Binaural speech enhancement and cue preservation algorithms

Prof. Dr. Simon Doclo

University of Oldenburg, Dept. of Medical Physics and Acoustics and
Cluster of Excellence Hearing4all

ELOBES Workshop – January 12, 2019

Introduction

- Hearing impaired suffer from a loss of speech understanding in adverse acoustic environments with competing speakers, background noise and reverberation

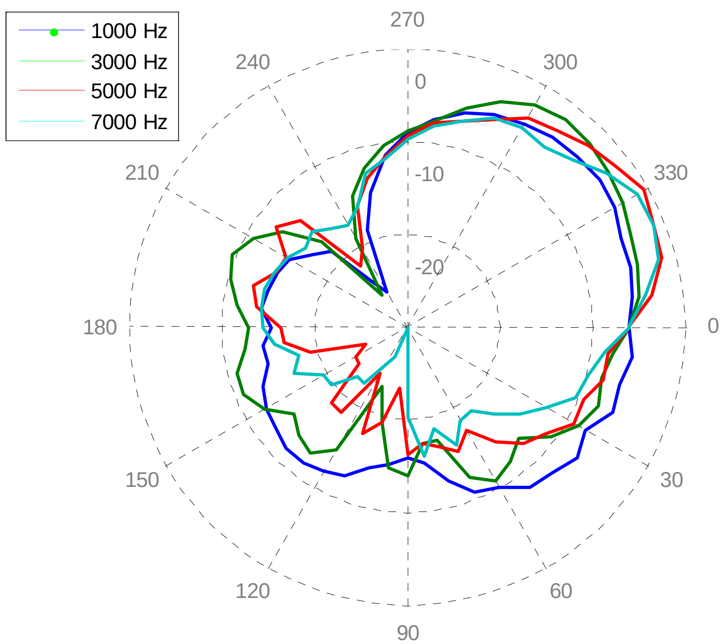
Apply **acoustic signal pre-processing techniques** in order to improve speech quality and intelligibility

Introduction

- Hearing impaired suffer from a loss of speech understanding in adverse acoustic environments with competing speakers, background noise and reverberation

Multiple microphones available → spatial + spectral processing

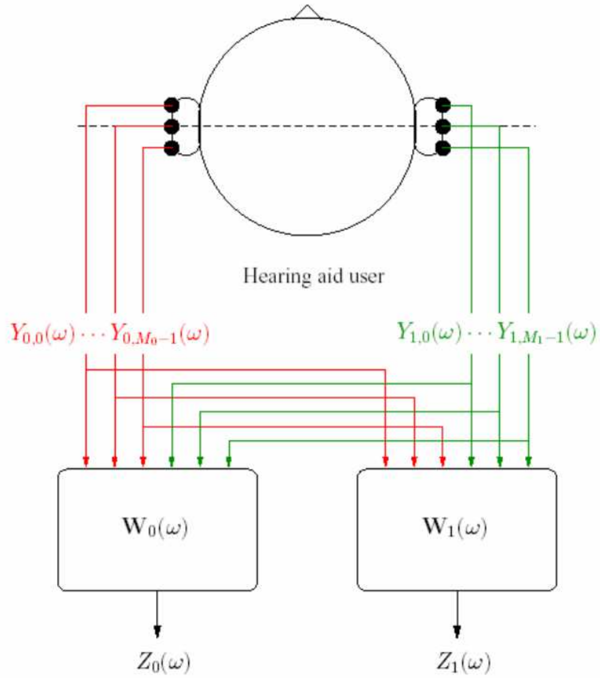
Monaural (2-3)



microphones

Introduction

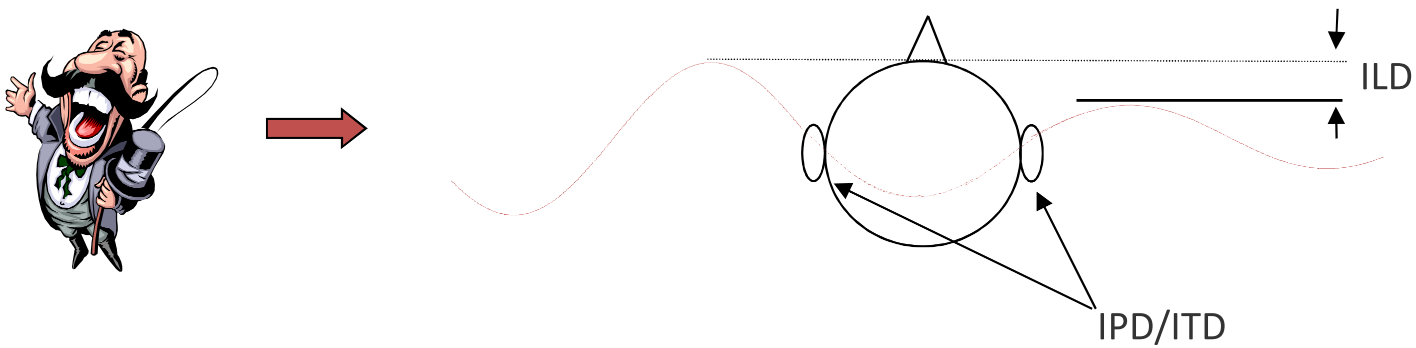
- **This presentation:**
  - **Binaural noise reduction algorithms** based on minimum variance distortionless response (MVDR) beamformer and multi-channel Wiener filter (MWF)
  - Integration with **external microphone**



- **Main objectives of algorithms:**
  - Improve speech intelligibility and avoid signal distortions
  - Preserve spatial awareness and directional hearing (binaural cues)

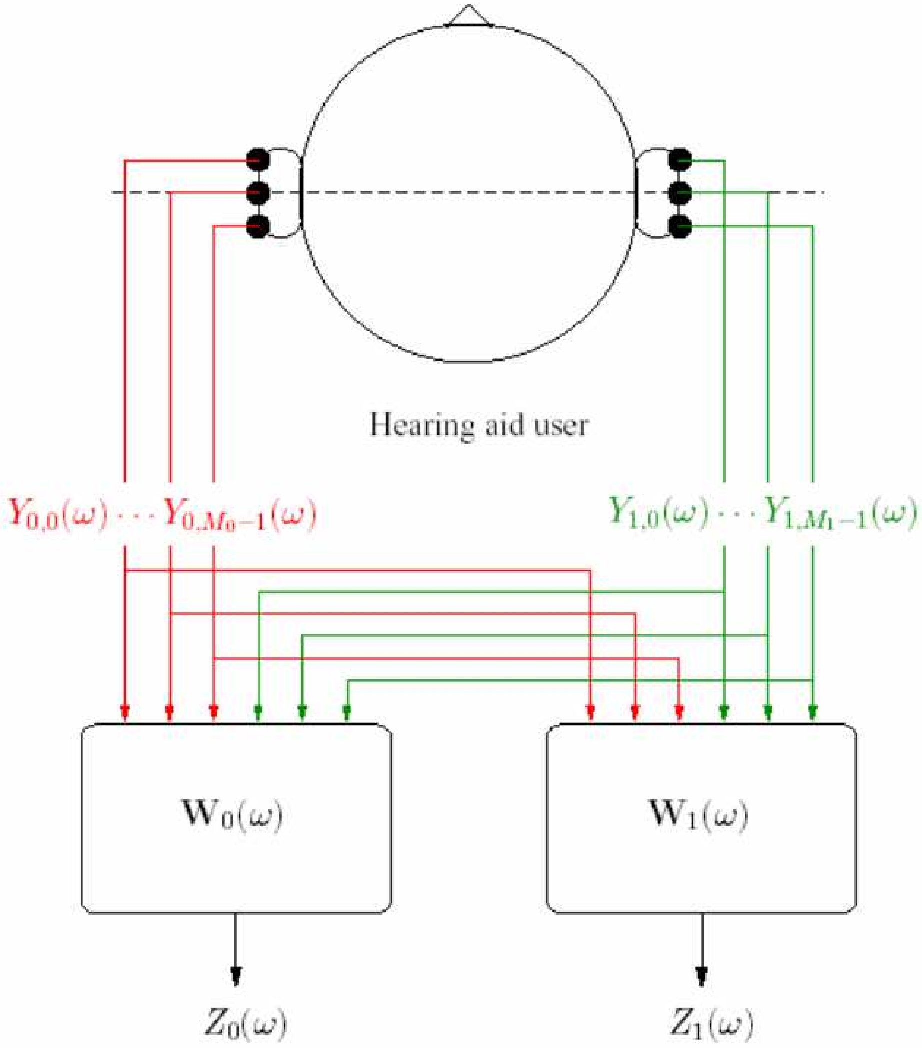
Binaural cues

- **Interaural Time/Phase Difference (ITD/IPD)**
  Interaural Level Difference (ILD)
  Interaural Coherence (IC)
  - ITD: f < 1500 Hz, ILD: f > 2000 Hz
  - IC: describes spatial characteristics, e.g. perceived width, of diffuse noise, and determines when ITD/ILD cues are *reliable*
- Binaural cues, in addition to spectro-temporal cues, play an important role in **auditory scene analysis** (source localization/segregation) and **speech intelligibility**



Binaural noise reduction

Binaural noise reduction: Configuration



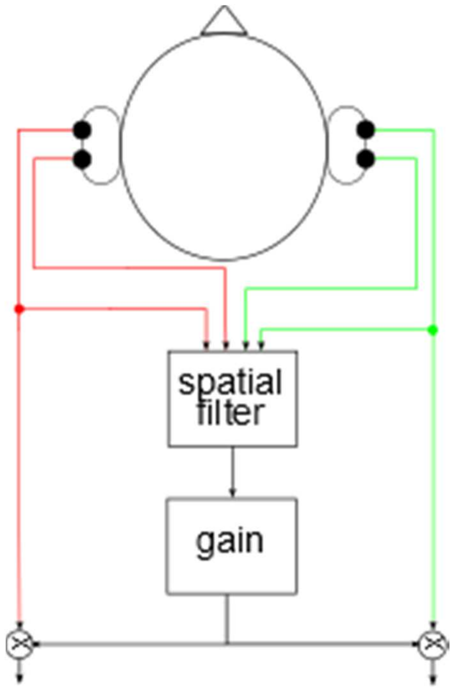
- Binaural hearing aid configuration:
  - Two hearing aids with in total *M* microphones
  - All microphone signals **Y** are assumed to be available at both hearing aids (perfect wireless link)
- Apply a filter $\mathbf{W}_0$ and $\mathbf{W}_1$ at the left and the right hearing aid, generating binaural output signals Z_0 and Z_1

$$Z_0(\omega) = \mathbf{W}_0^H(\omega)\mathbf{Y}(\omega), \quad Z_1(\omega) = \mathbf{W}_1^H(\omega)\mathbf{Y}(\omega)$$

Binaural noise reduction: Two main paradigms

Spectral post-filtering (based on multi-microphone noise reduction)

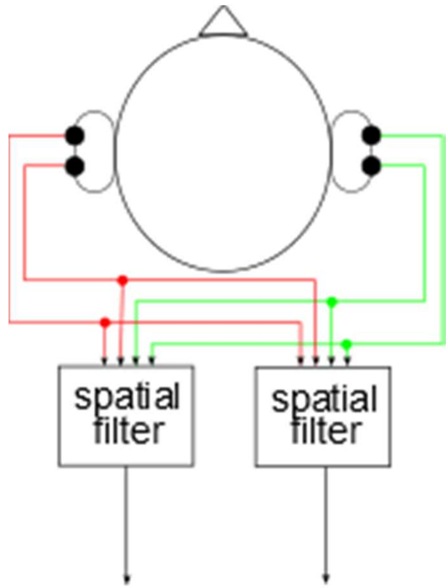
[Wittkop 2003, Lotter 2006, Rohdenburg 2008, Grimm 2009, Kamkar-Parsi 2011, Reindl 2013, Baumgärtel 2015, Enzner 2016]



- ⊕ Binaural cue preservation
- ⊖ Possible single-channel artifacts

Binaural spatial filtering techniques

[Welker 1997, Aichner 2007, Doclo 2010, Cornelis 2012, Hadad 2015-2016, Marquardt 2015-2018, Koutrouvelis 2017-2019]



- ⊕ Larger noise reduction performance
- ⊕ Merge spatial and spectral post-filtering
- ⊖ Binaural cue preservation not guaranteed

Binaural MVDR and MWF

Minimum-Variance-Distortionless-Response (MVDR) beamformer

Goal: minimize output noise power without distorting speech component in reference microphone signals

$$\min_{\mathbf{W}_0} \mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 \quad \text{subject to} \quad \mathbf{W}_0^H \mathbf{A} = A_0$$

$$\min_{\mathbf{W}_1} \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1 \quad \text{subject to} \quad \mathbf{W}_1^H \mathbf{A} = A_1$$

noise reduction — distortionless constraint

Requires estimate/model of noise coherence matrix (e.g. diffuse) and estimate/model of relative transfer function (RTF) of target speech source

Multi-channel Wiener Filter (MWF)

Goal: estimate speech component in reference microphone signals + trade off noise reduction and speech distortion

$$J_{\mathrm{MWF}}(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_0 - \mathbf{W}_0^H \mathbf{X} \\ X_1 - \mathbf{W}_1^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \mathbf{W}_0^H \mathbf{V} \\ \mathbf{W}_1^H \mathbf{V} \end{bmatrix} \right\|^2 \right\}$$

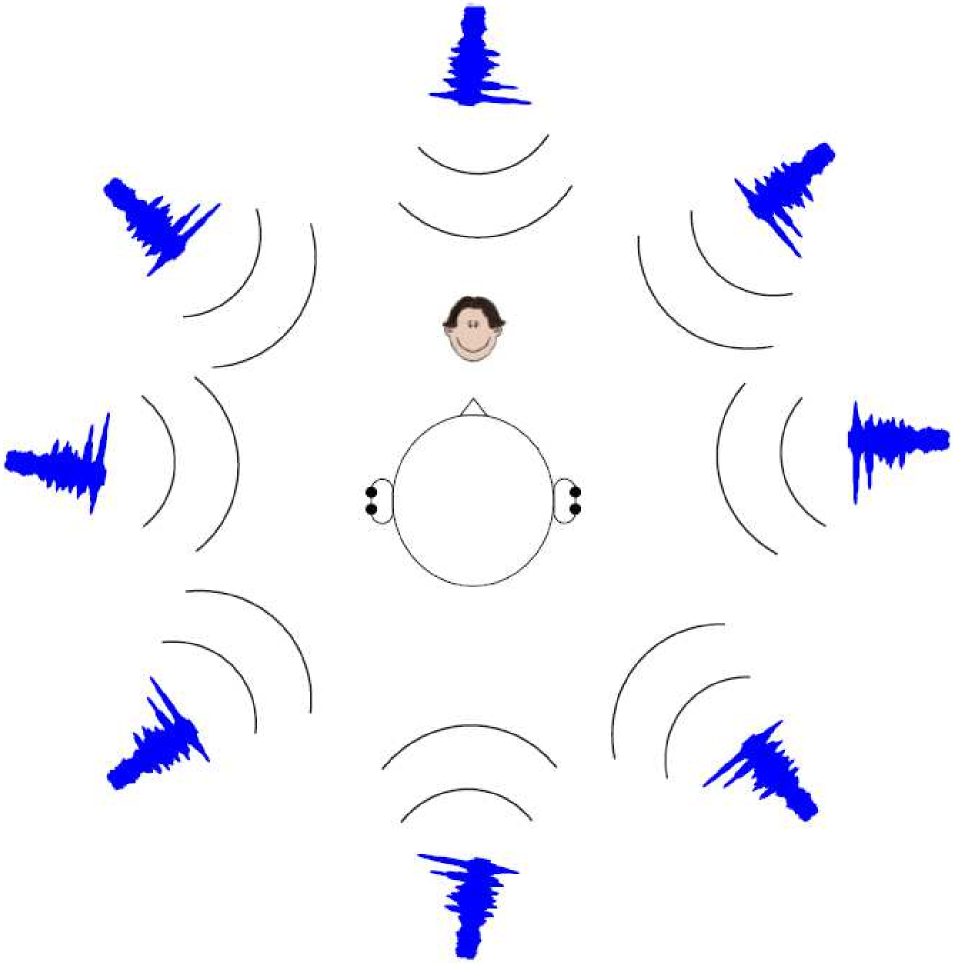
speech distortion — noise reduction

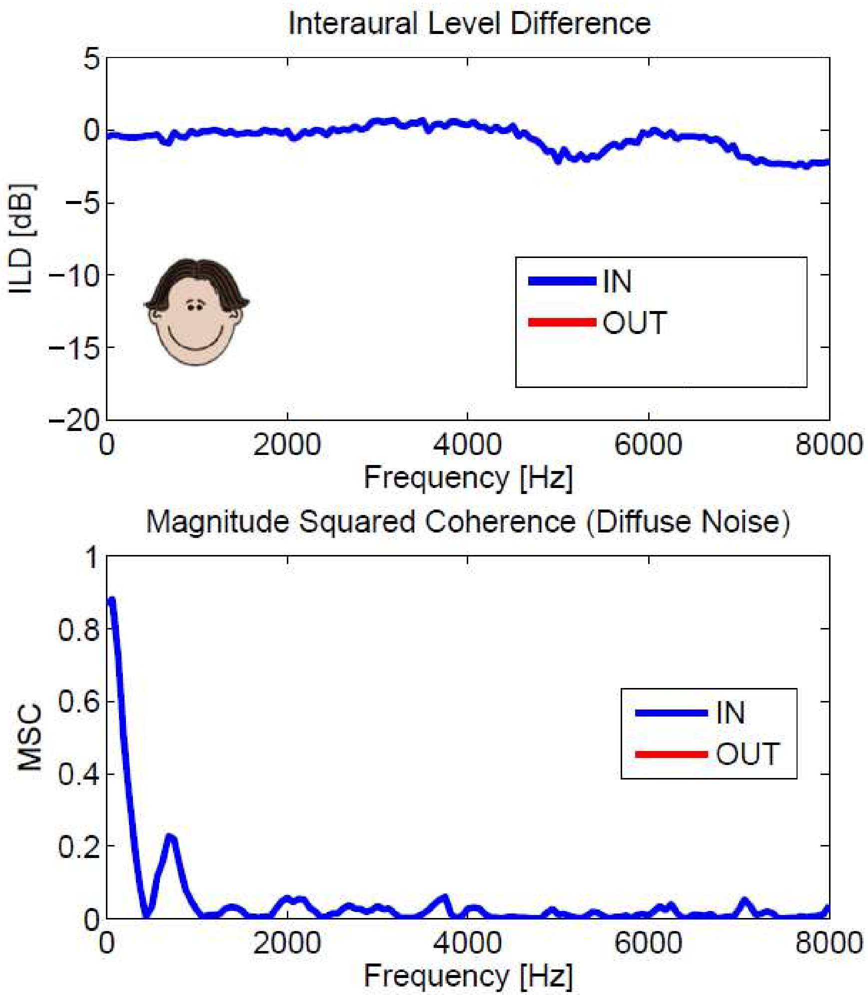
Requires estimate of speech and noise covariance matrices, e.g. based on VAD

Can be decomposed as binaural MVDR beamformer and spectral postfilter

Good noise reduction performance, what about binaural cues ?

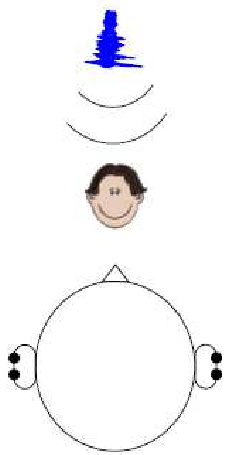
Binaural MVDR and MWF
Binaural cues (diffuse noise)



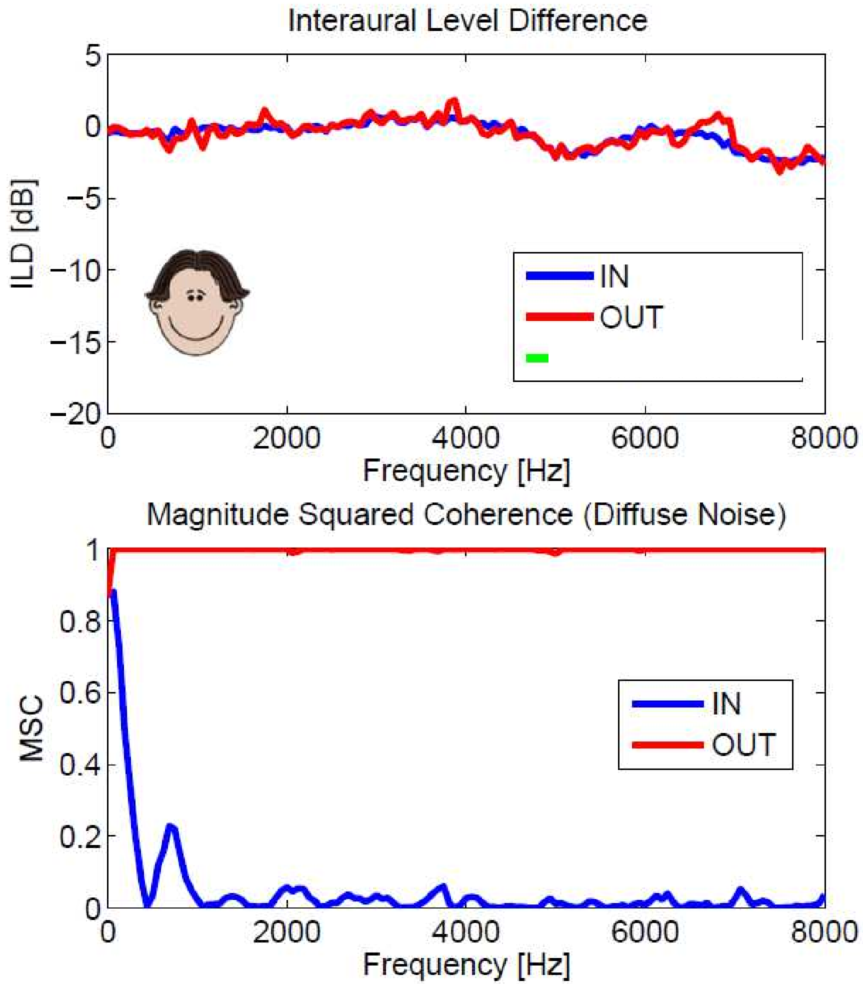


Note: MSC = Magnitude Squared Coherence

Binaural MVDR and MWF
Binaural cues (diffuse noise)



Binaural cues for residual noise and interference in binaural MVDR/MWF are not preserved



Binaural noise reduction
Extensions for diffuse noise

Binaural MWF: Extensions for diffuse noise

Binaural MWF

- ⊕ SNR improvement
- ⊕ Binaural cues of speech source
- ⊖ Binaural cues of noise

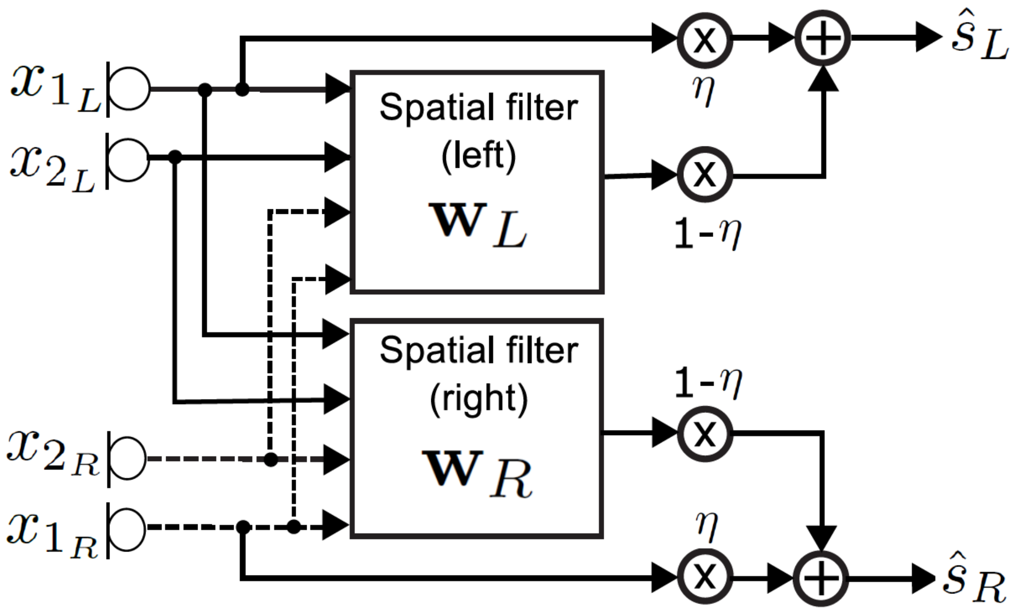
Interaural coherence preservation (MWF-IC)	Partial noise estimation (MWF-N)
$J_{MWF-IC}(\mathbf{W}) = J_{MWF}(\mathbf{W}) + \lambda \left| \frac{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_1}{\sqrt{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1}} - IC_v^{des} \right|^2$	$J_{\mathrm{MWF-N}}(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_0 - \mathbf{W}_0^H \mathbf{X} \\ X_1 - \mathbf{W}_1^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \eta V_0 - \mathbf{W}_0^H \mathbf{V} \\ \eta V_1 - \mathbf{W}_1^H \mathbf{V} \end{bmatrix} \right\|^2 \right\}$
⊖ No closed-form solution, iterative optimization procedures required	⊕ Closed-form solution

⊜ **Trade-off** between SNR improvement and binaural cue preservation, depending on **parameters** (η and λ)

[Cornelis et al., *IEEE/ACM TASLP*, 2010] [Marquardt et al., *IEEE/ACM TASLP*, 2015] [Marquardt and Doclo, *IEEE/ACM TASLP*, 2018]

Binaural MWF-N: Partial noise estimation

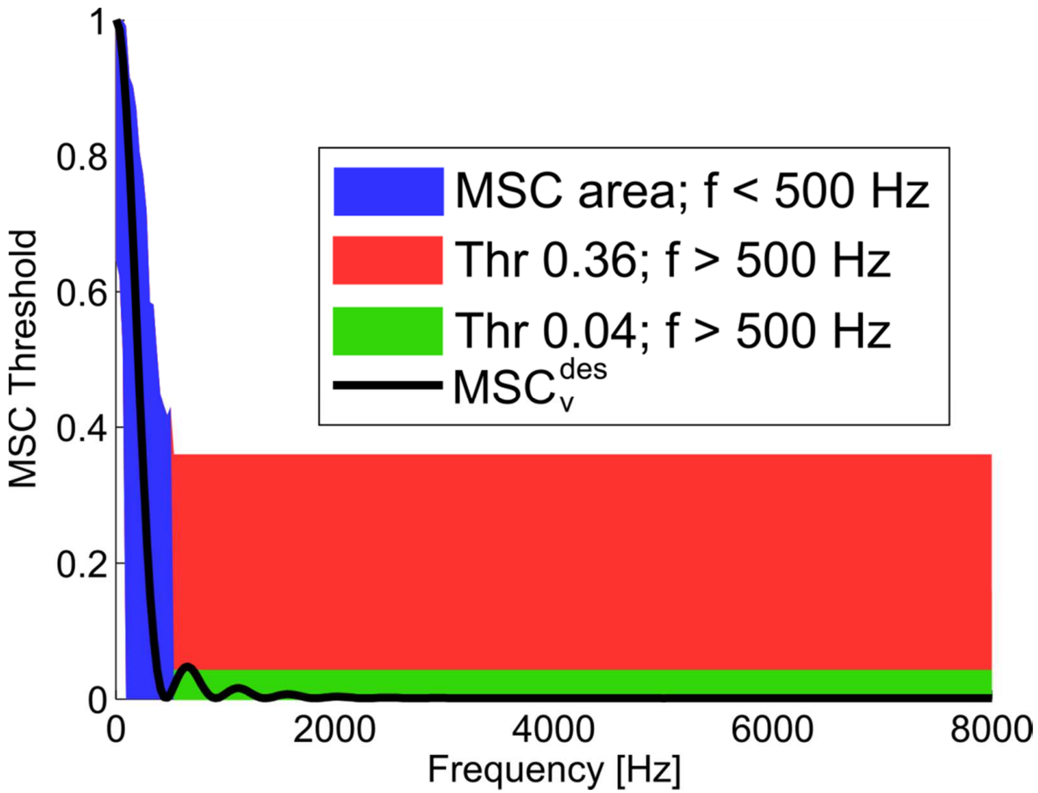
- **Closed-form filter expression** → mixing of binaural MWF output signals and reference microphone signals



- $\eta = 0$: binaural MWF (optimal noise reduction, but no cue preservation)
- $\eta = 1$: reference microphone signals (perfect cue preservation, but no noise reduction)

Binaural MWF-N: Trade-off parameter η

- **Fixed broadband values** (η = 0.1 … 0.3)
- **Frequency-dependent values** based on IC discrimination ability of human auditory system for diffuse noise



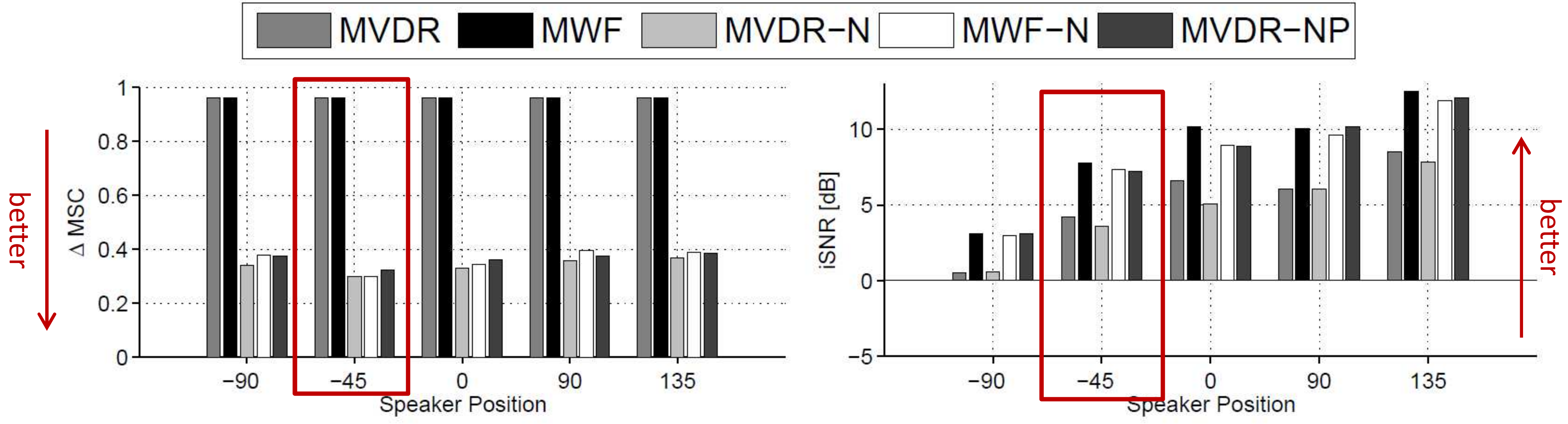
- IC discrimination ability depends on magnitude of reference IC
- **Boundaries on Magnitude Squared Coherence** ($MSC=|IC|^2$) :
  - For f < 500 Hz ("large" IC): frequency-dependent MSC boundaries (**blue**)
  - For f > 500 Hz ("small" IC): fixed MSC boundary, e.g. 0.36 (**red**) or 0.04 (**green**)

Binaural MWF-N: Trade-off parameter η

- **Fixed broadband values** (η = 0.1 … 0.3)
- **Frequency-dependent values** based on IC discrimination ability of human auditory system for diffuse noise
  - **Trade-off parameter** η achieving desired MSC:
    - **MWF-N** : exhaustive search for optimal trade-off parameter
    - **MVDR-N** (i.e. special case of MWF-N) :
      - Closed-form expression for optimal trade-off parameter
      - No spectral filtering as in MWF-N
    - **MVDR-N + spectral postfilter**
      - Not equivalent to MWF-N, but combining spatial and spectral filtering with closed-form expression for both filter and trade-off parameter

[Cornelis et al., *IEEE/ACM TASLP*, 2010] [Marquardt and Doclo, *IEEE/ACM TASLP*, 2018]

Binaural MWF-N: Instrumental evaluation



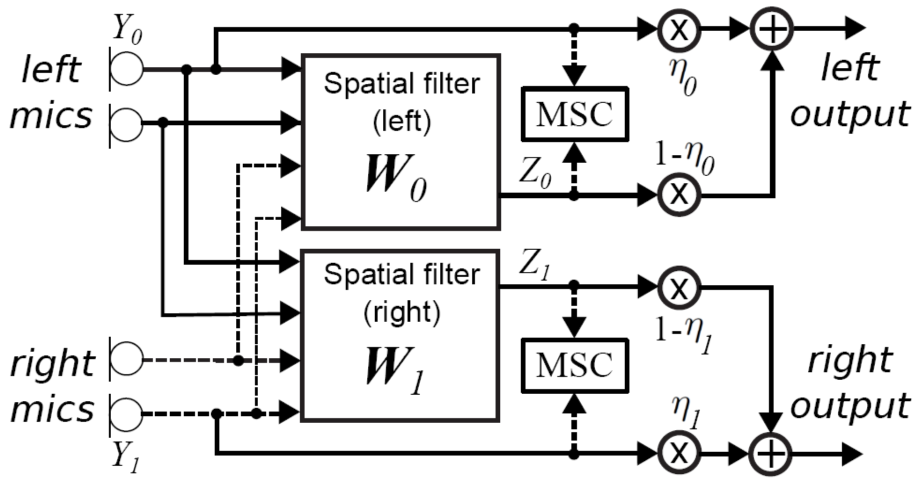
(a) $\Delta\mathrm{MSC_v}$ (b) $\mathrm{iSNR}_{\mathrm{be}}^{\mathrm{out}}$

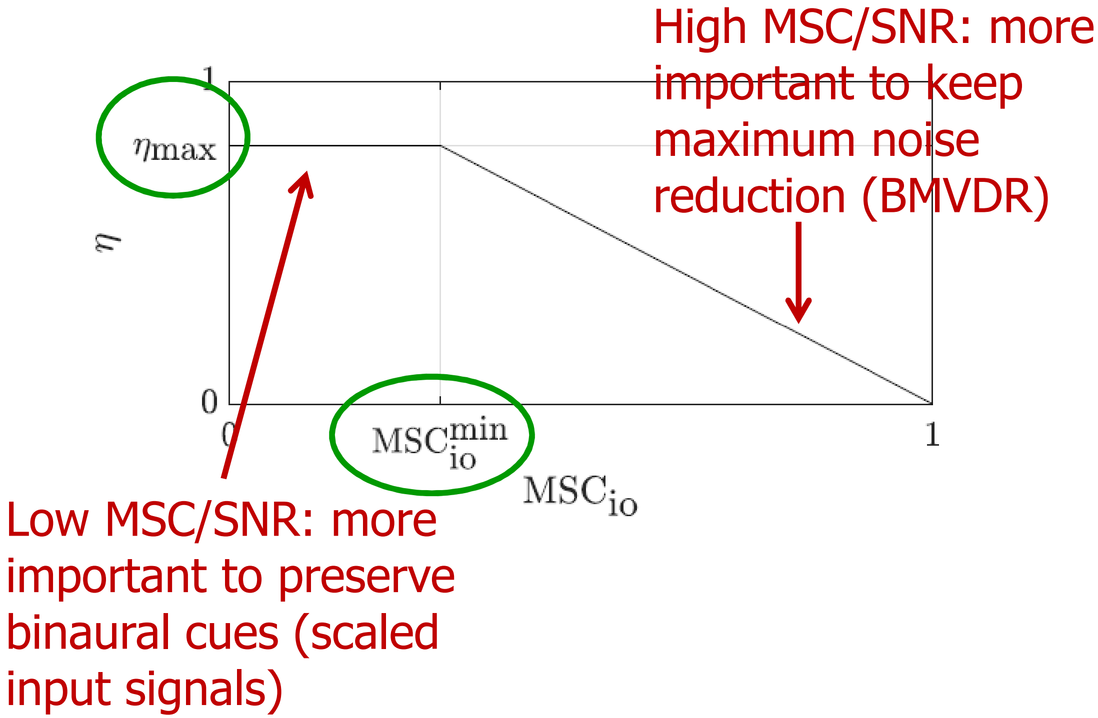
Input	MVDR	MWF	MVDR-N	MWF-N	MVDR-NP

Office ($T_{60} \approx$ 700ms), M=4 (binaural RIRs), recorded ambient noise, target speaker at -45°, 0 dB input iSNR (left hearing aid)
MVDR: anechoic RTF of target speaker (DOA known), diffuse spatial coherence matrix (from anechoic ATFs) / MWF = MVDR + postfilter (SPP-based)

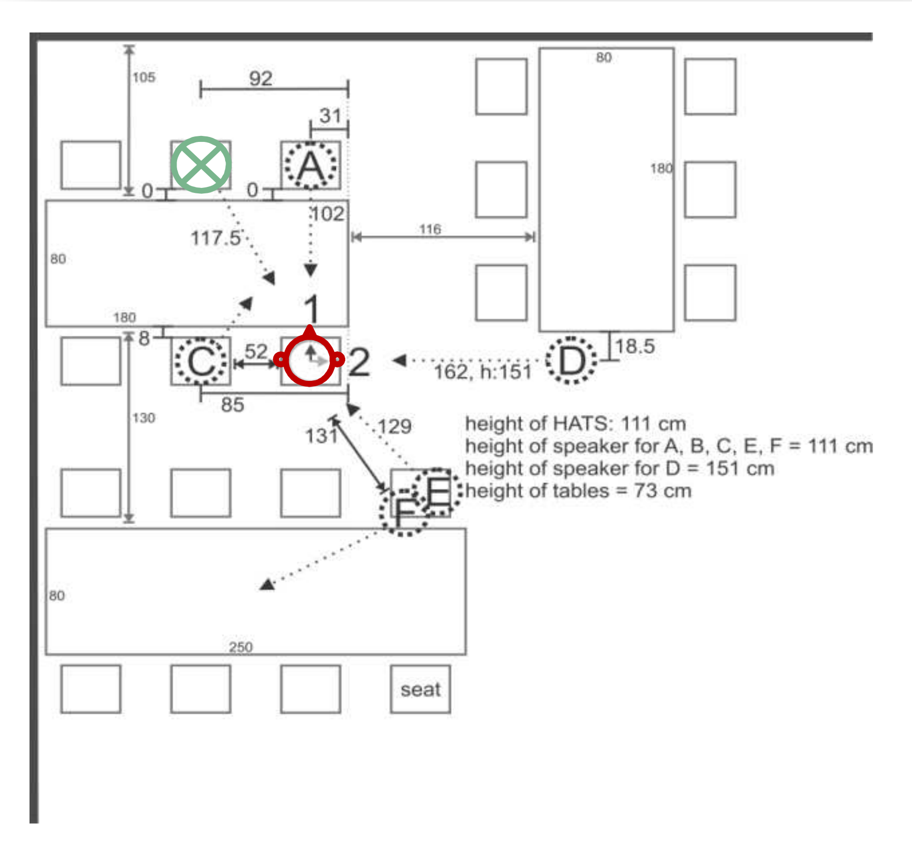
Binaural MWF-N: Trade-off parameter η

- **Fixed broadband values** (η = 0.1 … 0.3)
- **Frequency-dependent values** based on IC discrimination ability of human auditory system for diffuse noise
- **Frequency-dependent function** of MSC between noisy reference microphone signals and output signals of BMVDR beamformer





Subjective Evaluation: Test setup



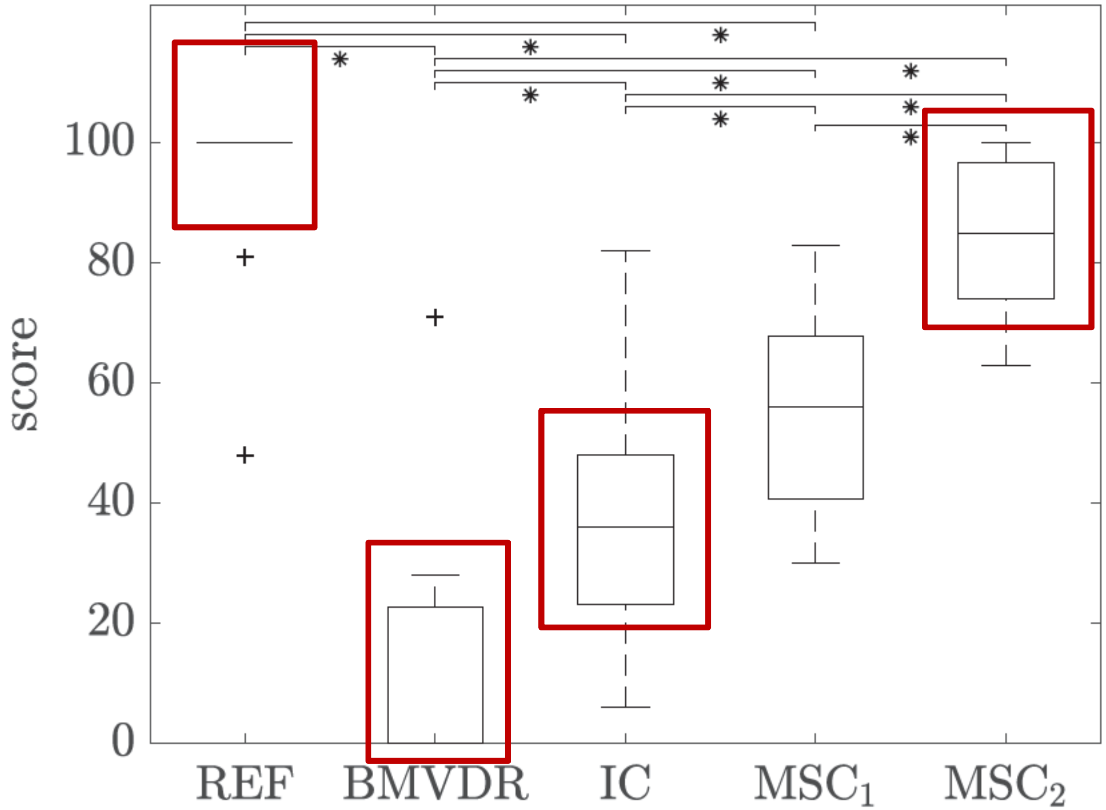
- Binaural hearing aid recordings (M=4 mics) in **cafeteria** (T_{60}≈1250 ms)
  - Target speaker at -35º
  - Realistic cafeteria ambient noise
- **Algorithms**: binaural MVDR and binaural MVDR-N with different trade-off parameters:
  - MVDR-IC
  - MVDR-MSC1: η_{max}=0.7, MSC_{min}=0
  - MVDR-MSC2: η_{max}=1.0, MSC_{min}=0.1
- **Subjective listening experiments**:
  - 11 normal-hearing subjects
  - **SRT** using Oldenburg Sentence Test (OLSA)
  - **Spatial quality (diffuseness)** using MUSHRA

Does binaural unmasking compensate for SNR decrease of cue preservation algorithms (MVDR-N) ?

[Klug, Marquardt, Doclo, *Proc. ITG Speech Communication*, 2018]

Subjective Evaluation: Spatial quality (MUSHRA)

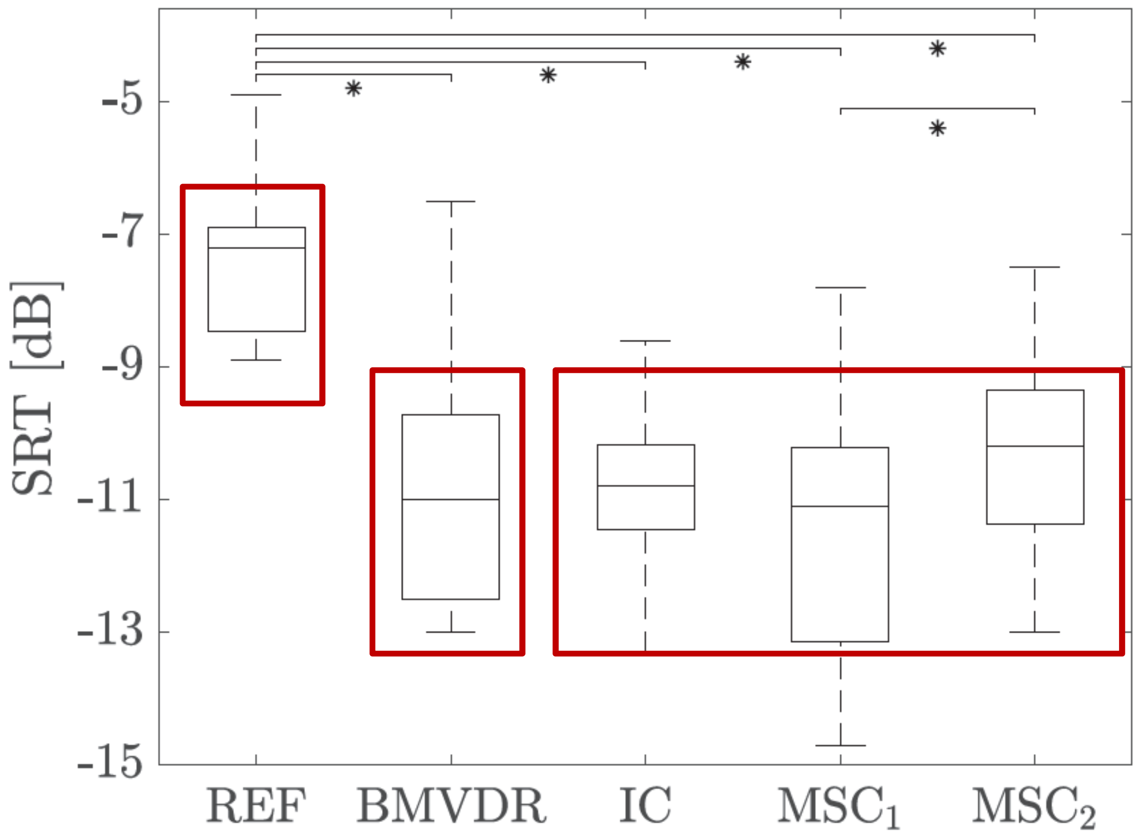
- Evaluate spatial difference between reference and output signal
- **MVDR-N (IC/MSC) outperforms BMVDR**
  - Trade-off parameters: MSC-based better than IC-based
  - Using MSC2 hardly any difference to input !



Binaural cue preservation for diffuse noise significantly improves spatial quality

Subjective Evaluation: Speech intelligibility (SRT)

- All algorithms show a highly significant speech reception threshold (SRT) improvement
- **No significant SRT difference between BMVDR and MVDR-N (IC/MSC)**



Binaural cue preservation for diffuse noise does not affect speech intelligibility

Binaural noise reduction

Extensions for interfering sources

Binaural MVDR: Extensions for interfering source

- ⊕ SNR improvement
- ⊕ Binaural cues of speech source
- ⊖ Binaural cues of interferer

Binaural MVDR

Relative transfer function (BMVDR-RTF)

$$\min_{\mathbf{W}_0,\mathbf{W}_1} \left\{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1\right\}$$

$$\text{s.t. } \mathbf{W}_0^H \mathbf{A} = A_0, \mathbf{W}_1^H \mathbf{A} = A_1, \frac{\mathbf{W}_0^H \mathbf{B}}{\mathbf{W}_1^H \mathbf{B}} = \frac{B_0}{B_1}.$$

Interference rejection (BMVDR-IR)

$$\min_{\mathbf{W}_0} \left\{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0\right\} \text{ s.t. } \mathbf{W}_0^H \mathbf{A} = A_0, \mathbf{W}_0^H \mathbf{B} = \eta B_0$$

$$\min_{\mathbf{W}_1} \left\{\mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1\right\} \text{ s.t. } \mathbf{W}_1^H \mathbf{A} = A_1, \mathbf{W}_1^H \mathbf{B} = \eta B_1$$

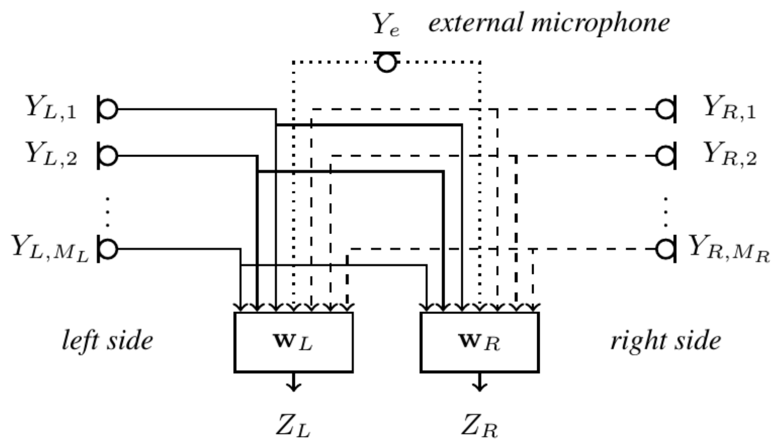
- ⊕ Binaural cues of speech source **and** interfering source preserved
- ⊕ Also binaural MWF-based versions (incl. spectral filtering) can be derived
- ⊖ MSC of background not exactly preserved, RTF estimation difficult

Binaural noise reduction

Integration with external microphone(s)

Binaural noise reduction: External microphone(s)

- Exploit the availability of one or more external microphones **(acoustic sensor network)** with hearing aids
  [Bertrand 2009, Szurley 2016, Yee 2017, Gößling 2018, Ali 2018]
- Integrating external microphone(s) with hearing aid microphones may lead to:
  - Improved **noise reduction** and **binaural cue preservation** performance
  - Low-complexity method to **estimate relative transfer function (RTF)** of target speaker

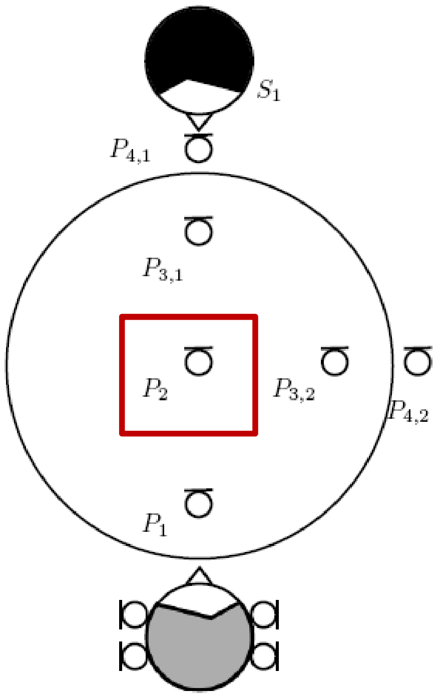


$$\mathbf{w}_\mathrm{L} = \frac{\mathbf{R}_\mathrm{n}^{-1}\mathbf{a}_\mathrm{L}}{\mathbf{a}_\mathrm{L}^H\mathbf{R}_\mathrm{n}^{-1}\mathbf{a}_\mathrm{L}}, \quad \mathbf{w}_\mathrm{R} = \frac{\mathbf{R}_\mathrm{n}^{-1}\mathbf{a}_\mathrm{R}}{\mathbf{a}_\mathrm{R}^H\mathbf{R}_\mathrm{n}^{-1}\mathbf{a}_\mathrm{R}}$$

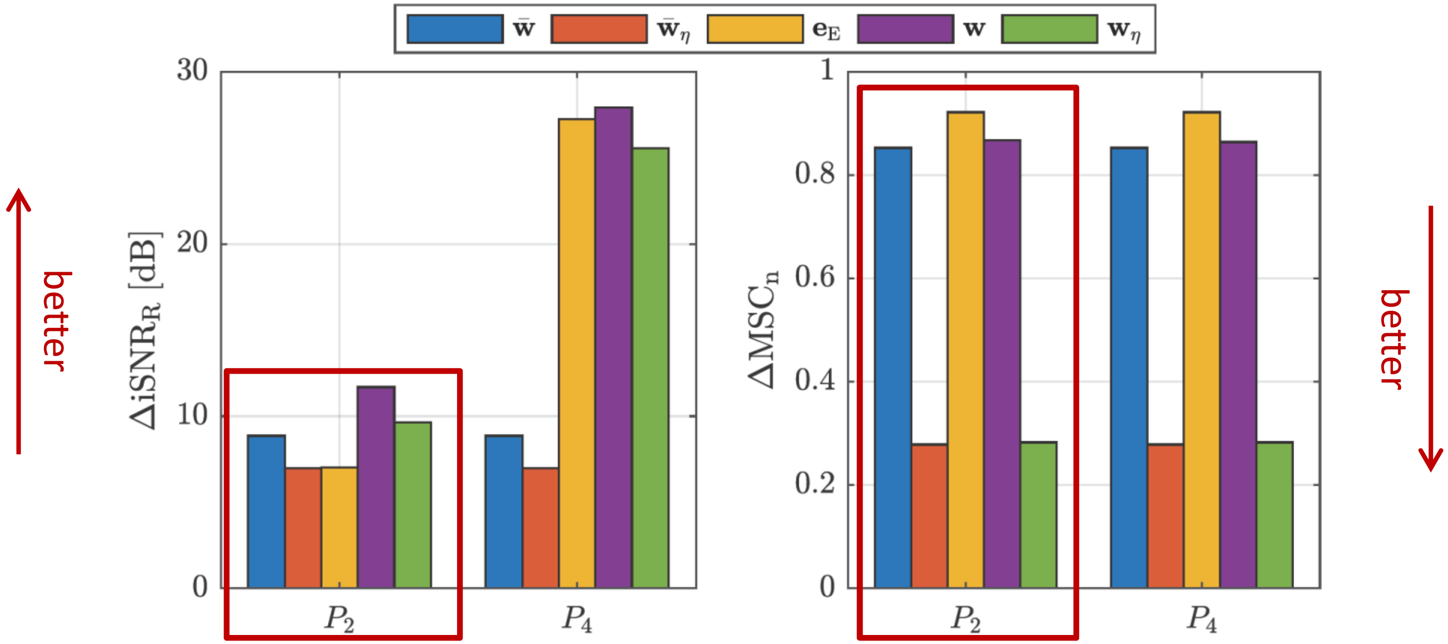
Binaural noise reduction: External microphone(s)

Hearing 4all — Carl von Ossietzky Universität Oldenburg

- Including external microphone in **binaural MVDR-N beamformer** leads to:
  - **Larger output SNR** for same trade-off parameter η
  - Same output SNR with larger trade-off parameter η → **better cue preservation**



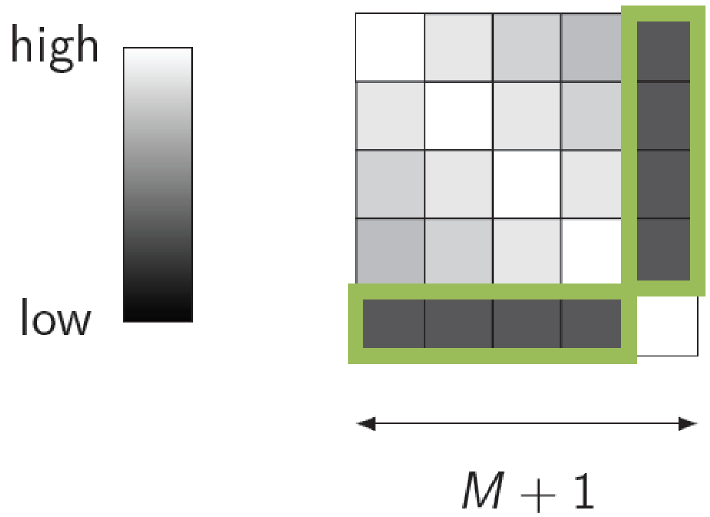
MVDR (HA) **MVDR-N (HA)** **External** **eMVDR (HA+ext)** **eMVDR-N (HA+ext)**



Starkey database with real-world recordings ($T_{60} \approx 620$ms), M=4, target speaker S_1, multi-talker babble noise, 0 dB input iSNR (right hearing aid)
MVDR: perfectly estimated noise correlation matrix, RTF of target speaker estimated using covariance whitening method

Binaural noise reduction: External microphone(s)

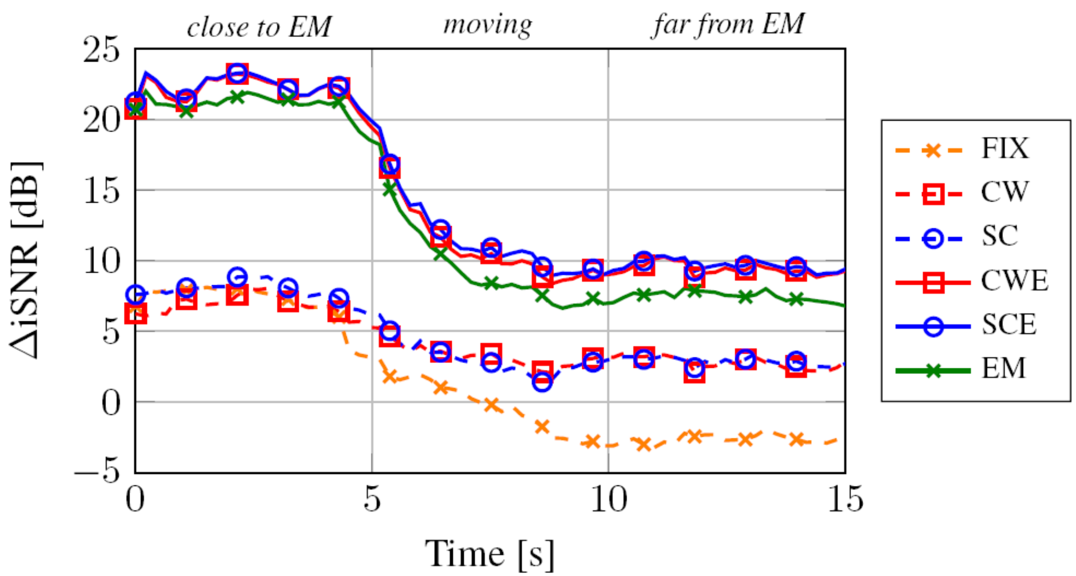
- **Estimate RTF of target speaker** to steer binaural MVDR beamformer
- **Spatial coherence (SC) method:** assume that *external microphone* is spatially separated from *HA microphones,* such that noise components in external microphone and HA microphones can be assumed to be uncorrelated

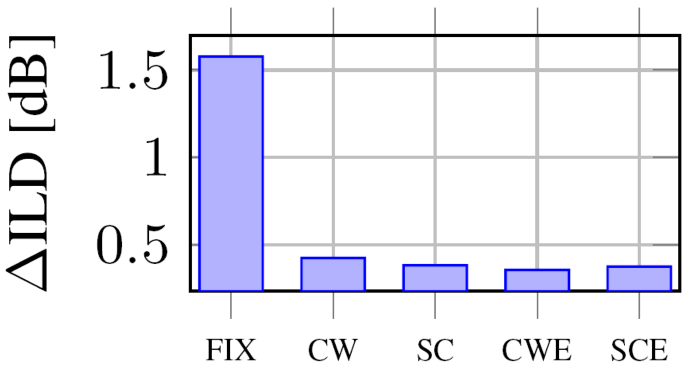


Low spatial coherence between noise components in local array signals and external microphone signal

$$\bar{\mathbf{a}}_{\mathrm{L}}^{\mathrm{SCE}} = \frac{\bar{\mathbf{R}}_{\mathrm{y}}\mathbf{e}_{\mathrm{E}}}{\mathbf{e}_{\mathrm{L}}^{T}\bar{\mathbf{R}}_{\mathrm{y}}\mathbf{e}_{\mathrm{E}}}, \quad \bar{\mathbf{a}}_{\mathrm{R}}^{\mathrm{SCE}} = \frac{\bar{\mathbf{R}}_{\mathrm{y}}\mathbf{e}_{\mathrm{E}}}{\mathbf{e}_{\mathrm{R}}^{T}\bar{\mathbf{R}}_{\mathrm{y}}\mathbf{e}_{\mathrm{E}}} \qquad \bar{\mathbf{w}}_{\mathrm{L}}^{\mathrm{SCE}} = \begin{bmatrix} \alpha \cdot [\mathbf{I}_{2M}, \mathbf{0}_{2M\times 1}]\,\bar{\mathbf{w}}_{\mathrm{L}} \\ \alpha(1+\beta)\cdot \mathbf{e}_{\mathrm{E}}^{T}\bar{\mathbf{w}}_{\mathrm{L}} \end{bmatrix}$$

Binaural noise reduction: External microphone(s)





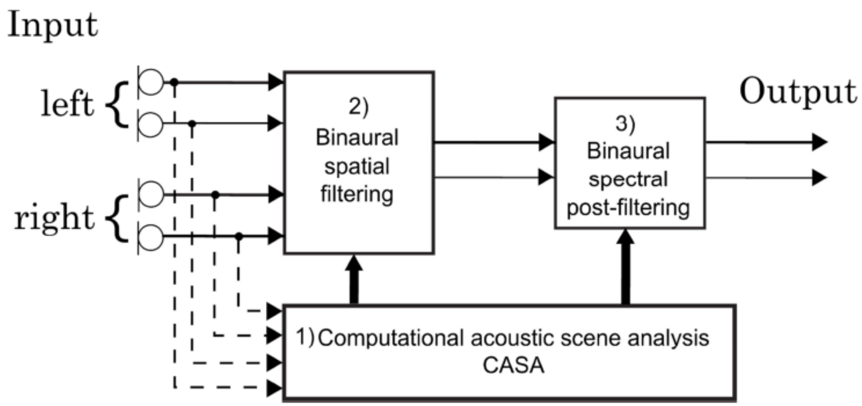
- MVDR with external microphone (SCE) leads to **better SNR** compared to MVDR using only HA microphones (SC,FIX) and external microphone (EM)
- MVDR using estimated RTFs (SCE, SC) **preserves binaural cues of target speaker** compared to fixed MVDR (FIX) and external microphone (EM)

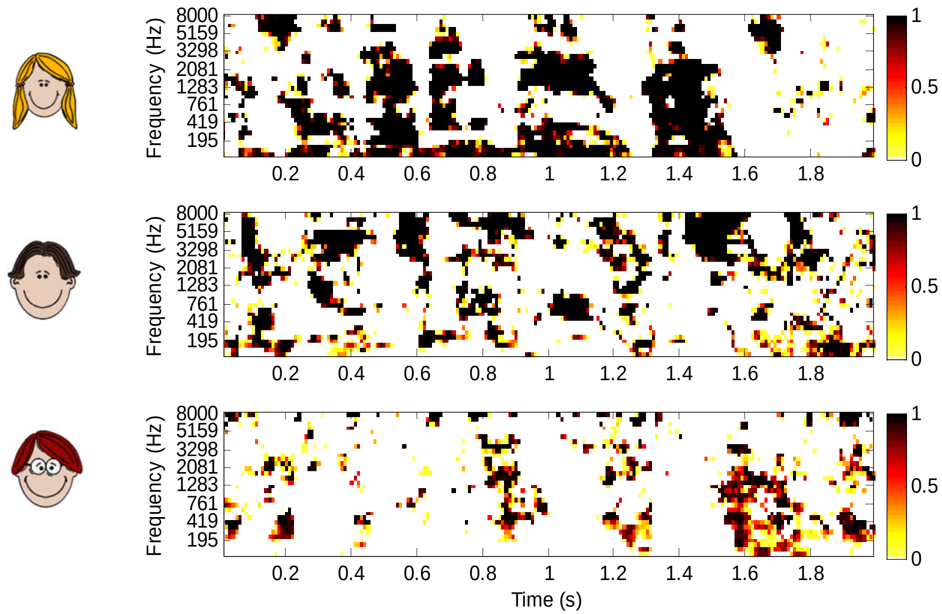
Oldenburg Varechoic Lab ($T_{60} \approx 350$ms), M=4 + 1 external mic (1.5m/0.5m), moving speaker, pseudo-diffuse babble noise, iSNR=0dB (right HA)
STFT: 32 ms, 50% overlap, sqrt-Hann; SPP in external microphone; smoothing: 100 ms (speech), 1 s (noise)

[Gößling, Doclo, *Proc. IWAENC*, 2018] [Gößling, Doclo, *submitted to ICASSP 2019*]

Current/future work

- Subjective evaluation of binaural speech enhancement algorithms with **HA/CI users** ongoing
- **Complex and time-varying scenarios**: incorporate computational acoustic scene analysis (CASA) into control path of developed algorithms
- Integration of **multiple external microphones** (acoustic sensor network)





Summary

- **Binaural noise reduction algorithms**: 2 main paradigms
  - Spectral post-filtering
  - "True" binaural spatial filtering
- **Extensions of binaural MVDR/MWF** for diffuse noise and interfering speaker, preserving binaural cues of residual noise/interference
- Evaluation of **binaural MVDR extensions (MVDR-N) for diffuse noise:**
  - Binaural cue preservation improves **spatial quality**
  - Binaural cue preservation does not/hardly affect **speech intelligibility**
- Extensions with **external microphone** possible

Acknowledgments

Dr. Daniel Marquardt

Nico Gößling

Jonas Klug

Dr. Elior Hadad

Prof. Sharon Gannot

- **Funding:**
  - Cluster of Excellence Hearing4all (DFG)
  - Research Unit Individualized Hearing Acoustics (DFG)
  - Joint Lower-Saxony Israel Project "Acoustic scene aware speech enhancement for binaural hearing aids" (Partner: Bar-Ilan University, Israel)

DFG Deutsche Forschungsgemeinschaft

VolkswagenStiftung

Recent publications

- **S. Doclo, S. Gannot, D. Marquardt, E. Hadad, "Binaural Speech Processing with Application to Hearing Devices", Chapter 18 in *Audio Source Separation and Speech Enhancement* (E. Vincent, T. Virtanen, S. Gannot, eds.), Wiley, 2018.**
- **S. Doclo, W. Kellermann, S. Makino, S. Nordholm, *Multichannel signal enhancement algorithms for assisted listening devices*, *IEEE Signal Processing Magazine,* vol. 32, no. 2, pp. 18-30, Mar. 2015.**
- D. Marquardt, V. Hohmann, S. Doclo, *Interaural Coherence Preservation in Multi-channel Wiener Filtering Based Noise Reduction for Binaural Hearing Aids*, *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 23, no. 12, pp. 2162-2176, Dec. 2015.
- D. Marquardt, E. Hadad, S. Gannot, S. Doclo, *Theoretical Analysis of Linearly Constrained Multi-channel Wiener Filtering Algorithms for Combined Noise Reduction and Binaural Cue Preservation in Binaural Hearing Aids*, *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 23, no. 12, pp. 2384-2397, Dec. 2015.
- E. Hadad, D. Marquardt, S. Doclo, S. Gannot, *Theoretical Analysis of Binaural Transfer Function MVDR Beamformers with Interference Cue Preservation Constraints*, *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 23, no. 12, pp. 2449-2464, Dec. 2015.
- E. Hadad, S. Doclo, S. Gannot, *The Binaural LCMV Beamformer and its Performance Analysis*, *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 24, no. 3, pp. 543-558, Mar. 2016.
- J. Thiemann, M. Müller, D. Marquardt, S. Doclo, S. van de Par, *Speech Enhancement for Multimicrophone Binaural Hearing Aids Aiming to Preserve the Spatial Auditory Scene*, *EURASIP Journal on Advances in Signal Processing*, 2016:12, pp. 1-11.
- N. Gößling, D. Marquardt, S. Doclo, *Performance analysis of the extended binaural MVDR beamformer with partial noise estimation in a homogeneous noise field*, in *Proc. Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, San Francisco, USA, Mar. 2017, pp. 1-5.
- D. Marquardt, S. Doclo, *Interaural Coherence Preservation in Binaural Hearing Aids using Partial Noise Estimation and Spectral Postfiltering*, *IEEE/ACM Trans. Audio, Speech and Language Processing,* vol. 26, no. 7, pp. 1257-1270, Jul. 2018.
- N. Gößling, D. Marquardt, I. Merks, T. Zhang, S. Doclo, *Optimal binaural LCMV beamforming in complex acoustic scenarios: theoretical and practical insights*, in *Proc. International Workshop on Acoustic Signal Enhancement*, Tokyo, Japan, Sep. 2018, pp. 381-385.
- N. Gößling, S. Doclo, *Relative transfer function estimation exploiting spatially separated microphones in a diffuse noise field*, in *Proc. International Workshop on Acoustic Signal Enhancement*, Tokyo, Japan, Sep. 2018, pp. 146-150.
- J. Klug, D. Marquardt, N. Gößling, S. Doclo, *Evaluation of Signal-Dependent Partial Noise Estimation Algorithms for Binaural Hearing Aids*, in *Proc. ITG Conference on Speech Communication*, Oldenburg, Germany, Oct. 2018, pp. 236-240.

http://www.sigproc.uni-oldenburg.de → Publications