Comparison of Generalized Sidelobe Canceller Structures Incorporating External Microphones for Joint Noise and Interferer Reduction

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

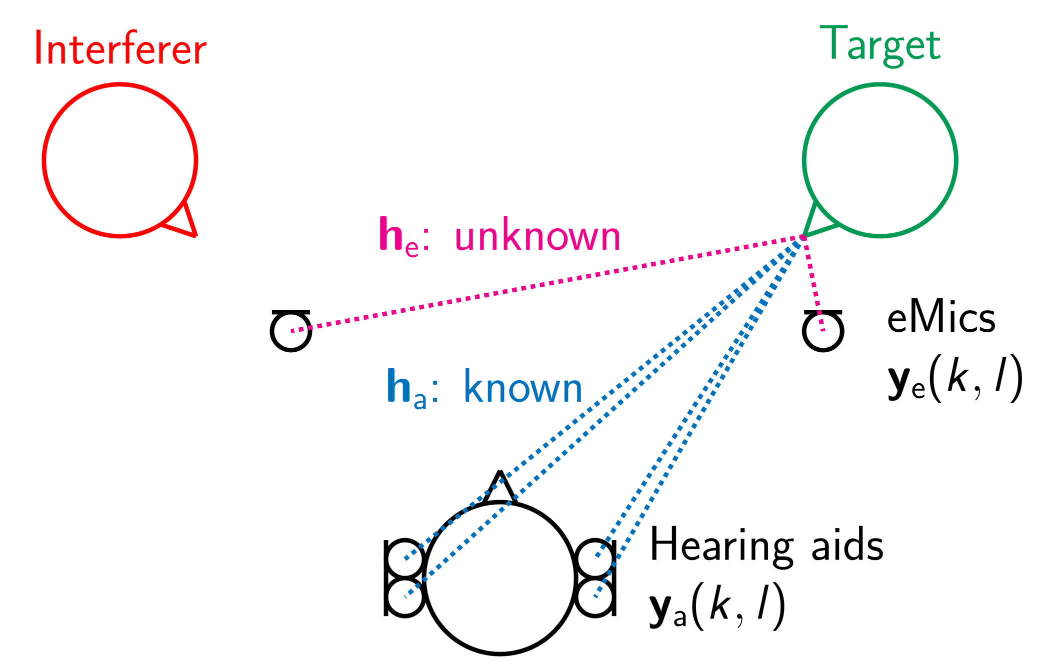
Problem Statement

- **Noise** and **interfering speaker** reduce intelligibility of target speaker
- Exploit **external microphones (eMics)** in conjunction with hearing aid microphones for speech enhancement
- MPDR-based beamforming to suppress undesired sources
  → **RTF vector of target speaker is required** to steer beamformer [1]
- **Blind estimation of target RTF vector** is difficult when interfering speaker is present
  → assume RTF vector for hearing aids to be known (e.g., frontal direction)
- **RTFs for eMics** are missing → need to be estimated

IN THIS POSTER

- Reduce noise and interferer by means of generalized sidelobe canceller (GSC) structures incorporating eMics [2]
- Pre-process local (and external) microphones to improve SIR and estimate external target RTFs more accurately
- Comparison of several GSC structures

Configuration and Notation



Signal model in STFT domain (hearing aid and external microphones):

$$\mathbf{y}(k,l) = \begin{bmatrix}\mathbf{y}_a(k,l)\\ \mathbf{y}_e(k,l)\end{bmatrix} = \underbrace{\mathbf{x}(k,l)}_{\text{desired}} + \underbrace{\mathbf{i}(k,l)}_{\text{interferer}} + \underbrace{\mathbf{n}(k,l)}_{\text{noise}}$$

Using relative transfer function (RTF) vectors: $\mathbf{x} = \mathbf{h}X_1, \quad \mathbf{i} = \mathbf{b}I_1, \quad \mathbf{h} = \begin{bmatrix}\mathbf{h}_a\\ \mathbf{h}_e\end{bmatrix}$

Accessibility of Information

- $\mathbf{R}_y = \phi_x \mathbf{h}\mathbf{h}^H + \phi_i \mathbf{b}\mathbf{b}^H + \mathbf{R}_n$ → assume that $\mathbf{R}_n$ can be estimated (e.g., VAD)
- Assume relative position of target speaker with respect to to hearing aids to be known:
  → local target RTF vector $\mathbf{h}_a$ known
  → external target RTF vector $\mathbf{h}_e$ and interferer RTF vector $\mathbf{b}$ unknown
  → to incorporate eMics in GSC structures $\mathbf{h}_e$ needs to be estimated

RTF Vector Estimation

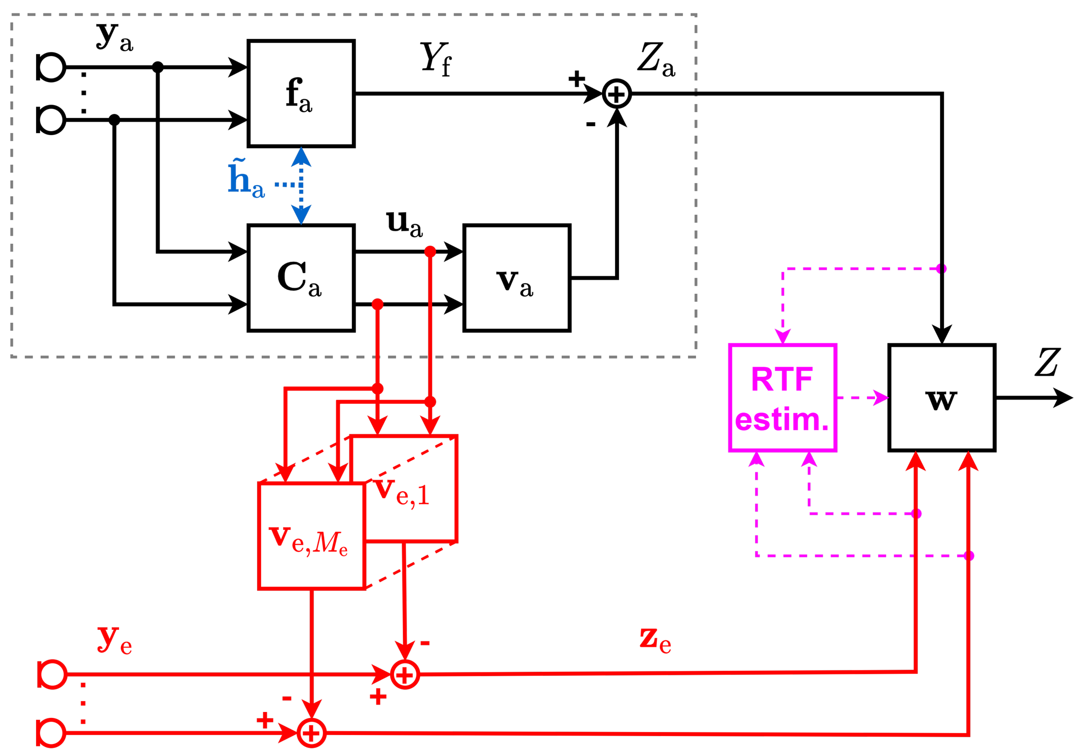
State-of-the-art RTF vector estimator **covariance whitening (CW)** [3]:

Whiten $(\mathbf{R}_y - \mathbf{R}_n)$: $\mathbf{R}_y^w = \phi_x \mathbf{R}_n^{-1/2}\mathbf{h}\mathbf{h}^H\mathbf{R}_n^{-H/2} + \phi_i \mathbf{R}_n^{-1/2}\mathbf{b}\mathbf{b}^H\mathbf{R}_n^{-H/2}$ → Principal eigenvector of $\mathbf{R}_y^w$: $\mathbf{v}_{\max} = \mathcal{P}\{\mathbf{R}_y^w\}$ → De-whiten and normalize: $\hat{\mathbf{h}}^{CW} = \frac{\mathbf{R}_n^{1/2}\mathbf{v}_{\max}}{\mathbf{e}_1^T\mathbf{R}_n^{1/2}\mathbf{v}_{\max}}$

Problem of Blind RTF Vector Estimation

$\mathbf{R}_y^w$ is rank-2 due to interfering speaker
→ CW will give **biased RTF vector estimate!**
→ Dependence on multi-channel signal-to-interferer ratio (SIR)

GSC Structures



1. Local GSC (L-GSC) [4, 5]

- Only uses hearing aid microphones (gray box)
- Exploits a-priori RTF vector $\tilde{\mathbf{h}}_a$

1. Fixed beamformer $\mathbf{f}_a$ → speech ref. Y_f
2. Blocking matrix $\mathbf{C}_a$ → noise-and-interferer refs. $\mathbf{u}_a$
3. Filter $\mathbf{v}_a$ → reduces correlation between Y_f and $\mathbf{u}_a$ to create output Z_a

2. GSC with External Speech References (GSC-ESR)

- **Novelty: Change MVDR [2] to MPDR implementation** to cancel interferer (complete diagram)
- Pre-process eMic signals $\mathbf{y}_e$ by noise-and-interferer refs. $\mathbf{u}_a$ and filters $\mathbf{v}_{e,m_e}$
- Enhanced local output Z_a and enhanced eMic signals $\mathbf{z}_e$ lead to higher mean SIR and better estimation of external RTF vector $\mathbf{h}_e$
  → used in joint beamformer $\mathbf{w}$

3. GSC with External References (GSC-ER)

- Simplified version of GSC-ESR (complete diagram without filters $\mathbf{v}_{e,m_e}$)
- No pre-processing of eMics
  → $\mathbf{v}_{e,m_e} = \mathbf{0}$
- Allows to assess benefit of pre-processing

Experimental Evaluation

- Reverberant recordings ($T_{60} \approx 350$ ms)
- 4 head-mounted microphones on a dummy head + 2 eMics
- Male target (35° to the right), female interferer (35° to the left)

Conditions

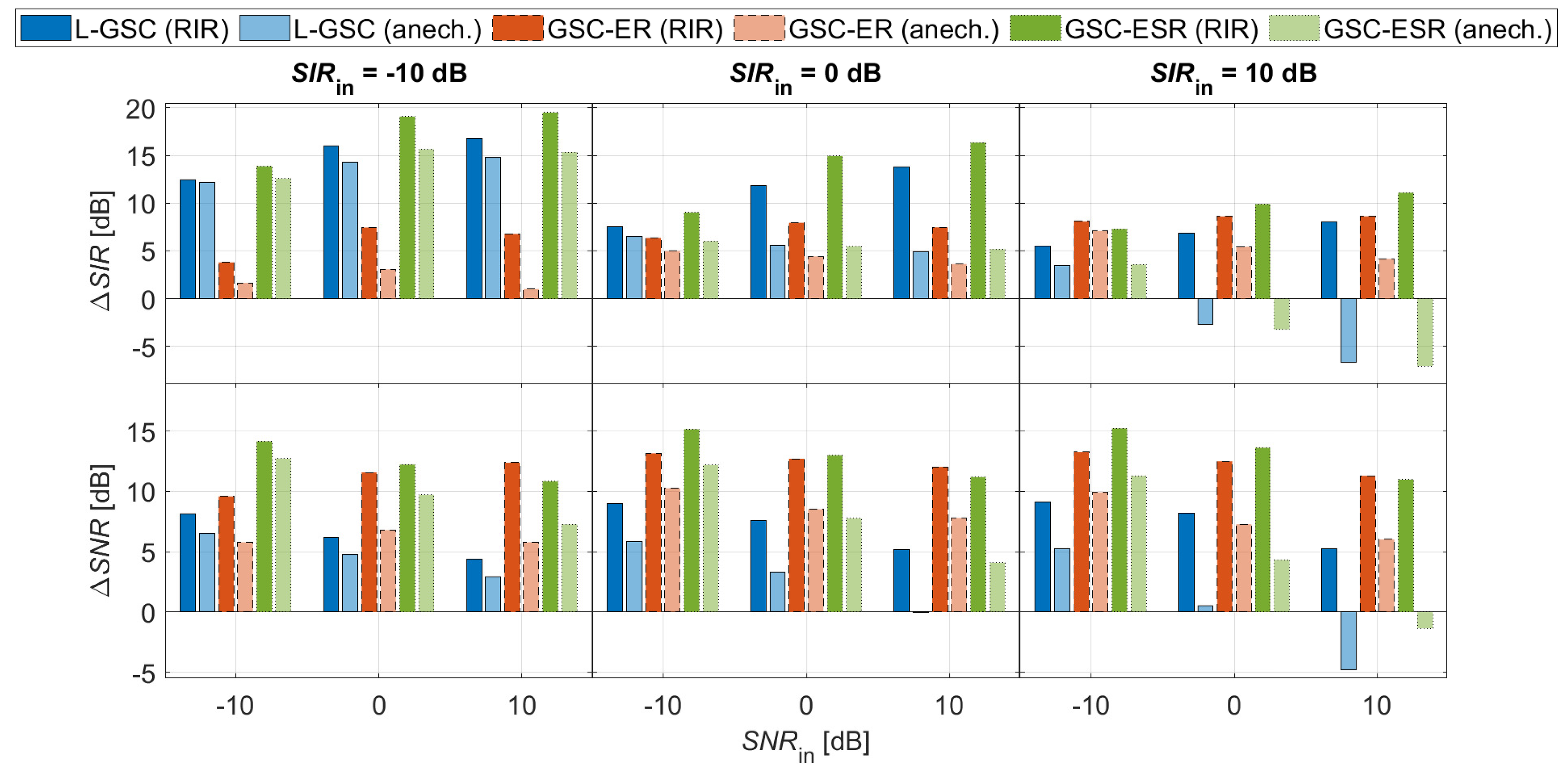
- $SIR_{in} = [-10, 0, 10]$ dB, $SNR_{in} = [-10, 0, 10]$ dB
- Two different a-priori RTF vectors $\tilde{\mathbf{h}}_a$:
  1. Reverberant RTF from measured target RIR
  2. Approximation from anechoic database [6]

Implementation and Framework

- Batch implementation
- 64 ms frame length with 50% overlap, sqrt-Hann window

Results

Evaluation of SNR improvement ΔSNR and SIR improvement ΔSIR



- Including eMics leads to better performance than processing hearing aids alone
- **Anechoic RTF vector** leads to overall lower scores than reverberant RTF vector
- **At high SIR:**
  - Using reverberant RTF vector: GSC-ESR and GSC-ER perform similarly (both better than L-GSC)
  - Using anechoic RTF vector: GSC-ESR performs worse than GSC-ER
  → target cancellation in eMic signals due to speech leakage in $\mathbf{u}_a$
- **At low SIR:** GSC-ESR outperforms L-GSC and GSC-ER in terms of noise and interferer reduction

Conclusions

✓ GSC-ESR outperforms L-GSC and GSC-ER in difficult conditions
  → Advantage of pre-processing eMics signals

✗ Sensitivity towards RTF vector mismatches
  → Especially for GSC-ESR at high SIR and SNR

Next Steps:
Analytical expression for performance of GSC structures

References

[1] 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.

[2] R. Ali, G. Bernardi, T. van Waterschoot, and M. Moonen, "Methods of extending a generalized sidelobe canceller with external microphones," IEEE/ACM Trans. on Audio, Speech, and Language Processing, vol. 27, pp. 1349–1364, Sep. 2019.

[3] S. Markovich-Golan and S. Gannot, "Performance analysis of the covariance-whitening and the covariance-subtraction methods for estimating the relative transfer function," in Proc. European Signal Processing Conference, Rome, Italy, Sep. 2018, pp. 544–548.

[4] L. J. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. on Antennas and Propagation, vol. 30, no. 1, pp. 27–34, Jan. 1982.

[5] 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.

[6] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel In-Ear and Behind-The-Ear Head-Related and Binaural Room Impulse Responses," Eurasip Journal on Advances in Signal Processing, vol. 2009, p. 10 pages, 2009.