Real-Time Evaluation of an RTF-Steered Binaural MVDR Beamformer Incorporating an External Microphone

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

Problem Statement

- Noise reduction:
  - Background noise reduces intelligibility of speech
  - Equal amplification of desired source and noise in hearing aids without filtering
- Cue preservation:
  - Awareness of acoustic scenes
  - Without cues mismatch between acoustic and visual information

In this Poster a real-time implementation of an RTF-steered **binaural MVDR** beamformer is presented and evaluated. A computationally cheap RTF estimator, which exploits an **external microphone**, is used [1, 2].

Signal Model

- Microphone signals stacked in one vector:

$$\mathbf{y} = [Y_{L1}(\omega), Y_{L2}(\omega), Y_{R1}(\omega), Y_{R2}(\omega)]^T, \qquad \bar{\mathbf{y}} = [\mathbf{y}^T, Y_E(\omega)]^T$$

- Noisy signal $\mathbf{y}$ decomposed into speech and noise:

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \quad \text{with} \quad \mathbf{x} = \mathbf{a}S$$

- Covariance matrices:

$$\mathbf{R}_y = \mathcal{E}\{\mathbf{y}\mathbf{y}^H\}, \quad \mathbf{R}_x = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}, \quad \mathbf{R}_n = \mathcal{E}\{\mathbf{n}\mathbf{n}^H\}$$

Binaural MVDR

Aims on minimizing output noise PSD while preserving the desired source:

$$\min_{\mathbf{w}} \mathbf{w}^H\mathbf{R}_n\mathbf{w}, \quad \text{s.t.} \quad \mathbf{w}^H\mathbf{h} = 1 \quad \Rightarrow \text{Solution:} \quad \mathbf{w} = \frac{\mathbf{R}_n^{-1}\mathbf{h}}{\mathbf{h}^H\mathbf{R}_n^{-1}\mathbf{h}}$$

Requirements:

- Two reference microphones (one at each side of the head) and two RTF vectors [3] with corresponding selection vectors $\mathbf{e}_{L/R}$
  $\Rightarrow$ Spatial perception
- Needs estimate of $\mathbf{R}_n$, obtained by SPP [4]

RTF Estimation Exploiting an External Microphone

Assuming sufficiently large distance between local array and external microphone and diffuse noise

$$\Rightarrow \quad \mathcal{E}\{\mathbf{n}N_E^*\} = \mathbf{0}$$

Extended covariance matrix:

$$\bar{\mathbf{R}}_y = \mathcal{E}\{\bar{\mathbf{y}}\bar{\mathbf{y}}^H\} = \begin{bmatrix} \mathbf{R}_y & \mathcal{E}\{\mathbf{y}Y_E^*\} \\ \mathcal{E}\{\mathbf{y}^H Y_E\} & \phi_{y,E} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_y & \mathcal{E}\{\mathbf{x}X_E^*\} \\ \mathcal{E}\{\mathbf{x}^H X_E\} & \phi_{y,E} \end{bmatrix}$$

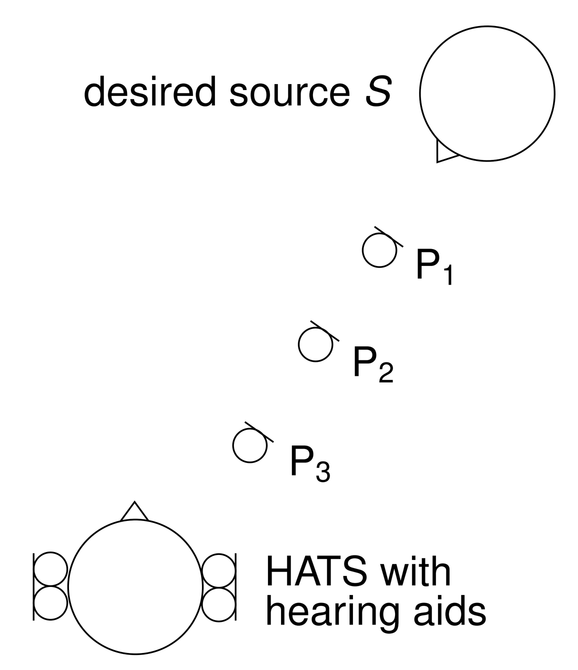
Adaptive estimation of RTF vector:

$$\mathbf{h} = \frac{\mathcal{E}\{\mathbf{y}Y_E^*\}}{\mathcal{E}\{Y_{ref}Y_E^*\}}$$

Algorithm 1: RTF Estimation Using Spatial Coherence
Input: $\bar{\mathbf{y}}(l)$, $\bar{\mathbf{R}}_y(l-1)$, $SPP(l)$
Output: $\mathbf{h}(l)$, $\bar{\mathbf{R}}_y(l)$
Parameter: α_y
for right and left **do**
  for all k **do**
    if $SPP \geq 0.6$ **then**
      $\bar{\mathbf{R}}_y(l) = \alpha_y\bar{\mathbf{R}}_y(l-1) + (1-\alpha_y)\bar{\mathbf{y}}(l)\bar{\mathbf{y}}(l)^H$;
    else
      $\bar{\mathbf{R}}_y(l) = \bar{\mathbf{R}}_y(l-1)$;
    $\mathbf{h}(l) = [\mathbf{I}, \mathbf{0}]\frac{\bar{\mathbf{R}}_y(l)\mathbf{e}}{\mathbf{e}_{L,R}^T\bar{\mathbf{R}}_y(l)\mathbf{e}}$;

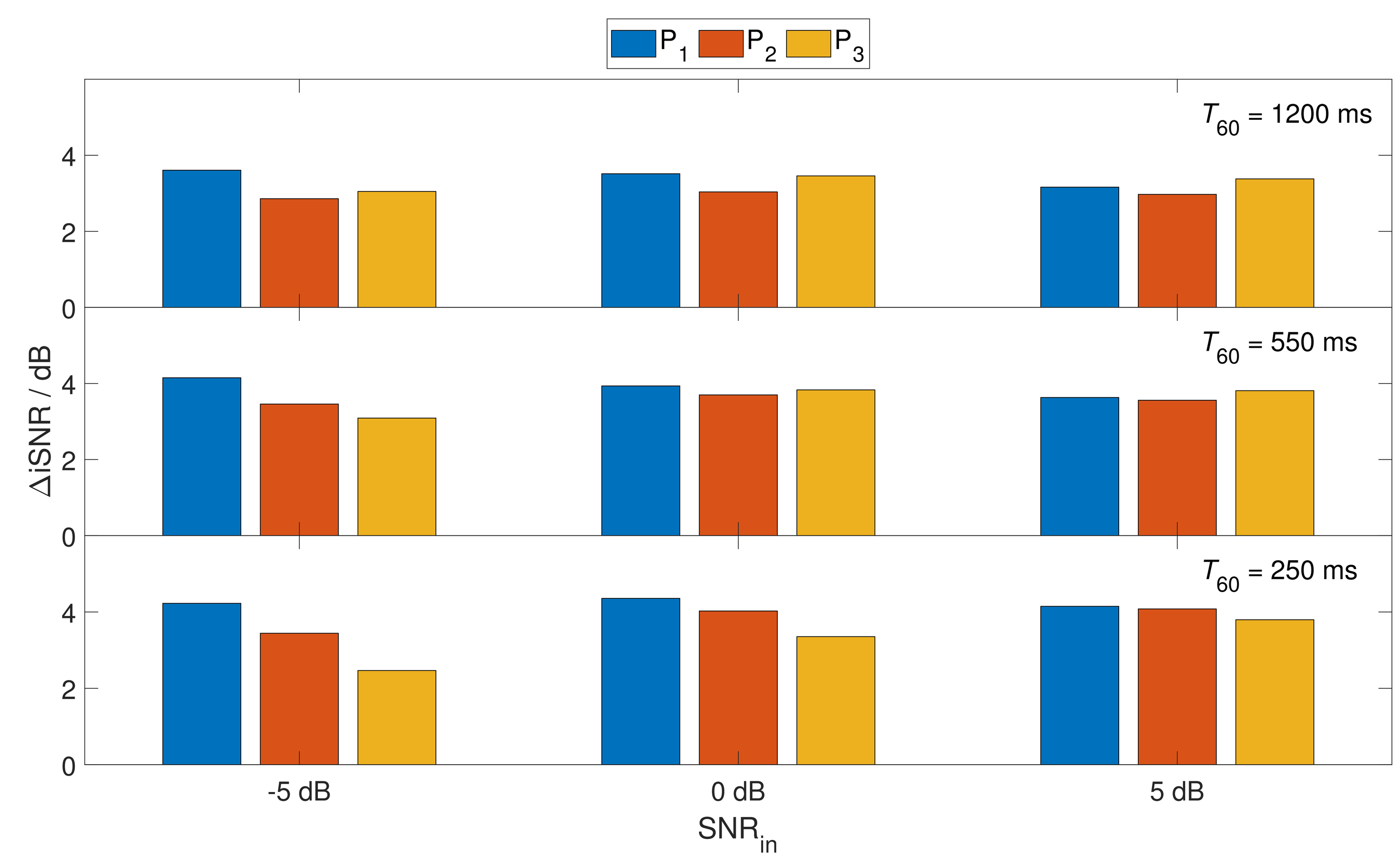
Experimental Setup

Scenario 1:
Speaker to the right (d_S = 2 m), external microphone to the right (d_E = 1.5 m), T_{60} = 600 ms, SNR_{in} = 0 dB
Scenario 2:
Speaker to the left (d_S = 2 m), external microphone to the right (d_E = 1 m), T_{60} = 600 ms, SNR_{in} = 0 dB
Scenario 3:
Moving speaker (from right to left), lapel microphone, T_{60} = 600 ms, $SNR_{in} \approx$ 0 dB



1. **Measurements & Setup**
   - HATS in *Variable Acoustics Laboratory* University Oldenburg
   - One active speaker in diffuse babble noise
   - T_{60} = {250, 550, 1200} ms
   - SNR_{in} = {-5, 0, 5} dB
   - Three different positions of external microphone
2. **WOLA Framework**
   - Sampling rate: f_s = 32 kHz
   - FFT length: n_{fft} = 1024
   - Sqrt. Hann Window, 50 % overlap
   - Ref. mics.: front on left and right side (channel 1 and 3)
3. **Performance measures**
   - Intelligibility weighted SNR improvement ΔiSNR

Results



- Applicability of SC-based RTF estimator in real-time framework
- Overall high stability in various acoustic scenarios
- More stable at high input SNR
- Better noise reduction performance at lower T_{60} and higher input SNR in external microphone

Conclusion & Outlook

In this real-time implementation, it has been shown that the RTF-steered MVDR beamformer using the SC method, leads to

✓ Good noise reduction performance
✓ Cue preservation of the desired source
✓ Low computational complexity

But

✗ Needs external microphone

- More external microphones with combined RTF estimates
- Post-filtering
- GSC implementation

References

[1] N. Gößling and S. Doclo. RTF-based binaural MVDR beamformer exploiting an external microphone in a diffuse noise field. In Proc. ITG Conference on Speech Communication, pages 106–110, Oldenburg, Germany, Oct. 2018.

[2] N. Gößling and S. Doclo. Relative transfer function estimation exploiting spatially separated microphones in a diffuse noise field. In Proc. International Workshop on Acoustic Signal Enhancement (IWAENC), pages 146–150, Tokyo, Japan, Sep. 2018.

[3] D. Marquardt. Development and Evaluation of Psychoacoustically Motivated Binaural Noise Reduction and Cue Preservation Techniques. PhD thesis, Carl von Ossietzky Universität Oldenburg, 2015.

[4] T. Gerkmann and R. C. Hendriks. Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay. IEEE Transactions on Audio, Speech, and Language Processing, 20(4):1383–1393, May 2012.