Bias Analysis of Spatial Coherence-Based RTF Vector Estimation for Acoustic Sensor Networks in a Diffuse Sound Field

Wiebke Middelberg and Simon Doclo
Dept. of Medical Physics and Acoustics and Cluster of Excellence Hearing4all,
University of Oldenburg, Germany

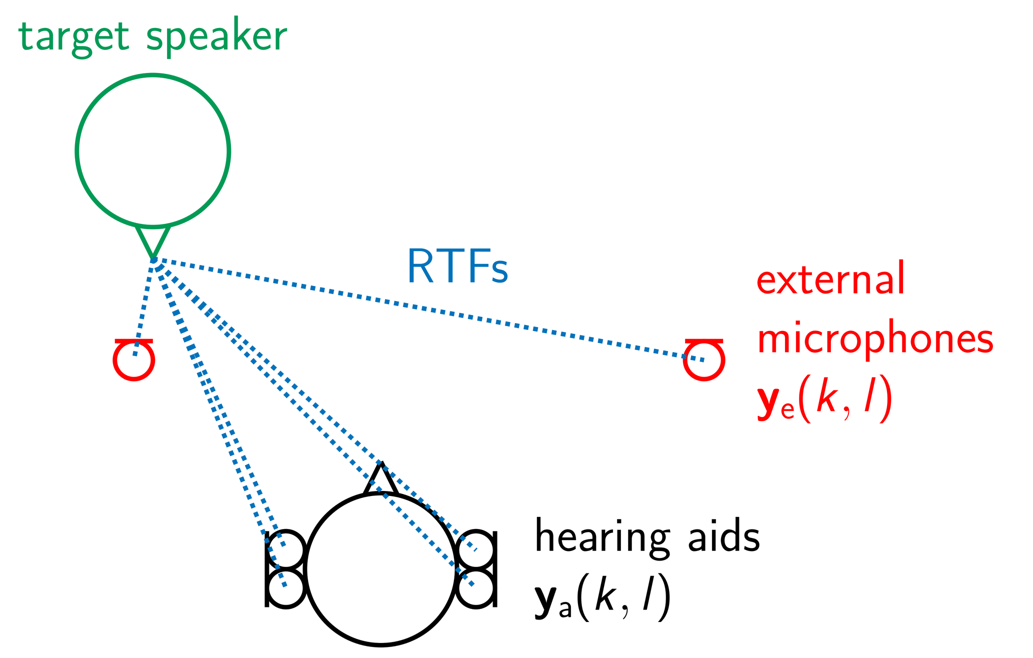
Problem Statement

- **Noise** reduces intelligibility of target speaker
- Exploit **external microphones** in conjunction with hearing aid microphones for speech enhancement
- MVDR beamformer to suppress undesired noise
  $\rightarrow$ **Relative transfer function (RTF) vector of target speaker is required** to steer beamformer [1, 2]
- **Blind estimation of target RTF vector** using external microphones [3, 4]

IN THIS POSTER

- Theoretical bias analysis of RTF vector estimation using **multiple** external microphones
- Comparison to bias analysis for RTF vector estimation using a **single** external microphone

Configuration and Notation



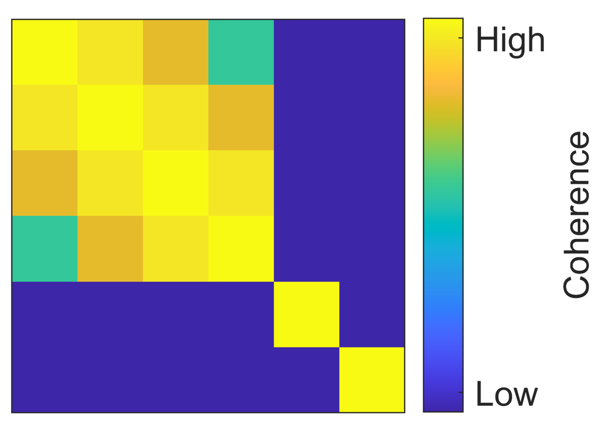
- Signal model in STFT domain (M_a hearing aid and M_e external microphones):
  $$\mathbf{y} = \mathbf{x} + \mathbf{n} \quad \text{with } \mathbf{y} = \begin{bmatrix}\mathbf{y}_a \\ \mathbf{y}_e\end{bmatrix}$$
- Covariance matrices:
  $\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_n$, with
  $\mathbf{R}_x = \phi_{x,1}\mathbf{h}\mathbf{h}^H$ rank-1 matrix spanned by the target RTF vector $\mathbf{h}$

Speech enhancement with external microphones

- Achieve noise reduction with MVDR beamformer: $\mathbf{w} = \frac{\mathbf{R}_n^{-1}\mathbf{h}}{\mathbf{h}^H\mathbf{R}_n^{-1}\mathbf{h}}$
- Target RTF vector $\mathbf{h}$ needs to be estimated:
  - State-of-the-art covariance whitening [5]
  - Spatial coherence (SC) method requiring external microphones [3, 4]

RTF Estimation Using Spatial Coherence

- **Assumption:**
  Noise between external microphones and all other microphones is uncorrelated [3, 4, 6]
- Spatial coherence method for **single** external microphone [3]
  $$\tilde{\mathbf{h}}^{SC}_{m_e} = \frac{\mathbf{R}_y\mathbf{e}_{e,m_e}}{\mathbf{e}_1^T\mathbf{R}_y\mathbf{e}_{e,m_e}}, \quad m_e \in \{1,\dots,M_e\}$$



- Perfect estimation except for entry corresponding to external microphone:
  Bias depends on SNR in external microphone
  $$\tilde{H}^{SC}_{e,m_e} = \left(1 + \frac{1}{\mathrm{SNR}_{e,m_e}}\right) H_{e,m_e}$$
- **Multiple** external microphones: multiple estimates of target RTF vector with bias in different entries $\rightarrow$ **linearly combine all biased SC RTF vector estimates**
  $$\tilde{\mathbf{h}}^{mSNR} = \tilde{\mathbf{H}}\boldsymbol{\alpha}$$
  $\rightarrow$ **mSNR approach**: determine weights $\boldsymbol{\alpha}$ by maximizing output SNR of MVDR beamformer **(mSNR approach)** [4]
  $$\max_{\boldsymbol{\alpha}} \frac{\boldsymbol{\alpha}^H \overbrace{\tilde{\mathbf{H}}^H\mathbf{R}_n^{-1}\mathbf{R}_y\mathbf{R}_n^{-1}\tilde{\mathbf{H}}}^{\mathbf{A}}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^H \underbrace{\tilde{\mathbf{H}}^H\mathbf{R}_n^{-1}\tilde{\mathbf{H}}}_{\mathbf{B}}\boldsymbol{\alpha}}, \quad \text{s.t. } \mathbf{1}^T\boldsymbol{\alpha} = 1$$
- Generalized eigenvalue decomposition (GEVD) as solution to optimization problem
  $$\boldsymbol{\alpha}^{GEVD} = \frac{\mathcal{P}\{\mathbf{B}^{-1}\mathbf{A}\}}{\mathbf{1}^T\mathcal{P}\{\mathbf{B}^{-1}\mathbf{A}\}}$$
- **Complex-valued** weights $\boldsymbol{\alpha}$ combine all SC RTF vector estimates

Bias Analysis of mSNR Approach

- **Assumption: Perfect estimation except for bias**
- Model for SC RTF vector estimates: $\tilde{\mathbf{H}} = \mathbf{H} + \mathbf{E}$
  $\rightarrow$ RTF matrix $\mathbf{H}$ contains true RTF vector: $\mathbf{H} = \mathbf{h}\mathbf{1}^T$

$$\rightarrow \text{Bias matrix: } \mathbf{E} = \begin{bmatrix} \mathbf{0}_{M_a\times 1} & \mathbf{0}_{M_a\times 1} & \dots & \mathbf{0}_{M_a\times 1} \\ \frac{H_{e,1}}{\mathrm{SNR}_{e,1}} & 0 & \dots & 0 \\ 0 & \frac{H_{e,2}}{\mathrm{SNR}_{e,2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \frac{H_{e,M_e}}{\mathrm{SNR}_{e,M_e}} \end{bmatrix}$$

Results of bias analysis

Theoretical model-based weights:

$$\boldsymbol{\alpha}^{model} = \frac{1}{\sum_{m_e'=1}^{M_e}\mathrm{SNR}_{e,m_e'}}\begin{bmatrix}\mathrm{SNR}_{e,1} \\ \vdots \\ \mathrm{SNR}_{e,M_e}\end{bmatrix}$$

- Real-valued weights
- RTF vector estimates weighted according to SNR in the respective external microphone

Bias for all external microphones:

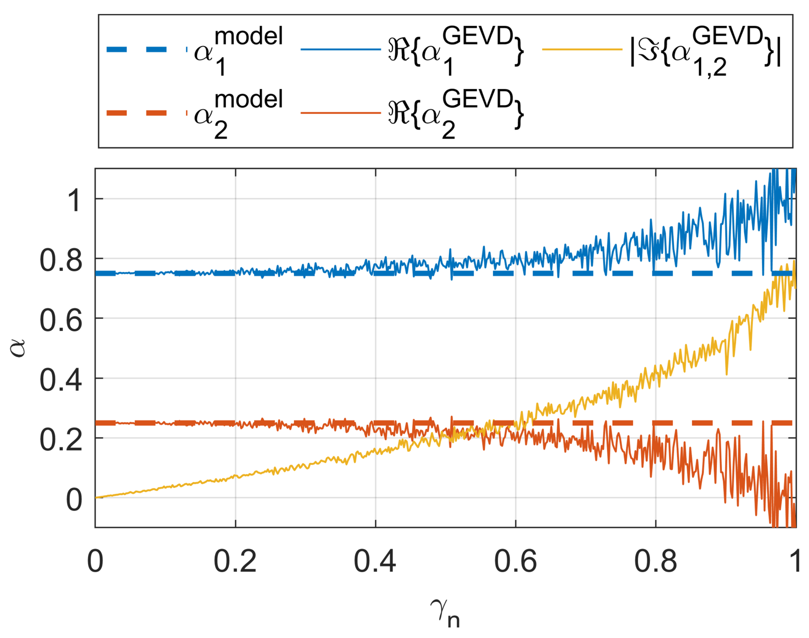
$$\tilde{H}^{mSNR}_{e,m_e} = \left(1 + \frac{1}{\sum_{m_e'=1}^{M_e}\mathrm{SNR}_{e,m_e'}}\right) H_{e,m_e}$$

- Bias always smaller than for single SC RTF vector estimates
- All entries of external microphones are biased

Validation and Evaluation

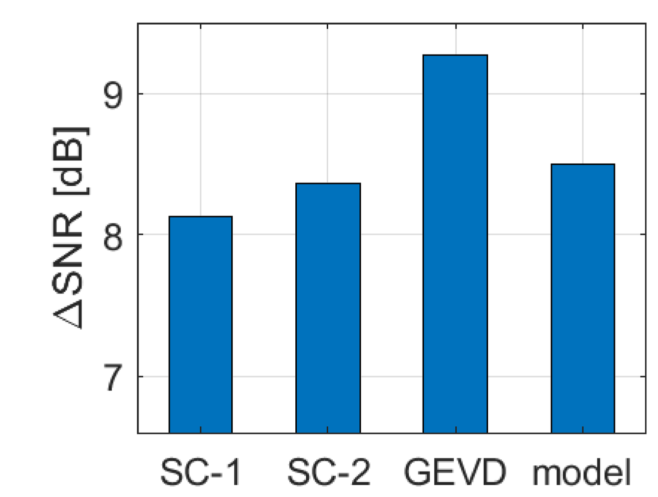
Validation

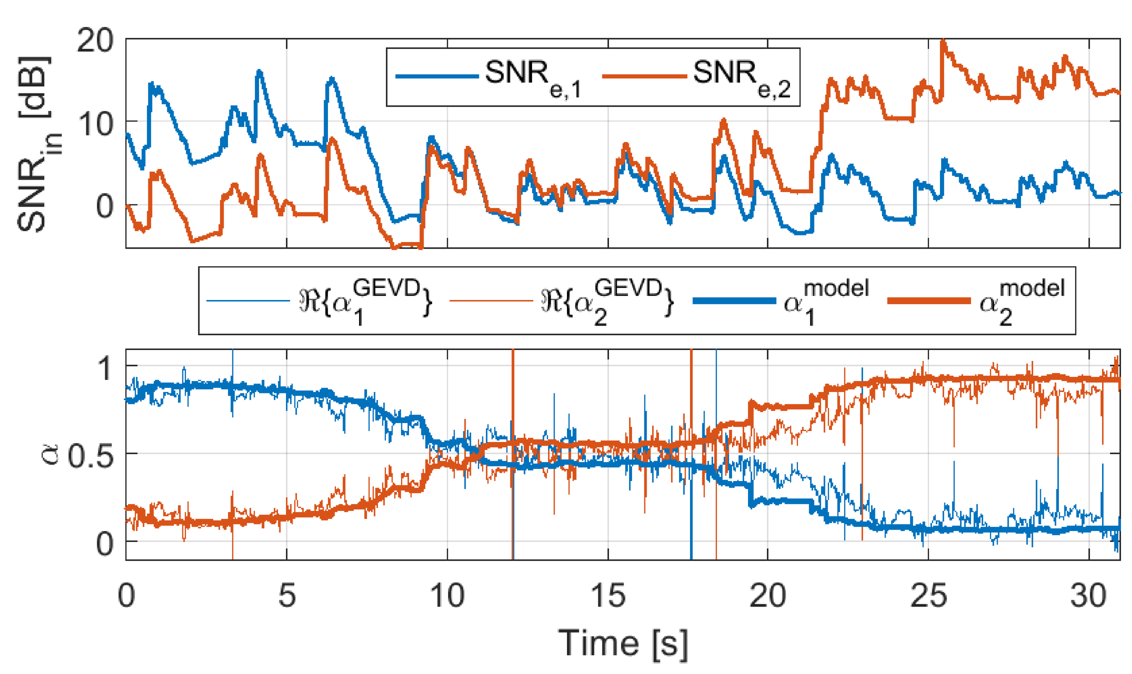
- Investigate deviations from spatial coherence assumption
- Model for data generation:
  $$\mathbf{R}_n = (1-\gamma_n)\boldsymbol{\Gamma}_n + \gamma_n\mathbf{b}\mathbf{b}^H$$
  $\rightarrow$ $\boldsymbol{\Gamma}_n$: fulfills SC assumption
  $\rightarrow$ $\mathbf{b}$: RTF vector of coherent noise
- If SC assumption is fulfilled ($\gamma_n = 0$): model-based weights are equivalent to GEVD-based weights (i.e., real-valued SNR weighting)
- **Violating assumption leads to deviations** between model- and GEVD-based weights



Evaluation

- Real-world recordings with $T_{60} \approx 400$ ms
- **Single moving speaker** and quasi-diffuse background noise
- 4 hearing aid microphones and 2 external microphones
- SNR_{in} in hearing aids varying from 0 - 6 dB
- Batch implementation in STFT domain with oracle covariance matrices





- **Model-based weights show good accordance with GEVD-based weights**
- **Better performance using GEVD-based weights**, presumably due to violation of SC and rank-1 assumption

Conclusions

- Analytical expression of weights in mSNR approach
  $\rightarrow$ **bias of mSNR approach always smaller than for single SC RTF vector estimates**
- Evaluation shows deviations between model-based weights and GEVD-based weights due to estimation errors and violation of assumptions

References

[1] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[2] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[3] N. Gößling and S. Doclo, "RTF-steered binaural MVDR beamforming incorporating an external microphone for dynamic acoustic scenarios," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK, May 2019, pp. 416–420.

[4] N. Gößling, W. Middelberg, and S. Doclo, "RTF-steered binaural MVDR beamforming incorporating multiple external microphones," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, Oct. 2019, pp. 368–372.

[5] S. Markovich-Golan and S. Gannot, "Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Brisbane, Australia, Apr. 2015, pp. 544–548.

[6] R. M. Corey and A. C. Singer, "Adaptive binaural filtering for a multiple-talker listening system using remote and on-ear microphones," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, Oct. 2021, pp. 1–5.

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2177/1 - Project ID 390895286 and Project ID 352015383 - SFB 1330 B2.