Digital Audio and Speech Processing (IN4182)

An Introduction to Multi-Microphone Speech Enhancement

Richard C. Hendriks

12/5/2020

TU Delft

Delft University of Technology

Speech Enhancement - Project

- Project is compulsory and carried out in groups of 2-3 students
- Evaluation is done during the exam (hand in report)

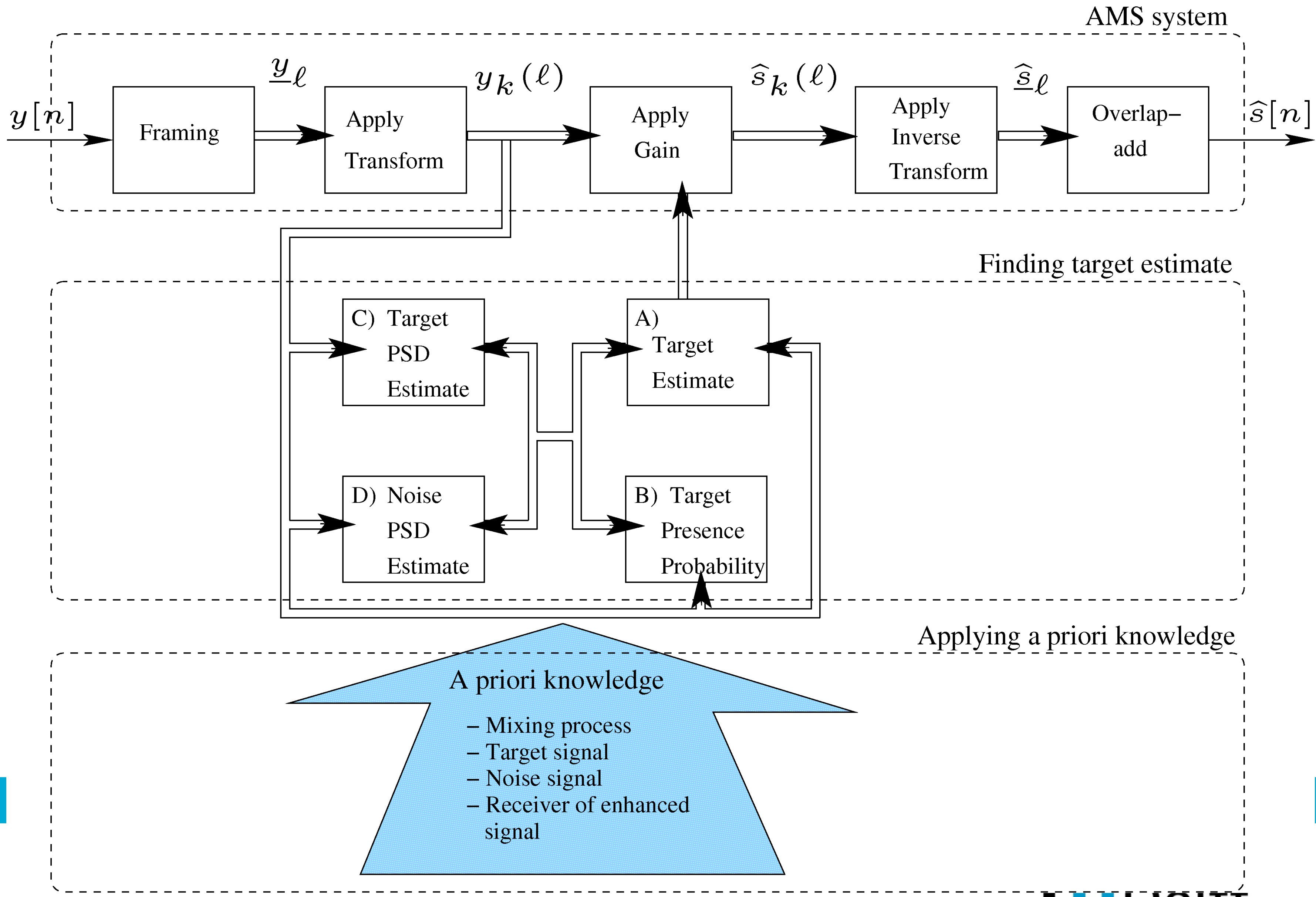
Project:

- Design and build at least a single-microphone speech enhancement system. You are free to extend this to a multi-microphone system.
  - Use matlab (or simulink)
  - The speech enhancement system should consist of a gain function, noise PSD estimator and speech PSD estimator.
  - Perform an evaluation of the speech enhancement system

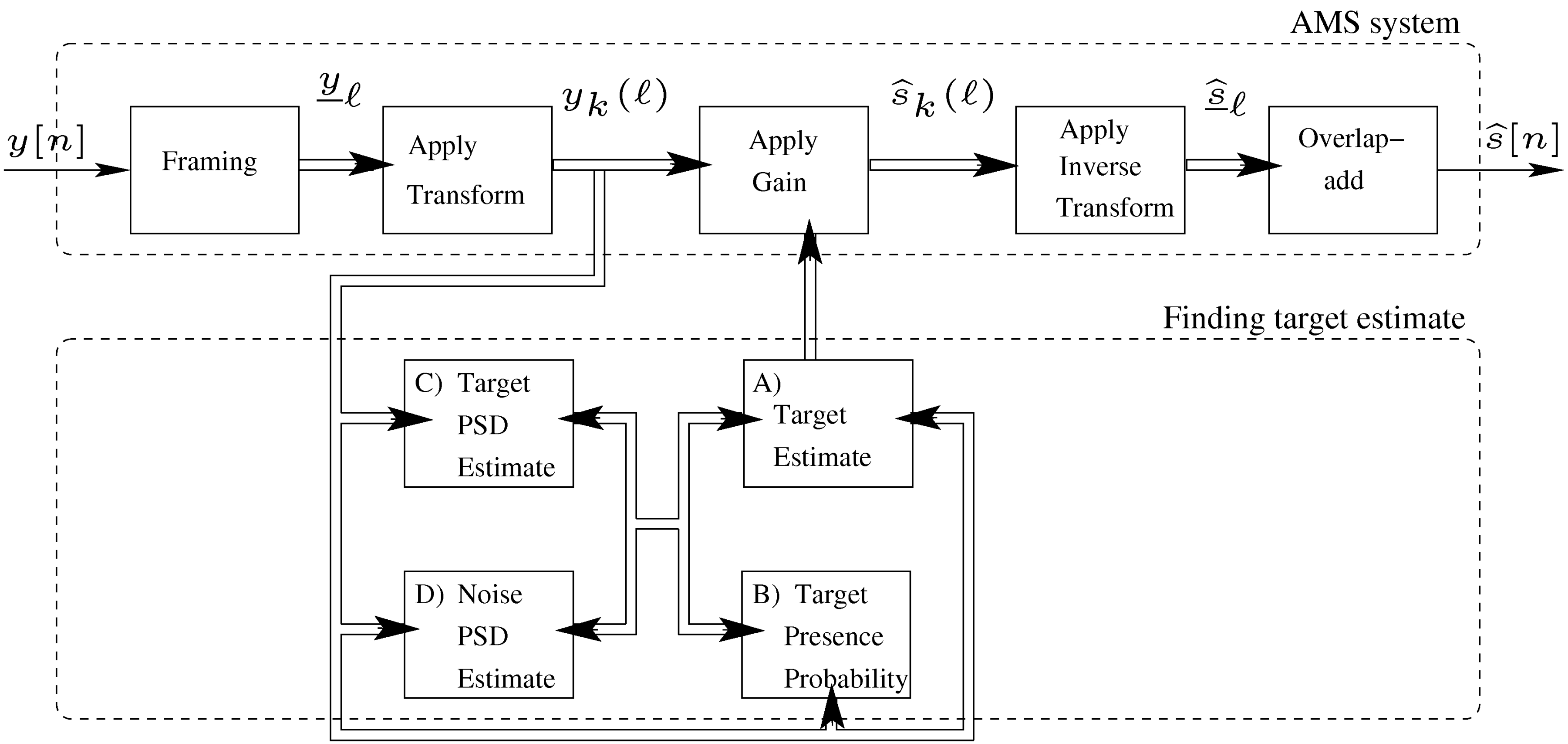
  Bonus:

  - Implement a multi-microphone system

Overview of single-channel NR algorithm



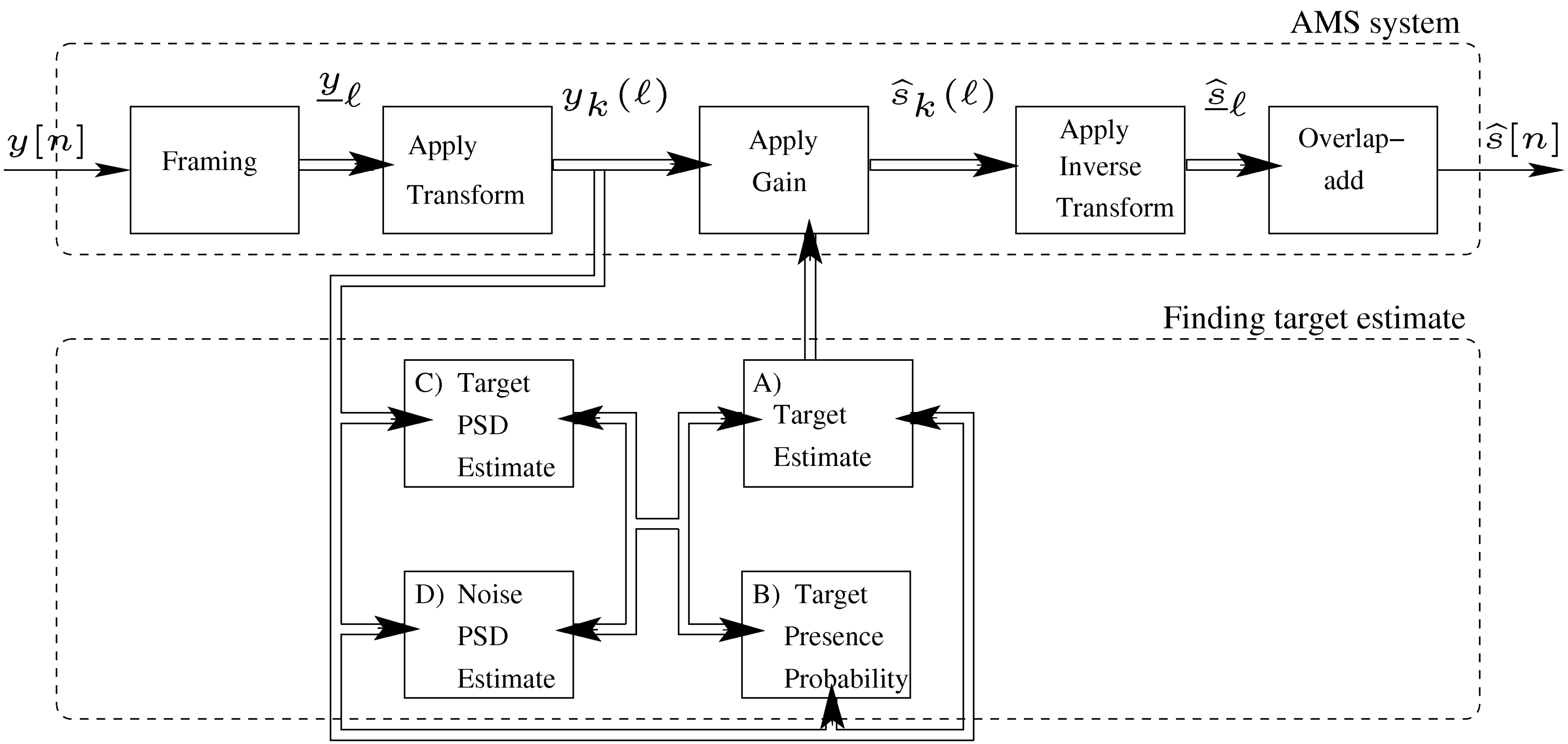
Overview of single-channel NR algorithm



Target Estimate

- Wiener gain: $\hat{s}_k(l) = \frac{\sigma_S^2}{\sigma_S^2+\sigma_N^2} y_k(l)$
- $\hat{s}_k(l) = E\left[S|y\right] = g\left(\sigma_N^2, \sigma_S^2, y, \nu, \gamma\right) y_k(l)$
- power spectral subtraction

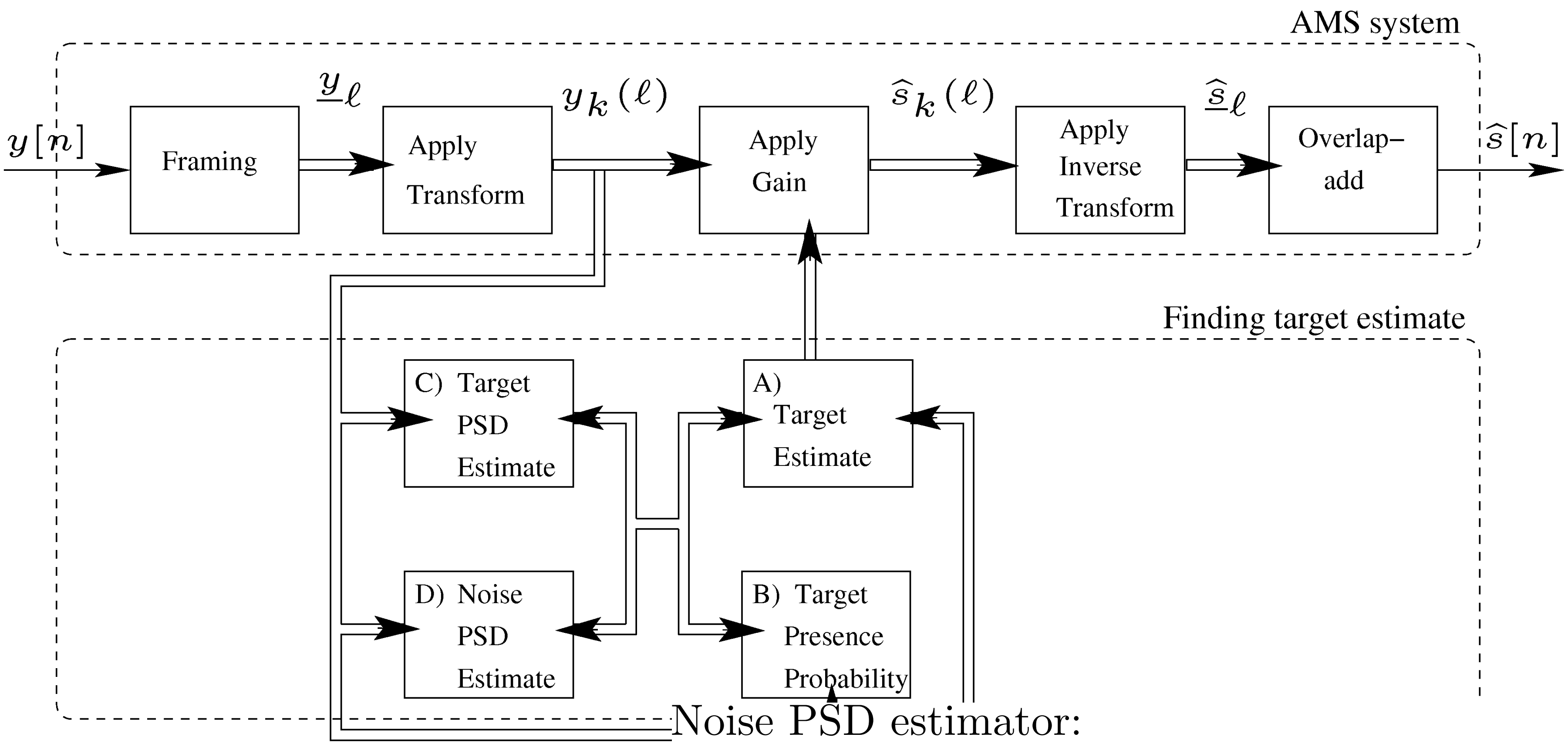
Overview of single-channel NR algorithm



Target (speech) PSD Estimator:

- Maximum likelihood (based on Bartlett estimate)
- Decision-directed approach

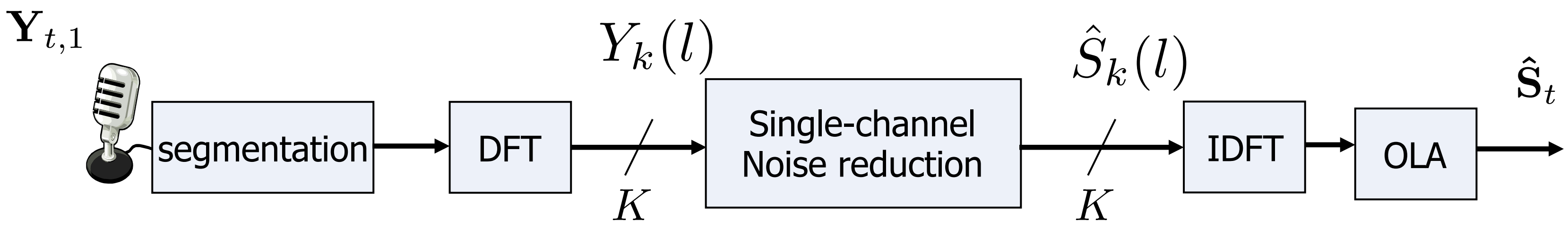
Overview of single-channel NR algorithm



Noise PSD estimator:

- Voice activity detector
- Minimum statistics
- MMSE based with speech presence uncertainty.

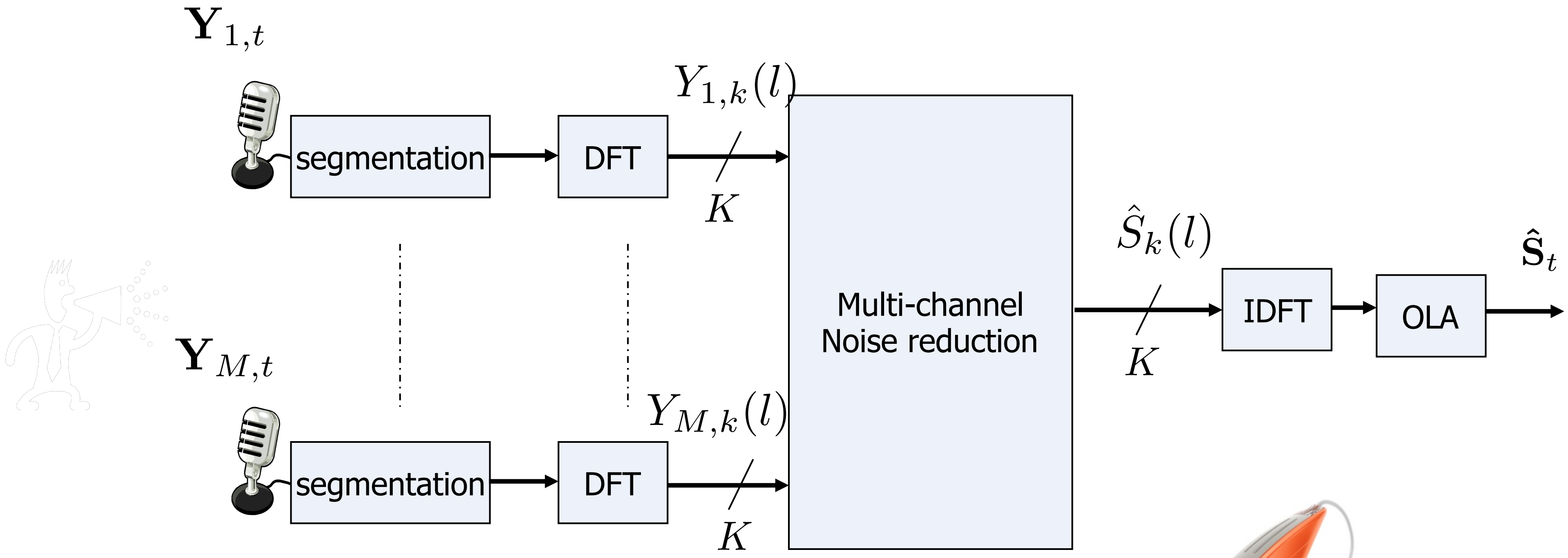
Single-Mic. Noise Reduction



- With a single-microphone noise reduction algorithm, filtering is limited to temporal-spectral filtering.

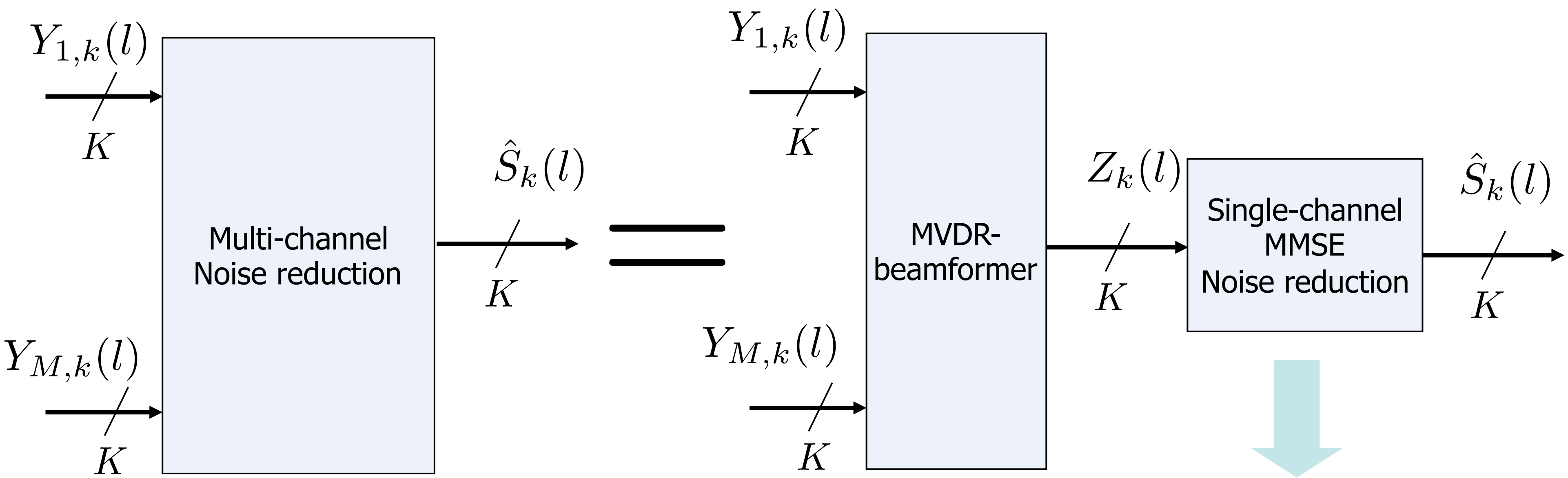
- If more microphones are available, it is also possible to sample the sound field spatially (rather than only temporal sampling). This in addition allows to perform spatial filtering.

Multi-Microphone Noise Reduction



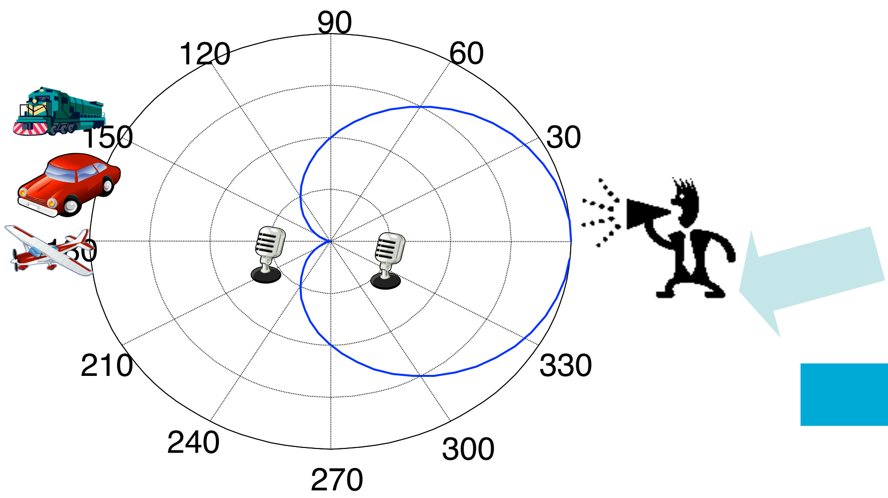
TU Delft

Multi-Microphone Noise Reduction



Temporal filtering

Spatial filtering



Beamformer response (similar to frequency response, but now a function of angle instead of frequency.)

Today:

- Concept of beamforming
- Signal models
- Derivation of multi-microphone noise reduction algorithms:
  - Sum & delay beamformer
  - MVDR beamformer
  - Multi-channel Wiener filter
  - LCMV beamformer
- Room impulse response and the acoustic transfer function

Spatial Sampling

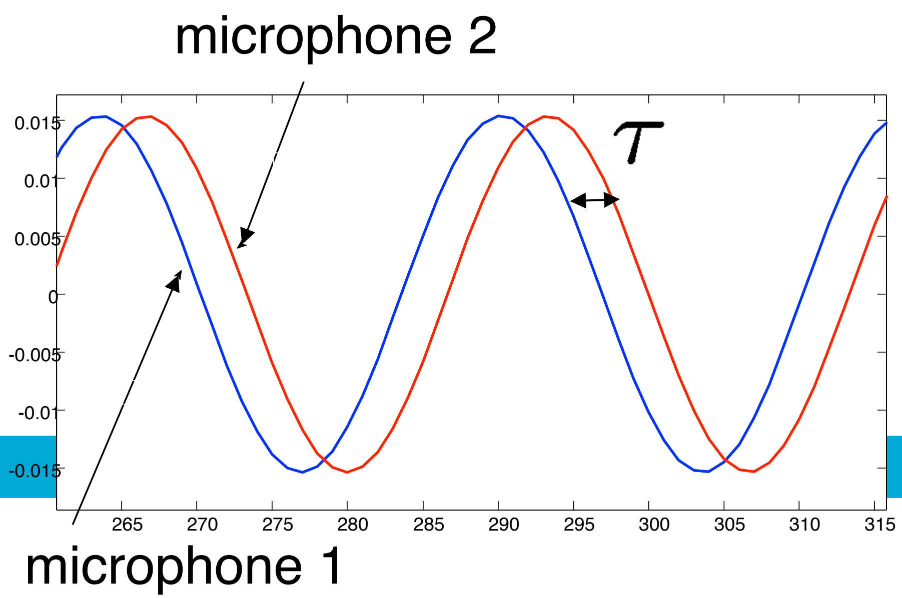
An important aspect of beamforming is the fact that by putting multiple microphones in a space, we sample the sound field.

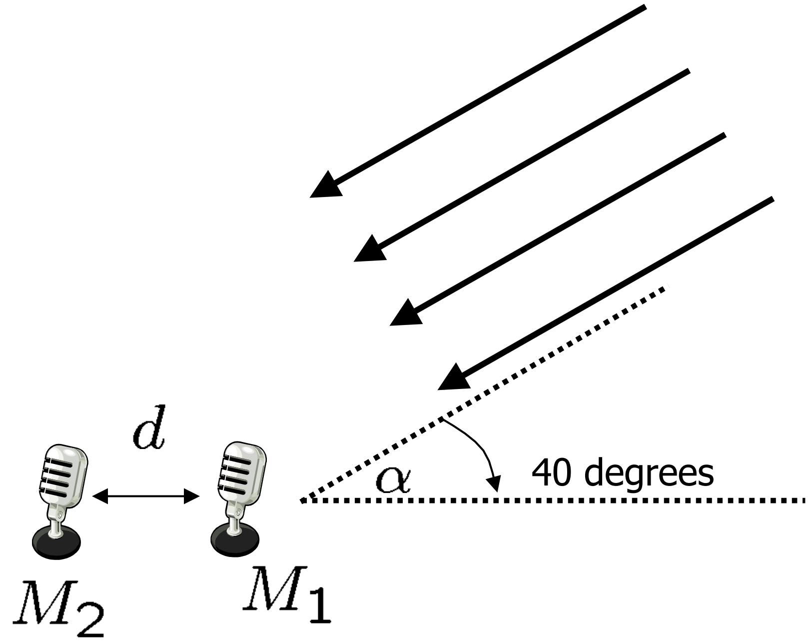
This is analogous to temporal sampling. Hence, we have to deal with similar concepts as with temporal filtering, e.g.,

- Spatial aliasing
- Beamformer response (frequency response)
- Correlation between samples
- ...

Concept of Beamforming

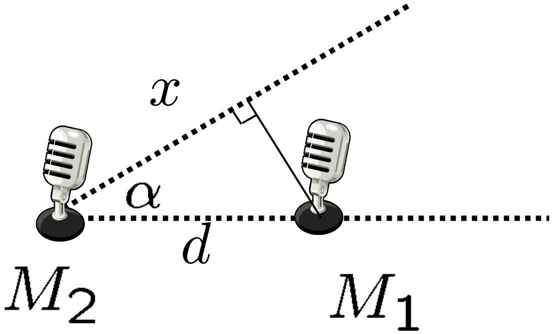
- Consider a sinusoidal source at 40 degrees of a dual microphone array (d=0.17 m).
- The sound source is in the far field (sound waves can be considered parallel)





TU Delft

Concept of Beamforming



$$x = \frac{\tau}{f_s} c$$

$$\tau = \cos(\alpha) \frac{d}{c} f_s$$

- $\alpha = 40$ degrees
- $f_s = 8000$ Hz
- $d = 0.17$ m
- $c = 340$ m/s

$\tau = 3.06$ samples

Non-integer shifts!

13

Concept of Beamforming – Freq. domain

Notice that delays are typically non-integer. To efficiently implement non-integer delays, beamformers can be implemented in the frequency domain. Notice:

- The response is frequency dependent. (In reality we do not have a single sinusoid, but broadband signals each experiencing a frequency dependent delay).
- We have to deal with spatial aliasing (the equivalent of temporal aliasing): $d < \frac{1}{2}\lambda_{min} < \frac{1}{2}\frac{c}{\frac{1}{2}f_s} = \frac{c}{f_s}$.

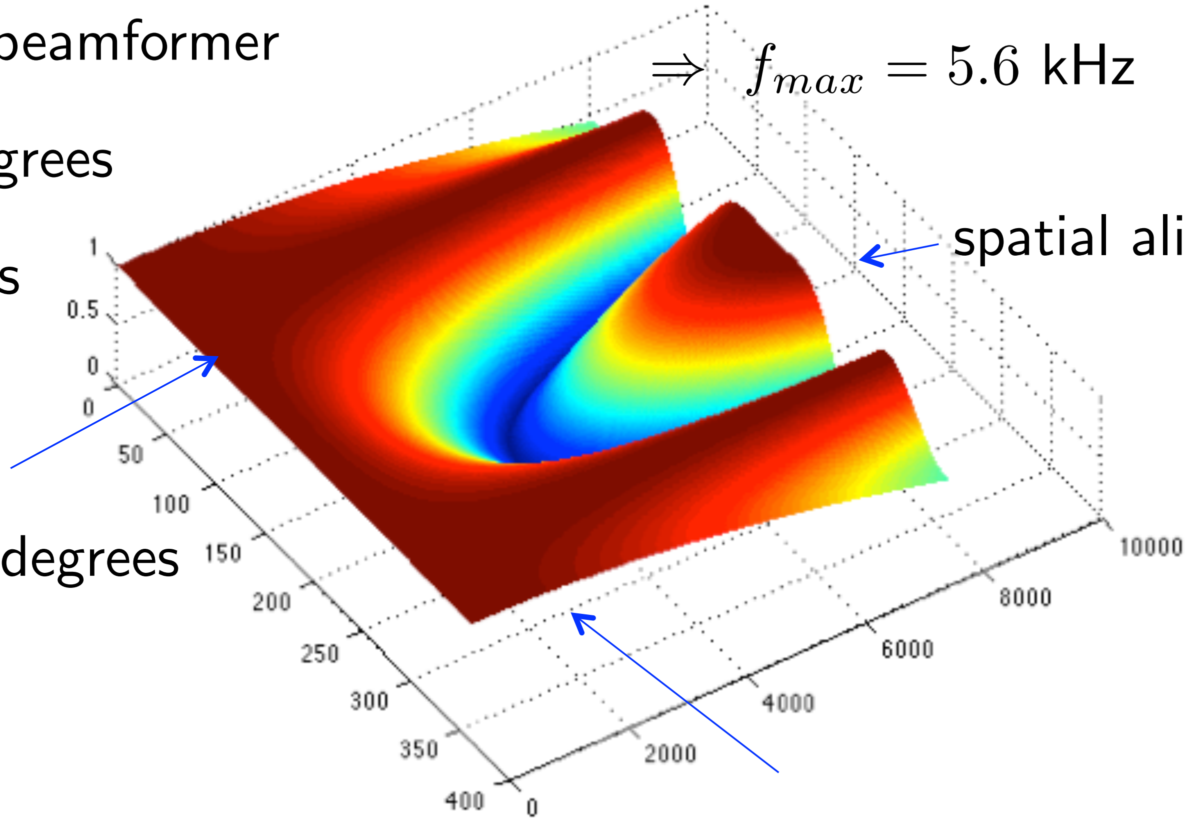
Concept of Beamforming – Freq. domain

Example:

$$d = 0.03 < \frac{1}{2}\lambda_{min} = \frac{1}{2}\frac{c}{f_{max}}.$$

$$\Rightarrow \quad f_{max} = 5.6 \text{ kHz}$$

- Sum and delay beamformer
- Target at 60 degrees
- two microphones
- $d = 0.03$



spatial aliasing

response of 1 at 60 degrees

For low frequencies, less spatial selectivity

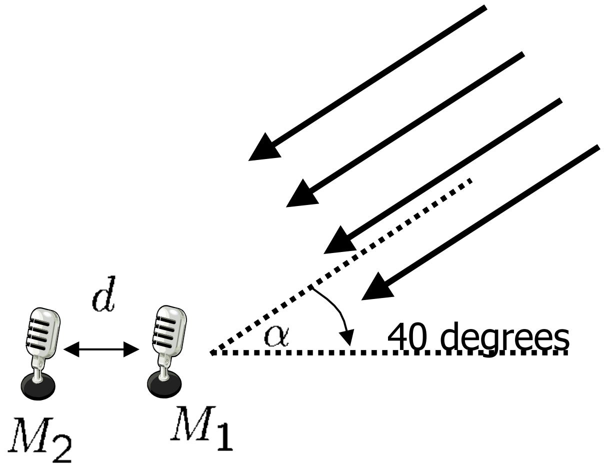
TU Delft

Concept of Beamforming – Freq. domain

Sampling a sound source with multiple microphones results in a measured source with different delay (and damping) on each microphone.

Let us initially assume the damping is the same for all microphones (source is thus in the far-field).

What is then the effect of a delay τ on the frequency domain description of the signals and how to exploit this?



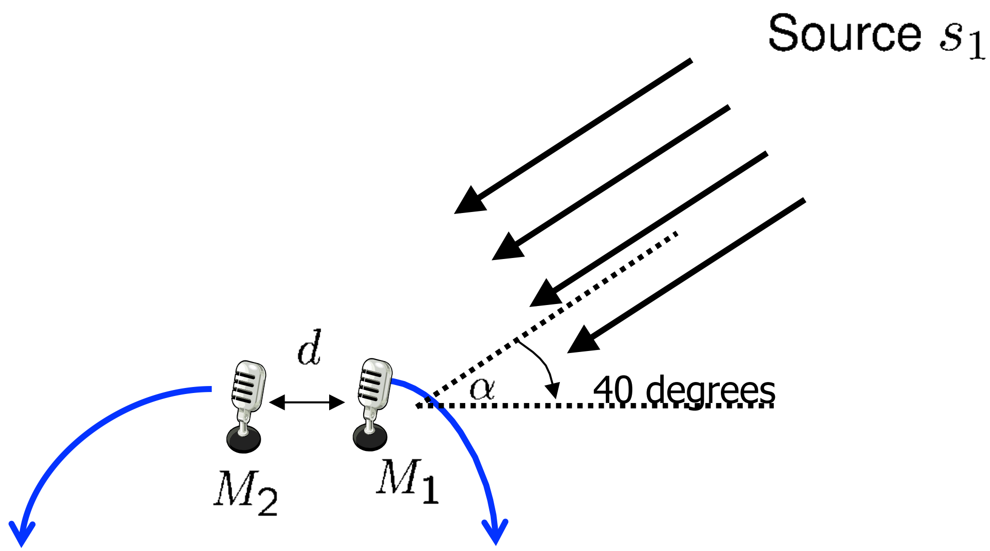
TU Delft

Concept of Beamforming – Freq. domain

Let us assume that microphone 1 is the reference microphone. The second microphone then experiences a delay τ.

A delay τ in time-domain means a multiplication in frequency domain with $e^{-j2\pi k\tau/N}$:

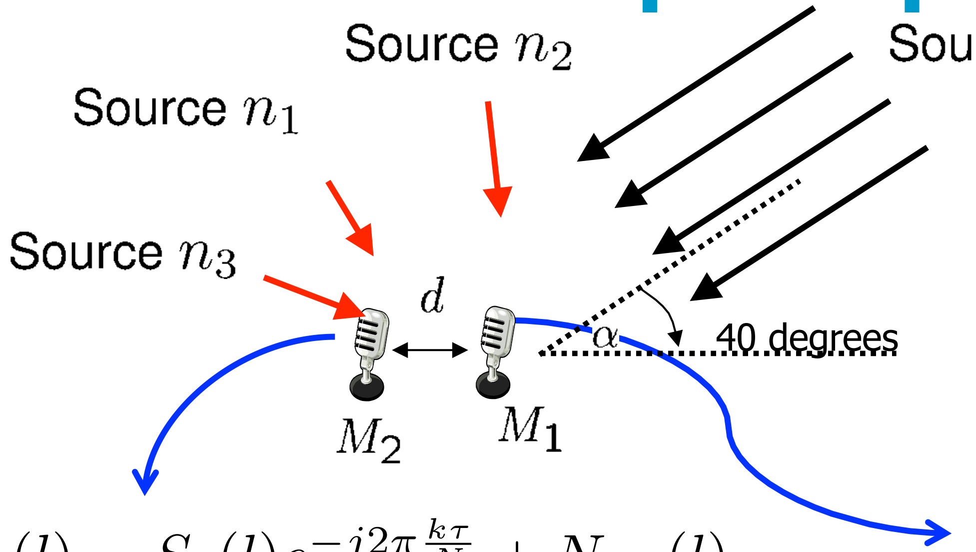
$$s(n-\tau) \Leftrightarrow S_k e^{-j2\pi k\tau/N}$$



$$S_{2,k}(l) = S_k(l)e^{-j2\pi k\tau/N} \qquad S_{1,k}(l) = S_k(l)$$

17
TU Delft

How to exploit spatial filtering?



$$Y_{2,k}(l) = S_k(l)e^{-j2\pi\frac{k\tau}{N}} + N_{2,k}(l)$$

$$Y_{1,k}(l) = S_k(l) + N_{1,k}(l)$$

How to obtain an estimate $\hat{S}_k(l)$?

Given that direction α is known (i.e., τ) compensate for delay:

$$\hat{S}_k(l) = \frac{Y_{1,k}(l)+Y_{2,k}(l)e^{j2\pi\frac{k\tau}{N}}}{2}$$

$$= \frac{S_k(l)+N_{1,k}(l)+S_k(l)e^{-j2\pi\frac{k\tau}{N}}e^{j2\pi\frac{k\tau}{N}}+N_{2,k}(l)e^{j2\pi\frac{k\tau}{N}}}{2} = S_k(l)+\frac{N_{1,k}(l)+N_{2,k}(l)e^{j2\pi\frac{k\tau}{N}}}{2}$$

18

TU Delft

How to exploit spatial filtering?

$$\hat{S}_k(l) = \frac{Y_{1,k}(l)+Y_{2,k}(l)e^{j2\pi\frac{k\tau}{N}}}{2} = S_k(l) + \frac{N_{1,k}(l)+N_{2,k}(l)e^{j2\pi\frac{k\tau}{N}}}{2}$$

- If the noise sources come from different angles as the speech source, the noise DFT coefficients $N_{1,k}(l)$ and $N_{2,k}(l)$ will be added destructively.
- If the noise is uncorrelated across microphones, i.e., $E[N_{1,k}(l)N^*_{2,k}(l)] = 0$, this operation involving two microphones will reduce the variance with a factor 2 (or three dB).
- This beam former is called the "delay and sum beamformer", after the two operations that are applied.

Signal models – near field

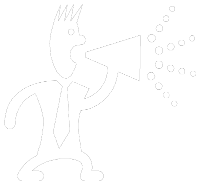
When sources travel to the microphones, the distance from source to each microphone influences the experienced damping and phase of the measured signal:

$$S_k(l) \Rightarrow S_k(l) a e^{-j2\pi \frac{k\tau}{N}}.$$

Depending on the size of the array and the distance of the array to the source, this gives rise to two different signal models:

- Near-field:
  - The source is close to the center of the array. The experienced damping is therefore different for every microphone.
  - Damping (a inversely proportional with distance) and phase differences τ are taken into account.

Signal models – far field



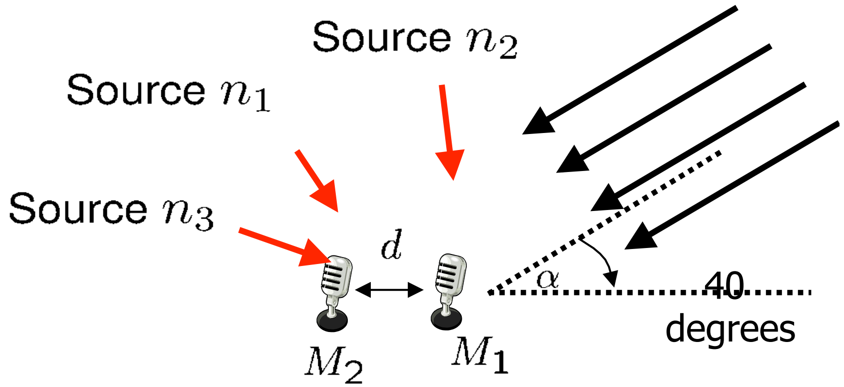
- Far-field:
  - The source is far away from the center of the array. The waves travel therefore parallel. The microphones experience no difference in damping.
  - Only phase differences τ are taken into account.

$$S_k(l) \Rightarrow S_k(l)e^{-j2\pi\frac{k\tau}{N}}.$$

TU Delft

Delay & Sum Beamformer

- Exploits the fact that the signal reaching the microphones are delayed with respect to each other.
- Changing the phase steers the beam towards the target.
- Takes only the direction of the target into account and neglects knowledge of the noise field completely.



TU Delft

Delay & Sum Beamformer

Assuming near-field, $\mathbf{Y}_k$ for a linear array consists of

$$\begin{aligned}\mathbf{Y}_k &= [Y_{1,k}(l), ..., Y_{M,k}(l)]^T \\ &= \left[S_{1,k}(l) + N_{1,k}(l), S_{1,k}(l)\frac{a_2 e^{-j2\pi\frac{k\tau_1}{N}}}{a_1} + N_{2,k}(l), ..., S_{1,k}(l)\frac{a_M e^{-j2\pi\frac{k\tau_M}{N}}}{a_1} + N_{M,k}(l)\right]^T.\end{aligned}$$

Choosing the first microphone as the reference, we can set $\mathbf{d}_k = [1, \frac{a_2 e^{-j2\pi\frac{k\tau_2}{N}}}{a_1}, ..., \frac{a_M e^{-j2\pi\frac{k\tau_M}{N}}}{a_1}]^T$. Notice we can use $\mathbf{d}_k$ to write

$$\mathbf{Y}_k = S_{1,k}(l)\mathbf{d}_k + \mathbf{N}_k$$

and calculate

$$\hat{S}_k(l) = \frac{\mathbf{d}^H \mathbf{Y}_k(l)}{\mathbf{d}^H \mathbf{d}}.$$

Near field: $\mathbf{w}_k(l) = \frac{\mathbf{d}}{\mathbf{d}^H \mathbf{d}}$

Far-field: $\mathbf{w}_k(l) = \frac{1}{M}\mathbf{d}$

Delay & Sum Beamformer

Delay and sum

- preserves the target.
- does not take explicit knowledge on the noise field into account.
- reduces the noise variance in most ideal case (uncorrelated noise across microphones) with a factor
$\frac{1}{M} = \frac{1}{2^p} \;\Rightarrow\; -p10\log_{10}(2) \approx -3p\ dB$

TU Delft

MVDR - beamformer

More advanced beamformers not only exploit position of target, but position of noise sources as well. Well-known adaptive beamformer is the minimum variance distortionless response (MVDR) beamformer

- Constrains the beamformer to have no change of magnitude and phase in direction of target source.
- Minimizes the variance of the beamformer output in all other directions.

MVDR - beamformer

MVDR - beamformer

Cost function: $J(\mathbf{w}_k) = \mathbf{w}_k^H(l)\mathbf{R}_{Y,k}(l)\mathbf{w}_k(l)$

Complex derivatives!

$$min_{\mathbf{w}_k} J$$
$$s.t. \mathbf{w}_k^H \mathbf{d}_k = 1.$$

Lagrange multiplier!

$$\frac{d}{d\mathbf{w}_k^H}\left\{J(\mathbf{w}_k) + \lambda(\mathbf{w}_k^H\mathbf{d}_k - 1)\right\} \quad = \quad \mathbf{R}_{Y,k}(l)\mathbf{w}_k(l) + \lambda\mathbf{d}_k$$

$$\mathbf{R}_{Y,k}(l)\mathbf{w}_k(l) + \lambda\mathbf{d}_k = 0 \;\; \Rightarrow, \; \mathbf{w}_k(l) = -\left(\mathbf{R}_{Y,k}(l)\right)^{-1}\lambda\mathbf{d}_k$$

MVDR - beamformer

Use the constraint: $\mathbf{d}_k^H \mathbf{w}_k(l) = 1 = -\mathbf{d}_k^H \left(\mathbf{R}_{Y,k}(l)\right)^{-1} \lambda \mathbf{d}_k$

$$\Rightarrow \quad \lambda = -\frac{1}{\mathbf{d}_k^H \left(\mathbf{R}_{Y,k}(l)\right)^{-1} \mathbf{d}_k} \quad \Rightarrow \quad \mathbf{w}_k(l) = \frac{\left(\mathbf{R}_{Y,k}(l)\right)^{-1} \mathbf{d}_k}{\mathbf{d}_k^H \left(\mathbf{R}_{Y,k}(l)\right)^{-1} \mathbf{d}_k}$$

MVDR - beamformer

$$\mathbf{w}_k(l) = \frac{(\mathbf{R}_{Y,k}(l))^{-1}\mathbf{d}_k}{\mathbf{d}_k^H (\mathbf{R}_{Y,k}(l))^{-1}\mathbf{d}_k}$$

The MVDR beamformer can also be written using the noise correlation matrix $\mathbf{R}_{N,k}(l)$ based on the matrix inversion lemma:

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}$$

Matrix $\mathbf{R}_{Y,k}(l)$ can be written as $\mathbf{R}_{Y,k}(l) = \mathbf{R}_{N,k}(l) + \mathbf{d}\mathbf{d}^H\sigma_{S,k}^2(l)$

$$\mathbf{w}_k(l) = \frac{\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k\left(1 - \frac{\mathbf{d}_k^H\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k\sigma_{S,k}^2(l)}{1+\mathbf{d}^H\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k\sigma_{S,k}^2(l)}\right)}{\mathbf{d}^H\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k\left(1 - \frac{\mathbf{d}_k^H\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k\sigma_{S,k}^2(l)}{1+\mathbf{d}^H\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k\sigma_{S,k}^2(l)}\right)} = \frac{\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k}{\mathbf{d}_k^H\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k}$$

MVDR - beamformer

$$\mathbf{w}_k(l) = \frac{\mathbf{R}_{Y,k}^{-1}(l)\mathbf{d}_k}{\mathbf{d}_k^H \mathbf{R}_{Y,k}^{-1}(l)\mathbf{d}_k} = \frac{\mathbf{R}_{N,k}^{-1}(l)\mathbf{d}_k}{\mathbf{d}_k^H \mathbf{R}_{N,k}^{-1}(l)\mathbf{d}_k}$$

This holds under the assumption that 1) target is rank-1 2) target and noise are uncorrelated and 3) target and noise are additive

TU Delft

MVDR – Spatially uncorrelated noise

$$\mathbf{w}_k(l) = \frac{\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k}{\mathbf{d}_k^H \mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k}$$

If the noise field is spatially uncorrelated, i.e., $\mathbf{R}_{N,k}(l) = \sigma^2_{N,k}(l)\mathbf{I}_M$, the MVDR equals the delay and sum beamformer

$$\mathbf{w}_k(l) = \frac{\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k}{\mathbf{d}_k^H \mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k} = \frac{\mathbf{d}_k}{\mathbf{d}_k^H \mathbf{d}_k}$$

(assuming far-field):

$$\mathbf{w}_k(l) = \frac{\mathbf{d}_k}{M}$$

Optimal Linear Multi-Channel Wiener

Signal model: $\mathbf{Y}_k(l) = S_{1,k}(l)\mathbf{d}_k(l) + \mathbf{N}_k(l)$

Cost function: $J_{MSE}(\mathbf{w}_k) = E[||S_{1,k}(l) - \mathbf{w}_k^H\mathbf{Y}_k(l)||_2^2]$

$$\frac{dJ_{MSE}(\mathbf{w}_k)}{d\mathbf{w}_k^H} = -E[S_{1,k}(l)^H\mathbf{Y}_k(l)]+\mathbf{R}_{Y_k}(l)\mathbf{w}_k = -\sigma^2_{S,k}\mathbf{d}_k(l)+\mathbf{R}_{Y_k}(l)\mathbf{w}_k$$

$$\mathbf{w}_k = \mathbf{R}_{Y_k}^{-1}(l)\sigma^2_{S,k}\mathbf{d}_k(l)$$

Optimal Linear Multi-Channel Wiener

Using again the Matrix inversion lemma, it can be shown that

$$\mathbf{w}_k = R_{\mathbf{Y}_k}^{-1}(l)\sigma_{S,k}^2\mathbf{d}_k(l)$$

can be written as

$$\mathbf{w}_k = \underbrace{\frac{\sigma_{S,k}^2(l)}{\sigma_{S,k}^2(l) + (\mathbf{d}_k^H(l)R_{\mathbf{N}_k}^{-1}\mathbf{d}_k(l))^{-1}}}_{\text{Single-channel Wiener}} \underbrace{\frac{R_{\mathbf{N}_k}^{-1}(l)\mathbf{d}_k(l)}{\mathbf{d}_k^H(l)R_{\mathbf{N}_k}^{-1}\mathbf{d}_k(l)}}_{MVDR}$$

TU Delft

Optimal Linear Multi-Channel Wiener

matrix inversion lemma:

$$(\mathbf{A}+\mathbf{u}\mathbf{v}^T)^{-1}=\mathbf{A}^{-1}-\frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}$$

Matrix $\mathbf{R}_{Y,k}(l)$ can be written as $\mathbf{R}_{Y,k}(l)=\mathbf{R}_{N,k}(l)+\mathbf{d}\mathbf{d}^H\sigma^2_{S,k}(l)$

$$\begin{aligned}
\mathbf{R}^{-1}_{Y,k}(l)\mathbf{d}_k\sigma^2_{S,k}(l) &= \left(\mathbf{R}_{N,k}(l)+\mathbf{d}_k\mathbf{d}_k^H\sigma^2_{S,k}(l)\right)^{-1}\mathbf{d}_k\sigma^2_{S,k}(l)\\
&= \mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k\sigma^2_{S,k}(l)-\mathbf{R}^{-1}_{N,k}(l)\mathbf{d}_k\frac{\sigma^2_{S,k}(l)\mathbf{d}_k^H\mathbf{R}^{-1}_{N,k}(l)\mathbf{d}_k}{1+\sigma^2_{S,k}(l)\mathbf{d}^H\mathbf{R}^{-1}_{N,k}(l)\mathbf{d}_k}\sigma^2_{S,k}(l)\\
&= \mathbf{R}^{-1}_{N,k}(l)\mathbf{d}_k\left(1-\frac{\sigma^2_{S,k}(l)\mathbf{d}_k^H\mathbf{R}^{-1}_{N,k}(l)\mathbf{d}_k}{1+\sigma^2_{S,k}(l)\mathbf{d}^H\mathbf{R}^{-1}_{N,k}(l)\mathbf{d}_k}\right)\sigma^2_{S,k}(l)
\end{aligned}$$

Optimal Linear Multi-Channel Wiener

$$
\begin{aligned}
&= \mathbf{R}_{N,k}^{-1}(l)\mathbf{d}_k \left(1 - \frac{\sigma_{S,k}^2(l)\mathbf{d}_k^H \mathbf{R}_{N,k}^{-1}(l)\mathbf{d}_k}{1+\sigma_{S,k}^2(l)\mathbf{d}^H\mathbf{R}_{N,k}^{-1}(l)\mathbf{d}_k}\right)\sigma_{S,k}^2(l) \\
&= \mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k \left(\frac{\sigma_{S,k}^2(l)}{1+\sigma_{S,k}^2(l)\mathbf{d}^H\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k}\right) \\
&= \frac{\mathbf{R}_{N,k}(l)^{-1}\mathbf{d}_k}{\mathbf{d}^H\mathbf{R}_{N,k}^{-1}(l)\mathbf{d}_k} \left(\frac{\mathbf{d}^H\mathbf{R}_{N,k}^{-1}(l)\mathbf{d}_k\sigma_{S,k}^2(l)}{1+\sigma_{S,k}^2(l)\mathbf{d}^H\mathbf{R}_{N,k}^{-1}(l)\mathbf{d}_k}\right) \\
&= \frac{\mathbf{R}_{N,k}^{-1}(l)\mathbf{d}_k}{\mathbf{d}^H\mathbf{R}_{N,k}^{-1}(l)\mathbf{d}_k} \left(\frac{\sigma_{S,k}^2(l)}{\left(\mathbf{d}^H\mathbf{R}_{N,k}^{-1}(l)\mathbf{d}_k\right)^{-1}+\sigma_{S,k}^2(l)}\right)
\end{aligned}
$$

Optimal Linear Multi-Channel Wiener

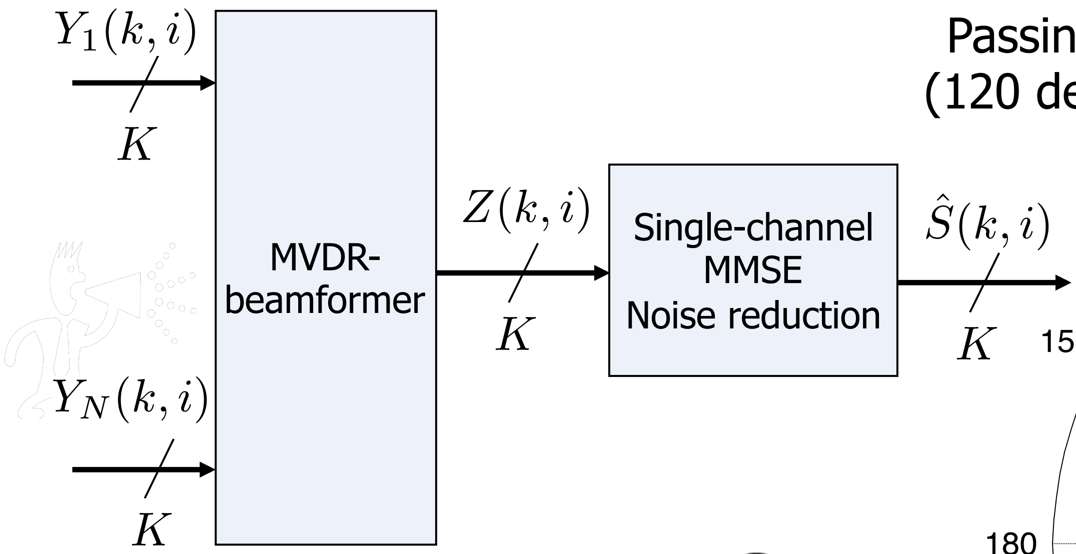
$$\mathbf{w}_k = \underbrace{\frac{\sigma^2_{S,k}(l)}{\sigma^2_{S,k}(l) + (\mathbf{d}_k^H(l) R_{\mathbf{N}_k}^{-1} \mathbf{d}_k(l))^{-1}}}_{\text{Single-channel Wiener}} \underbrace{\frac{R_{\mathbf{N}_k}^{-1}(l)\mathbf{d}_k(l)}{\mathbf{d}_k^H(l) R_{\mathbf{N}_k}^{-1} \mathbf{d}_k(l)}}_{MVDR}$$

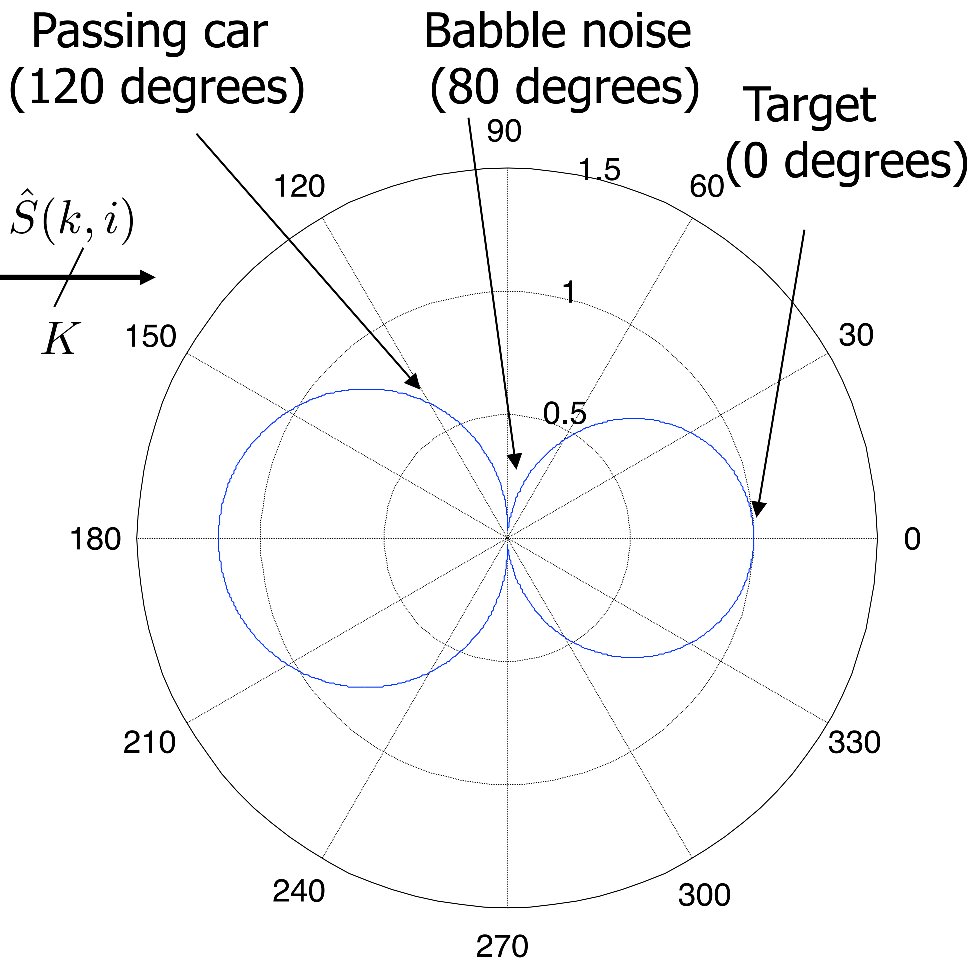
The multi-channel Wiener filter can thus be seen as a concatenation of two filters:

- An MVDR as spatial filter
- Single-Channel Wiener filter as post-processor where noise variance is set to the remaining noise PSD after beamforming: $\mathbf{d}_k^H(l) R_{\mathbf{N}_k}^{-1} \mathbf{d}_k(l)$

TU Delft

Example: Multi-Channel Noise Reduction





- Noisy input at mic. 1
- MVDR-beamformer output
- Output of total system

LCMV - beamformer

Remember the MVDR: $J(\mathbf{w}_k) = \mathbf{w}_k^H(l)\mathbf{R}_{Y,k}(l)\mathbf{w}_k(l)$

$$min_{\mathbf{w}_k} J$$
$$s.t. \mathbf{w}_k^H \mathbf{d}_k = 1.$$

- The MVDR imposes one constraint.
- This can be generalised to having d constraints.

LCMV - beamformer

Cost function: $J(\mathbf{w}_k) = \mathbf{w}_k^H(l)\mathbf{R}_{Y,k}(l)\mathbf{w}_k(l)$

$$min_{\mathbf{w}_k} J$$
$$s.t. \mathbf{w}_k^H \mathbf{\Lambda}_k = \mathbf{f}^H.$$

with $\mathbf{\Lambda} \in \mathbb{C}^{M \times d}$

When $d < M$, there is a closed form solution:

$$\mathbf{w}_k = \mathbf{R_{Y,k}^{-1}\Lambda_k}\left(\mathbf{\Lambda_k^H R_{Y,k}^{-1}\Lambda_k}\right)^{\mathbf{-1}}\mathbf{f}.$$

LCMV - beamformer

$$\mathbf{w}_k = \mathbf{R}_{\mathbf{Y},\mathbf{k}}^{-1}\mathbf{\Lambda_k}\left(\mathbf{\Lambda_k^H}\mathbf{R}_{\mathbf{Y},\mathbf{k}}^{-1}\mathbf{\Lambda_k}\right)^{-1}\mathbf{f}.$$

How to use the multiple constraints?

- To steer zeros in the direction of certain noise sources.
- To maintain the signal from certain directions.
- To maintain the spatial cues of for hearing aids.

Notice that the more constraints are used, less degrees of freedom are left to control the noise reduction.

Overview of Discussed filters

- Delay and sum beamformer

$$\mathbf{w}_k(l) = \frac{\mathbf{d}_k}{\mathbf{d}_k^H \mathbf{d}_k}$$

- MVDR beamformer

$$\mathbf{w}_k(l) = \frac{\left(\mathbf{R}_{Y,k}(l)\right)^{-1} \mathbf{d}_k}{\mathbf{d}_k^H \left(\mathbf{R}_{Y,k}(l)\right)^{-1} \mathbf{d}_k} = \frac{\left(\mathbf{R}_{N,k}(l)\right)^{-1} \mathbf{d}_k}{\mathbf{d}_k^H \left(\mathbf{R}_{N,k}(l)\right)^{-1} \mathbf{d}_k}$$

- Multi-Channel Wiener

$$\mathbf{w}_k = \underbrace{\frac{\sigma^2_{S,k}(l)}{\sigma^2_{S,k}(l) + (\mathbf{d}_k^H(l) R^{-1}_{\mathbf{N}_k} \mathbf{d}_k(l))^{-1}}}_{\text{Single-channel Wiener}} \underbrace{\frac{R^{-1}_{\mathbf{N}_k}(l)\mathbf{d}_k(l)}{\mathbf{d}_k^H(l) R^{-1}_{\mathbf{N}_k} \mathbf{d}_k(l)}}_{MVDR}$$

Overview of Discussed filters

- LCMV beamformer

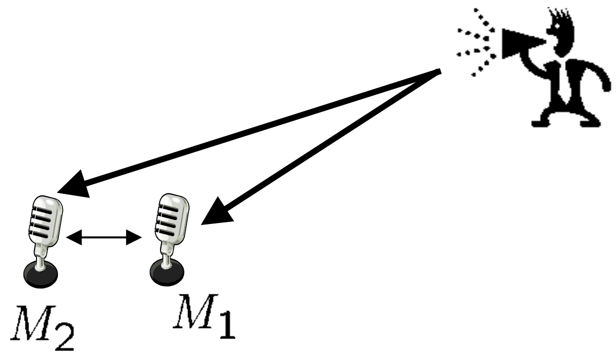
$$\mathbf{w}_k = \mathbf{R}_{\mathbf{Y},\mathbf{k}}^{-\mathbf{1}} \mathbf{\Lambda}_{\mathbf{k}} \left(\mathbf{\Lambda}_{\mathbf{k}}^{\mathbf{H}} \mathbf{R}_{\mathbf{Y},\mathbf{k}}^{-\mathbf{1}} \mathbf{\Lambda}_{\mathbf{k}} \right)^{-\mathbf{1}} \mathbf{f}.$$

Acoustic Transfer Function

Remember that the steering vector (for an arbitrary setup) and free field, $\mathbf{d}_k$ is given by

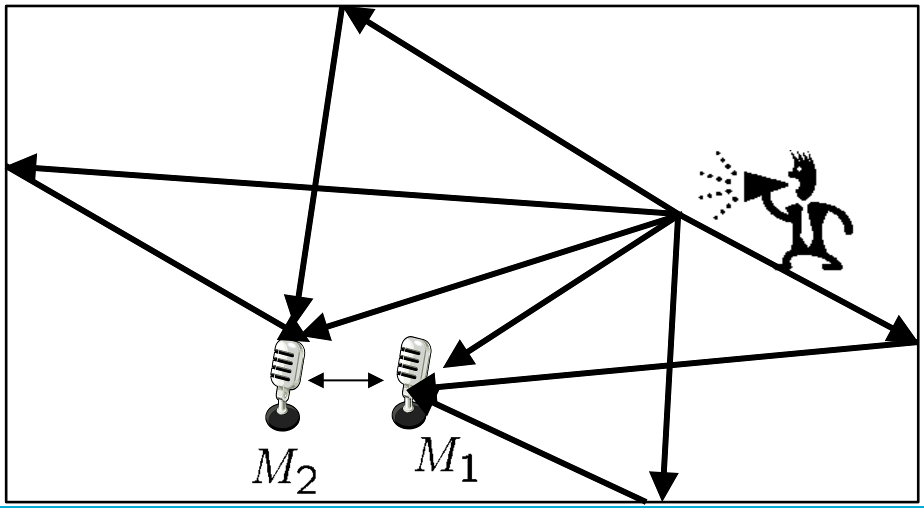
$$\mathbf{d}_k = [1, \frac{a_2 e^{-j2\pi\frac{k\tau_2}{N}}}{a_1}, ..., \frac{a_M e^{-j2\pi\frac{k\tau_M}{N}}}{a_1}]^T.$$

- The τ_m's in $\mathbf{d}_k$ essentially model the delays that the source experiences from source position to microphone.
- Remember that a delay τ in time, implies a phase change $e^{-j2\pi\frac{k\tau}{N}}$ in DFT domain.



Acoustic Transfer Function

- In general, if there is no free field, there is not only a direct path, but also reflections.
- In general, $\mathbf{d}_{m,k}$ is known as the DFT transform of the room impulse response $h(n)$, also known as the acoustic transfer function.



TU Delft