Comparison of RTF Estimation Methods between a Head-Mounted Binaural Hearing Device and an External Microphone

Nico Gößling, Daniel Marquardt and Simon Doclo

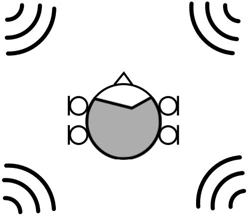
Department of Medical Physics and Acoustics
Signal Processing Group
Carl von Ossietzky University of Oldenburg

August 19, 2017, Stockholm, Sweden

Motivation

Motivation

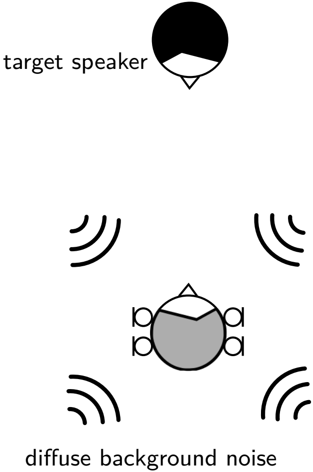
- Improve speech intelligibility in noisy scenarios



diffuse background noise

Motivation

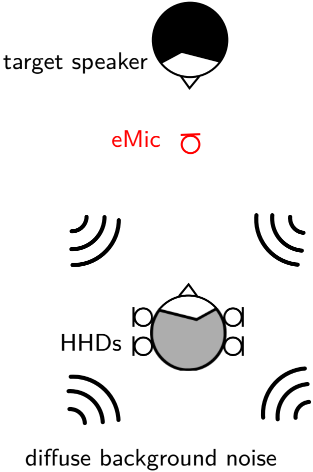
- Improve speech intelligibility in noisy scenarios
- Preserve binaural cues to assure spatial awareness



Motivation

- Improve speech intelligibility in noisy scenarios
- Preserve binaural cues to assure spatial awareness
- Adding an **external microphone** (eMic) to **head-mounted hearing devices** (HHDs) improves algorithm performance

[Szurley2016], [Gößling2017]

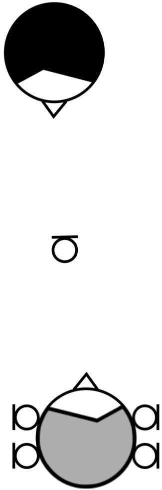


Motivation

- Fixed beamforming is possible for head-mounted microphones

Motivation

- Fixed beamforming is possible for head-mounted microphones
- **Position of eMic is unknown**

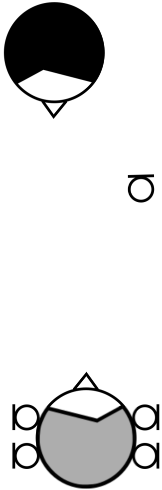


Motivation

- Fixed beamforming is possible for head-mounted microphones
- **Position of eMic is unknown**

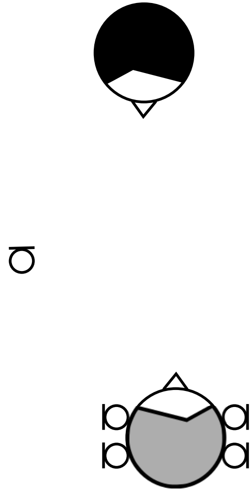
Motivation

- Fixed beamforming is possible for head-mounted microphones
- **Position of eMic is unknown**



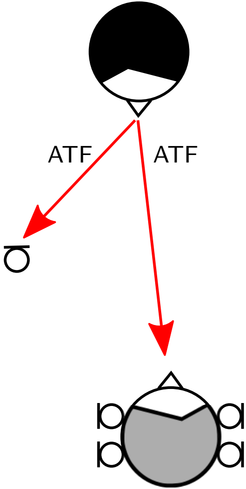
Motivation

- Fixed beamforming is possible for head-mounted microphones
- **Position of eMic is unknown**



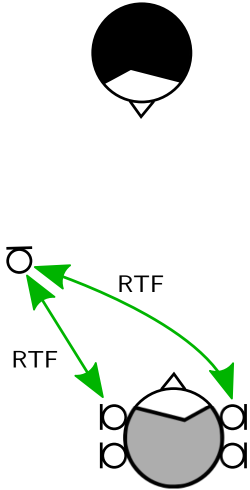
Motivation

- Fixed beamforming is possible for head-mounted microphones
- **Position of eMic is unknown**
- Estimate the **acoustic transfer functions** (ATFs) or...



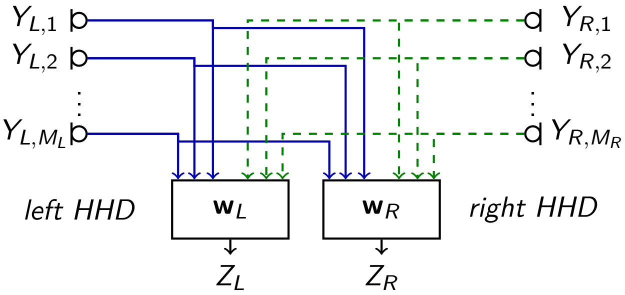
Motivation

- Fixed beamforming is possible for head-mounted microphones
- **Position of eMic is unknown**
- Estimate the **acoustic transfer functions** (ATFs) or...
- ... estimate the **relative transfer functions** (RTFs) between eMic and HHDs

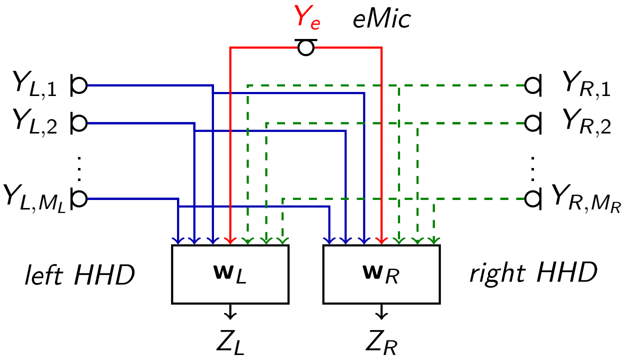


Binaural Noise Reduction

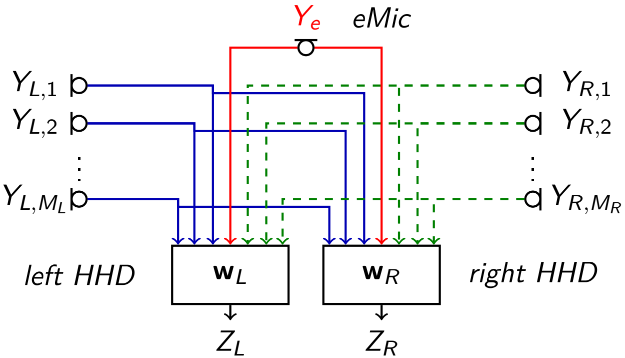
Extended Binaural Noise Reduction System



Extended Binaural Noise Reduction System



Extended Binaural Noise Reduction System



Input signal vector:
$\mathbf{y}(k,l) = \mathbf{x}(k,l) + \mathbf{n}(k,l)$

Single speech source:
$\mathbf{x}(k,l) = \mathbf{a}(k,l)S(k,l)$

Relative Transfer Functions (RTFs)

The relative transfer function (RTF) vectors, relating the ATF vector to the reference microphones, are defined as:

$$\mathbf{h}_L = \frac{\mathbf{a}}{A_L} = \left[1,\ \frac{A_{L,2}}{A_L},\ \ldots,\ \frac{A_{L,M_L}}{A_L},\ \frac{A_{R,1}}{A_L},\ \ldots,\ \frac{A_{R,M_R}}{A_L},\ \frac{A_e}{A_L}\right]^T$$

$$\mathbf{h}_R = \frac{\mathbf{a}}{A_R} = \left[\frac{A_{L,1}}{A_R},\ \ldots,\ \frac{A_{L,M_L}}{A_R},\ 1,\ \frac{A_{R,2}}{A_R},\ \ldots,\ \frac{A_{R,M_R}}{A_R},\ \frac{A_e}{A_R}\right]^T$$

Relative Transfer Functions (RTFs)

The relative transfer function (RTF) vectors, relating the ATF vector to the reference microphones, are defined as:

$$\mathbf{h}_L = \frac{\mathbf{a}}{A_L} = \left[1,\ \frac{A_{L,2}}{A_L},\ \ldots,\ \frac{A_{L,M_L}}{A_L},\ \frac{A_{R,1}}{A_L},\ \ldots,\ \frac{A_{R,M_R}}{A_L},\ \frac{A_e}{A_L}\right]^T$$

$$\mathbf{h}_R = \frac{\mathbf{a}}{A_R} = \left[\frac{A_{L,1}}{A_R},\ \ldots,\ \frac{A_{L,M_L}}{A_R},\ 1,\ \frac{A_{R,2}}{A_R},\ \ldots,\ \frac{A_{R,M_R}}{A_R},\ \frac{A_e}{A_R}\right]^T$$

Last entry relates the eMic to the (head-mounted) reference microphones and needs to be estimated

Binaural MVDR Beamforming

- Minimizing output noise power
- Preserving speech component in reference microphone signals

Optimization problem for left filter

$$\mathbf{w}_L = \arg\min_{\mathbf{w}} \mathcal{E}\left\{\left|\mathbf{w}^H\mathbf{n}\right|^2\right\} \quad \text{s.t.} \quad \mathbf{w}^H\mathbf{x} = X_L$$

Solution for left filter

$$\mathbf{w}_L = \frac{\mathbf{R}_n^{-1}\mathbf{a}}{\mathbf{a}^H\mathbf{R}_n^{-1}\mathbf{a}} A_L^* = \frac{\mathbf{R}_n^{-1}\mathbf{h}_L}{\mathbf{h}_L^H\mathbf{R}_n^{-1}\mathbf{h}_L}$$

Binaural MVDR Beamforming

- Minimizing output noise power
- Preserving speech component in reference microphone signals

Optimization problem for left filter

$$\mathbf{w}_L = \arg\min_{\mathbf{w}} \mathcal{E}\left\{\left|\mathbf{w}^H\mathbf{n}\right|^2\right\} \quad \text{s.t.} \quad \mathbf{w}^H\mathbf{x} = X_L$$

Solution for left filter

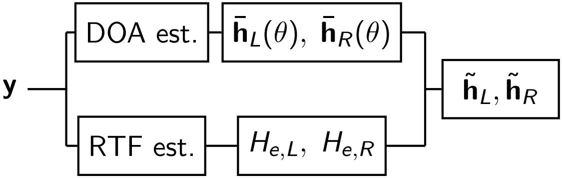
$$\mathbf{w}_L = \frac{\mathbf{R}_n^{-1}\mathbf{a}}{\mathbf{a}^H\mathbf{R}_n^{-1}\mathbf{a}} A_L^* = \frac{\mathbf{R}_n^{-1}\mathbf{h}_L}{\mathbf{h}_L^H\mathbf{R}_n^{-1}\mathbf{h}_L}$$

ATFs are not required to steer the beamformer!

RTF estimation methods

RTF vector construction

- **Fixed beamforming** for head-mounted microphones, e.g, based on direction-of-arrival θ (DOA) estimation
- RTFs can be measured or simulated in advance
- Not possible for eMic $\Rightarrow$ **RTF estimation is needed**



RTF estimation methods

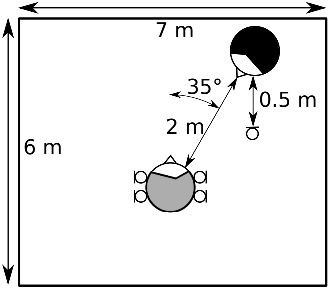
Ground truth:

$$\boxed{H_{e,L} = \frac{\mathbf{e}_e^T \mathbf{R}_x \mathbf{e}_L}{\mathbf{e}_L^T \mathbf{R}_x \mathbf{e}_L} = \frac{A_e}{A_L}}$$

- Biased approach: $H_{e,L}^{\mathrm{b}} = \frac{\mathbf{e}_e^T \mathbf{R}_y \mathbf{e}_L}{\mathbf{e}_L^T \mathbf{R}_y \mathbf{e}_L}$
- MVDR pre-processing approach: $H_{e,L}^{\mathrm{pp}} = \frac{\mathbf{e}_e^T \mathbf{R}_y \mathbf{w}_{\mathrm{H},L}}{\mathbf{w}_{\mathrm{H},L}^H \mathbf{R}_y \mathbf{w}_{\mathrm{H},L}}$
- Covariance whitening approach: $H_{e,L}^{\mathrm{cw}} = \frac{\mathbf{e}_e^T \mathbf{R}_n^{1/2} \mathbf{v}_{\max}}{\mathbf{e}_L^T \mathbf{R}_n^{1/2} \mathbf{v}_{\max}}$

Experimental results

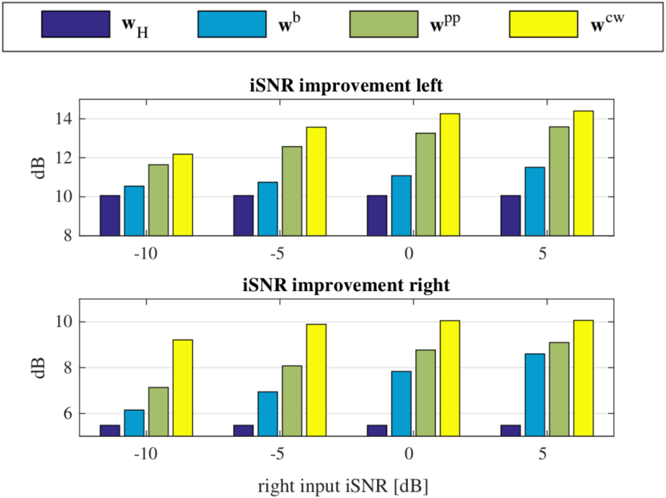
Experimental setup



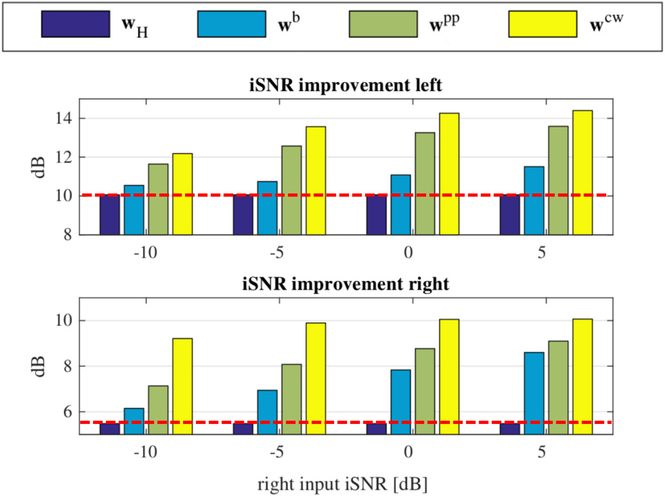
- Male target speaker
- Diffuse babble noise
- $M_L = M_R = 2$
- T60 = 350 ms
- 16 kHz @ 16 ms (50% OL)

- Assuming $\theta = 35^\circ$
- $\mathbf{R}_n$ estimation during 2 s noise-only initialization
- $\mathbf{R}_y$ estimation during 18 s speech-plus-noise
- Batch filtering

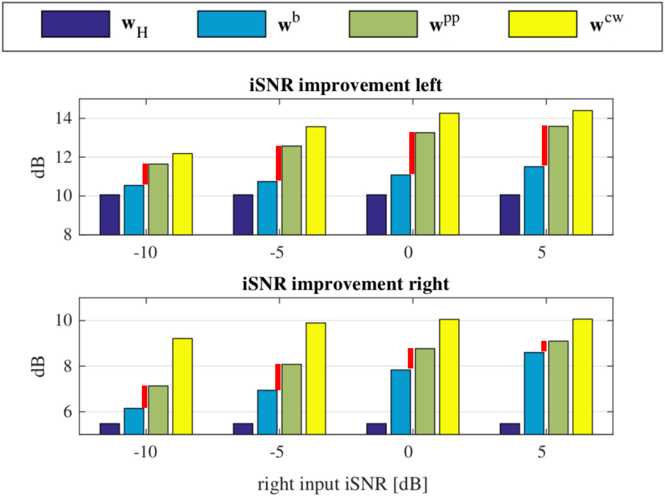
Experimental results



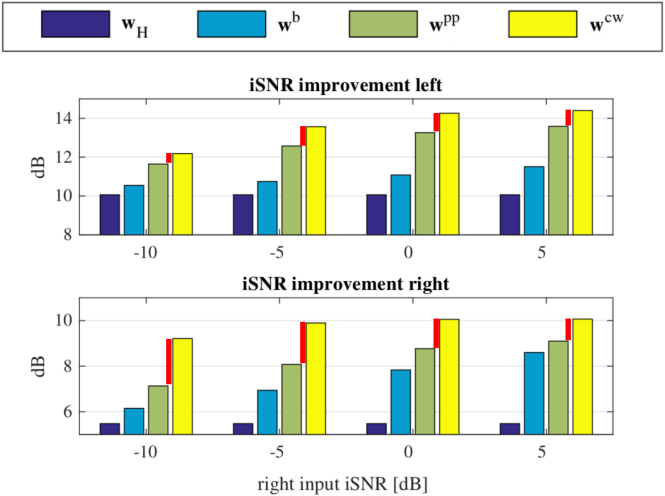
Experimental results



Experimental results



Experimental results



Conclusion and Outlook

Conclusion:

1. It is possible to estimate RTFs and incorporate an eMic into a binaural MVDR beamformer in an experimental scenario based on real-world signals
2. Pre-processing proved beneficial compared to baised approach
3. Covariance whitening outperformed the other approaches

Outlook:

1. Directional noise sources (extended BLCMV beamformer)
2. Practical problems, e.g. synchronization (influence of clock drift/offset), online implementations and dynamic scenarios

Conclusion and Outlook

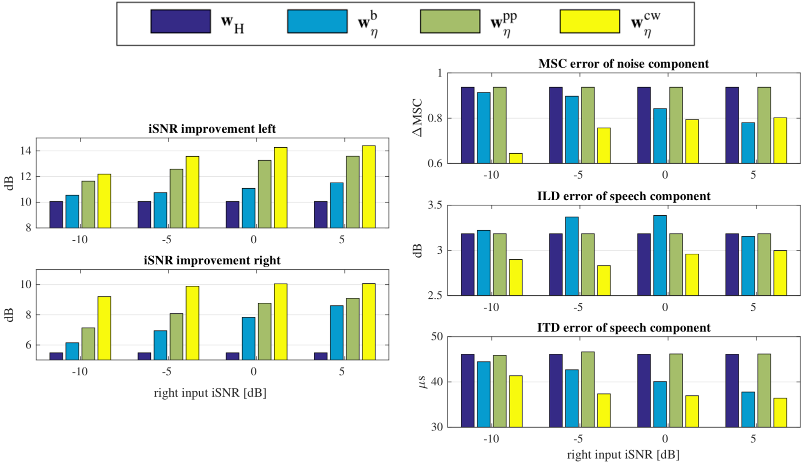
Conclusion:

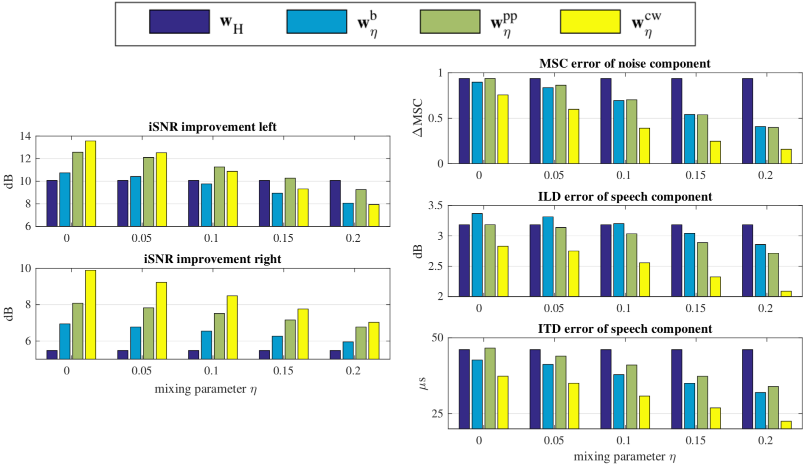
1. It is possible to estimate RTFs and incorporate an eMic into a binaural MVDR beamformer in an experimental scenario based on real-world signals
2. Pre-processing proved beneficial compared to baised approach
3. Covariance whitening outperformed the other approaches

Outlook:

1. Directional noise sources (extended BLCMV beamformer)
2. Practical problems, e.g. synchronization (influence of clock drift/offset), online implementations and dynamic scenarios

Thanks for your attention!


$\mathbf{w}_H$
$\mathbf{w}_\eta^b$
$\mathbf{w}_\eta^{pp}$
$\mathbf{w}_\eta^{cw}$
iSNR improvement left
dB
14
12
10
8
-10
-5
0
5
iSNR improvement right
dB
10
8
6
right input iSNR [dB]
MSC error of noise component
ΔMSC
1
0.8
0.6
ILD error of speech component
3.5
3
2.5
ITD error of speech component
μs
50
40
30
right input iSNR [dB]


$\mathbf{w}_H$
$\mathbf{w}_\eta^b$
$\mathbf{w}_\eta^{pp}$
$\mathbf{w}_\eta^{cw}$
iSNR improvement left
dB
14
12
10
8
6
0
0.05
0.1
0.15
0.2
iSNR improvement right
dB
10
8
6
mixing parameter η
MSC error of noise component
Δ MSC
1
0.5
0
ILD error of speech component
dB
3.5
3
2.5
2
ITD error of speech component
μs
50
25
mixing parameter η