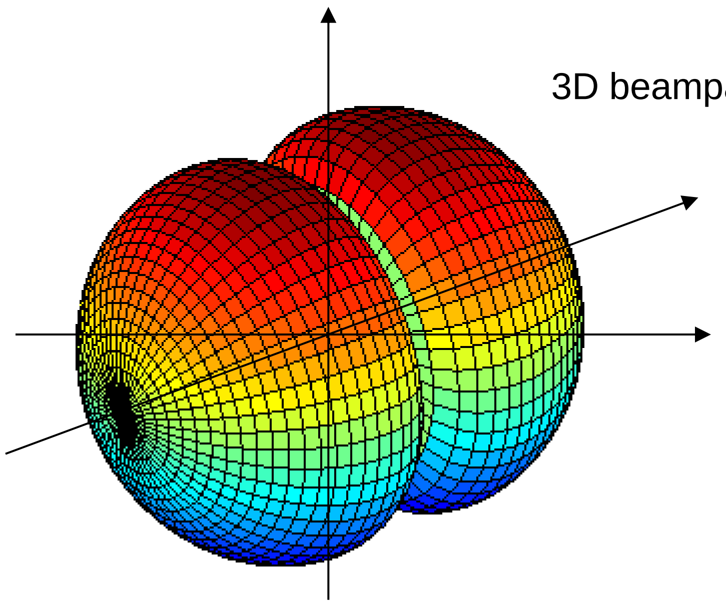
Differential beamformer: 3D beampattern

- **Rotation symmetry:**

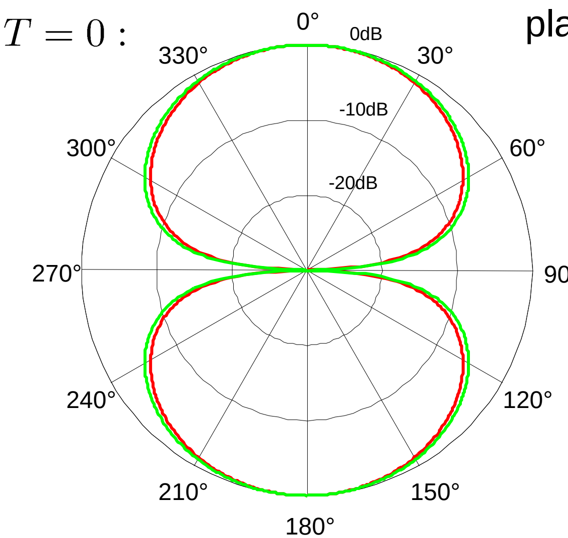
$$\Phi_{\mathrm{dB}}(\Omega,\boldsymbol{r}) = 20\,\log_{10}\left|\frac{\sin\left(\Omega\, f s\,(\frac{d}{c}\cos(\phi)\cos(\theta)+T)\right)}{\sin\left(\Omega\, f_s\,(\frac{d}{c}+T)\right)}\right|$$



3D beampattern



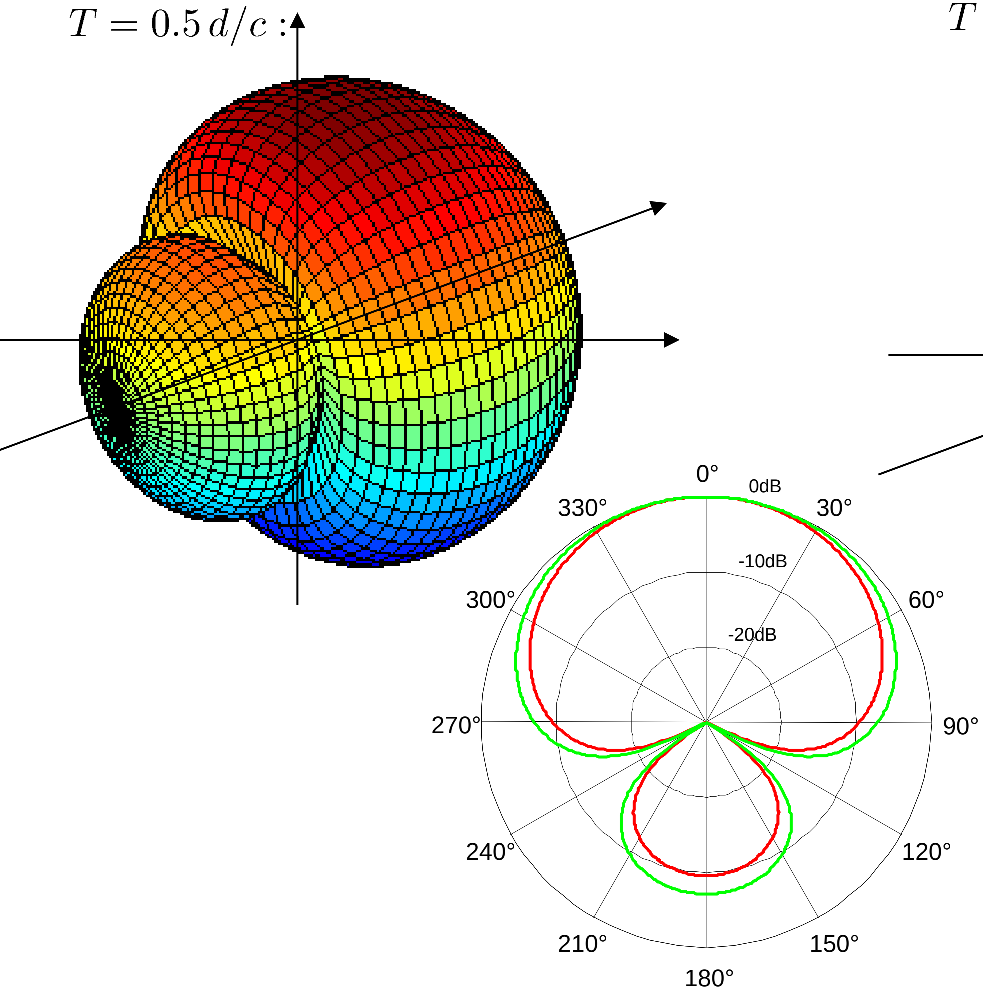
$T = 0:$



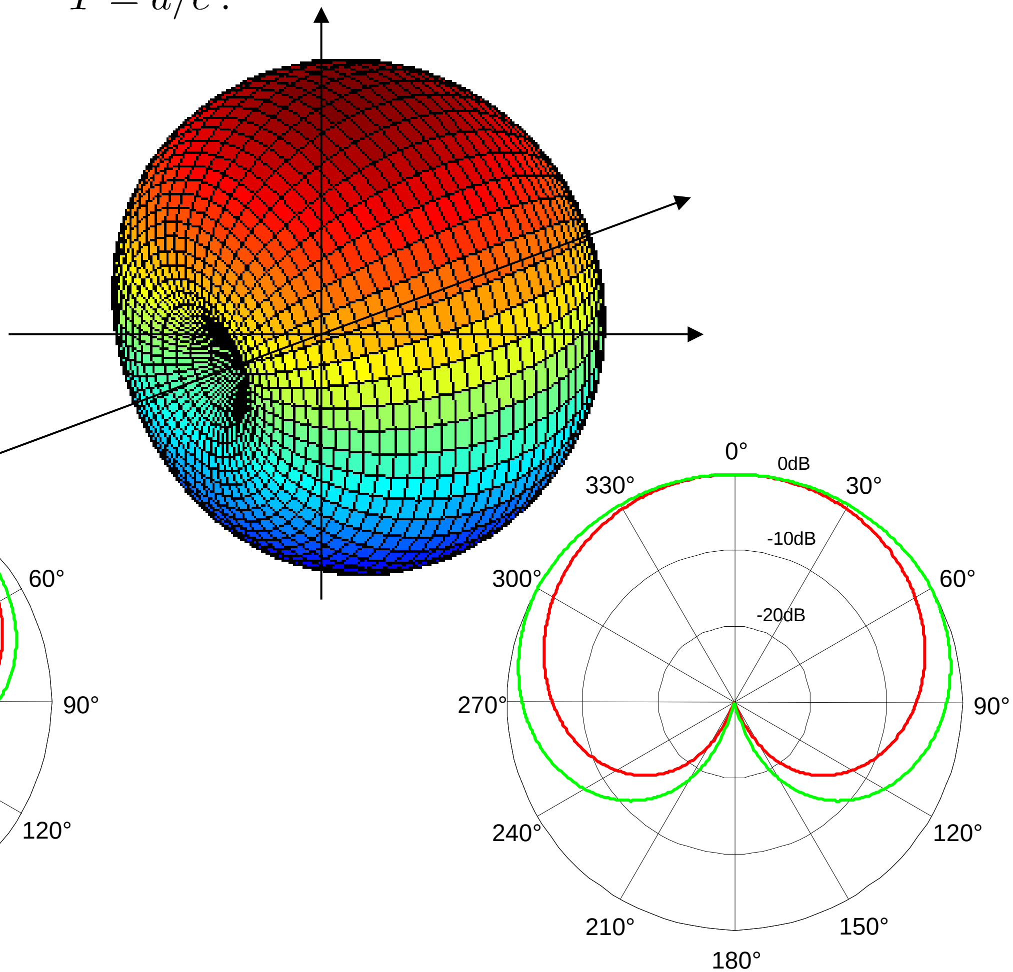
2D beampattern in the plane where $\theta = 0$

Differential beamformer: 3D beampattern

$T = 0.5\, d/c:$

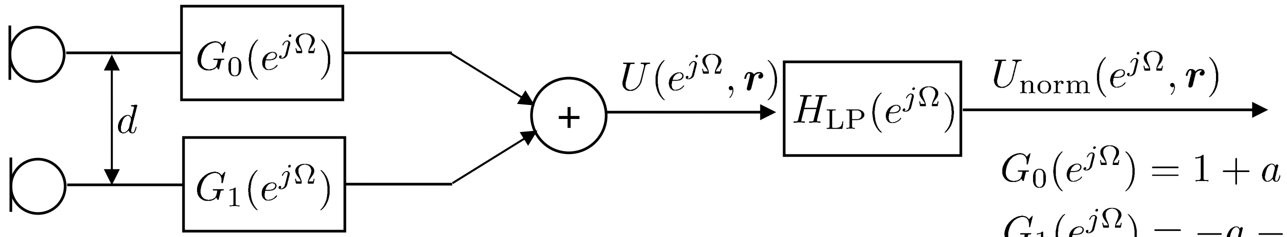


$T = d/c:$



Modified differential beamformer: Setup

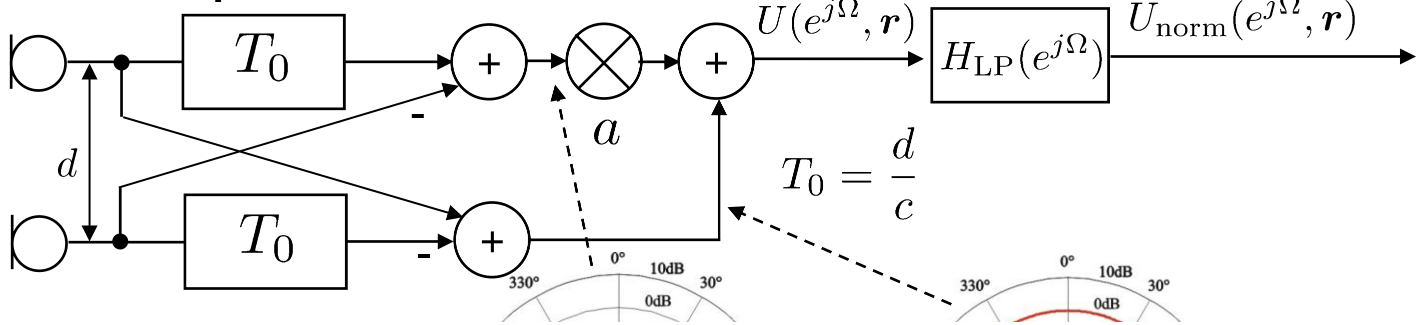
- **Modified approach with fixed delay and beam steering by weighting:**



$$G_0(e^{j\Omega}) = 1 + a\, e^{-j\,\Omega\, T_0 f_s}$$

$$G_1(e^{j\Omega}) = -a - e^{-j\,\Omega\, T_0 f_s}$$

- **Equivalent setup:**

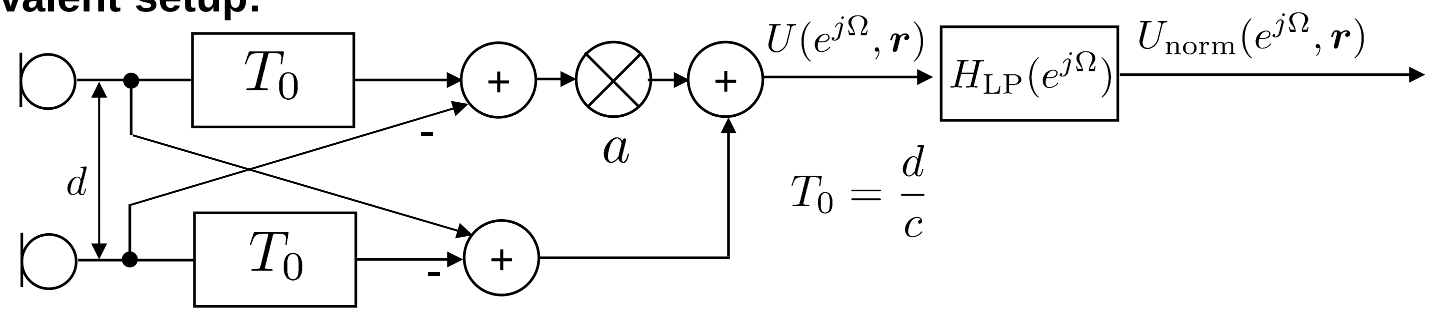


$$T_0 = \frac{d}{c}$$

- **Modification: compared to the solution know so far:**
  - Fixed internal delays T_0 : no modification necessary when beam should be switched during operation.
  - This is an advantage since the delays are usually fractions of the sampling delay and have to be realized by filters (slide 43).
  - The beampattern is controlled by the weighting factor: a

Modified differential beamformer: Spatial freq. response

- **Equivalent setup:**



$$U(e^{j\Omega}, \boldsymbol{r}) = V(e^{j\Omega}) \left[1 - e^{-j\Omega\, fs\, (\frac{d}{c}\cos(\phi)+T_0)} + a\, e^{-j\Omega\, fs\, T_0} \left(1 - e^{-j\Omega\, fs\, (\frac{d}{c}\cos(\phi)-T_0)}\right)\right]$$

$$U(e^{j\Omega}, \boldsymbol{r}(\phi = 0)) = V(e^{j\Omega}) \left[1 - e^{-j\Omega\, fs\, 2T_0}\right]$$

- Target signal normalization is independent of beampattern: $H_{\mathrm{LP}}(e^{j\Omega}) = \dfrac{1}{1 - e^{-j\Omega\, f_s\, 2T_0}}$

- Normalized output spectrum:

$$U_{\mathrm{norm}}(e^{j\Omega}, \boldsymbol{r}) = V(e^{j\Omega}) \frac{1 - e^{-j\Omega\, fs\, T_0\, (\cos(\phi)+1)} + a\, e^{-j\Omega\, fs\, T_0} \left(1 - e^{-j\Omega\, fs\, T_0\, (\cos(\phi)-1)}\right)}{1 - e^{-j\Omega\, fs\, 2T_0}}$$

- Spatial frequency response:

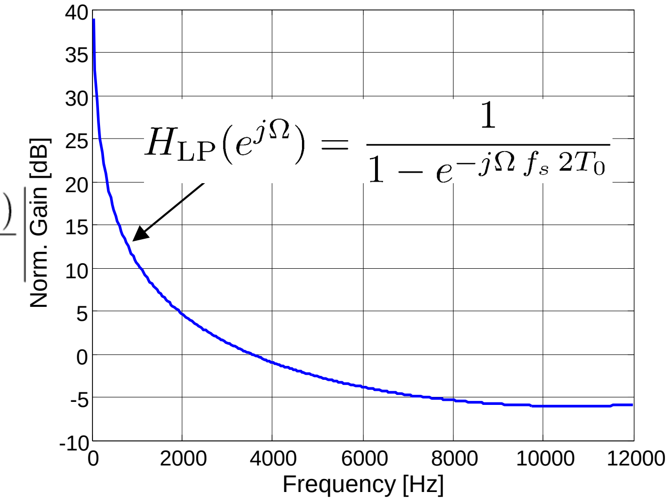
$$G_{\mathrm{BF}}(e^{j\Omega}, \boldsymbol{r}) = \frac{1 - e^{-j\Omega\, fs\, T_0\, (\cos(\phi)+1)} + a\, e^{-j\Omega\, fs\, T_0} \left(1 - e^{-j\Omega\, fs\, T_0\, (\cos(\phi)-1)}\right)}{1 - e^{-j\Omega\, fs\, 2T_0}}$$

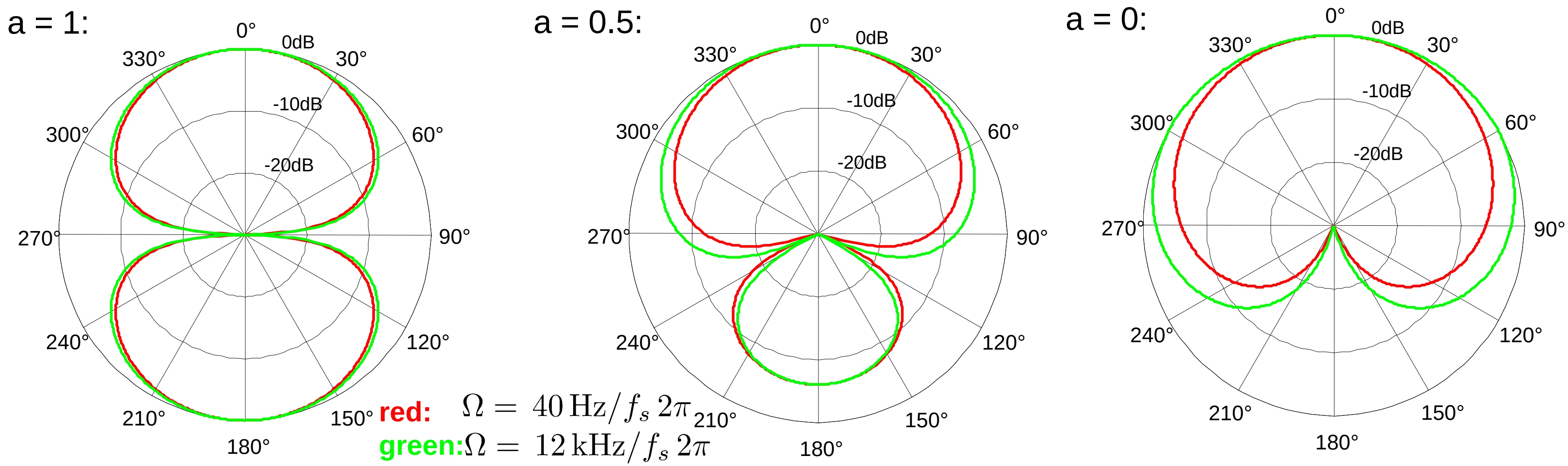
Modified differential beamformer: Beampattern

TECHNISCHE UNIVERSITÄT DARMSTADT

❑ Beampattern and normalization:

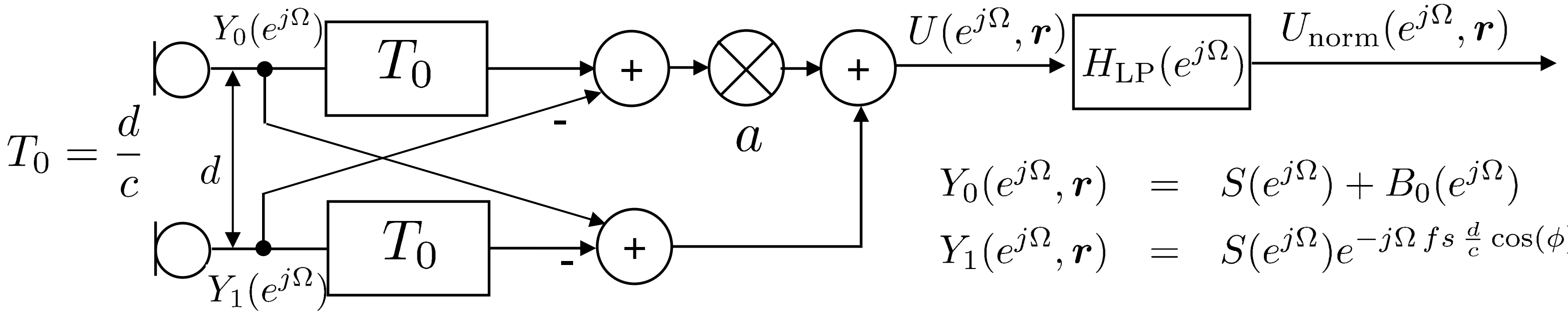
$$\Phi_{\mathrm{dB}}(\Omega, \boldsymbol{r}) \;=\; 20\,\log_{10}\left|\frac{1-e^{-j\Omega\, fs\, T_0\,(\cos(\phi)+1)}+a\,e^{-j\Omega\, fs\, T_0}\left(1-e^{-j\Omega\, fs\, T_0\,(\cos(\phi)-1)}\right)}{1-e^{-j\Omega\, fs\, 2T_0}}\right|$$





Differential beamformer: White noise gain problem

- Typically microphones generate noise which is amplified by the differential beamformer



$$
\begin{aligned}
Y_0(e^{j\Omega}, \boldsymbol{r}) &= S(e^{j\Omega}) + B_0(e^{j\Omega}) \\
Y_1(e^{j\Omega}, \boldsymbol{r}) &= S(e^{j\Omega}) e^{-j\Omega\, fs\, \frac{d}{c}\cos(\phi)} + B_1(e^{j\Omega})
\end{aligned}
$$

- Desired signal component of the output signal $U(e^{j\Omega}, \boldsymbol{r})$ (before low-pass filtering):

$$U_s(e^{j\Omega}, \boldsymbol{r}(\phi = 0)) = S(e^{j\Omega}) \left[1 - e^{-j\Omega\, fs\, 2T_0}\right]$$: Spectrum

$$S_{U_s U_s}(\Omega) = S_{ss}(\Omega) \left|1 - e^{-j\Omega\, f_s\, 2T_0}\right|^2$$: Power spectral density

- Microphone noise signal component of the output signal $U(e^{j\Omega}, \boldsymbol{r})$ (before low-pass filtering):

$$U_b(e^{j\Omega}) = B_0(e^{j\Omega}) \left[1 + a e^{-j\Omega\, f_s\, T_0}\right] - B_1(e^{j\Omega}) \left[a + e^{-j\Omega\, f_s\, T_0}\right]$$: Spectrum

$$S_{U_b U_b}(\Omega) = S_{b_0 b_0}(\Omega) \left|1 + a e^{-j\Omega\, f_s\, T_0}\right|^2 + S_{b_1 b_1}(\Omega) \left|a + e^{-j\Omega\, f_s\, T_0}\right|^2$$: Power spectral density

Differential beamformer: White noise gain problem

- PSD of the desired signal component after beamforming:

$$S_{U_sU_s}(\Omega) \quad = \quad S_{ss}(\Omega) \left|1 - e^{-j\Omega\, f_s\, 2T_0}\right|^2$$

- PSD of the microphone noise component after beamforming:

$$S_{U_bU_b}(\Omega) \quad = \quad S_{b_0b_0}(\Omega) \left|1 + a e^{-j\Omega\, f_s\, T_0}\right|^2 + S_{b_1b_1}(\Omega) \left|a + e^{-j\Omega\, f_s\, T_0}\right|^2$$

Assumptions: Microphone noise power spectral density identical at both microphones

$$S_{b_0b_0}(\Omega) = S_{b_1b_1}(\Omega) = S_{bb}(\Omega)$$

$$\frac{S_{U_bU_b}(\Omega)}{S_{U_sU_s}(\Omega)} \quad = \quad \frac{S_{bb}(\Omega)}{S_{ss}(\Omega)} \, \frac{\left|1 + a e^{-j\Omega\, f_s\, T_0}\right|^2 + \left|a + e^{-j\Omega\, f_s\, T_0}\right|^2}{\left|1 - e^{-j\Omega\, f_s\, 2T_0}\right|^2}$$

$$\frac{S_{U_bU_b}(\Omega)}{S_{U_sU_s}(\Omega)} \quad = \quad \frac{S_{bb}(\Omega)}{S_{ss}(\Omega)} \, \frac{a^2 + 1 + 2a\, \cos(\Omega\, f_s\, T_0)}{1 - \cos(2\Omega\, f_s\, T_0)}$$

Gain of microphone noise
due to differential beamforming

Mic. noise to desired signal
power after beamforming

Mic. noise to desired signal
power before beamforming

Differential Beamformer:
Sub-band processing for better signal suppression

- **Microphone noise problem:**
  => Full band processing: strong amplification of low-frequency components of microphone noise

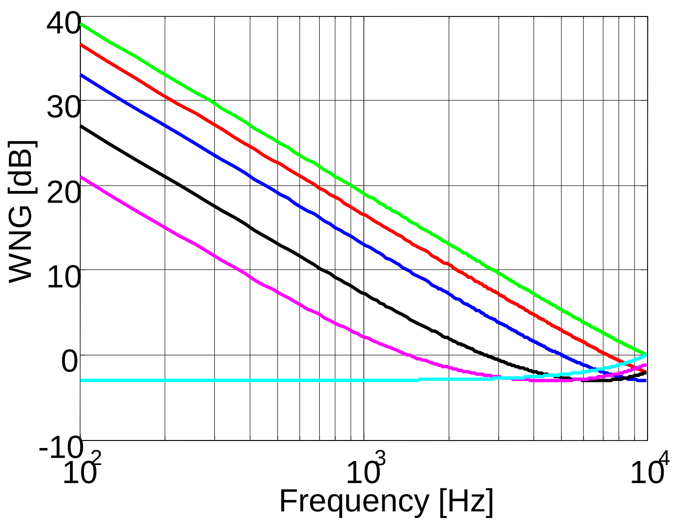
$$WNG(\Omega) = \frac{a^2 + 1 + 2a\,\cos(\Omega\, f_s\, T_0)}{1 - \cos(2\Omega\, f_s\, T_0)}$$

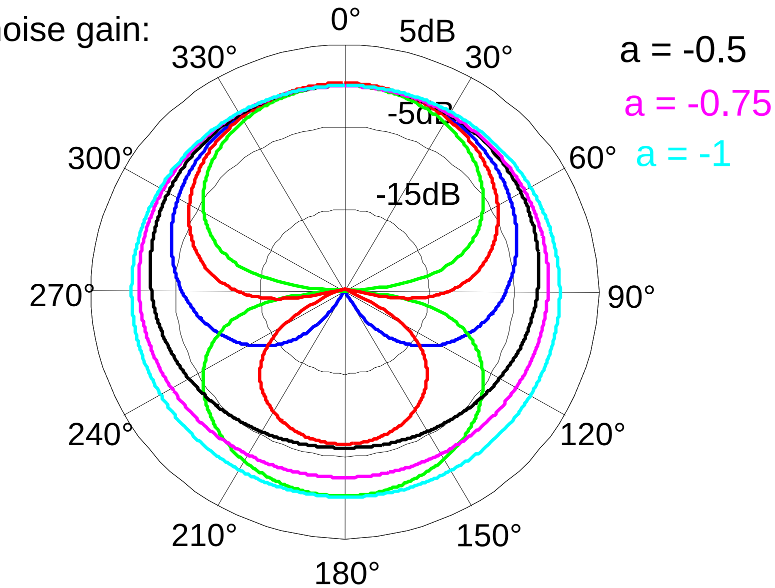
WNG:
white noise gain

- **Apparent solution:**
  - Signal decomposition in sub-bands; directional processing only in upper bands, else omni-dir
  - Problem: worst-case restrictions necessary

- **Alternative solution:** Choose negative values for steering parameter "*a*" down to *a = -1*
  => fading towards omni

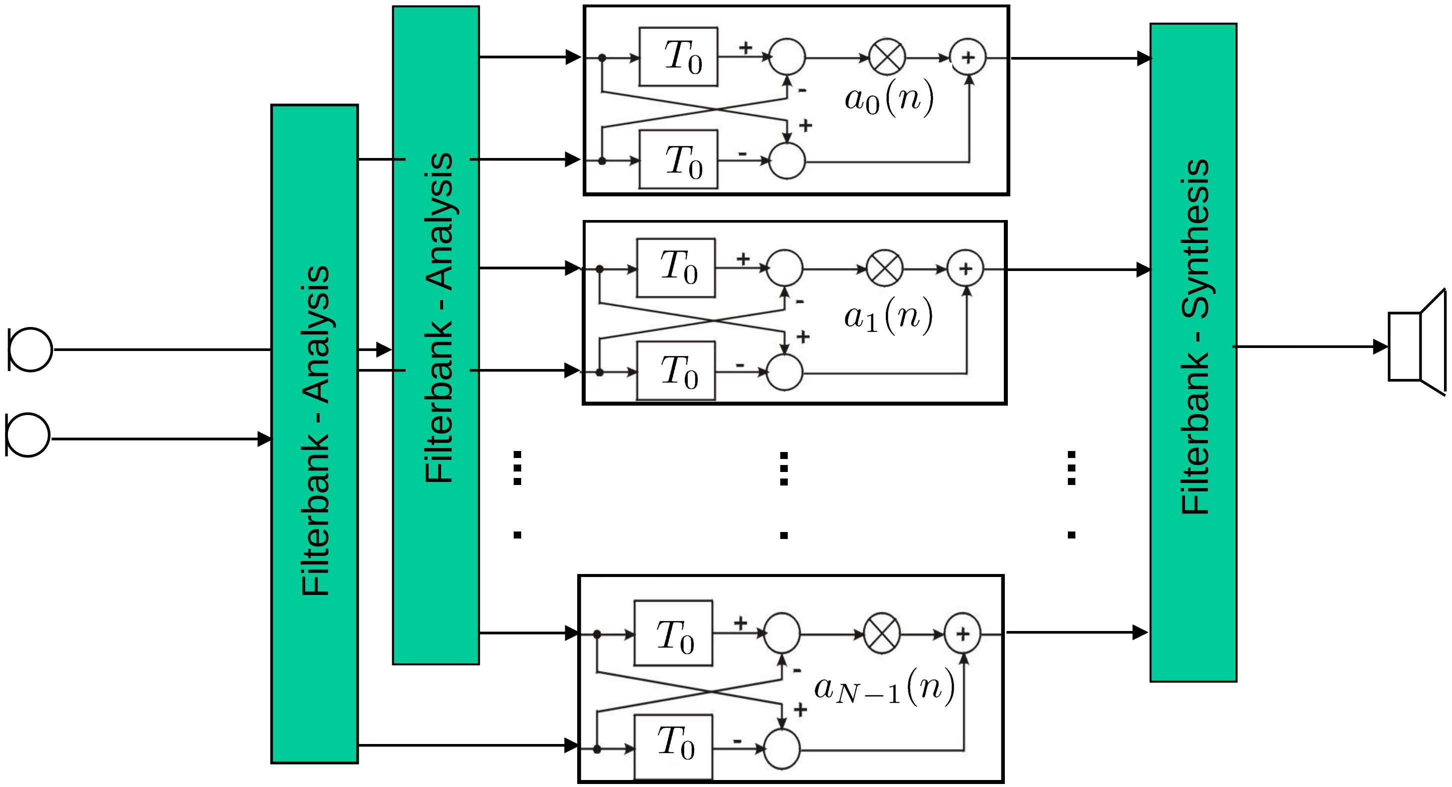
- Relation between directional performance & microphone noise gain:





Differential Beamformer: Sub-band processing

- Decomposition of microphone signals in frequency sub-bands
- Independent beamforming processing in the sub-bands

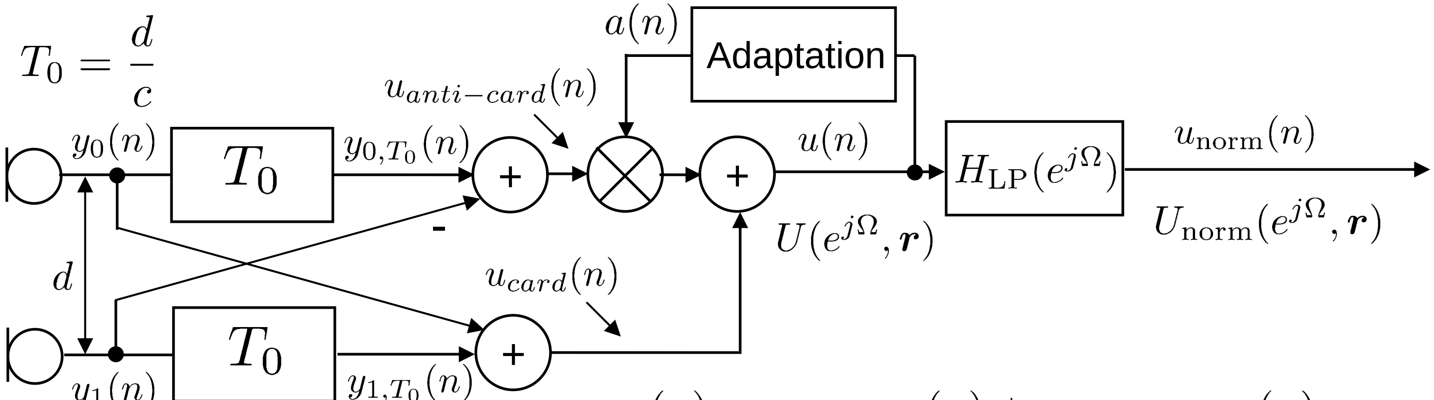


Adaptive differential beamformer (ADBF)

TECHNISCHE UNIVERSITÄT DARMSTADT

- The beam control with the weighting factor allows for a simple adaptation:

$$T_0 = \frac{d}{c}$$



$$\begin{aligned} u(n) &= u_{card}(n) + a\, u_{anti-card}(n) \\ &= y_0(n) - y_{1,T_0}(n) + a\left[-y_1(n) + y_{0,T_0}(n)\right] \end{aligned}$$

- The anti-card signal does not contain any target signal components
  => can be interpreted as “blocking path” [1]

- => Minimization of the power of $u(n)$ minimizes the interference power, i.e.
  no adaptation constraints are required.

[1] G.W. Elko and A.N. Pong: A simple adaptive first-order differential microphone, Proc. 1995 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 169-172, 1995

Adaptation procedure

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Minimization criterion:**

$$\mathrm{E}\{|u(n)|^2\} \underset{a=a_{\mathrm{opt}}}{\longrightarrow} \min$$

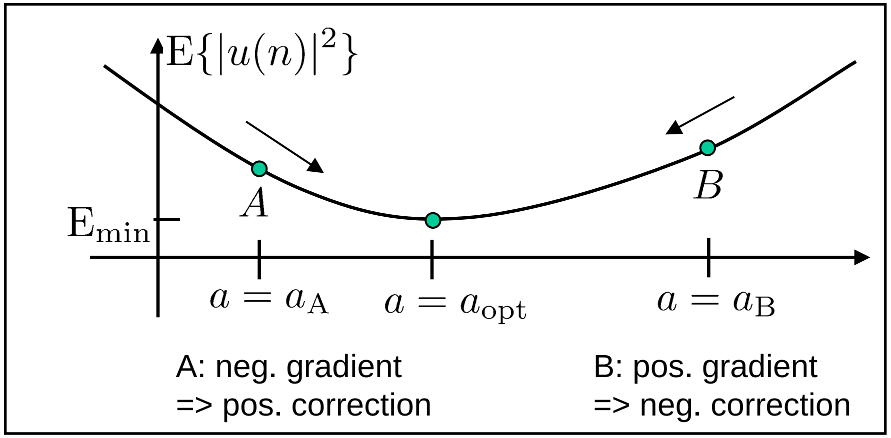
with:

$$u(n) = y_0(n) - y_{1,T_0}(n) + a\left[-y_1(n) + y_{0,T_0}(n)\right]$$

- **Iterative adaptation procedure:**

Correction in the direction of the negative gradient:

$$a(n+1) = a(n) - \frac{\mu}{2}\,\frac{\partial \mathrm{E}\{|u(n)|^2\}}{\partial a}$$



- Determine the gradient:

$$\frac{\partial \mathrm{E}\{|u(n)|^2\}}{\partial a} = 2\,\mathrm{E}\Big\{\left[-y_1(n) + y_{0,T_0}(n)\right] u(n)\Big\}$$

- Least mean square (LMS) procedure: Using instantaneous gradient values:

$$a(n+1) = a(n) + \mu\left[y_1(n) - y_{0,T_0}(n)\right] u(n)$$

- **Normalized LMS (NLMS) procedure** by the smoothed anti-card signal power:

$$\boxed{a(n+1) = a(n) + \mu\,\frac{y_1(n) - y_{0,T_0}(n)}{\sigma^2_{\mathrm{ac}}(n)}\,u(n)}$$

with: $y_{\mathrm{ac}}(n) = y_1(n) - y_{0,T_0}(n)$

$$\sigma^2_{\mathrm{ac}} = \overline{|y_{\mathrm{ac}}(n)|^2}$$

Adaptation to an overall noise minimum

- **Flexible solution:**
  Allow adaptation of parameter "a" down to -1:

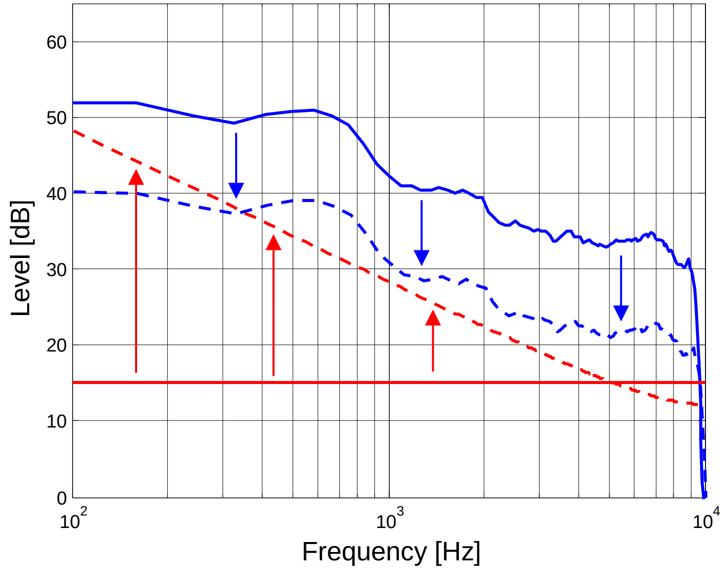
$$a(n+1) = a(n) + \mu \, \frac{y_1(n) - y_{0,T_0}(n)}{\sigma^2_{\mathrm{ac}}(n)} \, u(n)$$

3 components at the output:
- Desired signal (unmodified, indep. of a!)
- Residual ambient noise (reduced by DBF)
- Microphone noise (increased by DBF)

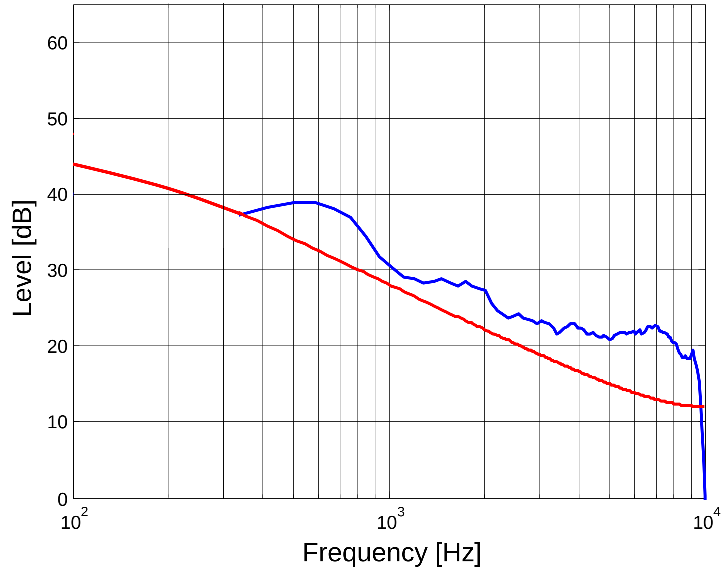
Adaptation automatically reaches the minimum of the sum of microphone and ambient noise at the BF output.

- Ambient noise power (blue) + microphone noise (red):

Before DBF (solid), after fixed DBF (dashed):



After adaptive differential beamformer (ADBF):

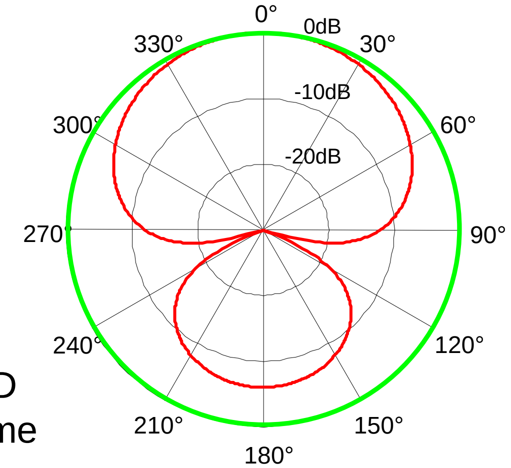


Beamforming: Characteristics

- **Directivity index (DI):**

$$\mathrm{DI}_{2\mathrm{D}}(e^{j\Omega}) = 10\,\log_{10}\left(\frac{\Phi(\Omega,\phi=0)}{\frac{1}{2\pi}\int_0^{2\pi}\Phi(\Omega,\phi)\,d\phi}\right)$$

Comparing the surface of the beampattern to the circle surface



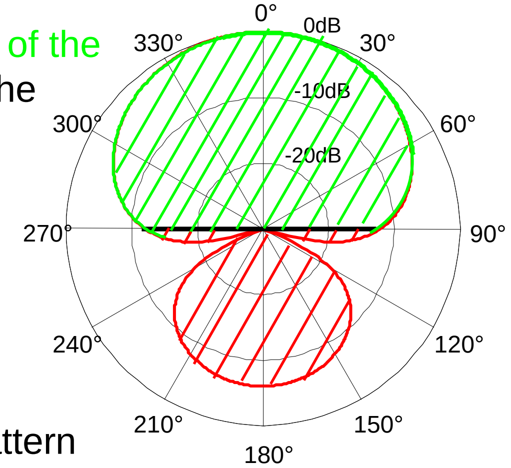
$$\mathrm{DI}_{3\mathrm{D}}(e^{j\Omega}) = 10\,\log_{10}\left(\frac{\Phi(\Omega,\phi=0,\theta=0)}{\frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}\Phi(\Omega,\phi,\theta)\sin(\theta)\,d\phi\,d\theta}\right)$$

Comparing the volume of the 3D beampattern to the sphere volume

- **Front-back ratio (FBR):**

$$\mathrm{FBR}_{2\mathrm{D}}(e^{j\Omega}) = 10\,\log_{10}\left(\frac{\int_{-\pi/2}^{\pi/2}\Phi(\Omega,\phi)\,d\phi}{\int_{\pi/2}^{3\pi/2}\Phi(\Omega,\phi)\,d\phi}\right)$$

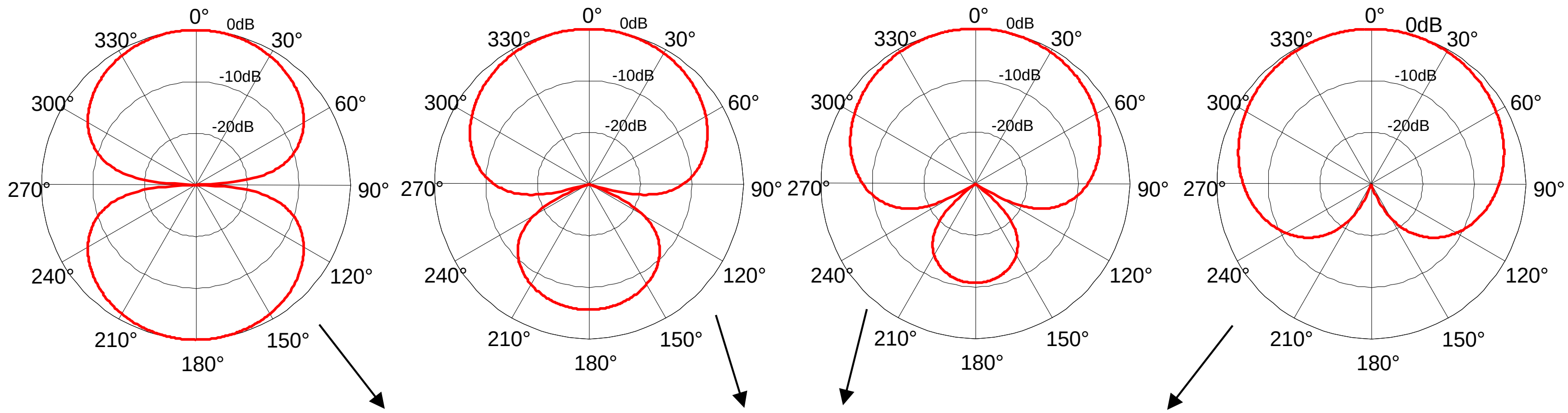
Comparing the surface of the fronal beampattern to the surface of the back beampattern



$$\mathrm{FBR}_{3\mathrm{D}}(e^{j\Omega}) = 10\,\log_{10}\left(\frac{\int_0^{\pi}\int_{-\pi/2}^{\pi/2}\Phi(\Omega,\phi,\theta)\sin(\theta)\,d\phi\,d\theta}{\int_0^{\pi}\int_{\pi/2}^{3\pi/2}\Phi(\Omega,\phi,\theta)\sin(\theta)\,d\phi\,d\theta}\right)$$

Comparing the volume of the frontal beampattern to the volume of the back beampattern

DI values of different differential beamformer setups



Name	Figure „8“	Hyper-cardioid	Super-cardioid	Cardioid
DI 2D	3.0 dB	4.6 dB	4.8 dB	4.3 dB
Di 3D	4.8 dB	6.0 dB	5.7 dB	4.8 dB
FBR 2D	0.0 dB	7.4 dB	12.1 dB	11.0 dB
FBR 3D	0.0 dB	8.5 dB	11.6 dB	8.6 dB
Delay ratio	0.0	0.333	0.577	1.0

AI-DI: Articulation index weighted DI

- In general the beampattern and the DI are frequency dependent values:

$$\mathrm{DI}_{3\mathrm{D}}(e^{j\Omega}) = 10\,\log_{10}\left(\frac{\Phi(\Omega,\phi=0,\theta=0)}{\frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}\Phi(\Omega,\phi,\theta)\sin(\theta)\,d\phi\,d\theta}\right)$$

- In order to calculate a frequency independent directional gain a mean value over the frequencies should be calculated.

  Since not all frequencies are equally important for the subjective perception, a weighting is performed:

- Continuous:

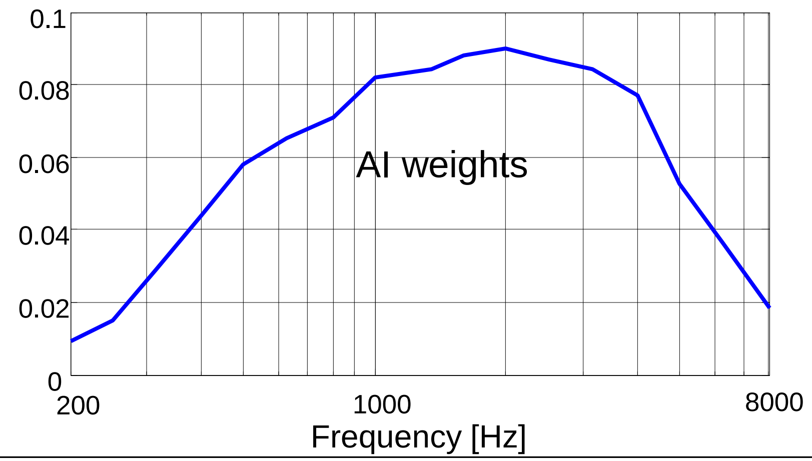
$$\text{AI-DI} = \int_0^{2\pi} w(\Omega)\,\mathrm{DI}(e^{j\Omega})\,d\Omega$$

with: $\int_0^{2\pi} w(\Omega)\,d\Omega = 1$

- Discrete:

$$\text{AI-DI} = \sum_{\mu=0}^{M} w(\Omega_\mu)\,\mathrm{DI}(e^{j\Omega_\mu})$$

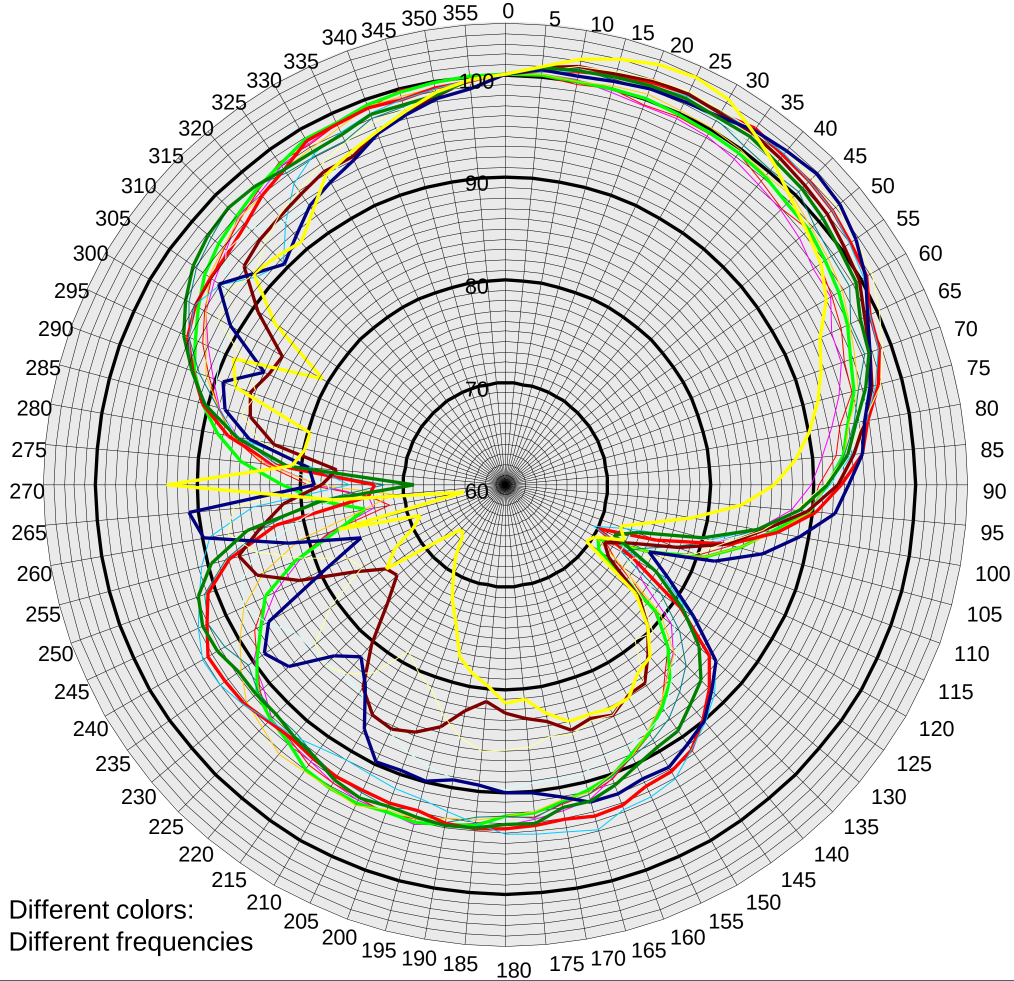
with: $\sum_{\mu=0}^{M} w(\Omega_\mu) = 1$

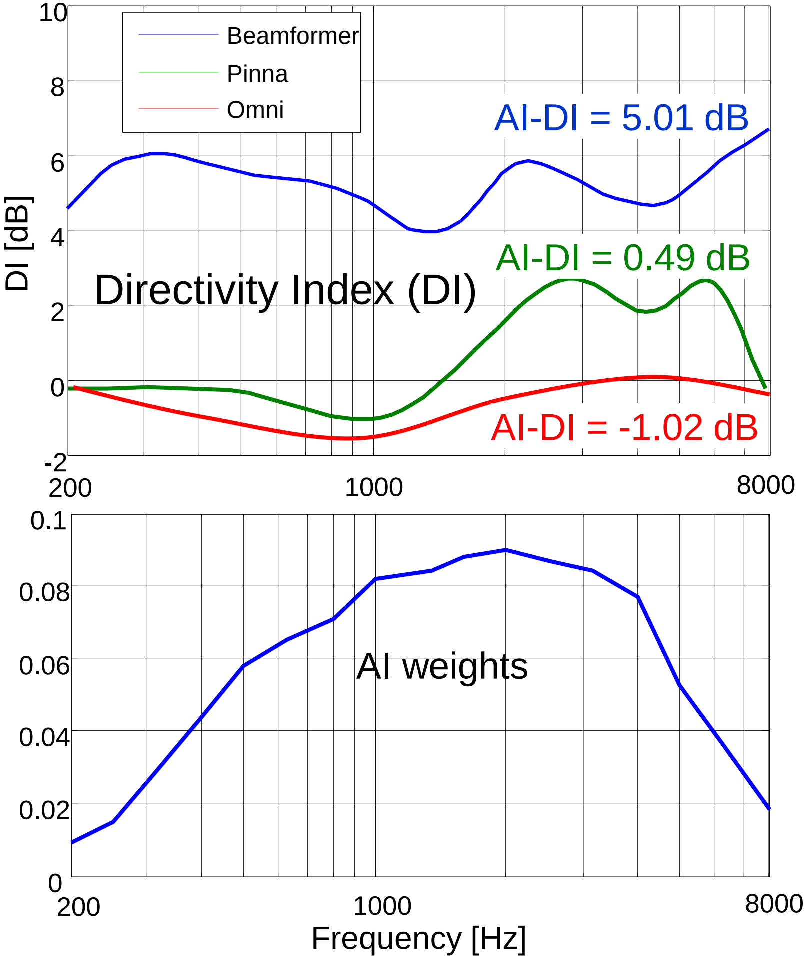


Differential Beamformer: Real measurements

- Frequency dependent beampattern, DI, and AI-DI:

AI-DI: the articulation index (AI) weighted frequency mean of the DI



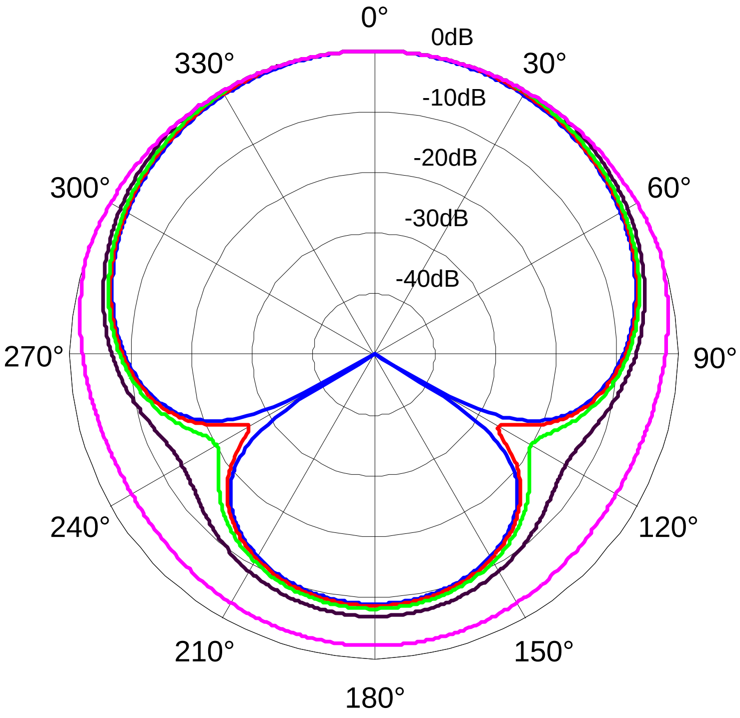


Differential Beamformer: Sensitivity to microphone mismatch

- So far, we assume that the microphones are equally sensitive, i.e. the result in the same output signal for the same acoustic excitation.
- In case of microphone mismatch (amplitude or phase) the directional performance reduces severely.
- Considering an amplitude mismatch:

Difference in microphone sensitivity:

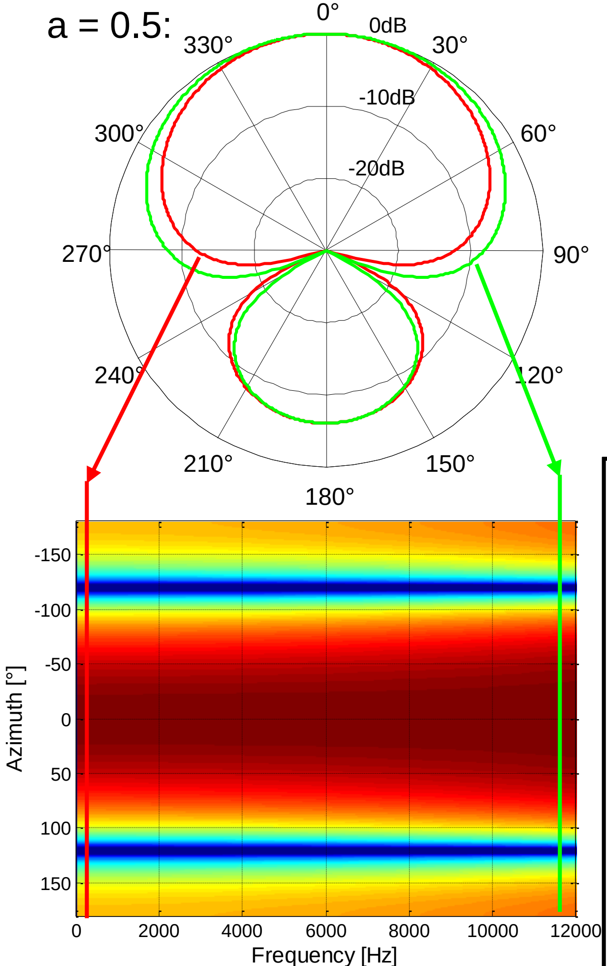
Blue:	0.0 dB
Red:	0.5 dB
Green:	1.0 dB
Black:	2.0 dB
Magenta:	5.0 dB

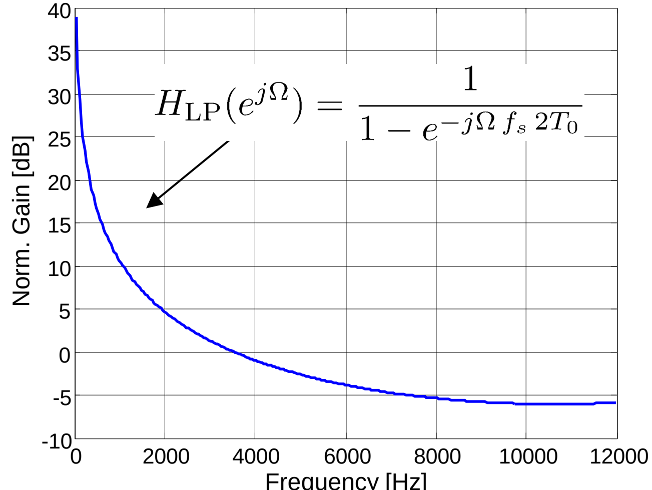


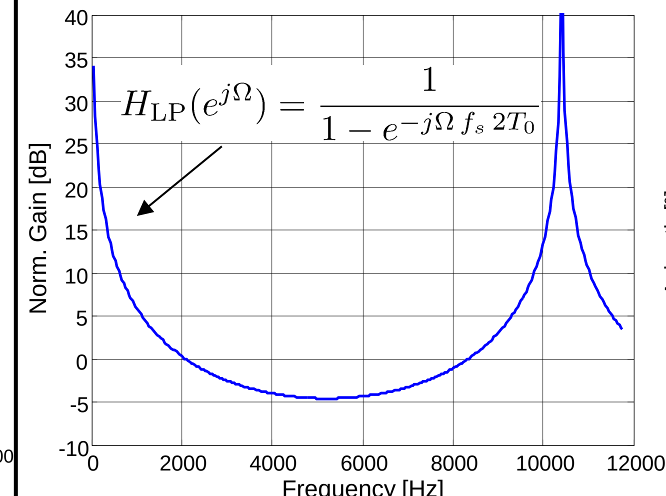
Differential Beamformer: Microphone distance limitation

d = 8 mm (fs = 24 kHz)

a = 0.5:

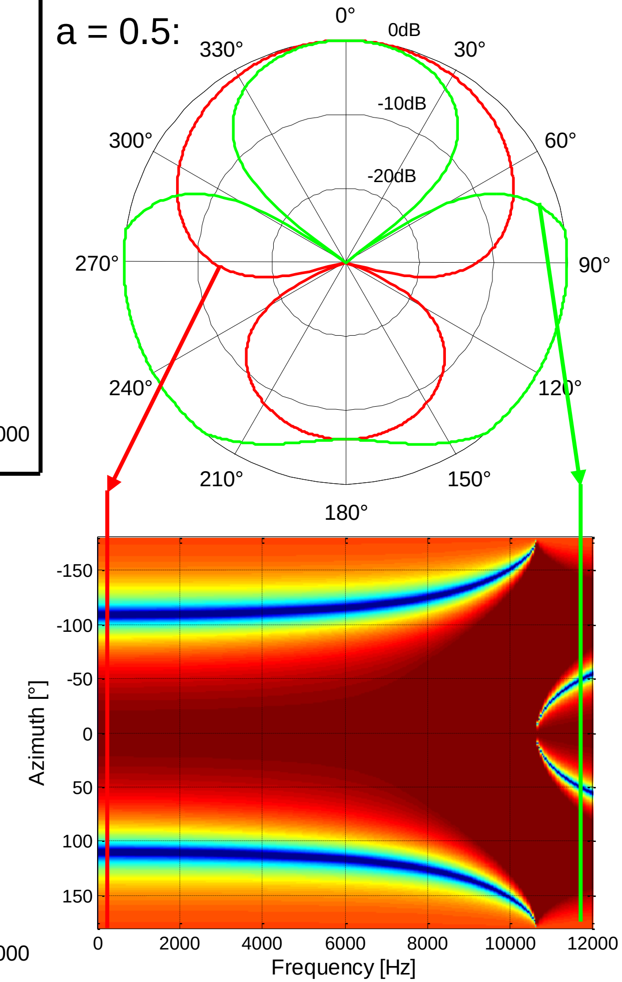






d = 16 mm (fs = 24 kHz)

a = 0.5:



Limitation of the differential beamformer

- As we have seen in the previous slide, dependent on the sampling frequency, there is a limitation of the microphone distance.

- For higher microphone distances additive beamformers (delay-and-sum) should be used.

- In simple terms: The distance should be smaller than half the wave length of the maximum frequency.

$$\Phi_{\mathrm{dB}}(\Omega,\boldsymbol{r}) = 20\,\log_{10}\left|\frac{\sin\left(\frac{1}{2}\Omega\,fs\,(\frac{d}{c}\cos(\phi)+T)\right)}{\sin\left(\frac{1}{2}\Omega\,f_s\,(\frac{d}{c}+T)\right)}\right|$$

Avoid zeros in the denominator of the directivity pattern:

$$\begin{aligned}
\frac{1}{2}\Omega_{\max}\,f_s\,(\frac{d}{c}+T_{\max}) &\leq \pi \\
\frac{1}{2}\pi f_s\,2\frac{d}{c} &\leq \pi \\
f_s\frac{d}{c} &\leq 1
\end{aligned}$$

$$\boxed{d \leq \frac{c}{f_s} = \frac{c}{2f_N} = \frac{\lambda_{\min}}{2}}$$

with: $T_{\max} = d/c$

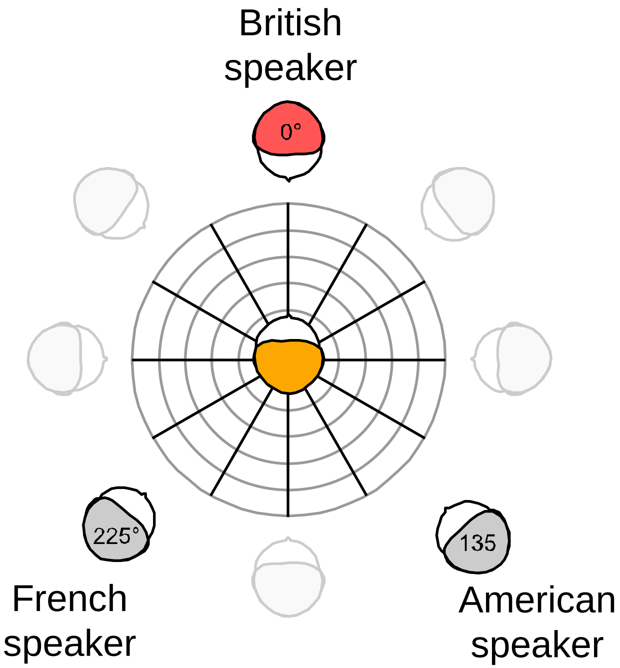
$\Omega_{\max} = \pi$

f_s : Sampling frequency

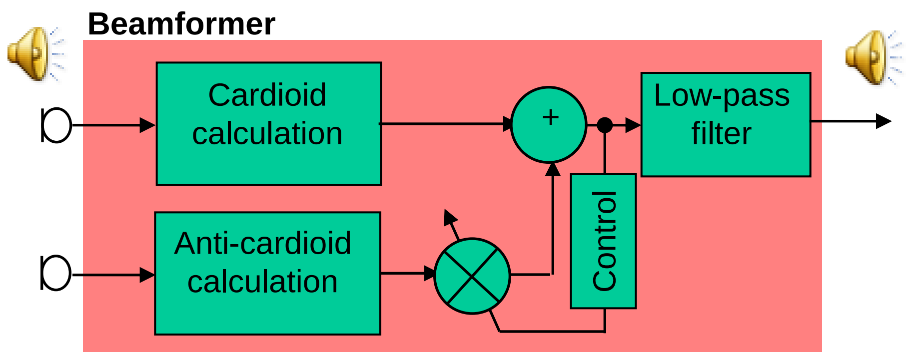
f_N : Nyquist frequency

Sound example

- **Setup:**



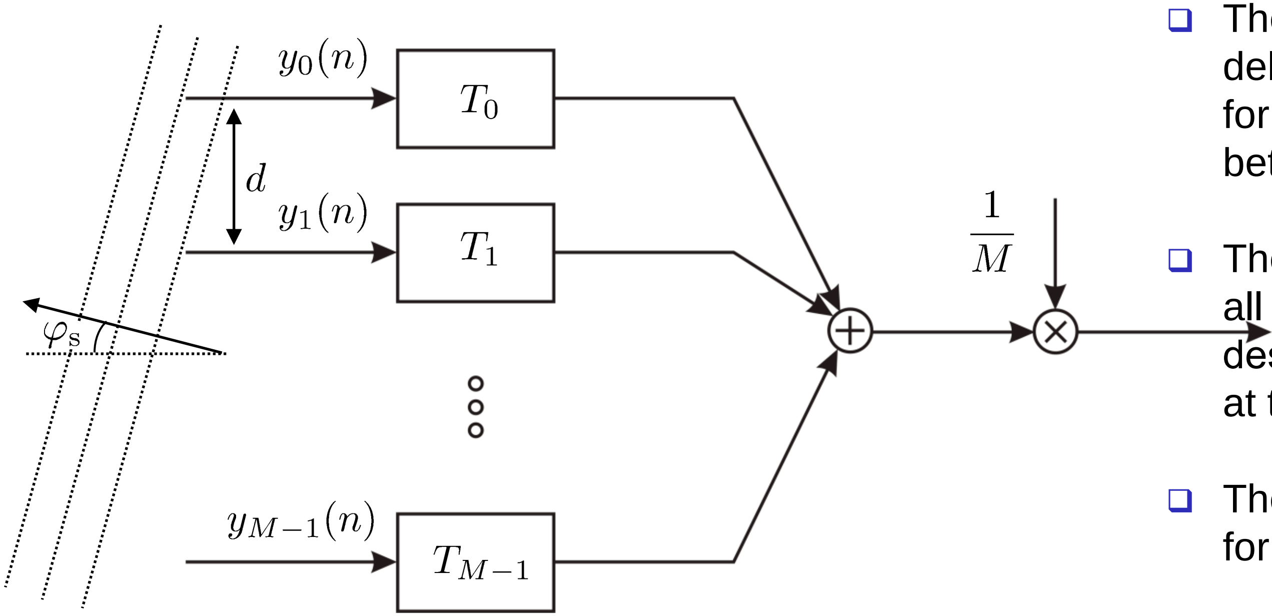
- **Sounds:**



Delay-and-sum beamformer

Delay-and-sum beamformer

Basic structure:



Delay compensation

- The microphone signals are delayed in order to compensate for the propagation delays between the microphones.
- The delays are chosen such that all microphone signals from a desired source are „in phase" at the summation point.
- The weighting (1/M) equalizes for the target signal level.
- Interference components arriving from other directions than the target signal are attenuated since they are not summed „in phase".

Spatial frequency response:

$$G_{\mathrm{BF}}(e^{j\Omega}, \boldsymbol{r}) = \frac{1}{M} \sum_{i=0}^{M-1} e^{-j\Omega\, fs\, \left(\frac{d\, i}{c}\, \sin(\varphi_{\mathrm{S}}) + T_i\right)}$$

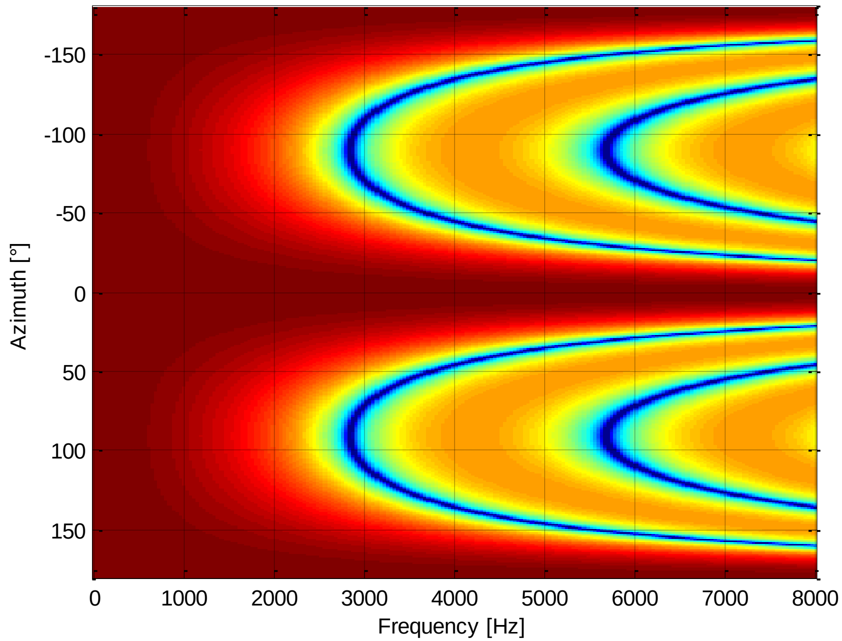
φ_{S}: angle of arrival of the desired source

Delay-and-sum beamformer

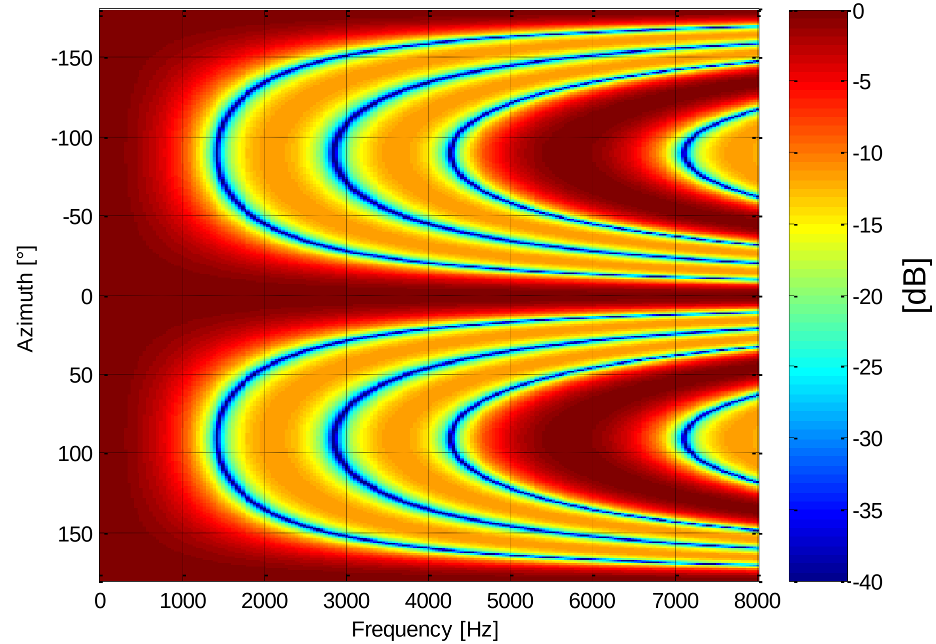
Examples for a beampattern (no delay compensation, broadside target signal direction):

$$\Phi_{\mathrm{dB}}(\Omega, \boldsymbol{r}) = 10\,\log_{10}\left|G_{\mathrm{BF}}(e^{j\Omega}, \boldsymbol{r})\right|^2 = 10\,\log_{10}\left|\frac{1}{4}\sum_{i=0}^{3} e^{-j\Omega\, f s\, \left(\frac{d\,i}{c}\,\sin(\varphi_{\mathrm{s}})\right)}\right|^2$$

d = 3 cm; 4 microphones with an omni-directional beampattern



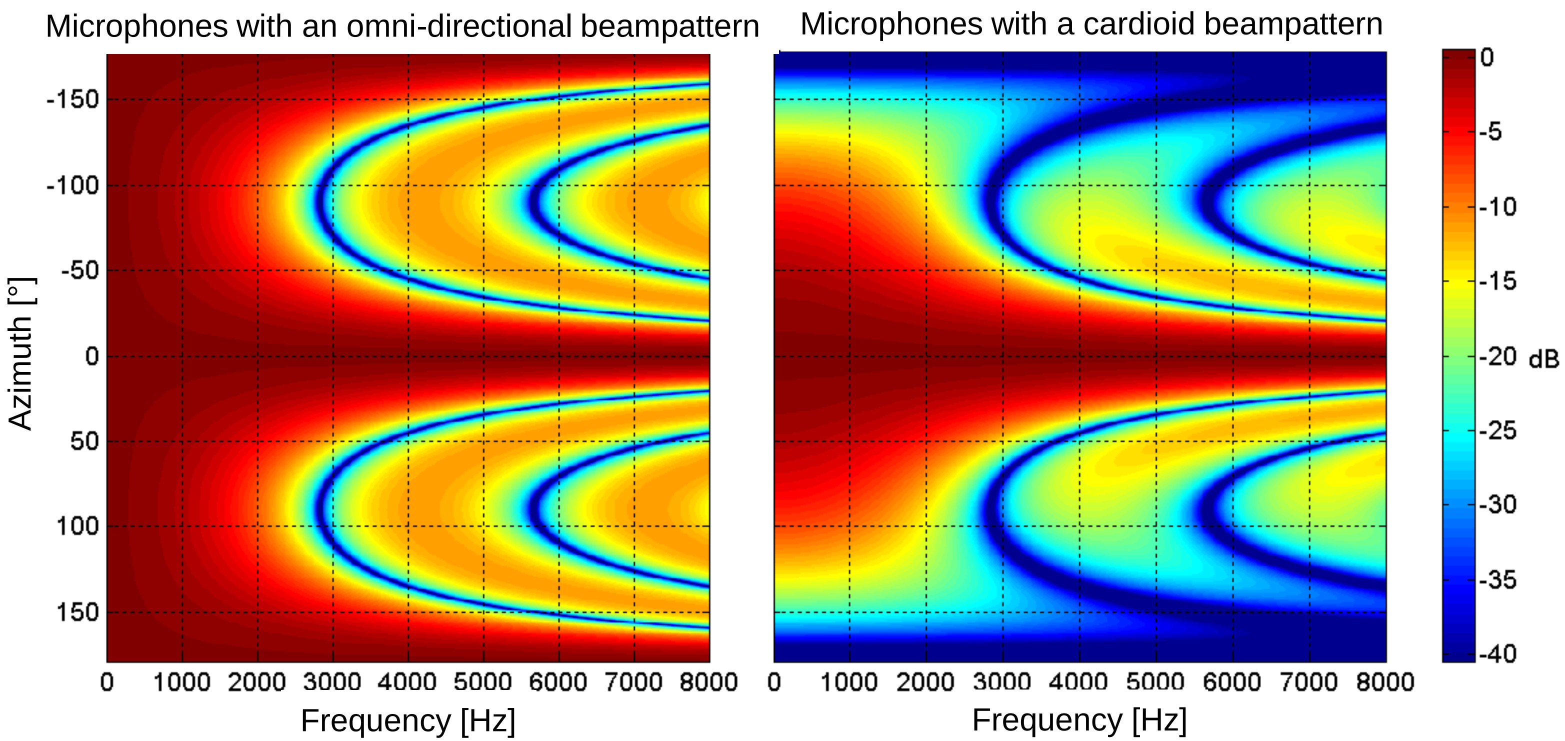
d = 6 cm; 4 microphones with an omni-directional beampattern



- Summation of 4 microphone signals with a distance of 3 or 6 cm each.
- A simple summation and weighting with ¼.

Delay-and-sum beamformer

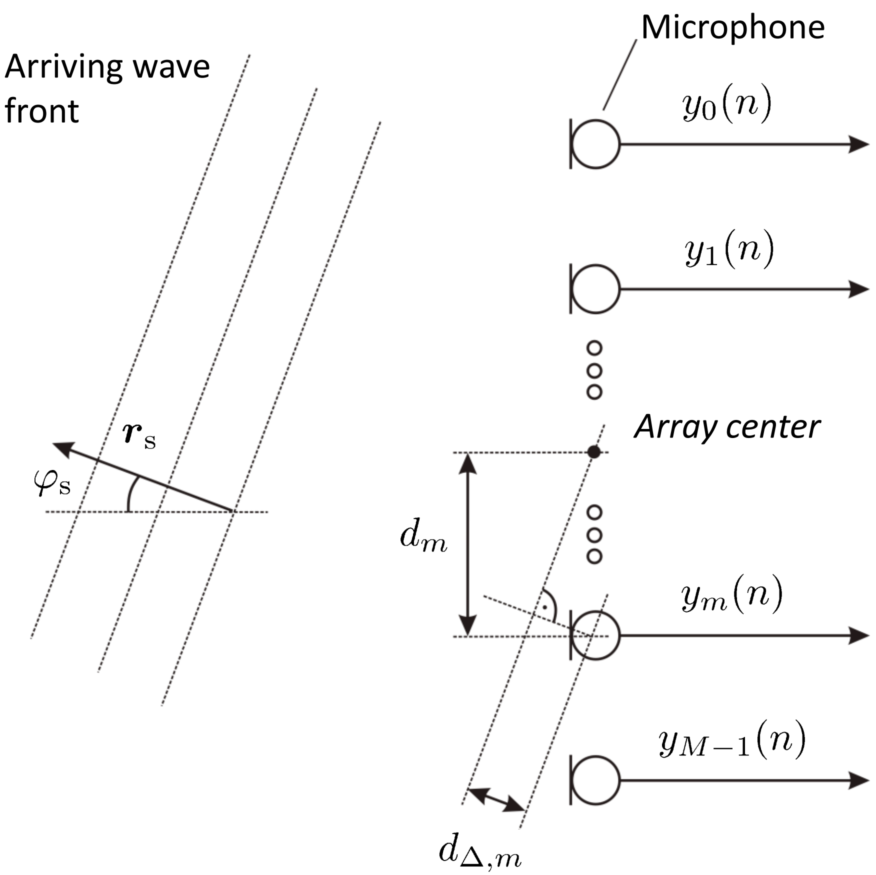
Examples for a beampattern (no delay compensation, broadside target signal direction):



- Summation of 4 microphone signals with a distance of 3 cm each.
- A simple summation and weighting with ¼.

Delay-and-sum beamformer

Calculation of the delay compensation:



- For a linear array with constant microphone distance d_Mic the distance of the m-th microphone to the center is following:

$$d_m = \|\boldsymbol{r}_m\| = \left| m - \frac{M-1}{2} \right| d_\mathrm{Mic}.$$

- The time delay of a plane wave front signal is following for the m-th microphone:

$$t_m = \frac{d_{\Delta,m}}{c} = \frac{d_m \, \sin(\varphi_\mathrm{s})}{c}.$$

- Resulting in the following latency (in samples):

$$\tau_m = t_m \, f_\mathrm{s}.$$

Optimal solution:

$$G_{\mathrm{opt},m}(e^{j\Omega}) \;=\; e^{-j\Omega\tau_m} \;\; \mathrm{mit} \;\; -\pi < \Omega \leq \pi \qquad \bullet\!\!-\!\!\!-\!\!\!-\!\!\circ \qquad g_{\mathrm{opt},\,m,\,i} \;=\; \frac{\sin\left(\pi(i-\tau_m)\right)}{\pi(i-\tau_m)}$$

Realization in the time domain (example):

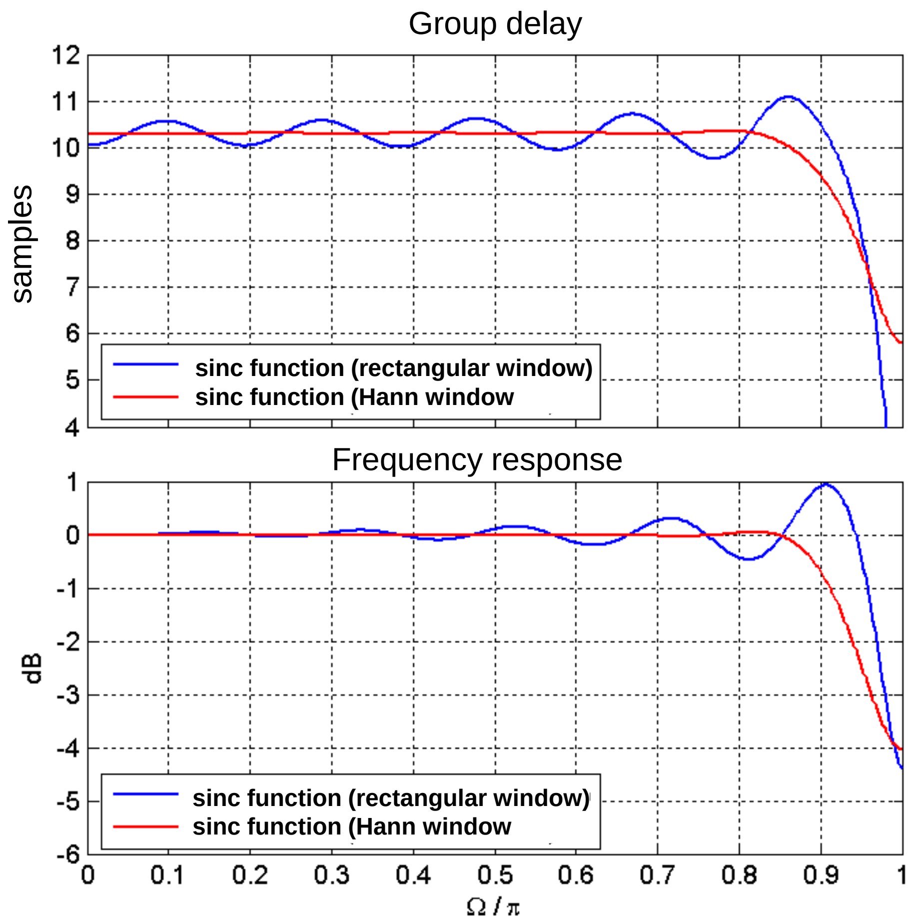
- Time shift of the impulse response (in order to obtain a causal solution) followed by a windowing:

$$g_{m,i} \;=\; g_{\mathrm{opt},m,\,i\,-\,N/2}\, w_i.$$

- A Hann window can be chosen:

$$w_i = \begin{cases} K\left[1+\sin\left(\frac{2\pi}{N}(i-\frac{N}{2})\right)\right], & \text{for } i \in \{0,\,...,\,N-1\}, \\ 0, & \text{else} \end{cases}$$

Realization in the time domain (example):

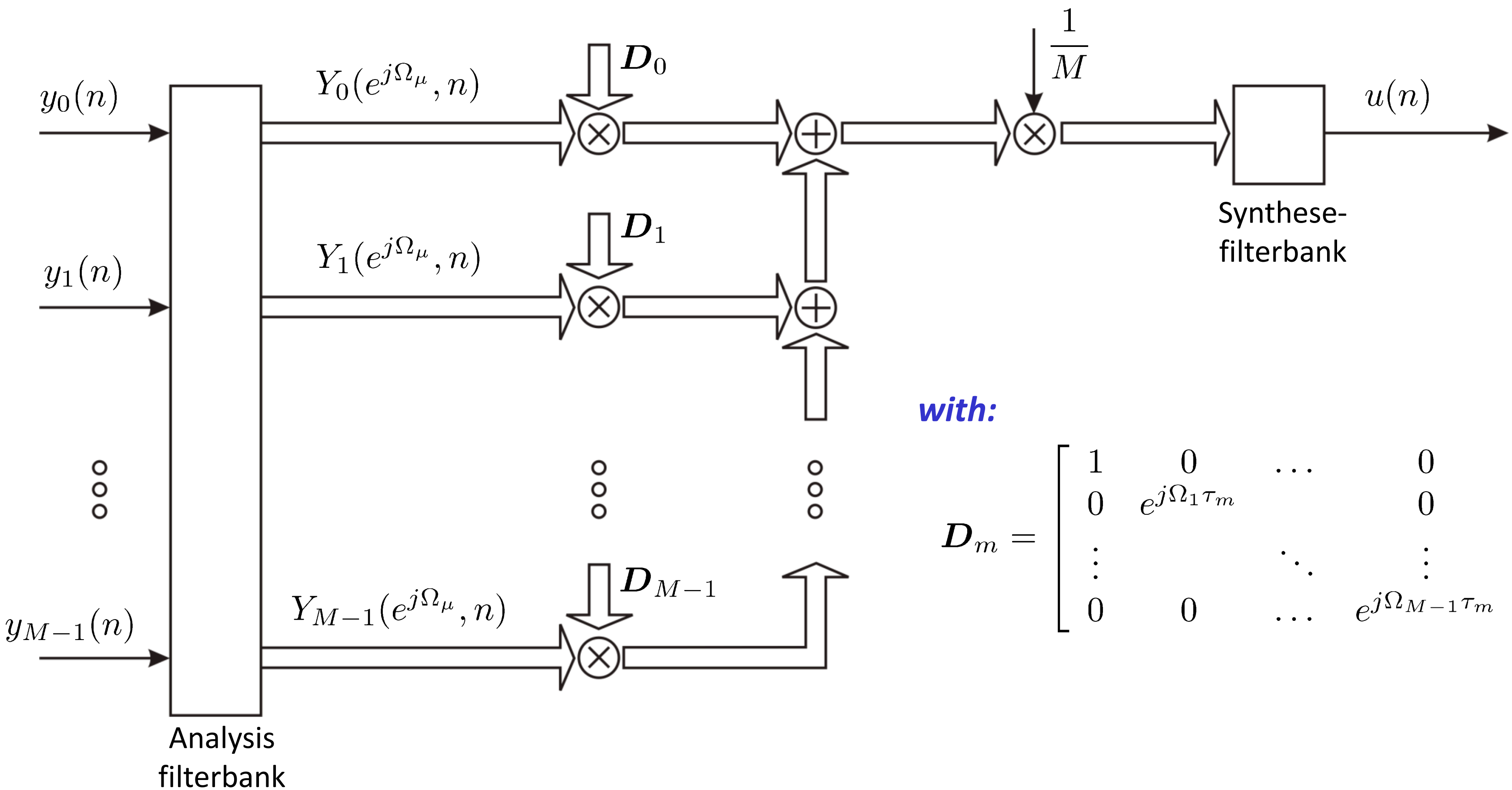


- Target: Design a filter with a group delay of 10.3 samples. (=> fractional delay filter)
- Constraint: 21 tap FIR filter.

Para trabalhar com um filtro desse tipo, a frequencia de amostragem deve ser maior do que a original, para que possamos descartar uma parte (0,75 para cima)

Delay-and-sum beamformer

Realization in the frequency domain:



with:

$$\boldsymbol{D}_m = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & e^{j\Omega_1 \tau_m} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & e^{j\Omega_{M-1} \tau_m} \end{bmatrix}$$

Summary & Outlook

TECHNISCHE UNIVERSITÄT DARMSTADT

Summary

- General setup of beamformers
- Endfire and broadside setups
- Differential beamformer (endfire)
  - Typical beamformer characteristics: Beampattern, Directivity index, Front-back-ratio, White noise gain
  - Adaptation possibilities.
- Delay & sum beamformer (broadside)
- Fractional delay compensation

Outlook to 2nd part:

- Beamformer Part II:
  - Filter-and-sum beamformer
  - Minimum Variance Distortionless Response (MVDR) beamformer
  - Multi-channel Wiener Filter
  - Interference compensation
  - Audio examples and results