Rate-Distributed Binaural LCMV Beamforming for Assistive Hearing in Wireless Acoustic Sensor Networks

Jie Zhang, Richard Heusdens, and Richard C. Hendriks

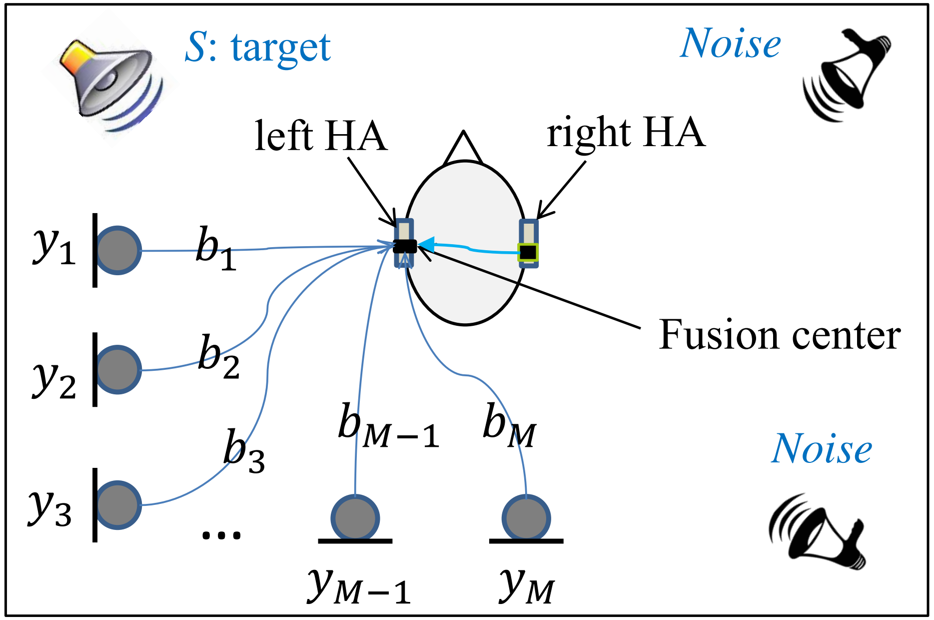
Circuits and Systems (CAS) Group, Delft University of Technology, the Netherlands

July 5, 2018

TU Delft

Introduction
Motivation
Fundamentals
BLCMV
RD-BLCMV
Simulations
Conclusion

Introduction



General binaural HAs configuration: HAs are a part of a bigger wireless acoustic sensor network.

Problem statement:

minimise	total transmission costs
subject to	noise reduction performance
	spatial cue preservation

TU Delft

Motivation

- Energy efficiency is essential in the design of algorithms in WASNs since usually each sensor has a limited energy budget (e.g., battery).
- In general, there are two ways to reduce energy costs:
  - Microphone subset selection
  - Bit-rate allocation
- Rate allocation is more general than sensor selection, i.e., sensor selection (hard/binary decision) can be seen as a special case of rate allocation (soft/multiple decision).

J. Zhang, S. P. Chepuri, R. C. Hendriks, and R. Heusdens, "Microphone subset selection for MVDR beamformer based noise reduction," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 8, pp. 550–563, 2018.
J. Zhang, R. Heusdens, and R. C. Hendriks "Rate-distributed spatial filtering for noise reduction in wireless acoustic sensor networks," *IEEE/ACM Trans. Acoustics, Speech, Language Processing, 2018. (to appear)*

Fundamentals

- Signal model in STFT domain:

$$\hat{\mathbf{y}} = \mathbf{y} + \mathbf{q} \in \mathbb{C}^M,$$

where

$$\begin{aligned} \mathbf{y} &= \mathbf{x} + \mathbf{z} + \mathbf{v} \\ &= \sum_{i=1}^{\mathcal{I}} \mathbf{a}_i s_i + \sum_{j=1}^{\mathcal{J}} \mathbf{h}_j u_j + \mathbf{v} \\ &= \mathbf{A}\mathbf{s} + \mathbf{H}\mathbf{u} + \mathbf{v}, \end{aligned}$$

where $\mathbf{a}_i = [a_{i1}, a_{i2}, \cdots, a_{iM}]^T$, $\mathbf{h}_j = [h_{j1}, h_{j2}, \cdots, h_{jM}]^T$, $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_{\mathcal{I}}]$, $\mathbf{s} = [s_1, \cdots, s_{\mathcal{I}}]^T$, $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_{\mathcal{J}}] \in \mathbb{C}^{M \times \mathcal{J}}$, $\mathbf{u} = [u_1, \cdots, u_{\mathcal{J}}]^T$.

TUDelft

Fundamentals

- Second-order statistics:

$$\mathbf{R}_{\mathbf{yy}} = \mathbb{E}\{\mathbf{y}\mathbf{y}^H\} = \mathbf{R}_{\mathbf{xx}} + \underbrace{\mathbf{R}_{\mathbf{zz}} + \mathbf{R}_{\mathbf{vv}}}_{\mathbf{R}_{\mathbf{nn}}} \in \mathbb{C}^{M \times M},$$

  where $\mathbf{R}_{\mathbf{xx}} = \sum_{i=1}^{\mathcal{I}} \mathbb{E}\{\mathbf{x}_i \mathbf{x}_i^H\}$ and $\mathbf{R}_{\mathbf{zz}} = \sum_{j=1}^{\mathcal{J}} \mathbb{E}\{\mathbf{z}_i \mathbf{z}_i^H\}$.

- Assumption: the target sources, interfering sources and quantisation noise are mutually uncorrelated, such that

$$\mathbf{R}_{\mathbf{n}+\mathbf{q}} = \mathbf{R}_{\mathbf{nn}} + \mathbf{R}_{\mathbf{qq}}.$$

- Uniform quantisation:

$$\mathbf{R}_{\mathbf{qq}} = \frac{1}{12}\mathrm{diag}\left(\left[\frac{\mathcal{A}_1^2}{4^{b_1}}, \frac{\mathcal{A}_2^2}{4^{b_2}}, ..., \frac{\mathcal{A}_M^2}{4^{b_M}}\right]\right),$$

  where $\mathcal{A}_k = \max\{|y_k|\}$ and $b_k, \forall k$ denotes the rate for the kth sensor node.

TU Delft

Fundamentals

- Transmission energy model:
  - SNR over communication channels: $\mathrm{SNR}_k = d_k^{-2} E_k / V_k$.
  - Channel capacity:
  $$b_k = \frac{1}{2}\log_2\left(1 + \mathrm{SNR}_k\right).$$
  - Transmission energy:
  $$E_k = d_k^2 V_k (4^{b_k} - 1),$$
  which holds under two conditions: 1) for spectrum-limited applications; 2) quantize at the channel capacity (i.e., the upper bound).

Binaural LCMV beamforming

- Binaural cue preservation: $\mathbf{w}_{\text{BLCMV}} = [\mathbf{w}_L^T \ \mathbf{w}_R^T]^T$
  - Preserving target sources:

$$\left.\begin{aligned} \mathbf{w}_L^H \mathbf{a}_i &= a_{iL} \\ \mathbf{w}_R^H \mathbf{a}_i &= a_{iR} \end{aligned}\right\} \Longrightarrow \text{ITF}_{\mathbf{x}_i}^{\text{in}} = \text{ITF}_{\mathbf{x}_i}^{\text{out}} \quad \Longrightarrow \mathbf{\Lambda}_1^H \mathbf{w} = \mathbf{f}_1$$

  - Preserving interfering sources:

$$\text{ITF}_{\mathbf{n}_j}^{\text{in}} = \text{ITF}_{\mathbf{n}_j}^{\text{out}} \quad \Rightarrow \quad \frac{h_{jL}}{h_{jR}} = \frac{\mathbf{w}_L^H \mathbf{h}_j}{\mathbf{w}_R^H \mathbf{h}_j}, \forall j,$$

$$\Longrightarrow \mathbf{w}_L^H \mathbf{h}_j h_{jR} - \mathbf{w}_R^H \mathbf{h}_j h_{jL} = 0 \quad \Longrightarrow \quad \mathbf{\Lambda}_2^H \mathbf{w} = \mathbf{f}_2$$

  - Binaural cues: $\text{ILD} = |\text{ITF}|^2, \quad \text{IPD} = \angle \text{ITF}.$

TU Delft

A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, “Relaxed binaural LCMV beamforming,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 1, pp. 137–152, 2017.

Binaural LCMV beamforming

- Binaural LCMV (BLCMV) beamforming for joint noise reduction and spatial cue preservation can be formulated as

$$\hat{\mathbf{w}}_{\mathrm{BLCMV}} = \arg\min_{\mathbf{w}} \ \mathbf{w}^H \tilde{\mathbf{R}}_{\mathbf{n+q}} \mathbf{w}, \quad \text{s.t.} \quad \mathbf{\Lambda}^H \mathbf{w} = \tilde{\mathbf{f}},$$

where

$$\tilde{\mathbf{R}}_{\mathbf{n+q}} = \begin{bmatrix} \mathbf{R}_{\mathbf{n+q}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{n+q}} \end{bmatrix} \in \mathbb{C}^{2M \times 2M},$$

$$\mathbf{\Lambda} = \left[\ \mathbf{\Lambda}_1 \ \middle| \ \mathbf{\Lambda}_2 \ \right] \in \mathbb{C}^{2M \times (2\mathcal{I}+\mathcal{J})},$$

$$\tilde{\mathbf{f}} = \left[\ \mathbf{f}_1^H \ \middle| \ \mathbf{f}_2^H \ \right]^T \in \mathbb{R}^{2\mathcal{I}+\mathcal{J}}.$$

- BLCMV beamformer: $\hat{\mathbf{w}}_{\mathrm{BLCMV}} = \tilde{\mathbf{R}}_{\mathbf{n+q}}^{-1} \mathbf{\Lambda} (\mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{n+q}}^{-1} \mathbf{\Lambda})^{-1} \tilde{\mathbf{f}}$.

TU Delft

A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Relaxed binaural LCMV beamforming," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 1, pp. 137–152, 2017.

Problem formulation

- **General problem formulation**:

$$\begin{aligned} \min_{\mathbf{w},\mathbf{b}} \quad & \sum_{k=1}^{M} d_k^2 V_k (4^{b_k} - 1) \\ \text{s.t.} \quad & \mathbf{w}^H \mathbf{R}_{\mathbf{n}+\mathbf{q}} \mathbf{w} \leq \frac{\beta}{\alpha} \\ & \mathbf{\Lambda}^H \mathbf{w} = \mathbf{f}, \\ & b_k \in \mathbb{Z}_+, \quad b_k \leq b_0, \forall k, \end{aligned} \tag{P1}$$

where

- β: minimum output noise power;
- $\alpha \in (0, 1]$: performance controller;
- b_0: maximum rate, e.g., 16 bits per sample.

Rate-distributed BLCMV beamforming

- Substitution of the BLCMV beamformer, we can obtain

$$\begin{aligned} \min_{\mathbf{b}} \quad & \sum_{k=1}^{M} d_k^2 V_k (4^{b_k} - 1) \\ \text{s.t.} \quad & \tilde{\mathbf{f}}^H (\mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{n}+\mathbf{q}}^{-1} \mathbf{\Lambda})^{-1} \tilde{\mathbf{f}} \leq \frac{\beta}{\alpha} \\ & b_k \in \mathbb{Z}_+, \quad b_k \leq b_0, \forall k. \end{aligned} \tag{P2}$$

- Convex optimisation:

$$\mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{n}+\mathbf{q}}^{-1} \mathbf{\Lambda} = \mathbf{Z}, \tag{1}$$

$$\tilde{\mathbf{f}}^H \mathbf{Z}^{-1} \tilde{\mathbf{f}} \leq \frac{\beta}{\alpha}, \tag{2}$$

where $\mathbf{Z} \in \mathbb{S}_+^{2\mathcal{I}+\mathcal{J}}$ is Hermitian.

Rate-distributed BLCMV beamforming

- Define a constant vector $\mathbf{e} = \left[\frac{12}{\mathcal{A}_1^2}, \cdots, \frac{12}{\mathcal{A}_M^2}\right]$ and introduce a variable change $t_k = 4^{b_k} \in \mathbb{Z}_+, \forall k$, such that $\mathbf{R}_{\mathbf{qq}}^{-1} = \text{diag}\left(\mathbf{e} \odot \mathbf{t}\right)$,

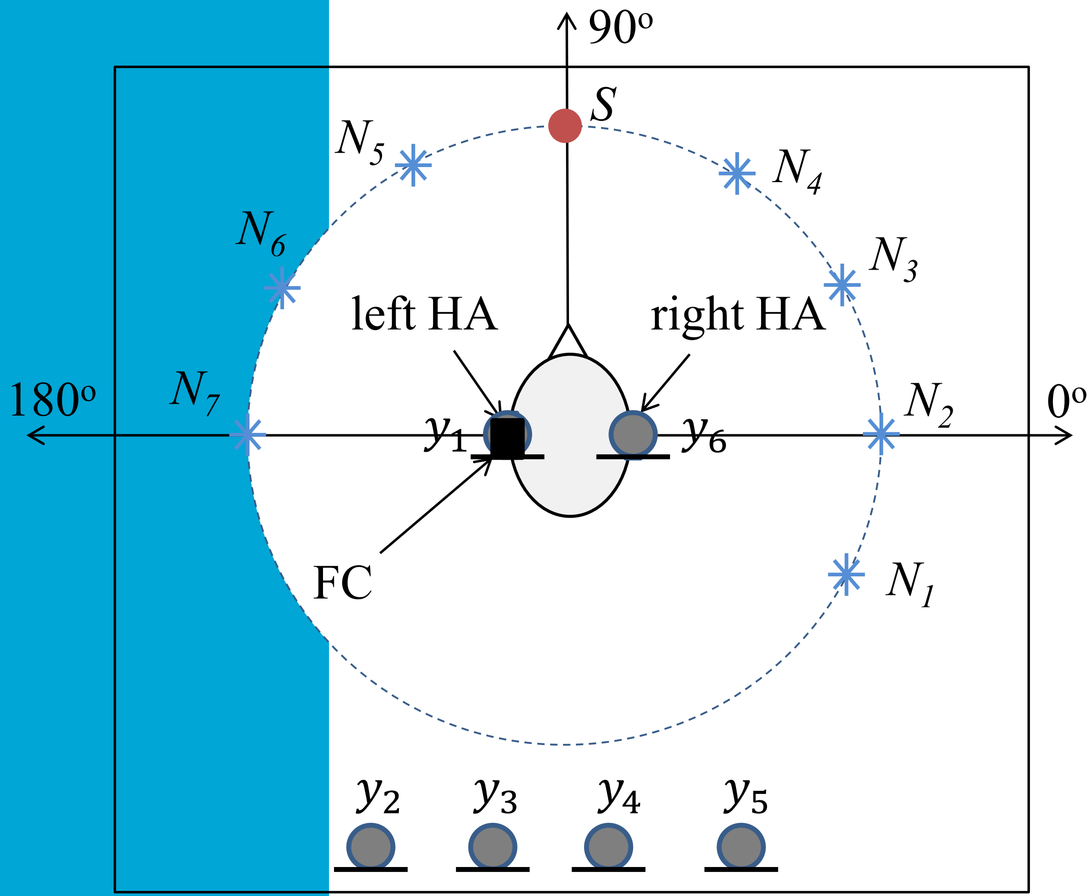
$$
\begin{aligned}
\min_{\mathbf{t},\mathbf{Z}} \quad & \sum_{k=1}^{M} d_k^2 V_k (t_k - 1) \\
\text{s.t.} \quad & \begin{bmatrix} \mathbf{Z} & \mathbf{f} \\ \mathbf{f}^H & \frac{\beta}{\alpha} \end{bmatrix} \succeq \mathbf{O}_{2\mathcal{I}+\mathcal{J}+1} \\
& \begin{bmatrix} \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} + \tilde{\mathbf{R}}_{\mathbf{qq}}^{-1} & \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1}\mathbf{\Lambda} \\ \mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} & \mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1}\mathbf{\Lambda} - \mathbf{Z} \end{bmatrix} \succeq \mathbf{O}_{2M+2\mathcal{I}+\mathcal{J}} \\
& 1 \leq t_k \leq 4^{b_0}, \forall k.
\end{aligned}
$$

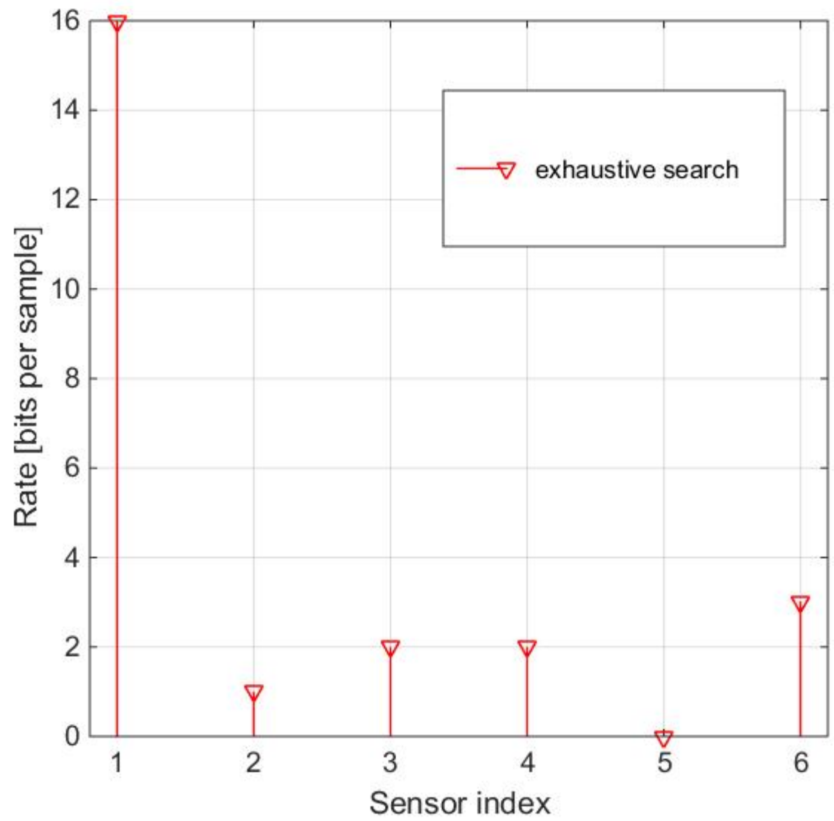
- The integer rates can be resolved by $b_k = \log_4 t_k, \forall k$ and randomised rounding technique.

TUDelft

Simulation results

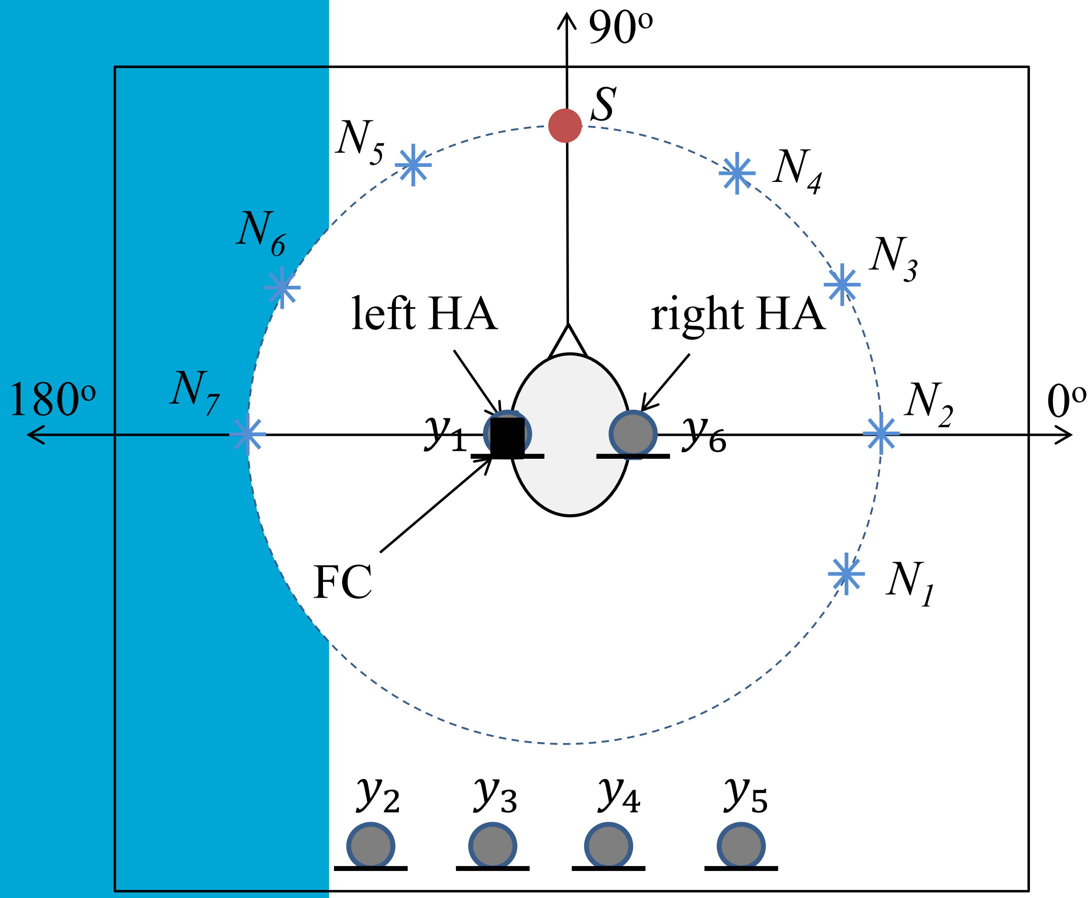
- Setting: 6 mics in (4×3)m 2D room, f_s=16kHz, T_{60}=200ms, $\alpha = 0.8$

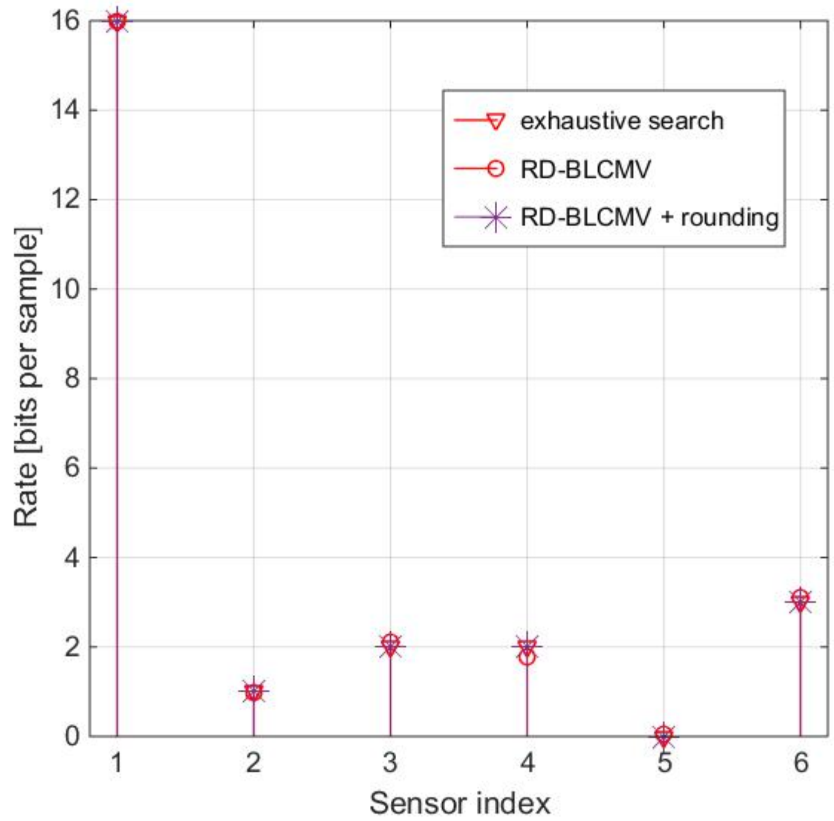




Simulation results

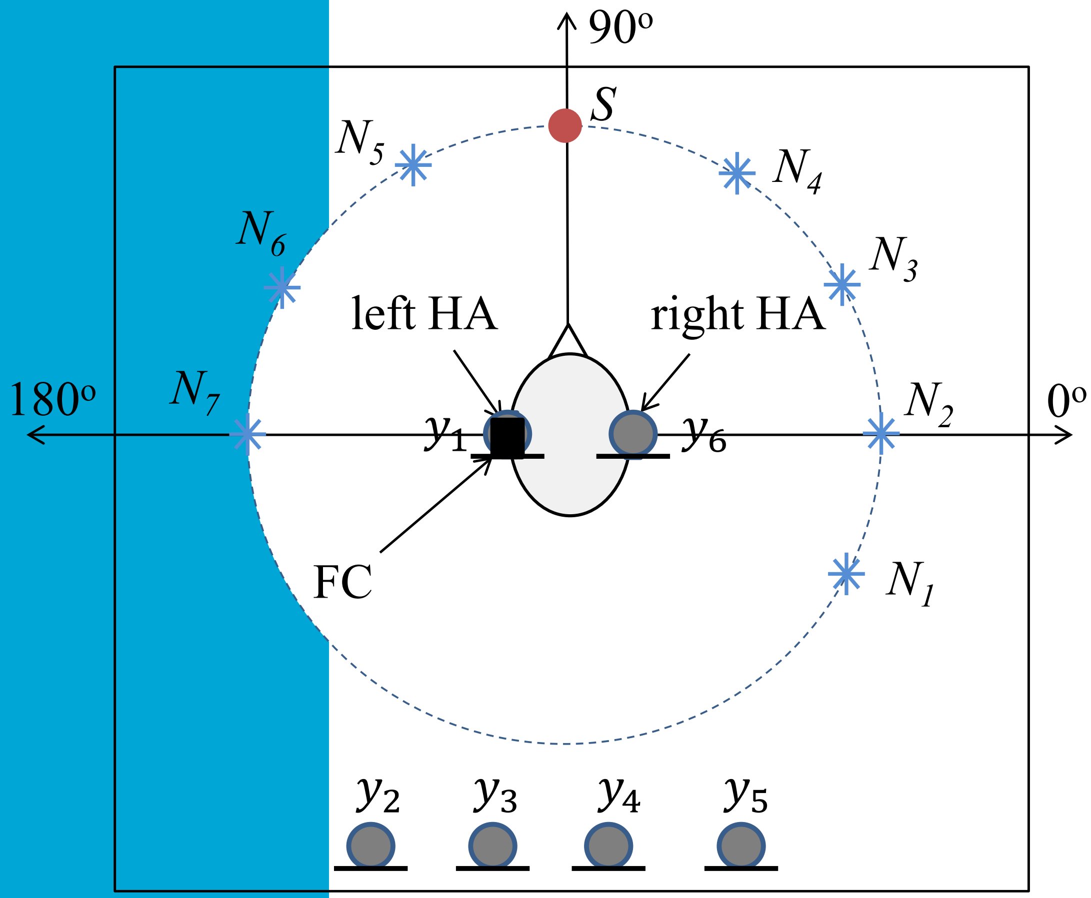
- Setting: 6 mics in (4×3)m 2D room, f_s=16kHz, T_{60}=200ms, $\alpha = 0.8$

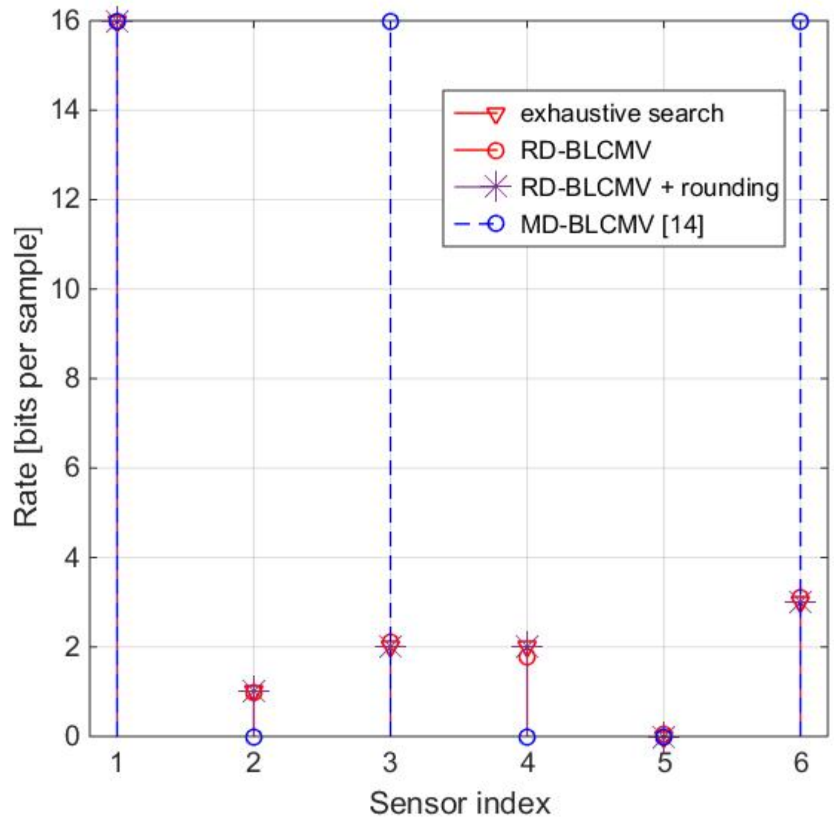




Simulation results

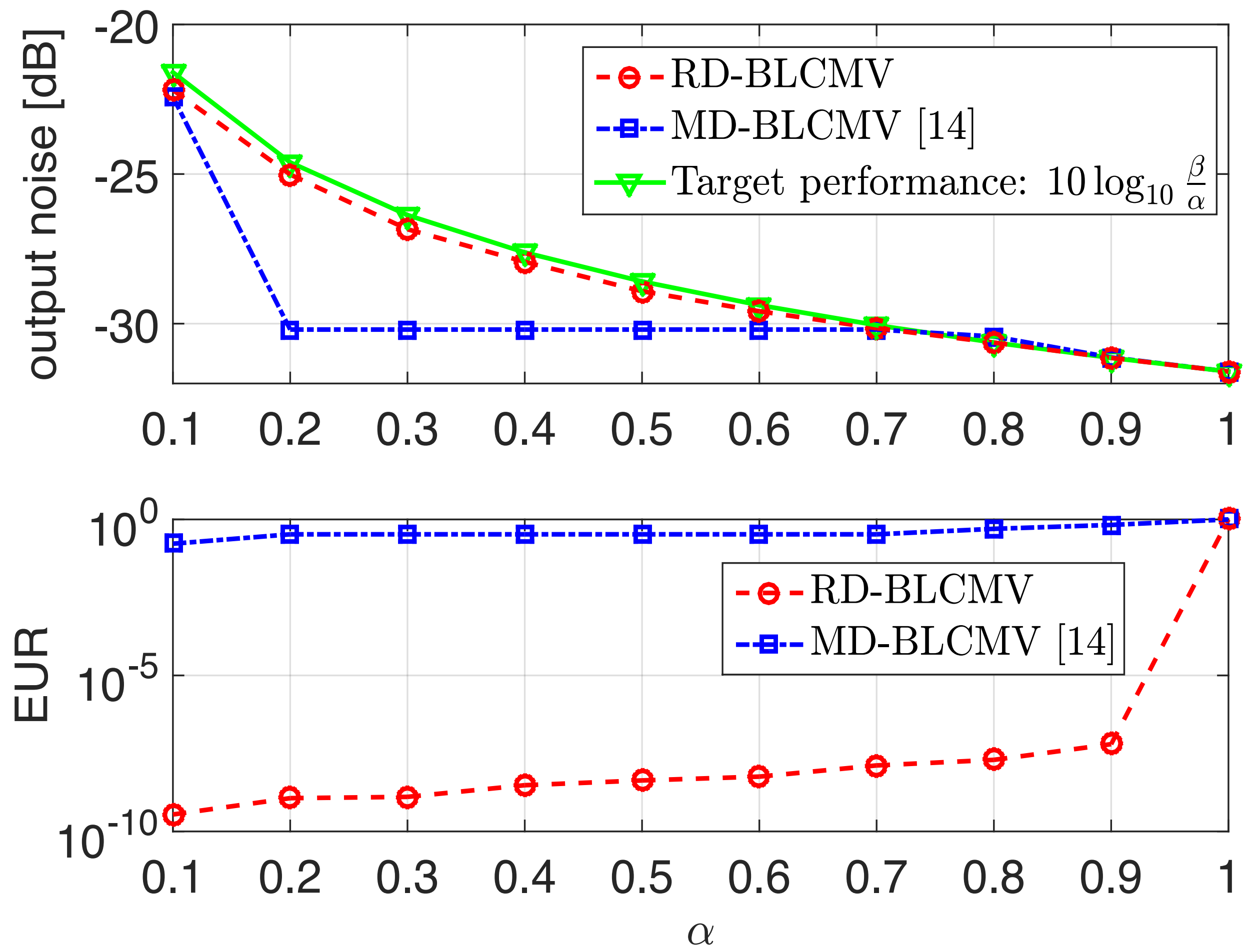
- Setting: 6 mics in (4×3)m 2D room, f_s=16kHz, T_{60}=200ms, $\alpha = 0.8$





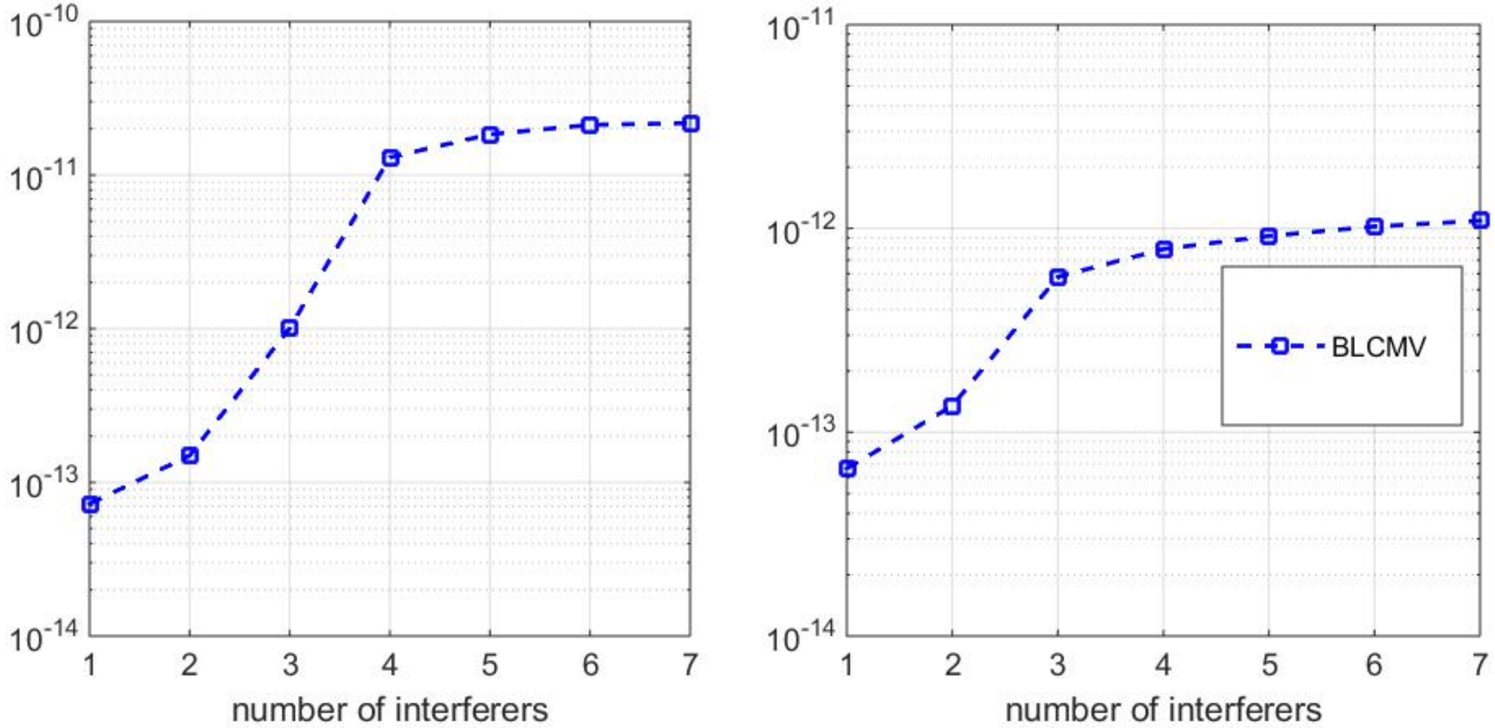
Simulation results

- Noise reduction performance and energy usage ratio (EUR):



Simulation results

- Performance of spatial cue preservation:

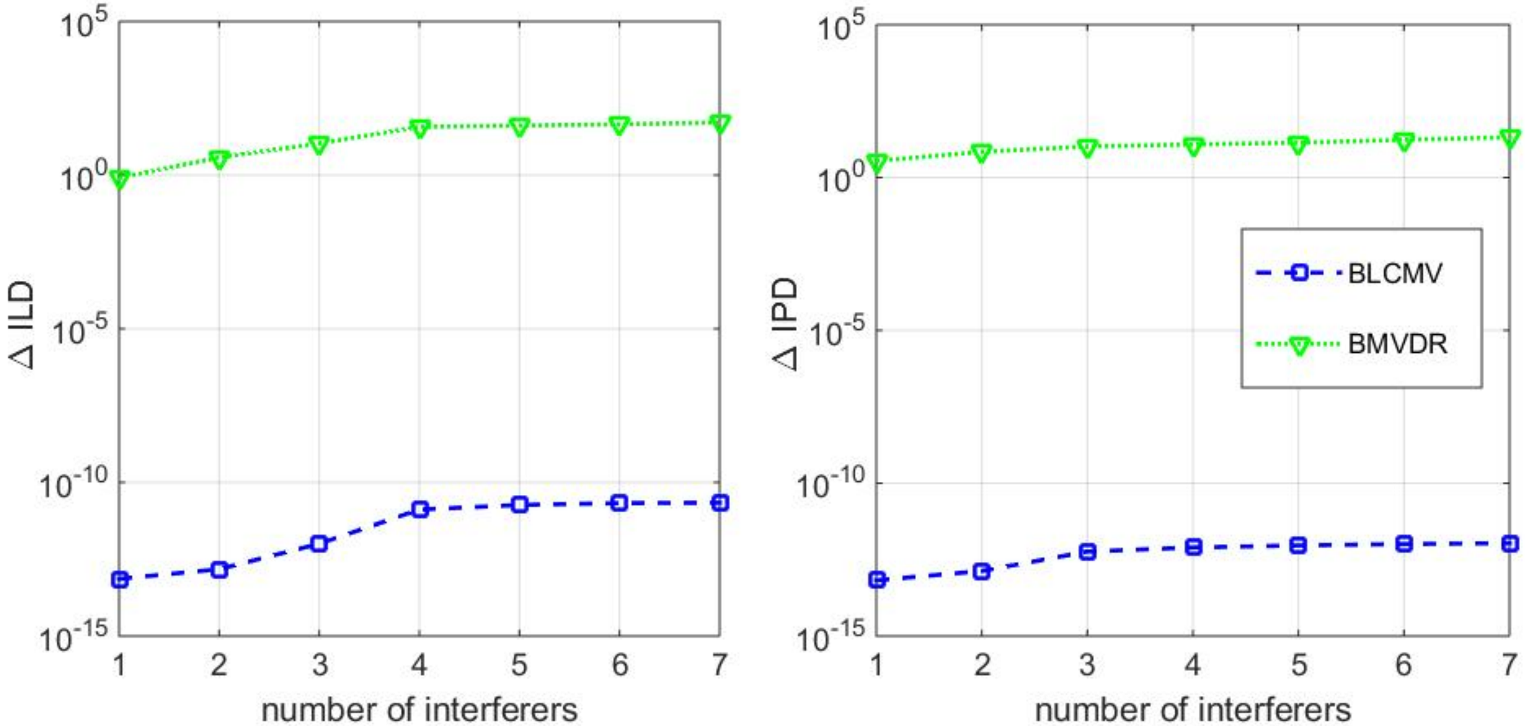


$$\Delta\mathrm{ILD} = \sum_{j=1}^{\mathcal{J}} \sum_{\omega} \left(\mathrm{ILD}_j(\omega) - \tilde{\mathrm{ILD}}_j(\omega)\right)^2; \qquad \Delta\mathrm{IPD} = \sum_{j=1}^{\mathcal{J}} \sum_{\omega} \left(\mathrm{IPD}_j(\omega) - \tilde{\mathrm{IPD}}_j(\omega)\right)^2$$

TUDelft

Simulation results

- Performance of spatial cue preservation:

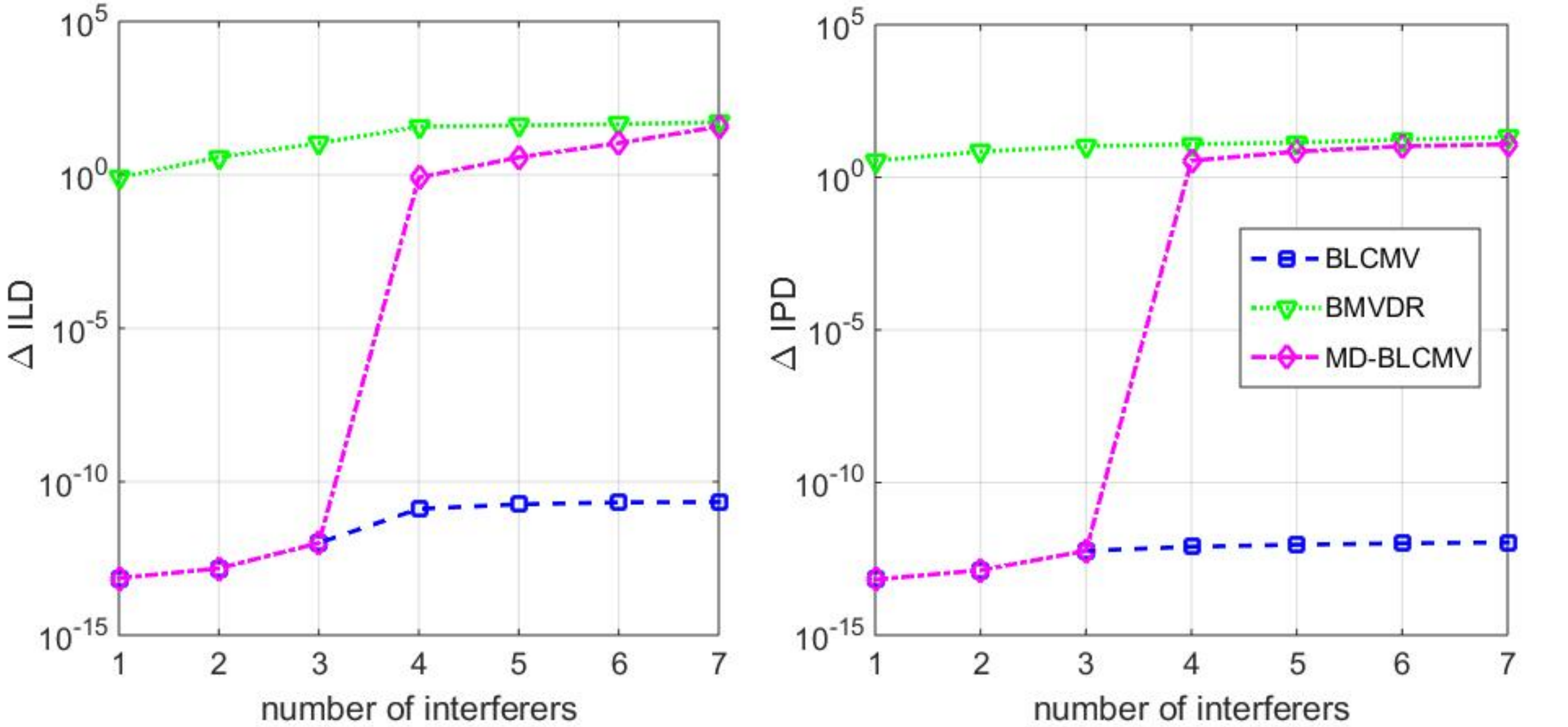


$$\Delta\mathrm{ILD} = \sum_{j=1}^{\mathcal{J}} \sum_{\omega} \left(\mathrm{ILD}_j(\omega) - \tilde{\mathrm{ILD}}_j(\omega) \right)^2 ; \qquad \Delta\mathrm{IPD} = \sum_{j=1}^{\mathcal{J}} \sum_{\omega} \left(\mathrm{IPD}_j(\omega) - \tilde{\mathrm{IPD}}_j(\omega) \right)^2$$

TU Delft

Simulation results

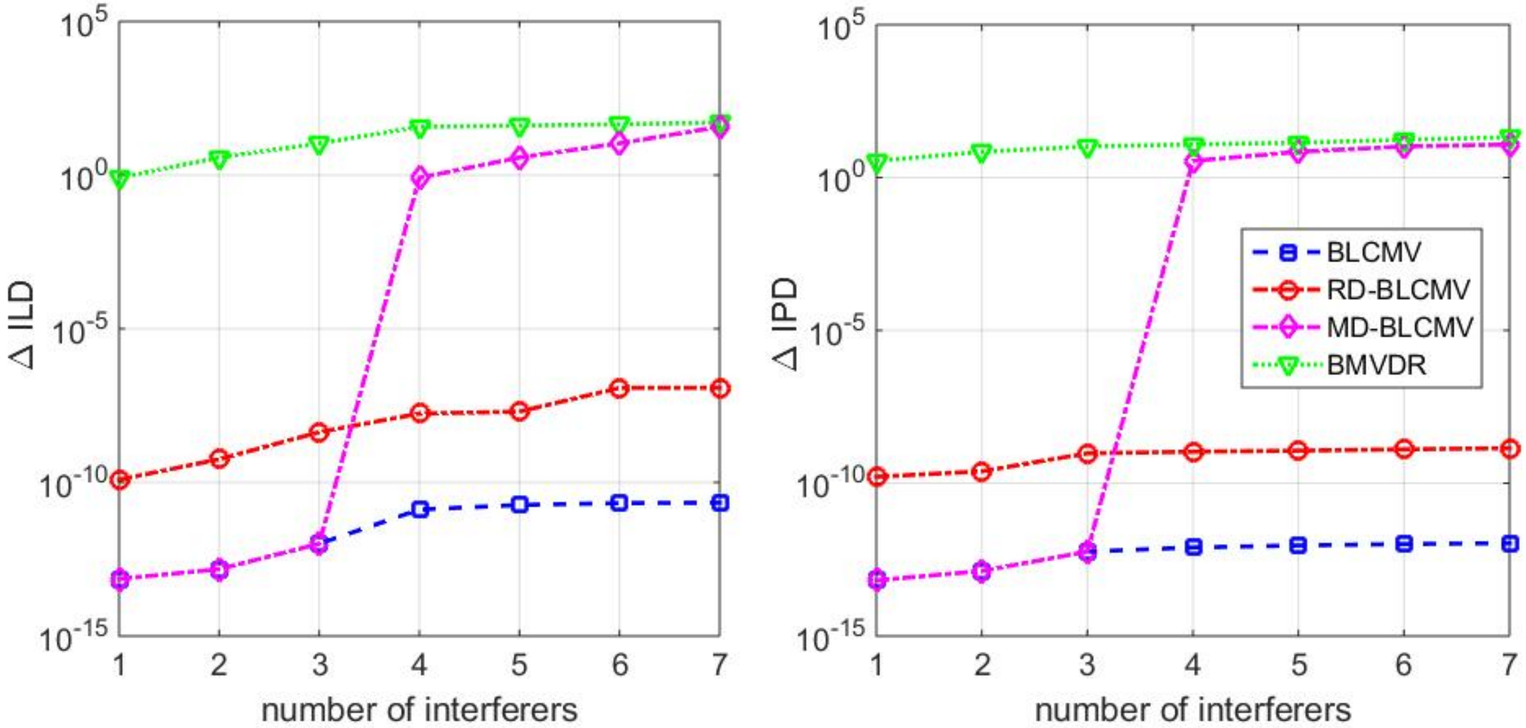
- Performance of spatial cue preservation:



$$\Delta\mathrm{ILD} = \sum_{j=1}^{\mathcal{J}} \sum_{\omega} \left(\mathrm{ILD}_j(\omega) - \widetilde{\mathrm{ILD}}_j(\omega)\right)^2; \qquad \Delta\mathrm{IPD} = \sum_{j=1}^{\mathcal{J}} \sum_{\omega} \left(\mathrm{IPD}_j(\omega) - \widetilde{\mathrm{IPD}}_j(\omega)\right)^2$$

Simulation results

- Performance of spatial cue preservation:



$$\Delta\text{ILD} = \sum_{j=1}^{\mathcal{J}} \sum_{\omega} \left(\text{ILD}_j(\omega) - \tilde{\text{ILD}}_j(\omega)\right)^2; \qquad \Delta\text{IPD} = \sum_{j=1}^{\mathcal{J}} \sum_{\omega} \left(\text{IPD}_j(\omega) - \tilde{\text{IPD}}_j(\omega)\right)^2$$

TU Delft

Conclusion

- We studied rate-distributed BLCMV beamforming for wireless binaural hearing aids. The proposed method was formulated by minimizing the energy usage and constraining the noise reduction performance.
- Under the utilization of a BLCMV beamformer, the problem was solved by semi-definite programming with the capability of joint noise reduction and binaural cue preservation.
- The proposed method can achieve better energy efficiency by distributing bit rates, and preserve more interferers' spatial cues by activating more sensors as compared to sensor selection approaches.

Conclusion

- We studied rate-distributed BLCMV beamforming for wireless binaural hearing aids. The proposed method was formulated by minimizing the energy usage and constraining the noise reduction performance.
- Under the utilization of a BLCMV beamformer, the problem was solved by semi-definite programming with the capability of joint noise reduction and binaural cue preservation.
- The proposed method can achieve better energy efficiency by distributing bit rates, and preserve more interferers' spatial cues by activating more sensors as compared to sensor selection approaches.

Thank you!

Appendix

- Matrix inversion lemma:

$$\tilde{\mathbf{R}}_{\mathbf{n+q}}^{-1} = (\tilde{\mathbf{R}}_{\mathbf{nn}} + \tilde{\mathbf{R}}_{\mathbf{qq}})^{-1} = \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} - \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1}(\tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} + \tilde{\mathbf{R}}_{\mathbf{qq}}^{-1})^{-1}\tilde{\mathbf{R}}_{\mathbf{nn}}^{-1},$$

where

$$\tilde{\mathbf{R}}_{\mathbf{nn}} = \begin{bmatrix} \mathbf{R}_{\mathbf{nn}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{nn}} \end{bmatrix}, \quad \tilde{\mathbf{R}}_{\mathbf{qq}} = \begin{bmatrix} \mathbf{R}_{\mathbf{qq}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{qq}} \end{bmatrix}.$$

- Convex relaxation:

$$\boldsymbol{\Lambda}^H \mathbf{R}_{\mathbf{n+q}}^{-1} \boldsymbol{\Lambda} \succeq \mathbf{Z}$$

$$\Longrightarrow \boldsymbol{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} \boldsymbol{\Lambda} - \mathbf{Z} \succeq \boldsymbol{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1}(\tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} + \tilde{\mathbf{R}}_{\mathbf{qq}}^{-1})^{-1}\tilde{\mathbf{R}}_{\mathbf{nn}}^{-1}\boldsymbol{\Lambda}$$

$$\Longrightarrow \begin{bmatrix} \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} + \tilde{\mathbf{R}}_{\mathbf{qq}}^{-1} & \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1}\boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} & \boldsymbol{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1}\boldsymbol{\Lambda} - \mathbf{Z} \end{bmatrix} \succeq \mathbf{O}_{2M+2\mathcal{I}+\mathcal{J}}.$$