Design and evaluation of binaural speech enhancement and cue preservation algorithms

Simon Doclo, Daniel Marquardt

University of Oldenburg, Dept. of Medical Physics and Acoustics and Cluster of Excellence Hearing4All

STADIUS Seminar– KU Leuven – March 8, 2016

Introduction

- Hearing impaired suffer from a loss of speech understanding in adverse acoustic environments (“cocktail-party effect”)

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

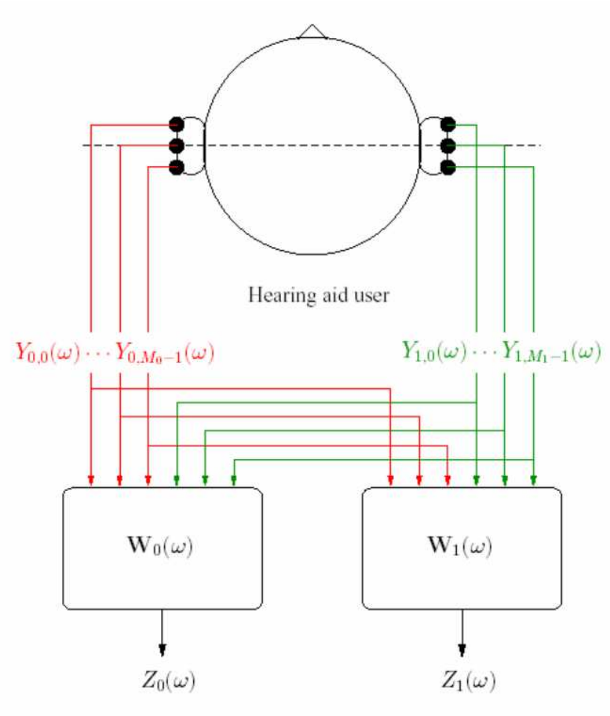
Introduction

- Digital hearing aids allow for advanced acoustical signal pre-processing
  - Multiple microphones available → spatial + spectral processing
  - Speech enhancement (noise reduction, beamforming), computational auditory scene analysis (source localisation, environment classification)

Monaural (2-3) *Binaural* *External microphones*

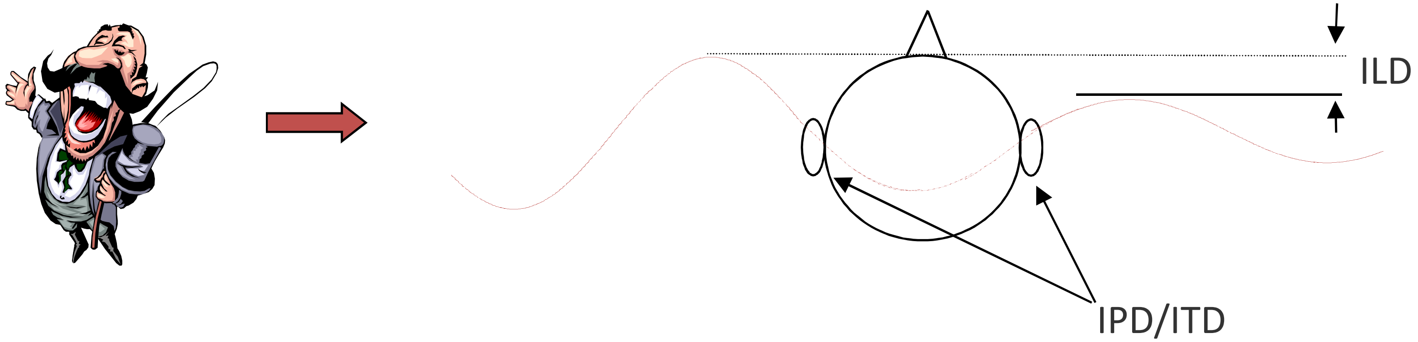
Introduction

- **This Presentation:**
  - Instrumental and subjective evaluation of recent binaural noise reduction algorithms based on MVDR/MWF
  - Two acoustic scenarios: diffuse noise and interfering speaker
- **Main Objectives:**
  - 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 segregation) and speech intelligibility



Binaural cues

❑ Spatial release from masking (BMLD):

- ❑ *Localized noise source* : large effect for NH listeners (especially in free-field)
- ❑ *Diffuse noise* : about 2-3 dB

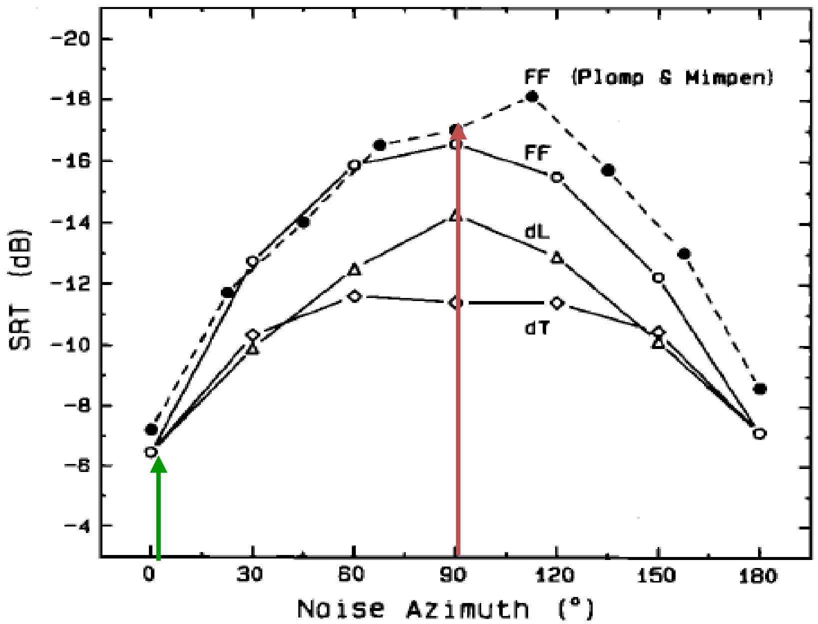
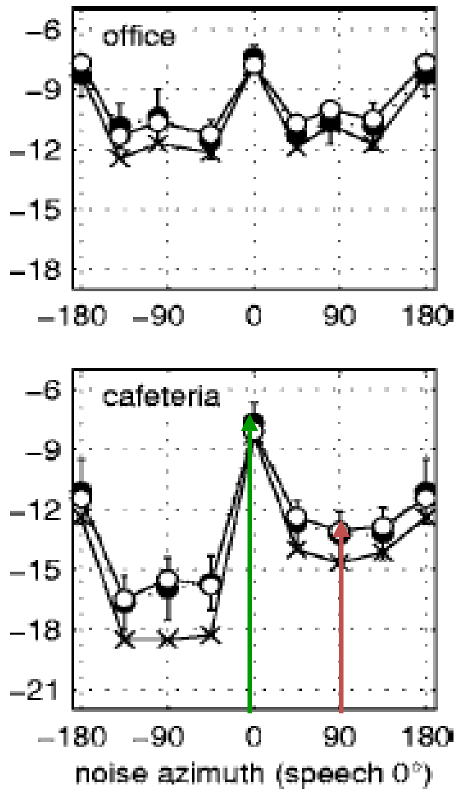


FIG. 5. Mean speech reception thresholds obtained in experiment I for three different noise types: FF (free field), dL (headshadow only), and dT (ITD only). The closed data points represent results of Plomp and Mimpen (1981) obtained in a free field.



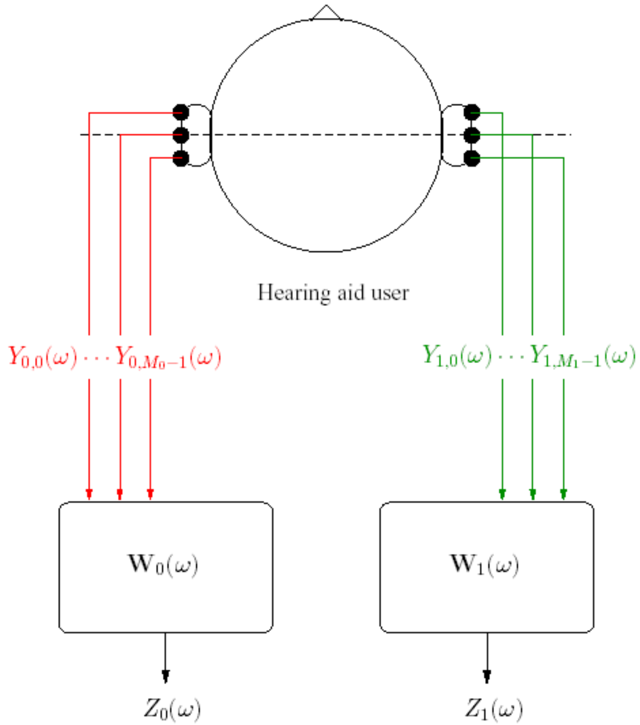
[Bronkhorst and Plomp, 1988]

[Beutelmann and Brand, 2006]

Binaural noise reduction algorithms

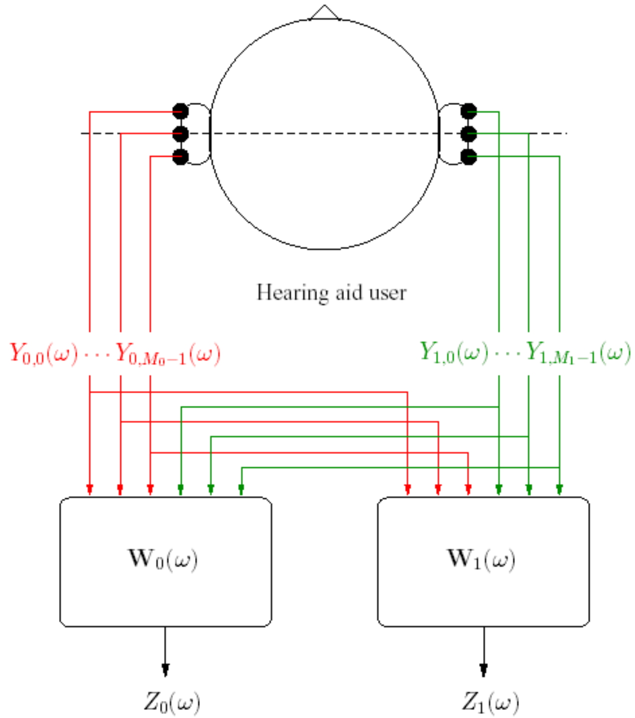
Binaural noise reduction: Configuration

Monaural/Bilateral system



⊖ **Independent** left/right processing:
- No cooperation (e.g. different environment classification)
- preservation of binaural cues ?

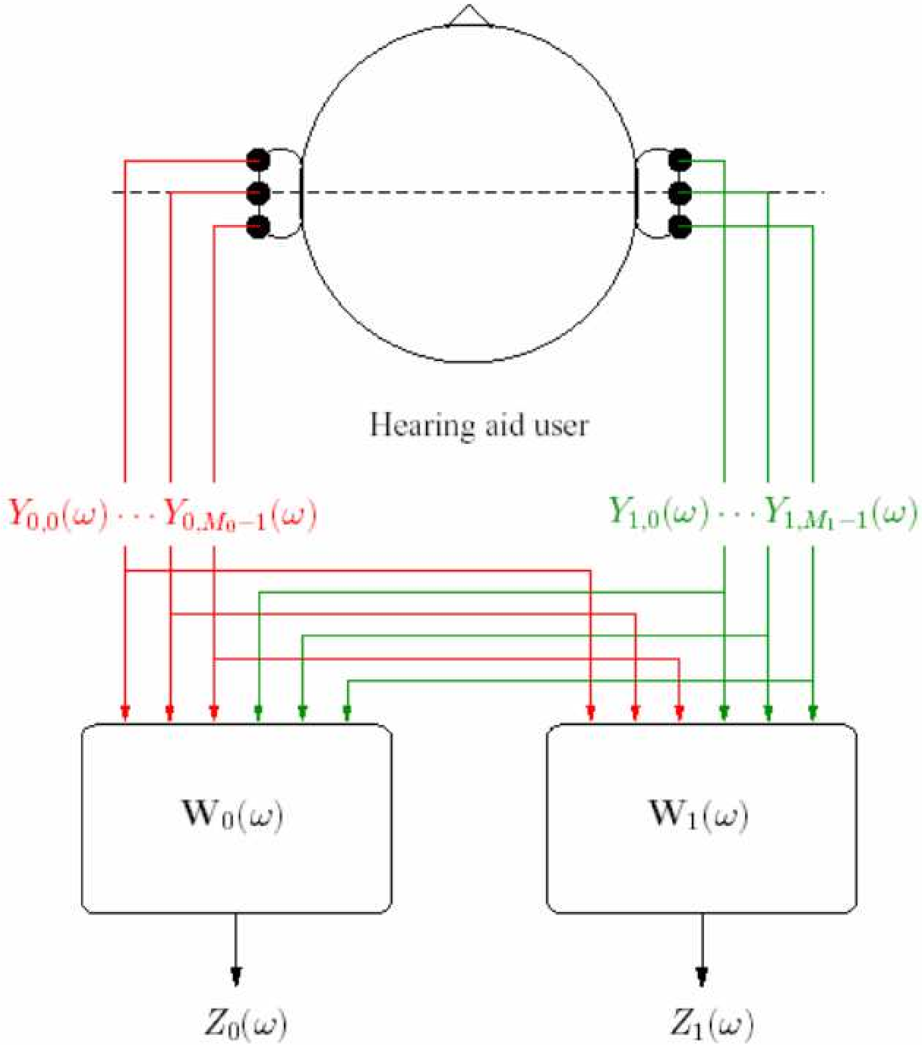
Binaural system



⊕ Exchange of:
- **parameters** (volume, environment)
- **signals** (cooperative processing for noise reduction, feedback, ...)

⊖ Need for wireless binaural link

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: Acoustic scenario

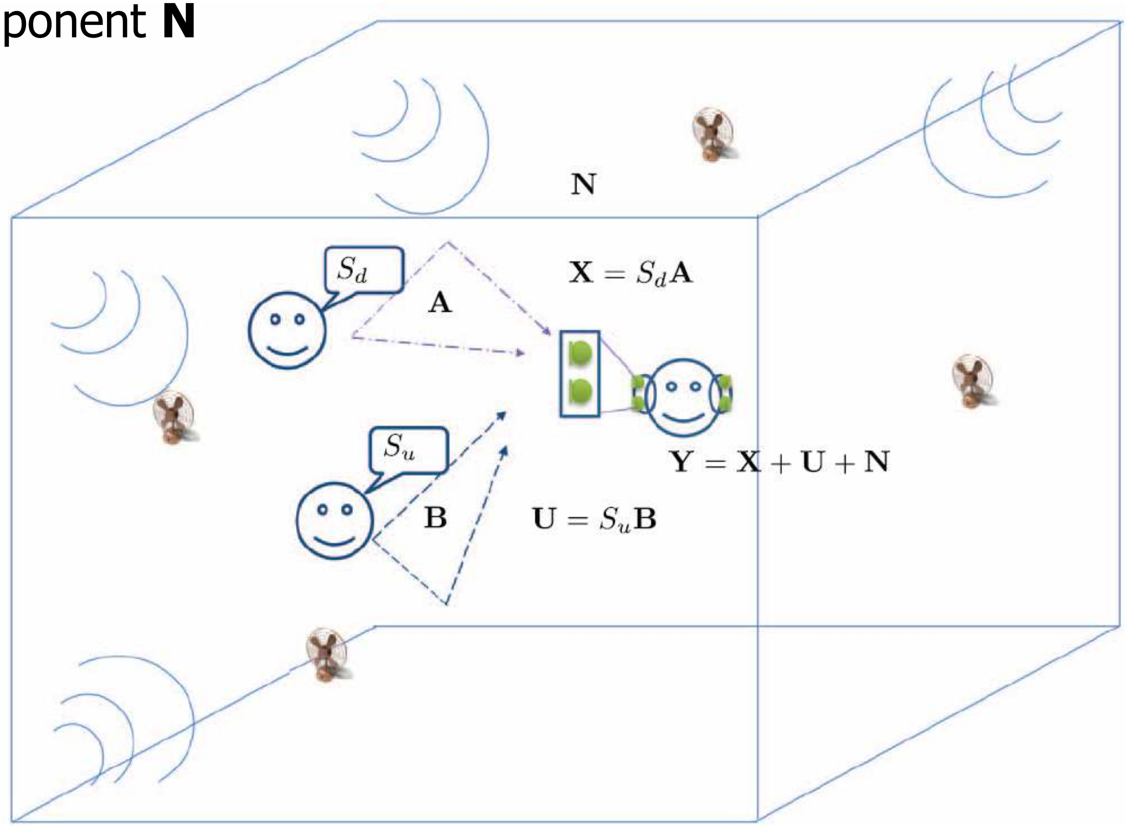
Acoustic Transfer Functions (ATFs)

- The microphone signals **Y** are composed of
  - (desired) speech component $\mathbf{X} = S_d \mathbf{A}$
  - (undesired) directional interference component $\mathbf{U} = S_u \mathbf{B}$
  - (undesired) background noise component **N**
- Correlation matrices:

$$\mathbf{R}_y = \mathbf{R}_x + \underbrace{\mathbf{R}_u + \mathbf{R}_n}_{\mathbf{R}_v}$$

$$\mathbf{R}_x = \mathcal{E}\left\{\mathbf{X}\mathbf{X}^H\right\} = P_s \mathbf{A}\mathbf{A}^H$$
$$\mathbf{R}_u = \mathcal{E}\left\{\mathbf{U}\mathbf{U}^H\right\} = P_u \mathbf{B}\mathbf{B}^H$$
$$\mathbf{R}_n = \mathcal{E}\left\{\mathbf{N}\mathbf{N}^H\right\},$$

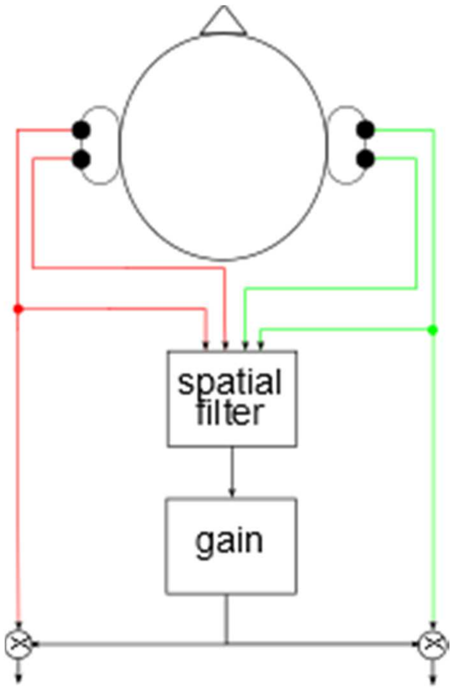
- All **binaural cues** can be written in terms of these matrices



Binaural noise reduction: Two main paradigms

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

[Doerbecker 1996, Wittkop 2003, Lotter 2006, Rohdenburg 2007, Grimm 2009, Reindl 2012]

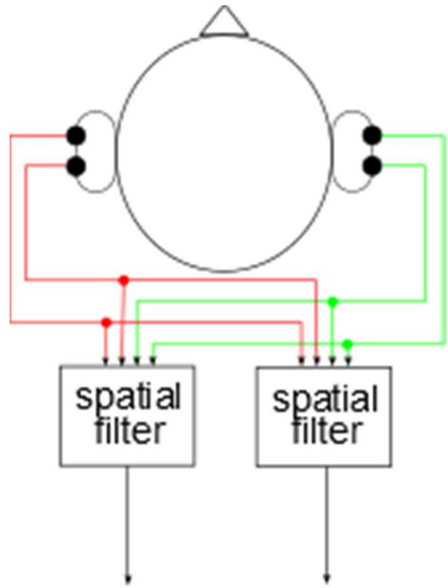


⊕ Binaural cue preservation

⊖ Possible single-channel artifacts

Binaural multi-microphone noise reduction techniques

[Welker 1997, Doclo 2010, Cornelis 2012, Hadad 2014-2016, Marquardt 2014-2016]



⊕ Larger noise reduction performance

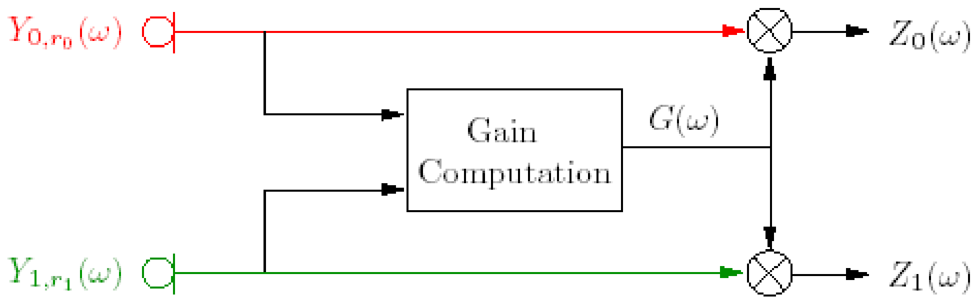
⊕ Merge spatial and spectral post-filtering

⊖ Binaural cue preservation not guaranteed

1) Spectral post-filtering

Time-frequency post-filtering/masking:

- Computation and application of **real-valued** spectral gain/mask
- Gain G(ω) based on **coherence, binaural cues (ITD/ILD)** and temporal/spectral cues (pitch, onset, modulation frequencies)



[Doerbecker 1996, Wittkop 2003, Grimm 2009, Martin 2015]

1) Spectral post-filtering

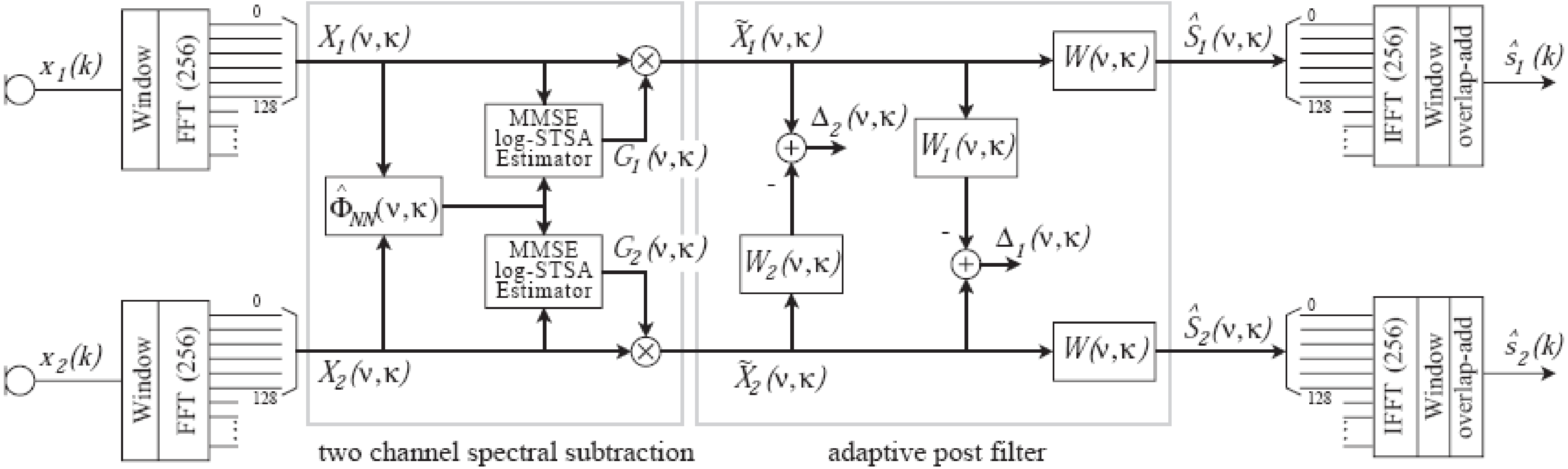


Figure 1: Proposed noise reduction system with two channel Spectral Subtraction and adaptive Wiener post-filtering
ν: frequency index, κ: index of decimated time

[Doerbecker 1996]

1) Spectral post-filtering

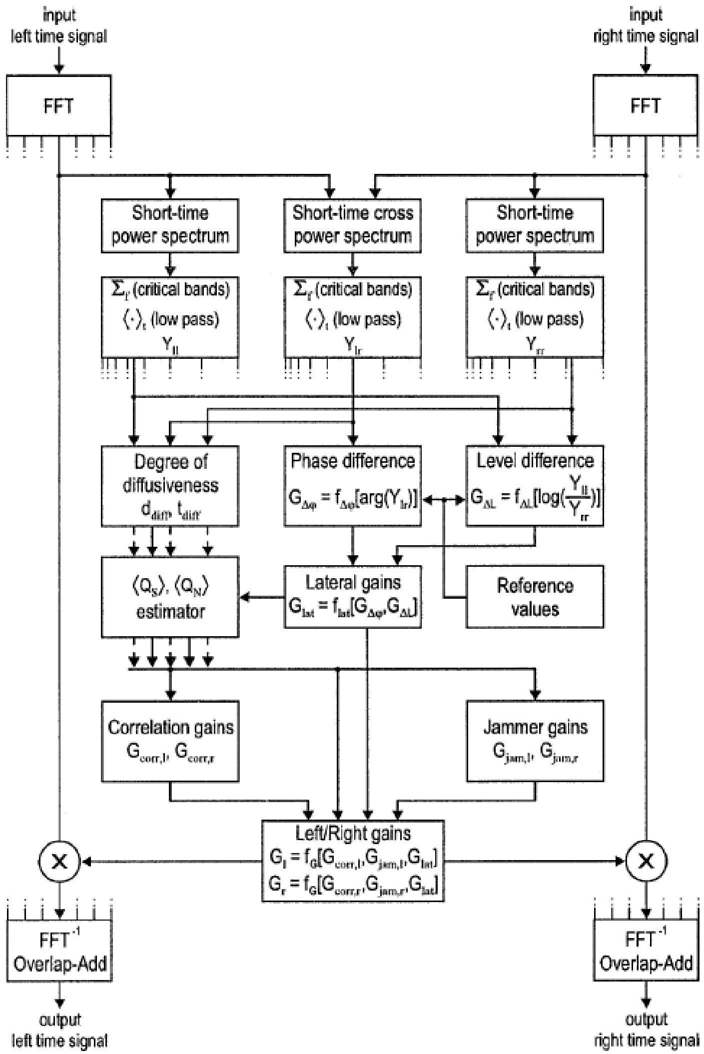


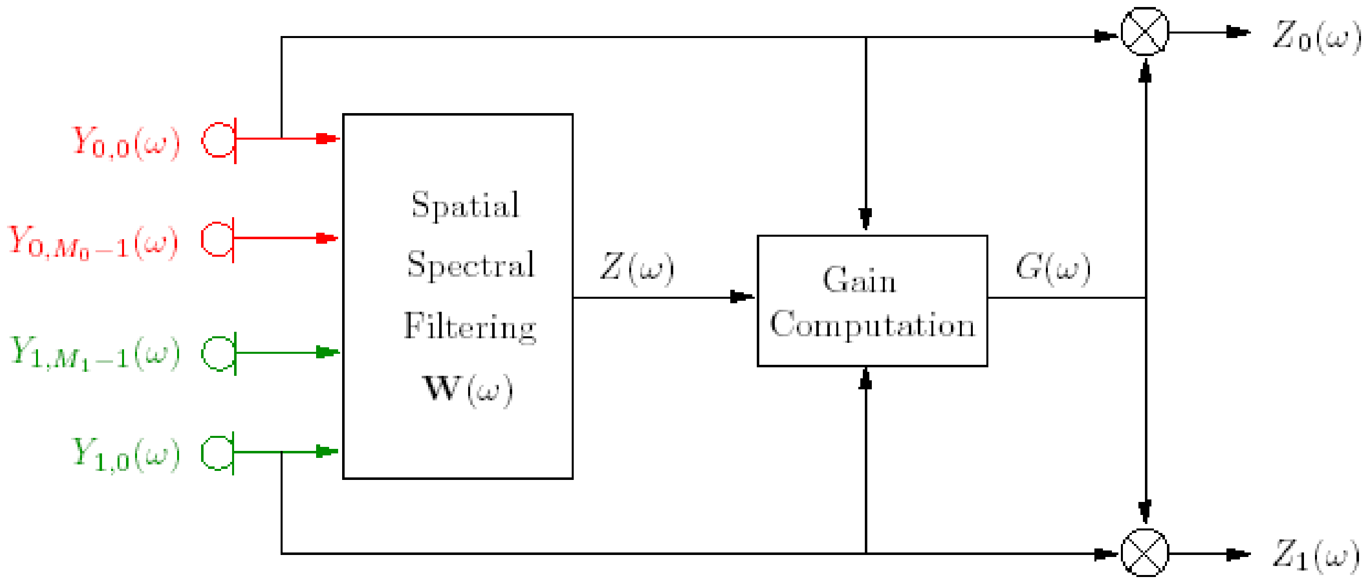
Fig. 8. Block diagram of the strategy-selective algorithm for dereverberation and suppression of lateral noise sources.

[Wittkop 2003]

1) Spectral post-filtering

Can be merged with multi-microphone noise reduction:

- E.g. based on fixed/adaptive beamforming or blind source separation
- **However**: still in principle single-channel noise reduction (noisy phase, possible artefacts)



[Lotter 2006, Wehr 2008, Rohdenburg 2009, Saruwatari 2010, Reindl 2012, Baumgaertel 2015]

1) Spectral post-filtering

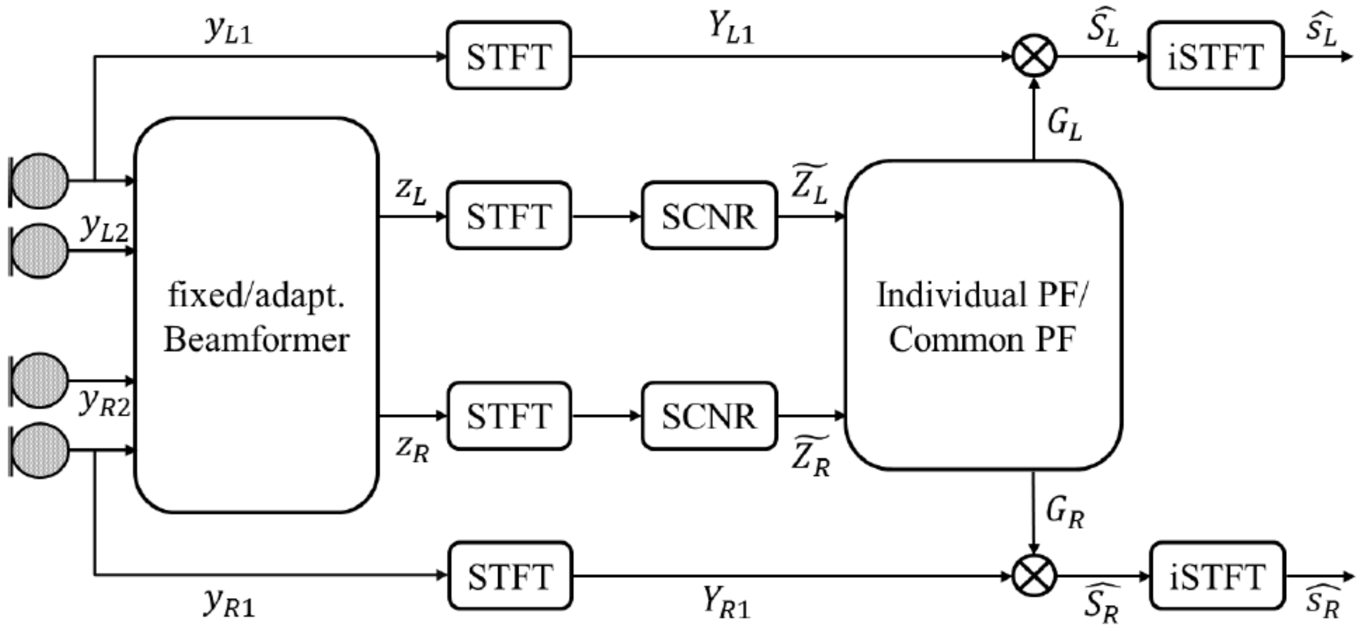


Figure 4. Block diagram illustrating the combination of a binaural MVDR beamformer with single-channel postfiltering. The indices n, k, and l were dropped for the sake of brevity.

[Baumgaertel 2015]

2) Binaural MVDR and MWF

Minimum-Variance-Distortionless-Response (MVDR) beamformer

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

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

$$\min_{\mathbf{W}_1} \underbrace{\mathbf{W}_1^H \mathbf{R}_\mathrm{v} \mathbf{W}_1}_{\text{noise reduction}} \quad \text{subject to} \quad \underbrace{\mathbf{W}_1^H \mathbf{A} = A_1}_{\text{distortionless constraint}}$$

$$\mathbf{W}_{\mathrm{MVDR},0} = \frac{\mathbf{R}_\mathrm{v}^{-1}\mathbf{A}}{\mathbf{A}^H \mathbf{R}_\mathrm{v}^{-1}\mathbf{A}} A_0^*$$

$$\mathbf{W}_{\mathrm{MVDR},1} = \frac{\mathbf{R}_\mathrm{v}^{-1}\mathbf{A}}{\mathbf{A}^H \mathbf{R}_\mathrm{v}^{-1}\mathbf{A}} A_1^*$$

Multi-channel Wiener Filter (MWF)

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

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

$$\mathbf{W}_{\mathrm{MWF},0} = (\mathbf{R}_x + \mu \mathbf{R}_v)^{-1} \mathbf{r}_{\mathrm{x},0}$$

$$\mathbf{W}_{\mathrm{MWF},1} = (\mathbf{R}_x + \mu \mathbf{R}_v)^{-1} \mathbf{r}_{\mathrm{x},1}$$

$$\mathbf{W}_{\mathrm{MWF},0} = \frac{\rho}{\mu+\rho} \frac{\mathbf{R}_\mathrm{v}^{-1}\mathbf{A}}{\mathbf{A}^H \mathbf{R}_\mathrm{v}^{-1}\mathbf{A}} A_0^* = \frac{\rho}{\mu+\rho} \mathbf{W}_{\mathrm{MVDR},0}$$

$$\mathbf{W}_{\mathrm{MWF},1} = \frac{\rho}{\mu+\rho} \frac{\mathbf{R}_\mathrm{v}^{-1}\mathbf{A}}{\mathbf{A}^H \mathbf{R}_\mathrm{v}^{-1}\mathbf{A}} A_1^* = \frac{\rho}{\mu+\rho} \mathbf{W}_{\mathrm{MVDR},1}$$

2) Binaural MVDR and MWF

Minimum-Variance-Distortionless-Response (MVDR) beamformer

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

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

$$\min_{\mathbf{W}_1} \underbrace{\mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1}_{\text{noise reduction}} \quad \text{subject to} \quad \underbrace{\mathbf{W}_1^H \mathbf{A} = A_1}_{\text{distortionless constraint}}$$

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

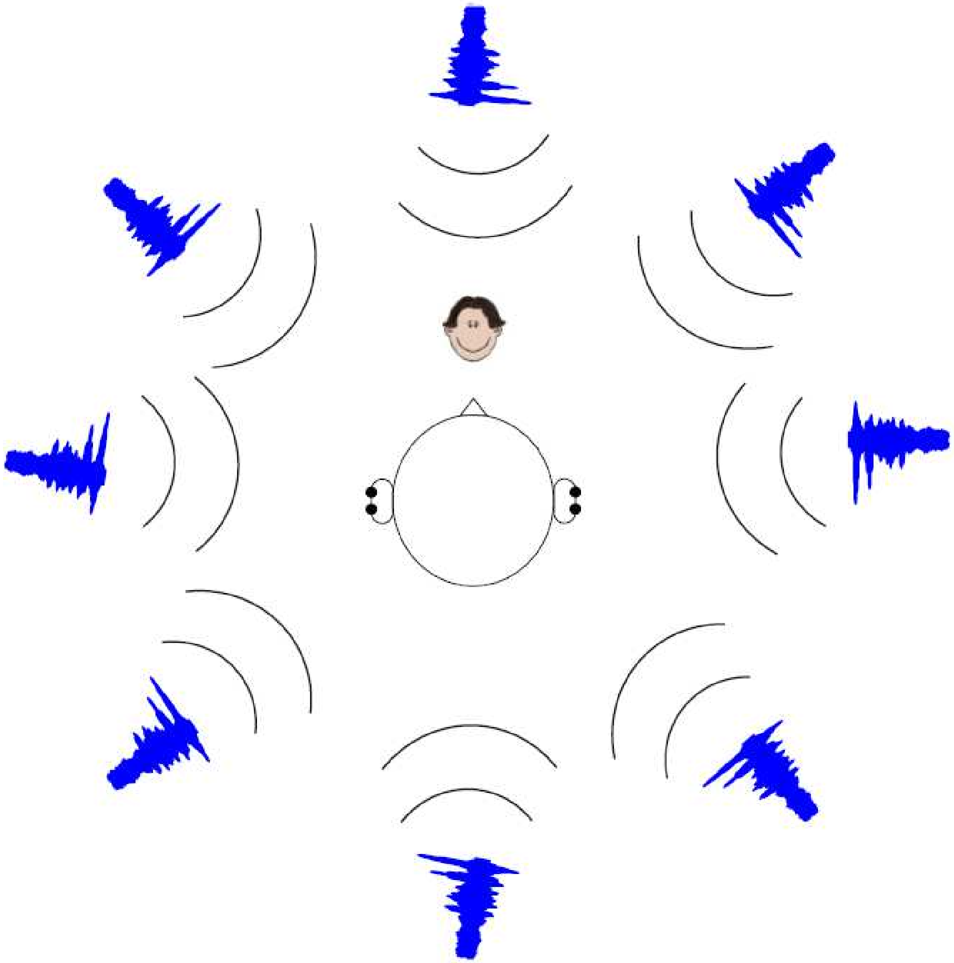
$$J_{\text{MWF}}(\mathbf{W}) = \mathcal{E}\left\{ \underbrace{\left\| \begin{bmatrix} X_0 - \mathbf{W}_0^H \mathbf{X} \\ X_1 - \mathbf{W}_1^H \mathbf{X} \end{bmatrix} \right\|^2}_{\text{speech distortion}} + \mu \underbrace{\left\| \begin{bmatrix} \mathbf{W}_0^H \mathbf{V} \\ \mathbf{W}_1^H \mathbf{V} \end{bmatrix} \right\|^2}_{\text{noise reduction}} \right\}$$

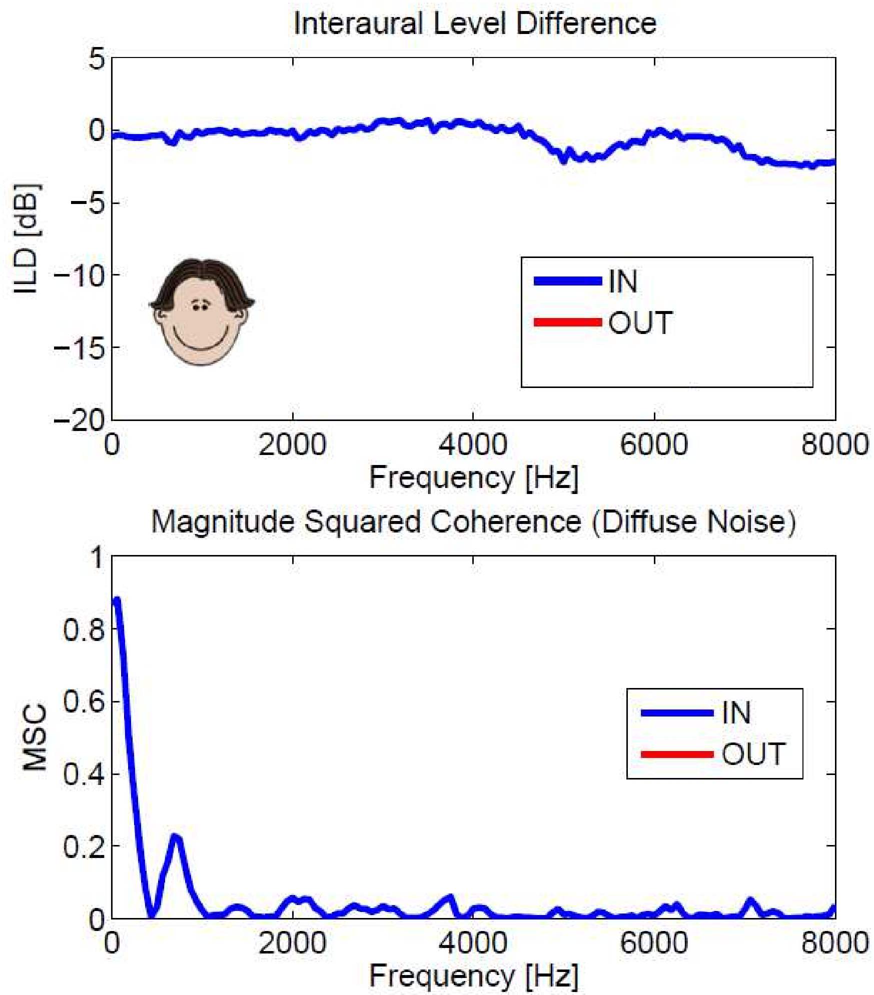
Requires estimate of speech and noise covariance matrices, e.g. based on VAD

Can be decomposed as binaural MVDR beamformer and spectral postfilter

Good noise reduction performance, what about binaural cues ?

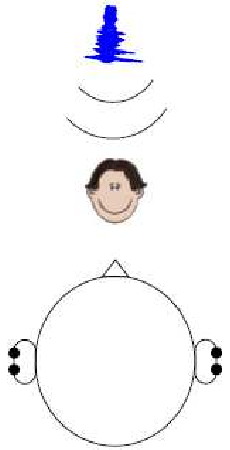
2) Binaural MVDR and MWF Binaural cues (diffuse noise)

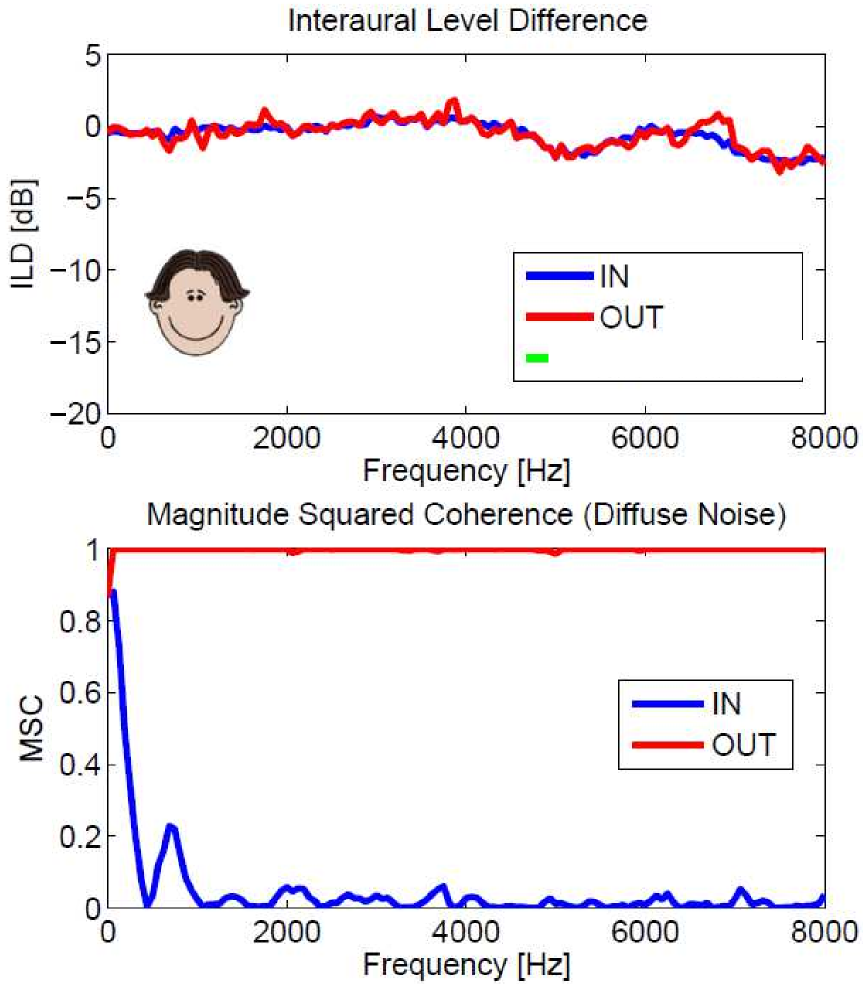




MSC = Magnitude Squared Coherence=$|IC|^2$

2) Binaural MVDR and MWF Binaural cues (diffuse noise)



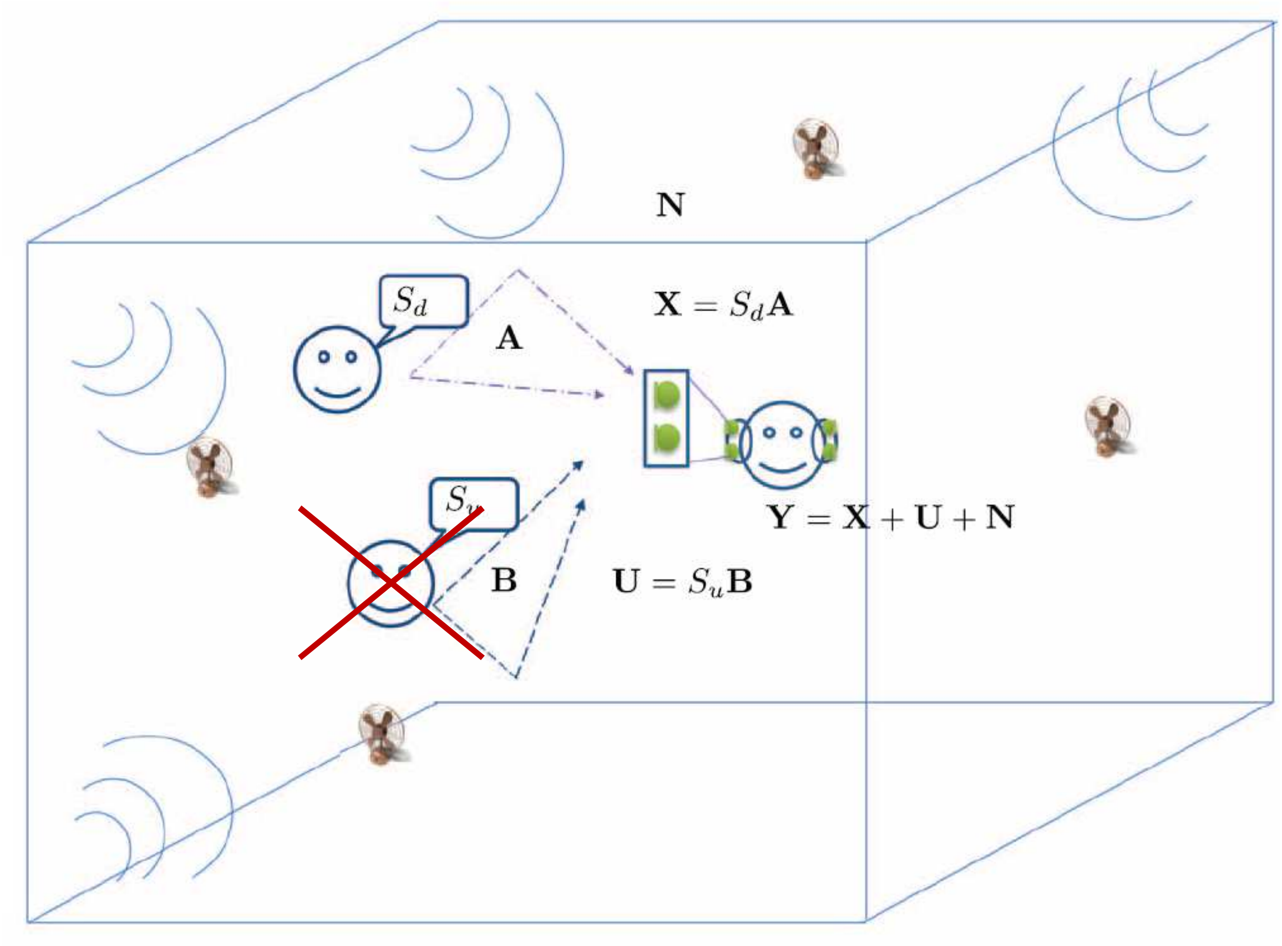


Binaural cues for residual noise/interference in binaural MVDR/MWF not preserved

Binaural cue preservation for diffuse noise

Acoustic scenario

- Speech source + background 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

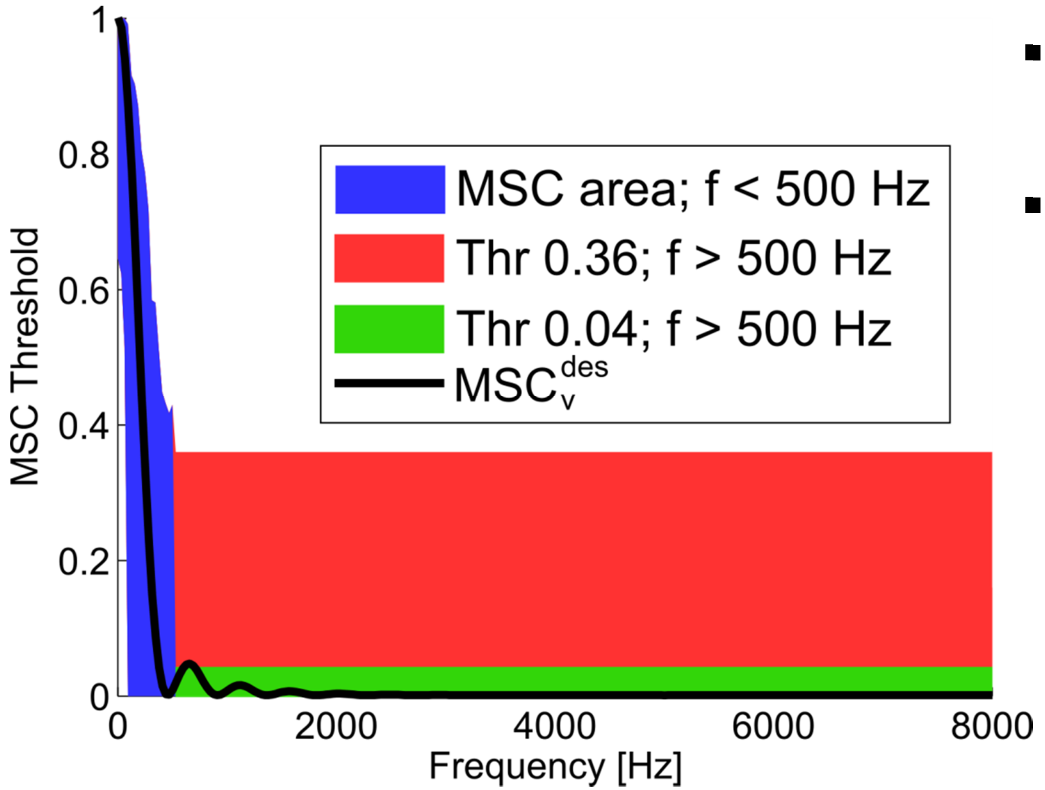
⊕ Closed-form solution (mixing with reference microphone signals)

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

[Marquardt 2013/2014/2015, Braun 2014] [Doclo 2010, Cornelis 2010/2012]

Binaural MWF: Extensions for diffuse noise

- **Determine (frequency-dependent) trade-off parameters based on psycho-acoustic criteria**
  - Amount of IC preservation based on subjective listening experiments evaluating the IC discrimination abilities of the human auditory system



- 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**)

[Marquardt 2014/2015]

Binaural MWF: Extensions for diffuse noise

- **MWF-IC: Interaural coherence preservation**
  - IC preservation cost function: minimize difference between IC of output noise component and desired IC

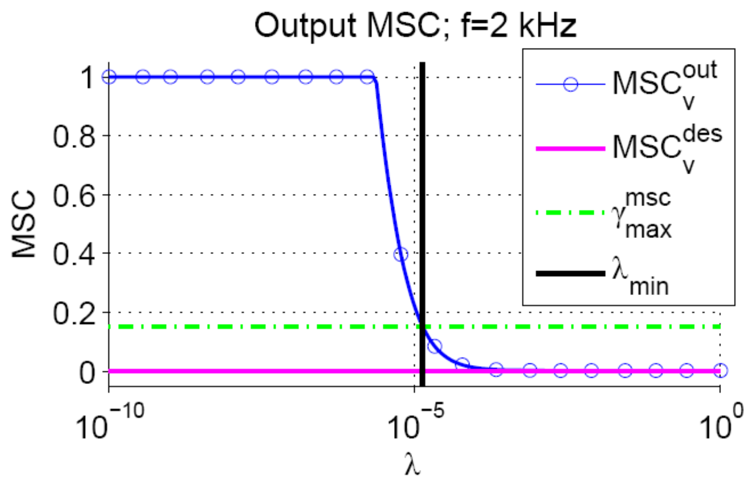
$$
\boxed{
\begin{aligned}
J_{\mathrm{IC}}(\mathbf{W}) &= \left| IC_{\mathrm{v}}^{\mathrm{out}} - IC_{\mathrm{v}}^{\mathrm{des}} \right|^2 , \\
&= \left| \frac{\mathbf{W}_0^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_1}{\sqrt{\left(\mathbf{W}_0^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_0\right)\left(\mathbf{W}_1^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_1\right)}} - IC_{\mathrm{v}}^{\mathrm{des}} \right|^2
\end{aligned}
}
$$

  - **Desired IC** can be computed using estimated input noise correlation matrix, or based on head model / measured HRTFs
  - **No closed-form filter expression** → iterative optimization procedures required

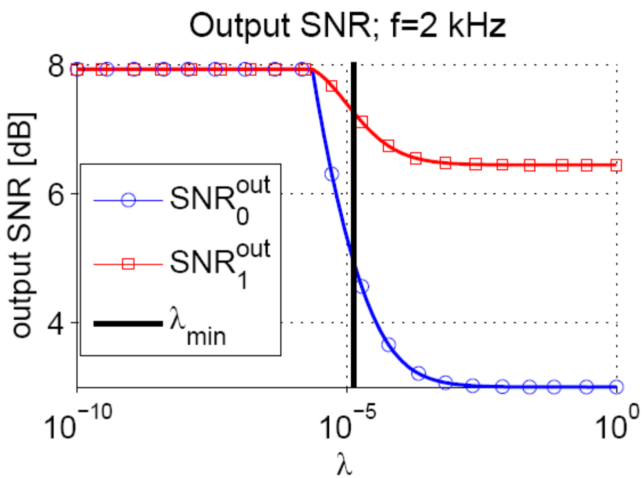
[Marquardt 2013-2015]

Binaural MWF: Extensions for diffuse noise

- **MWF-IC: Interaural coherence preservation**
  - **Trade-off parameter λ** trades off MSC error and output SNR



(a) Output MSC



(b) Output SNR

  - Exhaustive / iterative search to determine optimal trade-off parameter λ, satisfying psycho-acoustically motivated MSC boundaries

$$\Lambda = \{\lambda \,|\, \gamma_{\min}^{\mathrm{msc}} \leq MSC_{\mathrm{v}}^{\mathrm{out}}(\mathbf{W}_{\mathrm{MWF-IC}}(\lambda)) \leq \gamma_{\max}^{\mathrm{msc}}\}$$

Binaural MWF: Extensions for diffuse noise

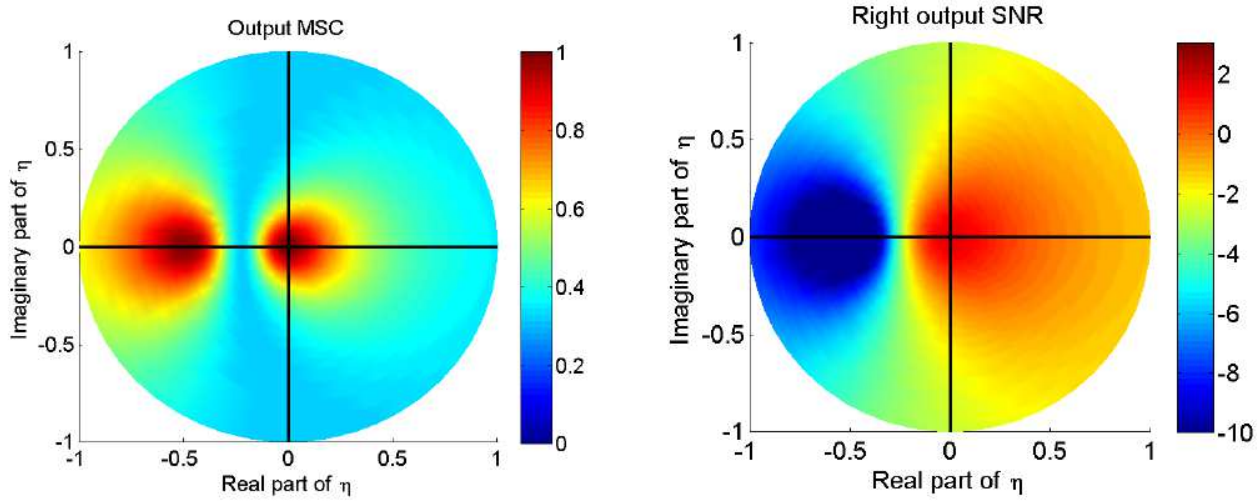
- **MWF-N: Partial noise estimation**
  - **Closed form filter expression** → mixing of binaural MWF output signals and reference microphone signals

$$\boxed{\begin{aligned}\mathbf{W}_{\mathrm{MWF-N},0} &= (1-\eta^*)\,\mathbf{W}_{\mathrm{MWF},0} + \eta^*\mathbf{e}_0 \\ \mathbf{W}_{\mathrm{MWF-N},1} &= (1-\eta^*)\,\mathbf{W}_{\mathrm{MWF},1} + \eta^*\mathbf{e}_1\end{aligned}}$$

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

Binaural MWF: Extensions for diffuse noise

- **MWF-N: Partial noise estimation**
  - **Trade-off parameter η** trades off MSC error and output SNR
    - **MWF-N** : exhaustive search for optimal trade-off parameter



(a) $MSC_{\mathrm{v}}^{\mathrm{out}}$ for the MWF-N (b) SNR_{1}^{out} for the MWF-N

$$\Upsilon = \{\eta \,|\, \gamma_{\min}^{\mathrm{msc}} \leq MSC_{\mathrm{v}}^{\mathrm{out}}(\eta) \leq \gamma_{\max}^{\mathrm{msc}}\} \qquad \eta_{\mathrm{opt}} = \underset{\eta \in \Upsilon}{\arg\max} \frac{SNR_{\mathrm{MWF-N},0}^{\mathrm{out}}(\eta) + SNR_{\mathrm{MWF-N},1}^{\mathrm{out}}(\eta)}{2}$$

Binaural MWF: Extensions for diffuse noise

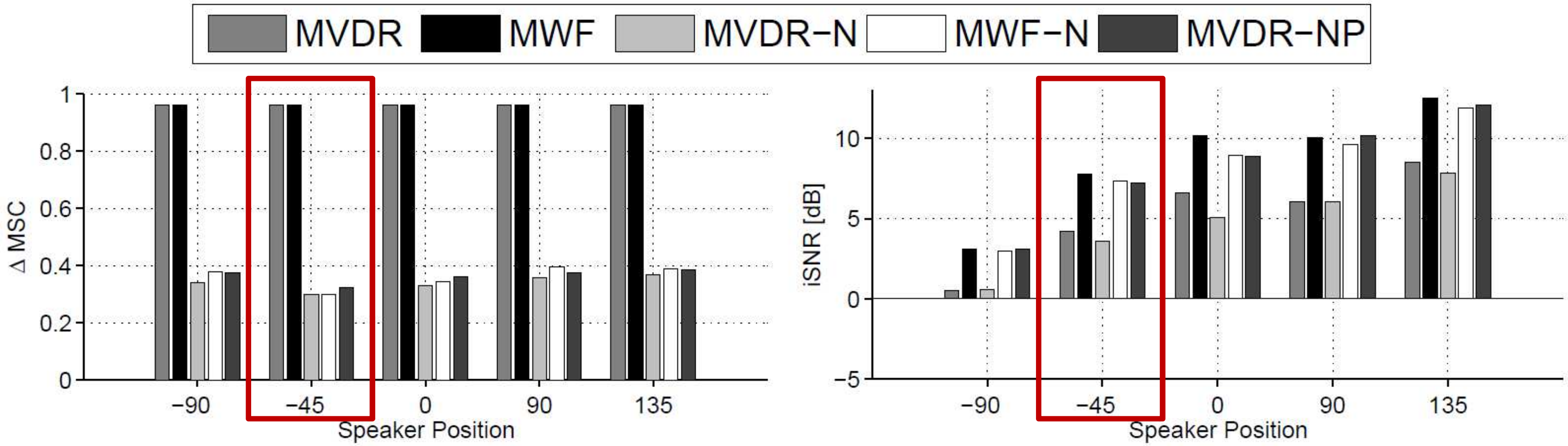
- **MWF-N: Partial noise estimation**
  - **Trade-off parameter η** trades off MSC error and output SNR
    - **MWF-N** : exhaustive search for optimal trade-off parameter
    - **MVDR-N** (i.e. special case of MWF-N with μ=0) :
      - Closed-form expression for optimal trade-off parameter
      - No 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
  - Note: main difference with previous work is **frequency-dependency and signal-dependency of trade-off parameter**
  - Note: also other criteria for determining trade-off parameter possible, e.g. based on **output SNR**

Table 1 Comparison of mixing factors for binaural beamforming algorithms

	MVDR	MVDR-N	SBB
α	1	$1-\eta$	$1 \quad \hat{\phi}_s(n)/\hat{\phi}_v(n) > 1$ $0 \quad$ otherwise
β	0	η	$0 \quad \hat{\phi}_s(n)/\hat{\phi}_v(n) > 1$ $\gamma \quad$ otherwise

Binaural MWF: Extensions for diffuse noise

❑ Instrumental evaluation / sound samples



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

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

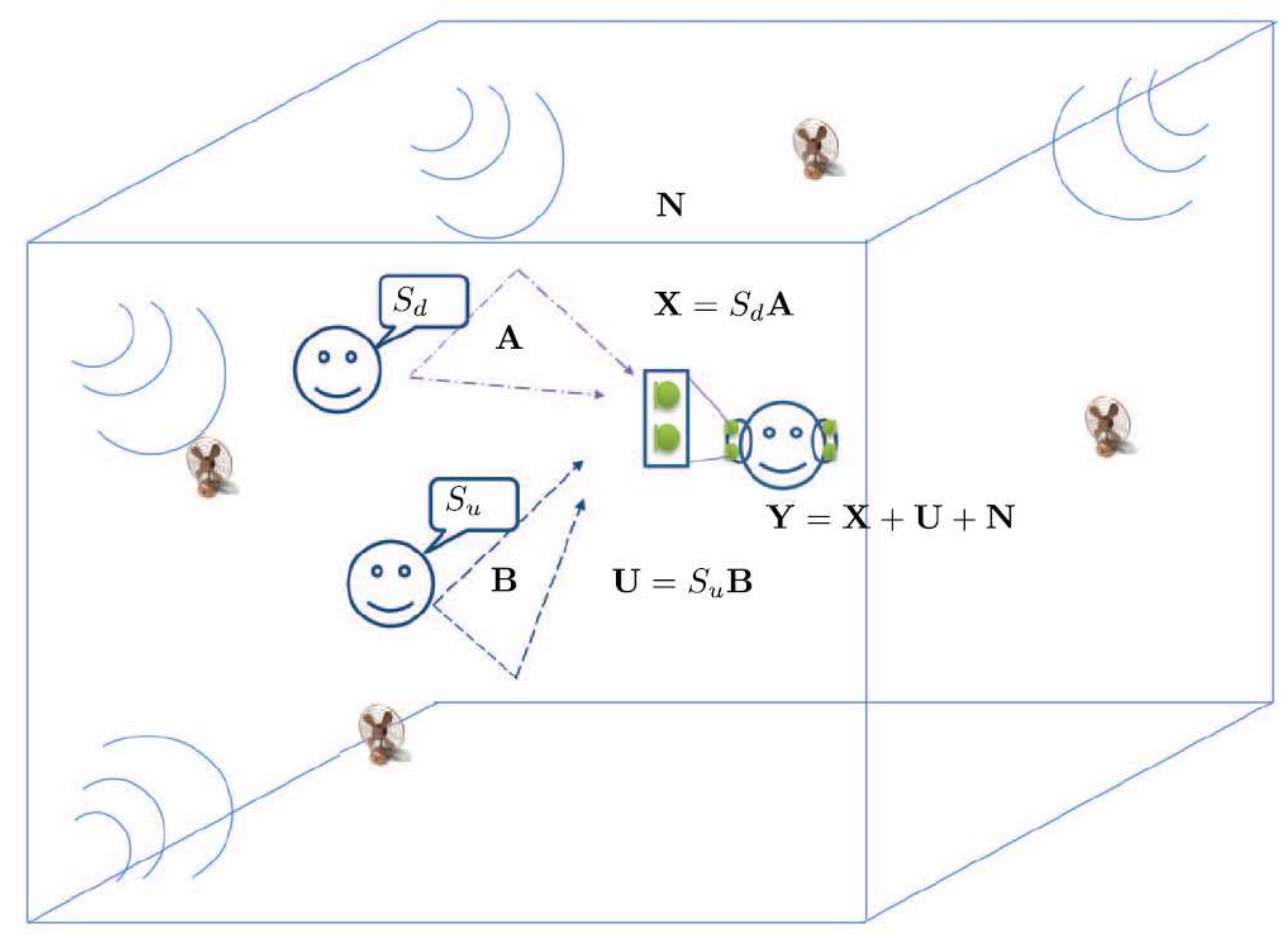
Cafeteria with recorded ambient noise, speaker at -35°, 0 dB input iSNR (left hearing aid)
MVDR: anechoic ATF, DOA known, spatial coherence matrix calculated from anechoic ATFs / MWF = MVDR + postfilter (SPP-based)

[Marquardt 2016 - unpublished]

Binaural cue preservation for interfering source

Acoustic scenario

- Speech source + interfering speaker + background noise



Binaural MVDR: Extensions for interfering source

Binaural MVDR

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

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
- ⊖ Background noise: MSC not exactly preserved, possible noise amplification

Binaural MVDR: Extensions for interfering source

❑ Comparison between BMVDR-RTF and BMVDR-IR

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

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

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

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

$$\begin{aligned}
\mathbf{W}_0 &= \frac{1}{\sigma_a}\left[A_0^* + \frac{(A_0+\alpha A_1)^*}{(1+|\alpha|^2)}\frac{\Sigma}{(1-\Sigma)}\right]\mathbf{R}_v^{-1}\mathbf{A} \\
&\quad - \frac{(A_0+\alpha A_1)^*}{(1+|\alpha|^2)\,\sigma_{ab}}\frac{\Sigma}{(1-\Sigma)}\mathbf{R}_v^{-1}\mathbf{B}, \\
\mathbf{W}_1 &= \frac{1}{\sigma_a}\left[A_1^* + \alpha\frac{(A_0+\alpha A_1)^*}{(1+|\alpha|^2)}\frac{\Sigma}{(1-\Sigma)}\right]\mathbf{R}_v^{-1}\mathbf{A} \\
&\quad - \alpha\frac{(A_0+\alpha A_1)^*}{(1+|\alpha|^2)\,\sigma_{ab}}\frac{\Sigma}{(1-\Sigma)}\mathbf{R}_v^{-1}\mathbf{B},
\end{aligned}$$

$$\begin{aligned}
\mathbf{W}_0 &= \left(\frac{A_0^*}{\sigma_a} - \frac{\eta B_0^* \sigma_{ab}}{\sigma_a\sigma_b}\right)\frac{\mathbf{R}_v^{-1}\mathbf{A}}{1-\Sigma} + \left(\frac{\eta B_0^*}{\sigma_b} - \frac{A_0^*\sigma_{ab}^*}{\sigma_a\sigma_b}\right)\frac{\mathbf{R}_v^{-1}\mathbf{B}}{1-\Sigma}, \\
\mathbf{W}_1 &= \left(\frac{A_1^*}{\sigma_a} - \frac{\eta B_1^* \sigma_{ab}}{\sigma_a\sigma_b}\right)\frac{\mathbf{R}_v^{-1}\mathbf{A}}{1-\Sigma} + \left(\frac{\eta B_1^*}{\sigma_b} - \frac{A_1^*\sigma_{ab}^*}{\sigma_a\sigma_b}\right)\frac{\mathbf{R}_v^{-1}\mathbf{B}}{1-\Sigma}.
\end{aligned}$$

Binaural MVDR: Extensions for interfering source

Comparison between BMVDR-RTF and BMVDR-IR

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

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

- **BMVDR-RTF:** one constraint (relative transfer function)
  BMVDR-IR: two constraints (interference reduction parameter η)
- BMVDR-RTF is a special case of BMVDR-IR for specific (typically complex-valued) parameter η
- **Implementation** requires:
  - estimate of noise correlation matrix $\mathbf{R}_v$
  - estimates of relative transfer function vectors for speech source and interfering source(s) **A** and **B** – estimation procedures available

[Hadad 2014/2015/2016, Marquardt 2014]

Binaural MVDR: Extensions for interfering source

- **Comparison between BMVDR-RTF and BMVDR-IR**

- **Signal-to-interference-and-noise ratio (SINR)**

$$\mathrm{SINR}^{\mathrm{out}}_{\mathrm{BMVDR}} \geq \mathrm{SINR}^{\mathrm{out}}_{\mathrm{BMVDR-RTF}} \geq \mathrm{SINR}^{\mathrm{out}}_{\mathrm{BMVDR-IR}}$$

- **Signal-to-interference ratio (SIR)**

$$\begin{aligned}\mathrm{SIR}^{\mathrm{in}} &= \mathrm{SIR}^{\mathrm{out}}_{\mathrm{BMVDR-IR}_{\eta=1}} \\ &\leq \mathrm{SIR}^{\mathrm{out}}_{\mathrm{BMVDR-IR}} \leq \mathrm{SIR}^{\mathrm{out}}_{\mathrm{BMVDR-IR}_{\eta=0}},\end{aligned}$$

$$\mathrm{SIR}^{\mathrm{out}}_{\mathrm{BMVDR}} \leq \mathrm{SIR}^{\mathrm{out}}_{\mathrm{BMVDR-RTF}}$$

- **Signal-to-noise ratio (SNR)**

$$\mathrm{SNR}^{\mathrm{out}}_{\mathrm{BMVDR}} \geq \mathrm{SNR}^{\mathrm{out}}_{\mathrm{BMVDR-RTF}}$$

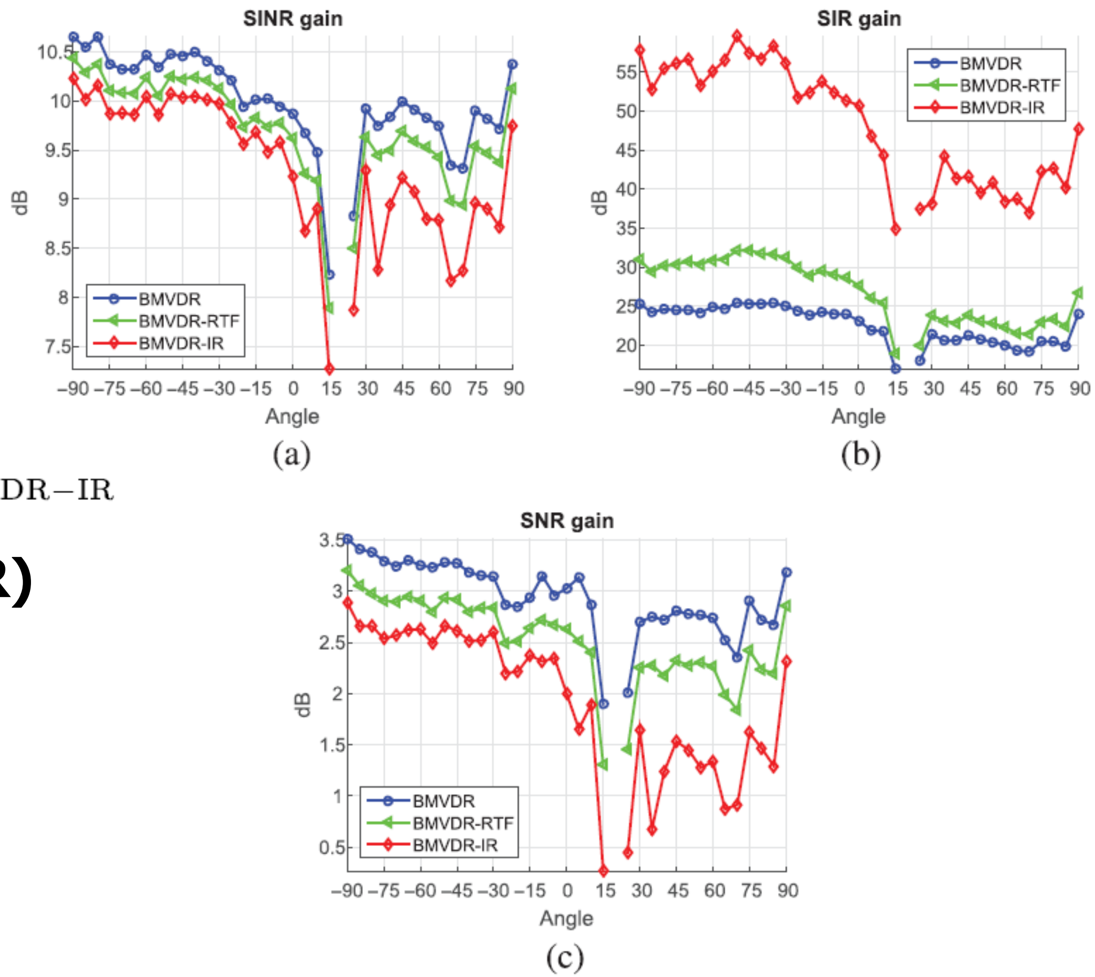


Fig. 4. Global binaural SINR, SIR and SNR gains for BMVDR, BMVDR-RTF and BMVDR-IR ($\eta = \eta_{opt}$) beamformers for different angles of the interfering source (Office environment, desired source at 20°, $M = 3$). (a) SINR gain (b) SIR gain (c) SNR gain.

Binaural MVDR: Extensions for interfering source

- Comparison between BMVDR-RTF and BMVDR-IR
  - **Beampattern**

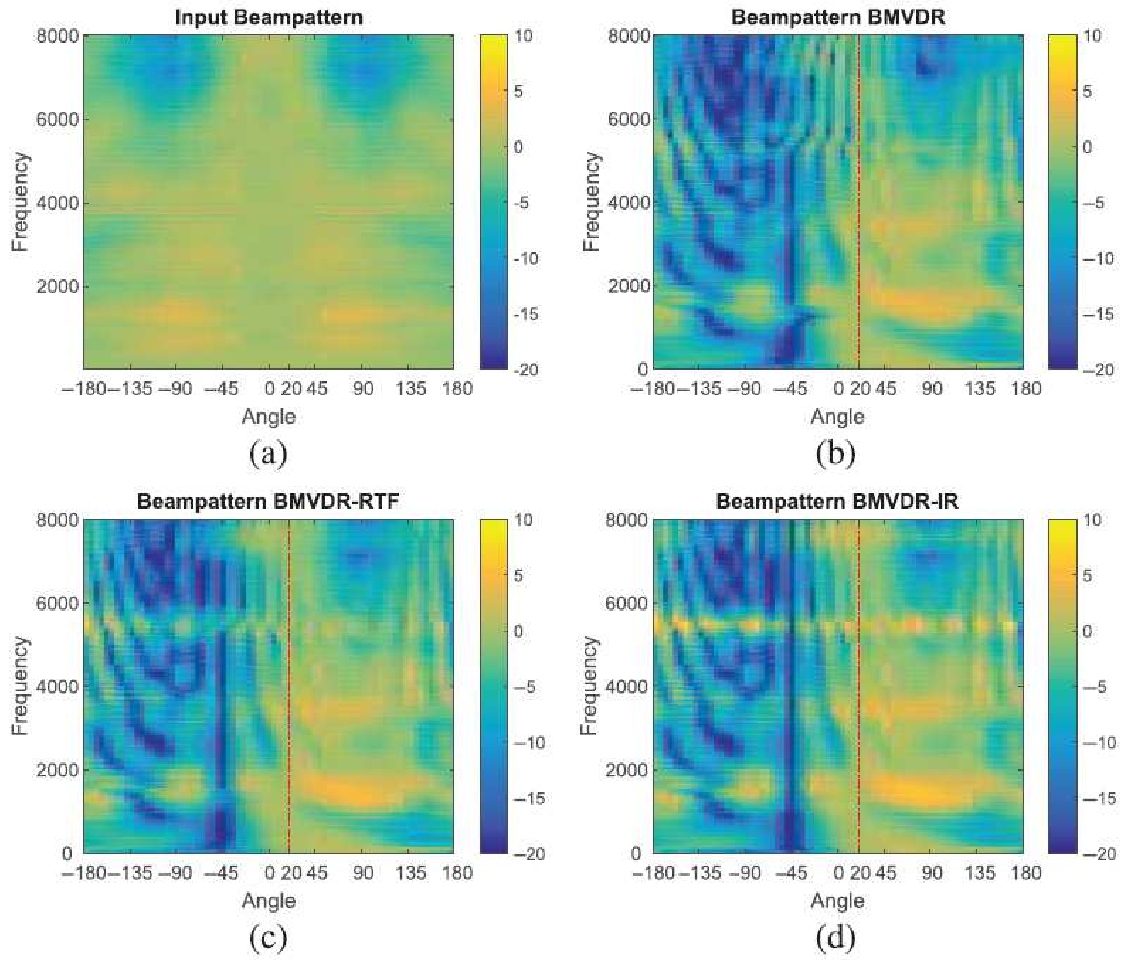


Fig. 3. Beampattern of the BMVDR, BMVDR-RTF and BMVDR-IR ($\eta = 0$) beamformers, together with the input beampattern (anechoic environment, desired source at 20°, interfering source at $-45°$, $M = 3$). (a) Input (b) BMVDR (c) BMVDR-RTF (d) BMVDR-IR.

Binaural MVDR: Extensions for interfering source

- Comparison between BMVDR-RTF and BMVDR-IR
  - **MSC of background noise**

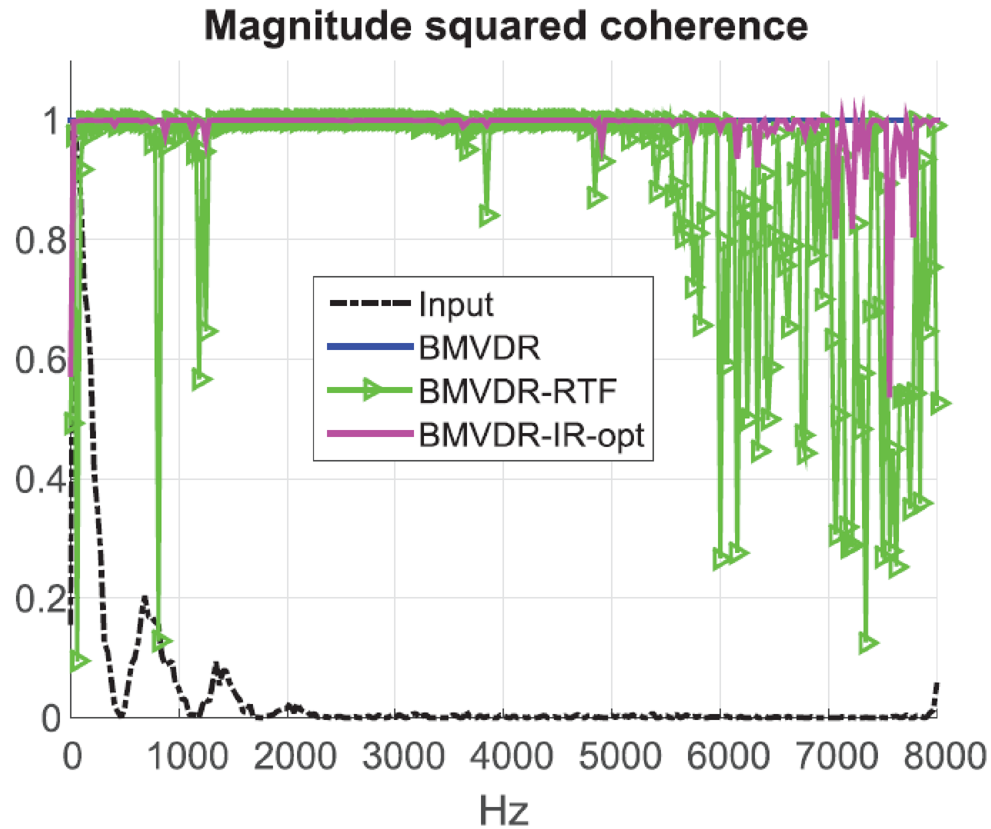
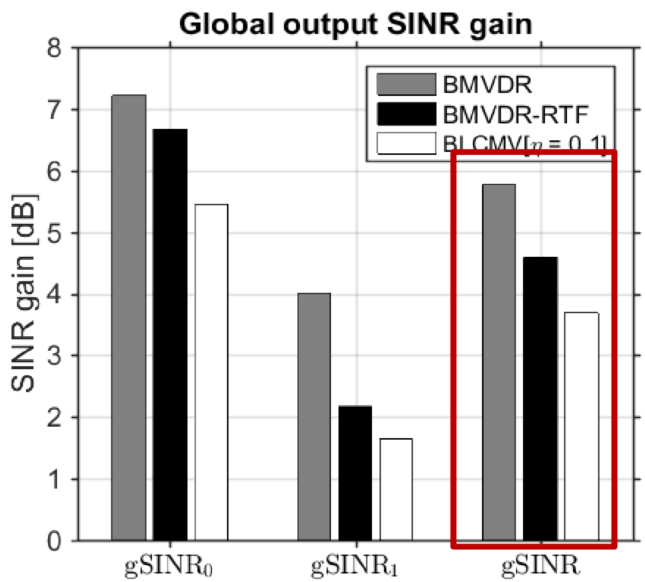


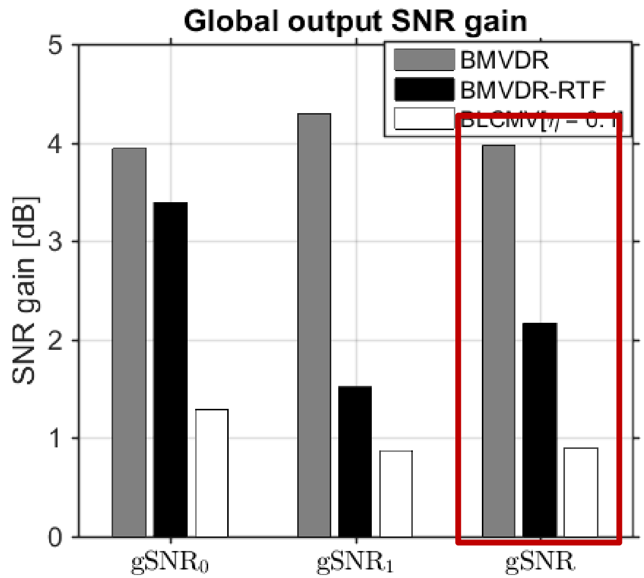
Fig. 6. The MSC for diffuse noise field at the input and at the output of the BMVDR, BMVDR-RTF and BMVDR-IR ($\eta = \eta_{opt}$) beamformers (Office, desired source at 20°, interfering source at $-45°$, $M = 3$).

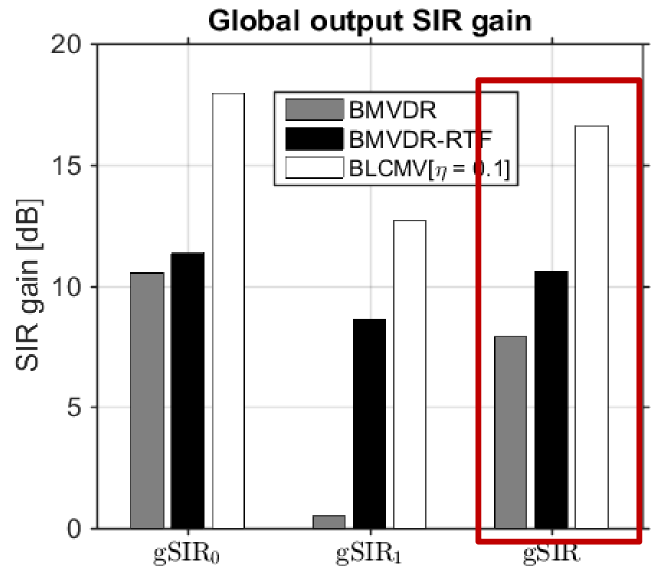
[Hadad 2015]

Binaural MVDR: Extensions for interfering source

❑ Instrumental evaluation / sound samples







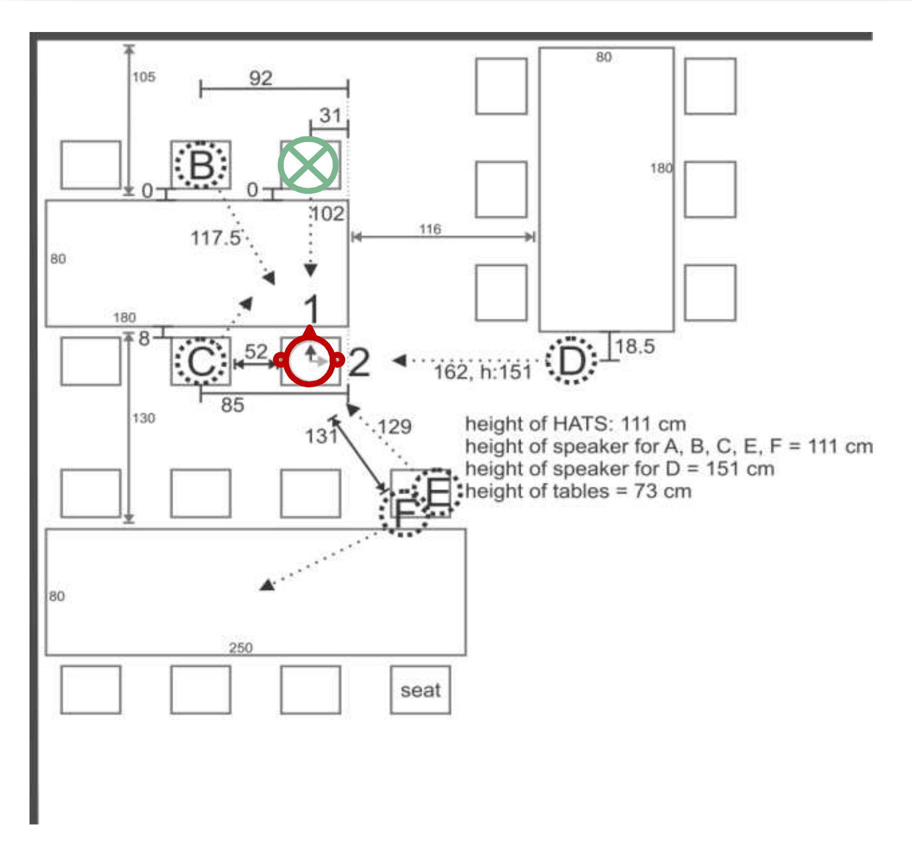
Input	BMVDR	BMVDR-RTF	BMVDR-IR (η = 0.1)

Cafeteria with recorded ambient noise, speaker at 0°, Interference at -45°, 0 dB input SIR and SNR (left hearing aid)
RTF calculated from correlation matrix (Rx and Ru), 3 microphones (2 left, 1 right)

[Marquardt 2014]

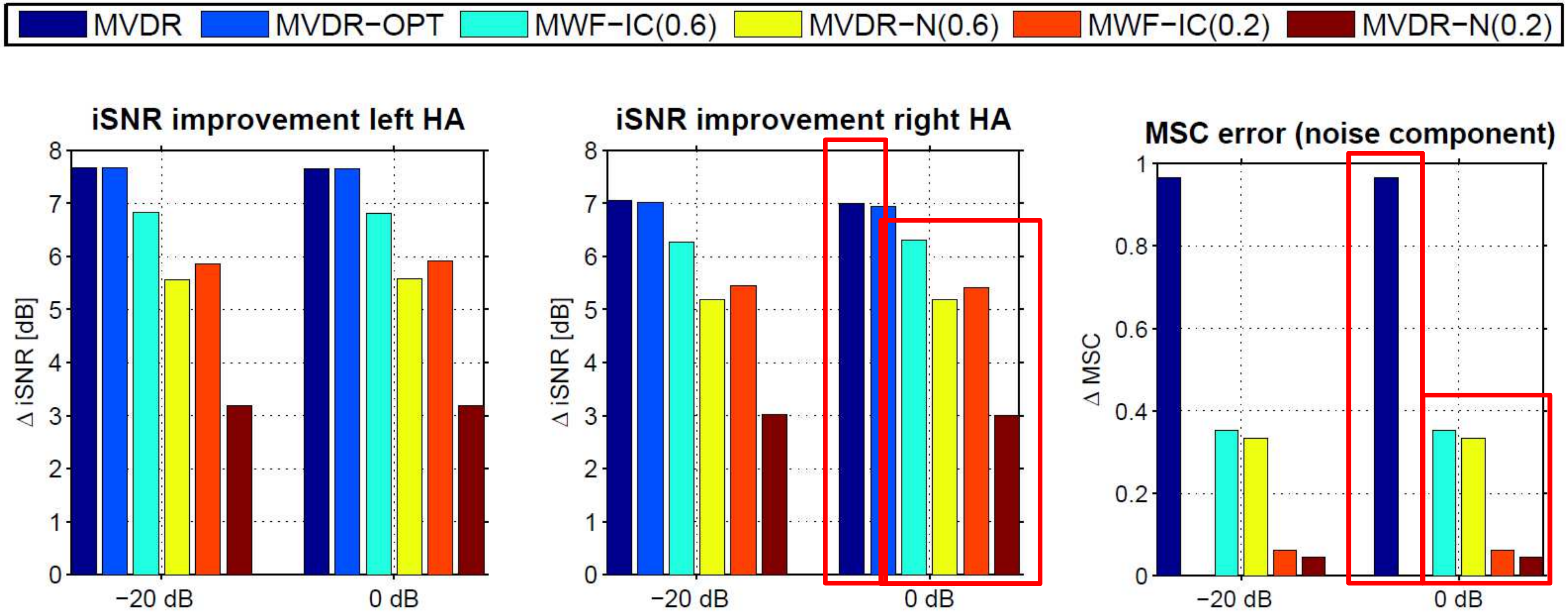
Evaluation Study 1 (NH)

Evaluation: Test setup



- Binaural hearing aid recordings (4 mics) in **cafeteria** (T_{60} ≈ 1250 ms) [Kayser 2009]
- **Noise:** realistic cafeteria ambient noise
- **Algorithms**: binaural MVDR + cue preservation extensions (MWF-IC, MVDR-N)
- **Instrumental measures**: MSC error and intelligibility-weighted SNR gain
- **Subjective listening experiments**:
  - 15 normal-hearing subjects
  - **SRT** using Oldenburg Sentence Test (OLSA)
  - **Spatial quality (diffuseness)** using MUSHRA

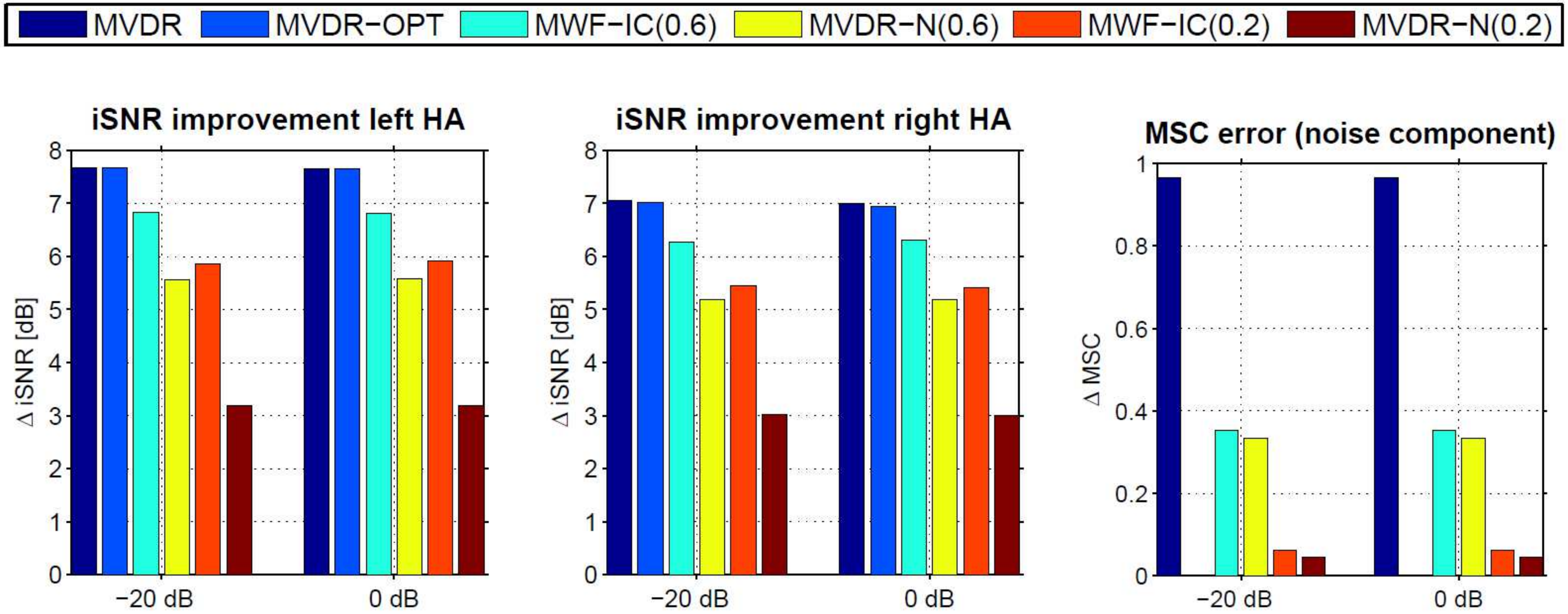
Evaluation: Instrumental measures



- Compared to MVDR, MSC error can be significantly reduced using MWF-IC and MVDR-N, but SNR improvement decreases (MWF-IC > MVDR-N)

Does binaural unmasking compensate for SNR decrease ?

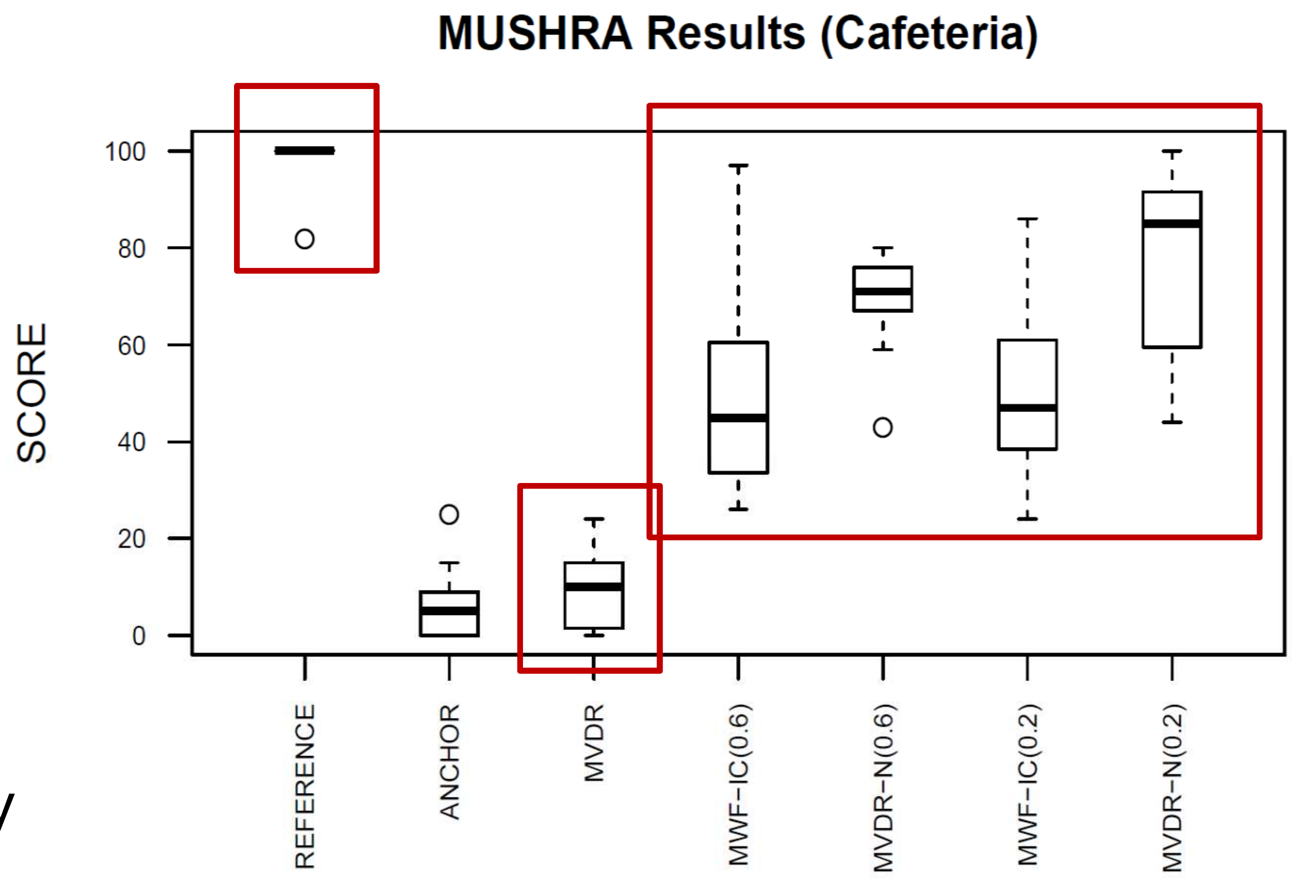
Evaluation: Audio Examples



Input	MVDR	MVDR-OPT	MWF-IC (0.6)	MVDR-N (0.6)	MWF-IC (0.2)	MVDR-N (0.2)

Evaluation: Spatial quality (MUSHRA)

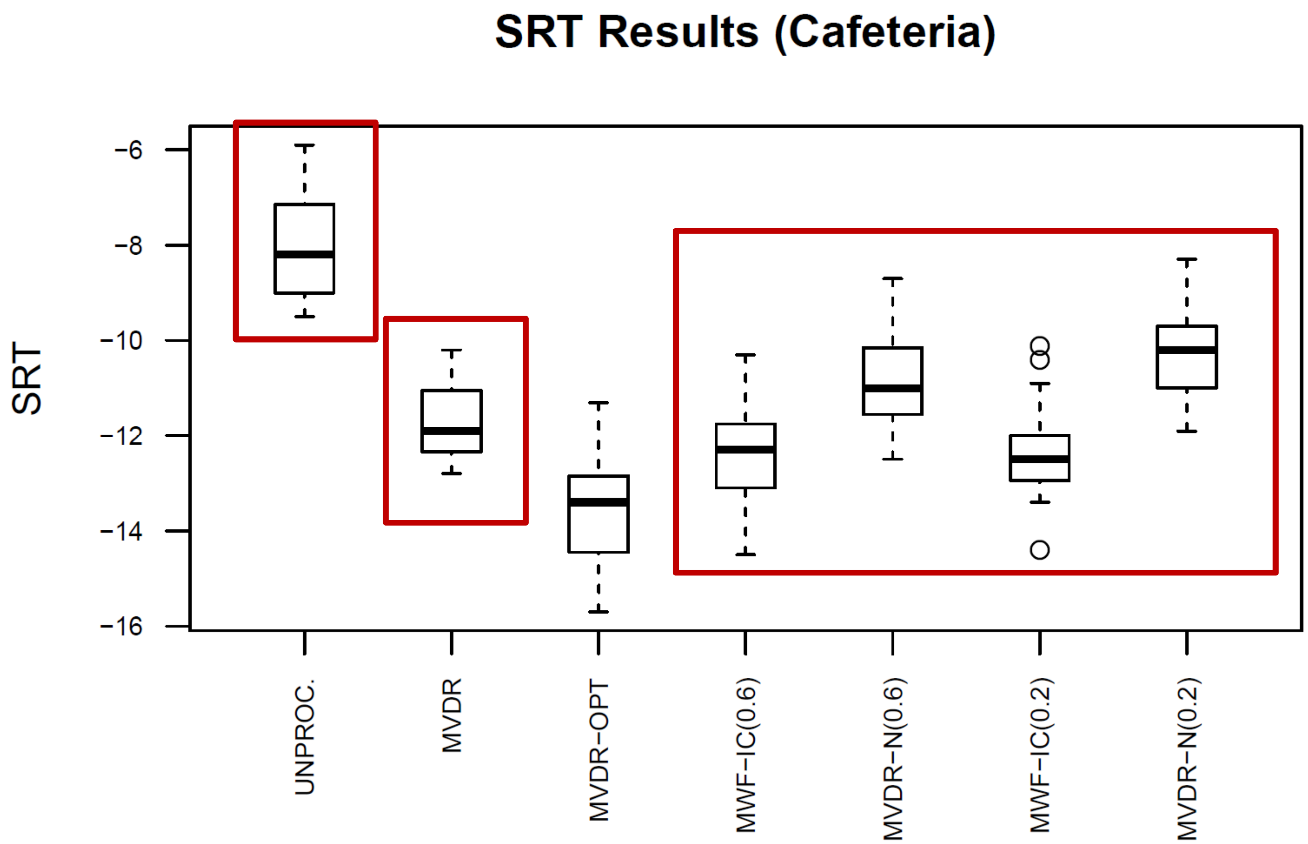
- Evaluate spatial difference between reference and output signal
- **MWF-IC and MVDR-N outperform MVDR**
    - MVDR-N shows better results than MWF-IC
    - Decreasing the MSC threshold slightly improves spatial quality



Binaural cue preservation for diffuse noise improves spatial quality

Evaluation: Speech intelligibility (SRT)

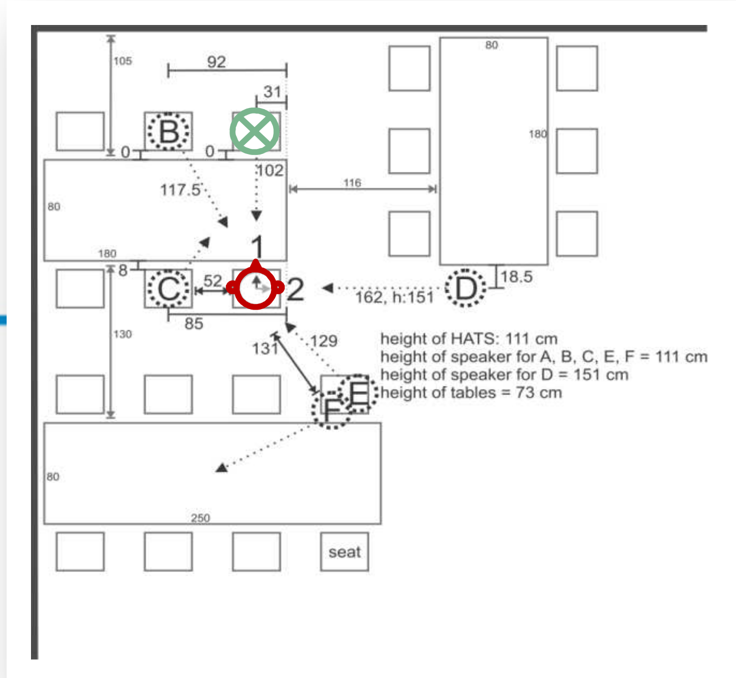
- All algorithms show a highly significant SRT improvement
- The SRT results mainly reflect the SNR differences between algorithms: MWF-IC outperforms MVDR-N
- **No significant SRT difference between MVDR and MWF-IC**



Binaural cue preservation for diffuse noise does not/hardly affect speech intelligibility

Evaluation Study 2 (CI)

Test Environment



- Cafeteria environment
  - T_{60} = 1250ms
- Speech material: Oldenburg sentence test (OLSA)
- Noise scenarios:
  - OlNoise (**OLN**):
    - Speech shaped noise, stationary, omnidirectional
  - 20 Talker-Babble (**20T**):
    - 20 Male talkers, EUROM1 speech corpus, omnidirectional
  - Cafeteria Ambient Noise (**CAN**):
    - KEMAR recording at UniOl cafeteria, omnidirectional
  - Single interfering talker (**SCT**):
    - 1 Male talker, EUROM1 speech corpus, directional (90°)

Algorithms

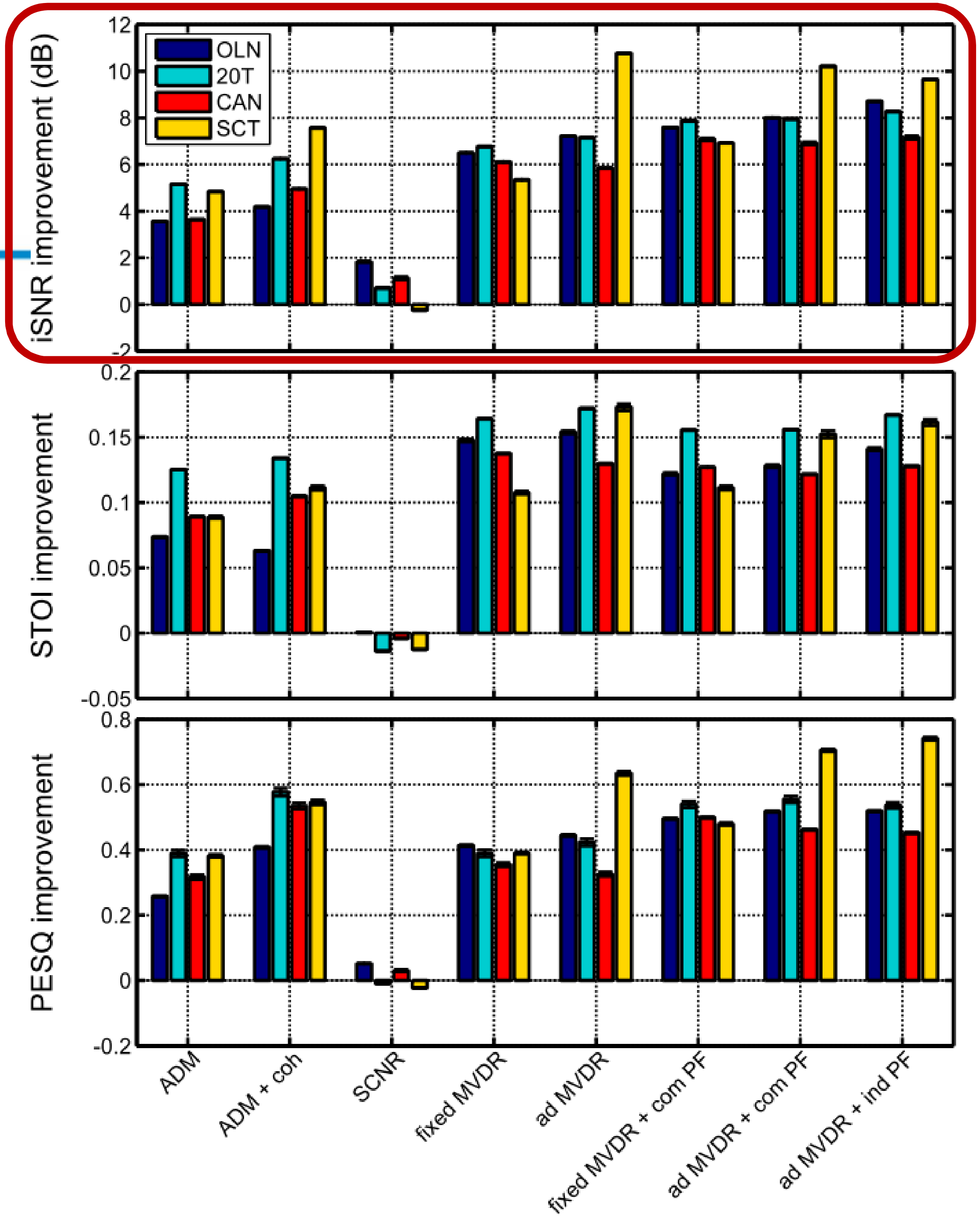
#	Abbreviation	Algorithm	monaural / binaural
1	NoPre	no preprocessing	
2	ADM	adaptive differential microphones	monaural
3	ADM + coh	adaptive differential microphone + coherence-based postfilter* (paradigm 1)	binaural
4	SCNR	single channel noise reduction **	monaural
5	fixed MVDR	fixed binaural MVDR beamformer (paradigm 2)	binaural
6	ad MVDR	adaptive binaural MVDR beamformer (paradigm 2)	binaural
7	fixed MVDR + com PF	common postfilter based on fixed binaural MVDR beamformer (paradigm 1)	binaural
8	ad MVDR + com PF	common postfilter based on adaptive binaural MVDR beamformer (paradigm 1)	binaural
9	ad MVDR + ind PF	Individual postfilter based on adaptive binaural MVDR beamformer (paradigm 1)	binaural

* Grimm et al. (2009)

** Gerkmann and Hendriks (2012), Breithaupt et al. (2008)

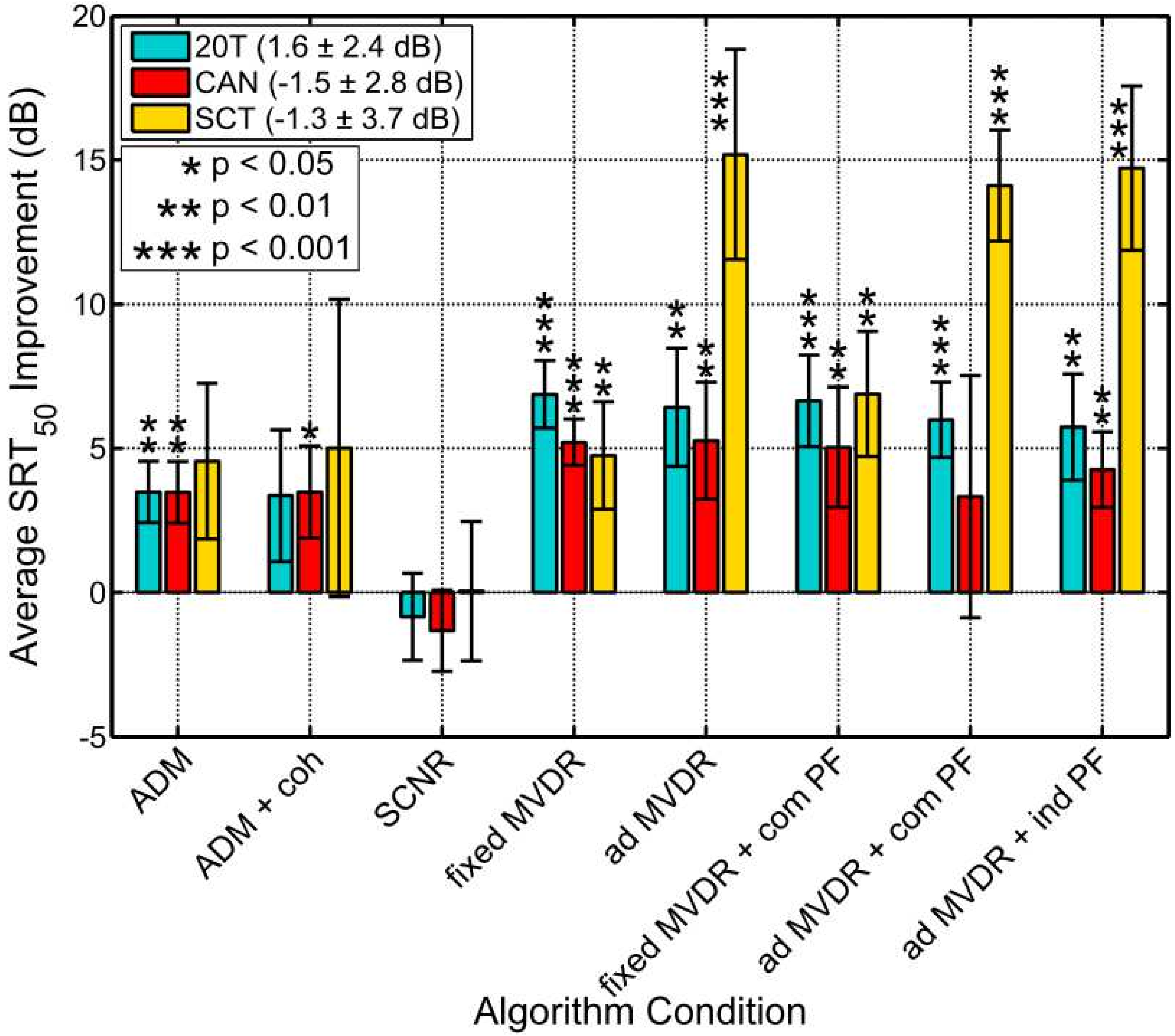
Instrumental Evaluation

- **Instrumental measures:**
  - Intelligibility weighted signal to noise ratio (iSNR)
  - STOI (speech intelligibility)
  - PESQ (speech quality)
- Input SNR: 0 dB
- **Results:**
  - SCNR: minor improvement
  - (Binaural) adaptive MVDR better than (bilateral) ADM
  - Adaptive MVDR similar to fixed MVDR, much better for single-competing-talker (SCT)
  - Post-filtering seems to improve quality (PESQ) but not intelligibility (STOI)
  - Individual post-filter slightly better as common post-filter



Perceptual Evaluation - Results

- **8 CI subjects**
- Substantial and significant SRT improvements
- Large inter-individual variability
- **Similar results as instrumental evaluation**
- Binaural MVDR performs especially well
- *Note: large SRT improvement in SCT due to simple scenario and "perfect" assumptions*



Conclusion

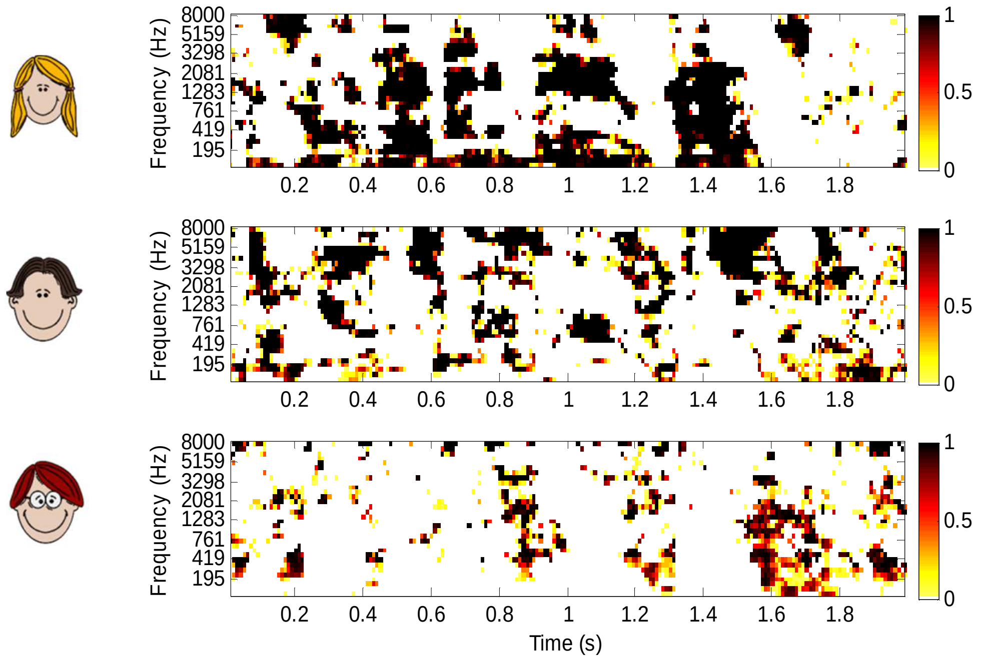
- **Binaural noise reduction algorithms**: 2 main paradigms
  - Spectral post-filtering
  - "True" binaural multi-microphone noise reduction
- **Extensions of binaural MVDR/MWF** for diffuse noise and interfering speaker
- Evaluation of **binaural MVDR extensions for diffuse noise (NH)**
  - Binaural cue preservation improves **spatial quality**
  - Binaural cue preservation does not/hardly affect **speech intelligibility**
  - MVDR-N : best spatial quality, MWF-IC : best SRT
- Evaluation of **several binaural algorithms for different noise conditions (CI)**
  - Binaural algorithms better as bilateral/monaural algorithms
  - No SRT improvement by (common/individual) postfiltering
  - Best performance for binaural adaptive MVDR beamformer

Current/Future work

- Binaural noise reduction algorithms for **interfering sources** (BMVDR-IR, BMVDR-RTF):
  - Subjective evaluation (incl. binaural cue preservation) for HA/CI users
  - Robustness against RTF estimation errors

- **Mixed noise fields and time-varying scenarios**: incorporate computational acoustic scene analysis (CASA) into developed algorithms

- Extend algorithms to include **external microphones (acoustic sensor networks)**



Acknowledgments

- **Funding:**
  - Cluster of Excellence Hearing4All (DFG)
  - Joint Lower-Saxony Israel Project "Acoustic scene aware speech enhancement for binaural hearing aids" (Partner: Bar-Ilan University, Israel)
  - German-Israeli Foundation Project "Signal Dereverberation Algorithms for Next-Generation Binaural Hearing Aids" (Partners: International Audiolabs Erlangen; Bar-Ilan University, Israel)

DFG Deutsche Forschungsgemeinschaft

VolkswagenStiftung

GIF German-Israeli Foundation for Scientific Research and Development

Recent publications

- 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.
- 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.
- 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.
- 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.
- 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.
- R. Baumgärtel, M. Krawczyk-Becker, D. Marquardt, C. Völker, H. Hu, T. Herzke, G. Coleman, K. Adiloglu, S. Ernst, T. Gerkmann, S. Doclo, B. Kollmeier, V. Hohmann, M. Dietz, *Comparing binaural pre-processing strategies I: Instrumental evaluation*, *Trends in Hearing*, vol. 19, pp. 1-16, 2015.
- R. Baumgärtel, H. Hu, M. Krawczyk-Becker, D. Marquardt, T. Herzke, G. Coleman, K. Adiloglu, K. Bomke, K. Plotz, T. Gerkmann, S. Doclo, B. Kollmeier, V. Hohmann, M. Dietz, *Comparing binaural pre-processing strategies II: Speech intelligibility of bilateral cochlear implant users*, *Trends in Hearing*, vol. 19, pp. 1-18, 2015.

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

Questions ?