Recent advances in noise reduction and dereverberation algorithms for binaural hearing aids

Prof. Dr. Simon Doclo

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

Erlangen Kolloquium – February 10, 2017

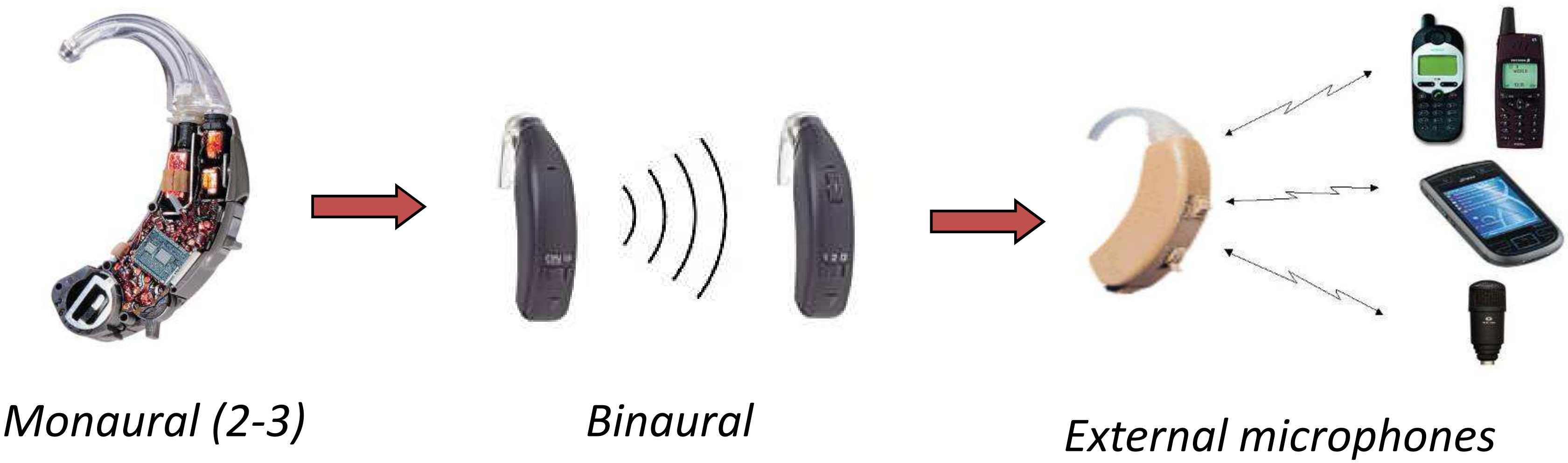
Introduction

- 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

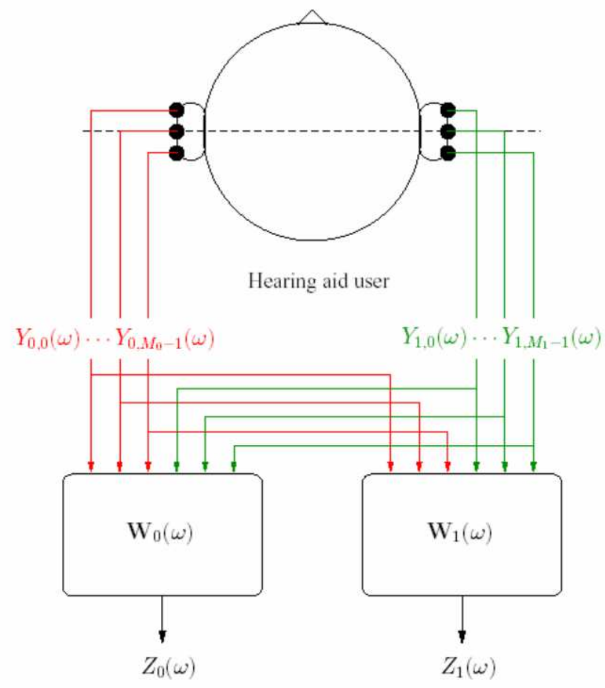
Introduction

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



Introduction

- **This presentation:**
  - Instrumental and subjective evaluation of recent **binaural noise reduction algorithms** based on MVDR/MWF
  - Recent advances in **blind multi-microphone dereverberation algorithms**

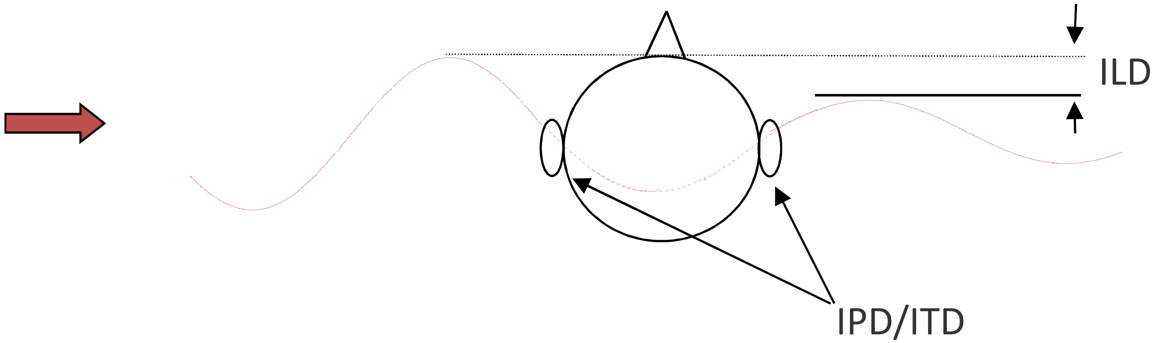


- **Main objectives of algorithms:**
  - Improve speech intelligibility and avoid signal distortions
  - Preserve spatial awareness and directional hearing (binaural cues)

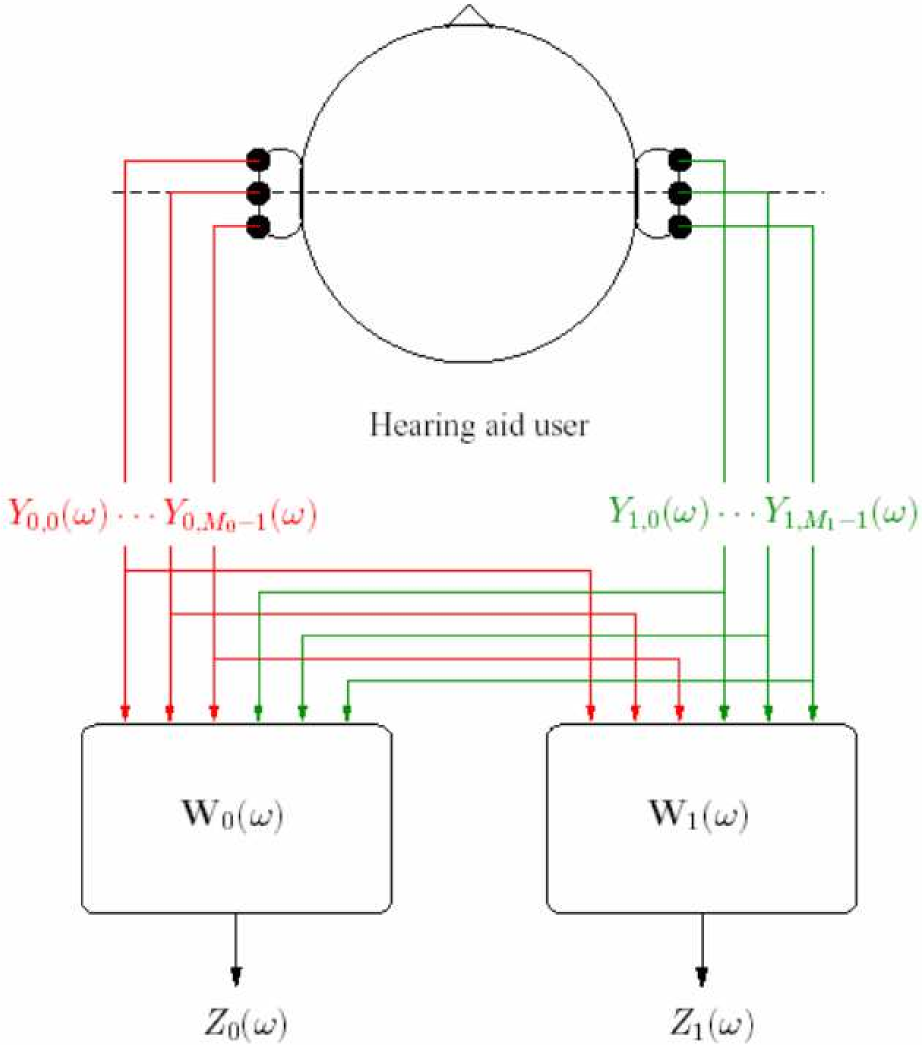
I. Binaural noise reduction

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 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

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

Binaural noise reduction: Acoustic scenario

- 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**

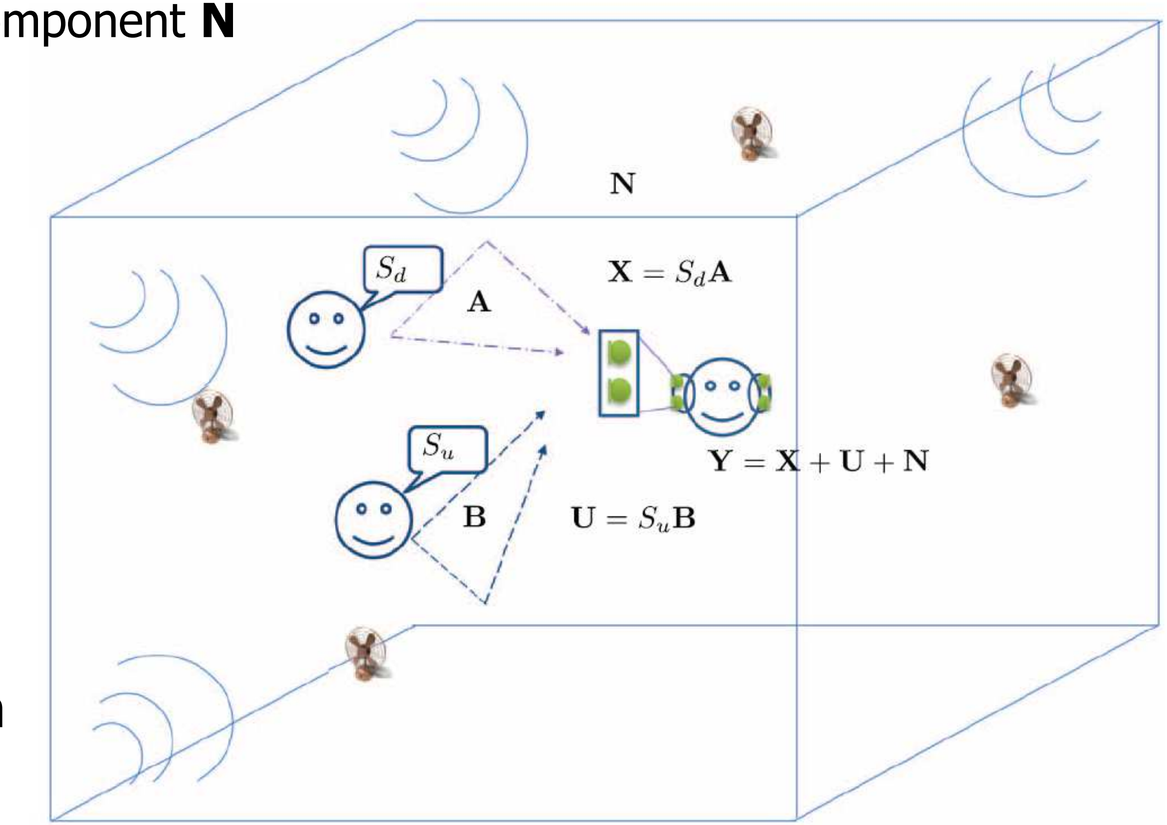
Acoustic Transfer Functions (ATFs)

- Correlation matrices:

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

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

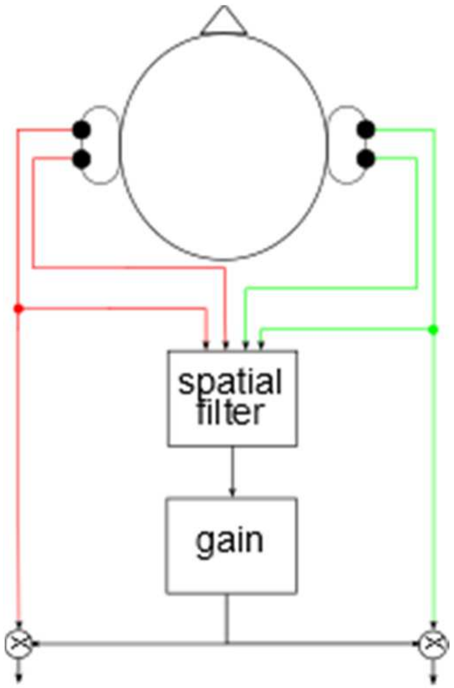
- All **binaural cues** can be written in terms of these matrices



Binaural noise reduction: Two main paradigms

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

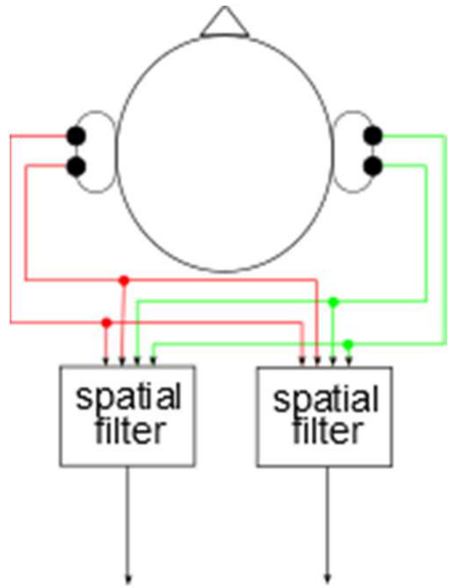
[Dörbecker 1996, Wittkop 2003, Lotter 2006, Rohdenburg 2008, Grimm 2009, Kamkar-Parsi 2011, Reindl 2013, Baumgärtel 2015]



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

Binaural spatial filtering techniques

[Merks 1997, Welker 1997, Aichner 2007, 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

Binaural MVDR and MWF

Minimum-Variance-Distortionless-Response (MVDR) beamformer

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

$$\min_{\mathbf{W}_0} \underbrace{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0}_{\text{noise reduction}} \quad \text{subject to} \quad \underbrace{\mathbf{W}_0^H \mathbf{A} = A_0}_{\text{distortionless constraint}}$$

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

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

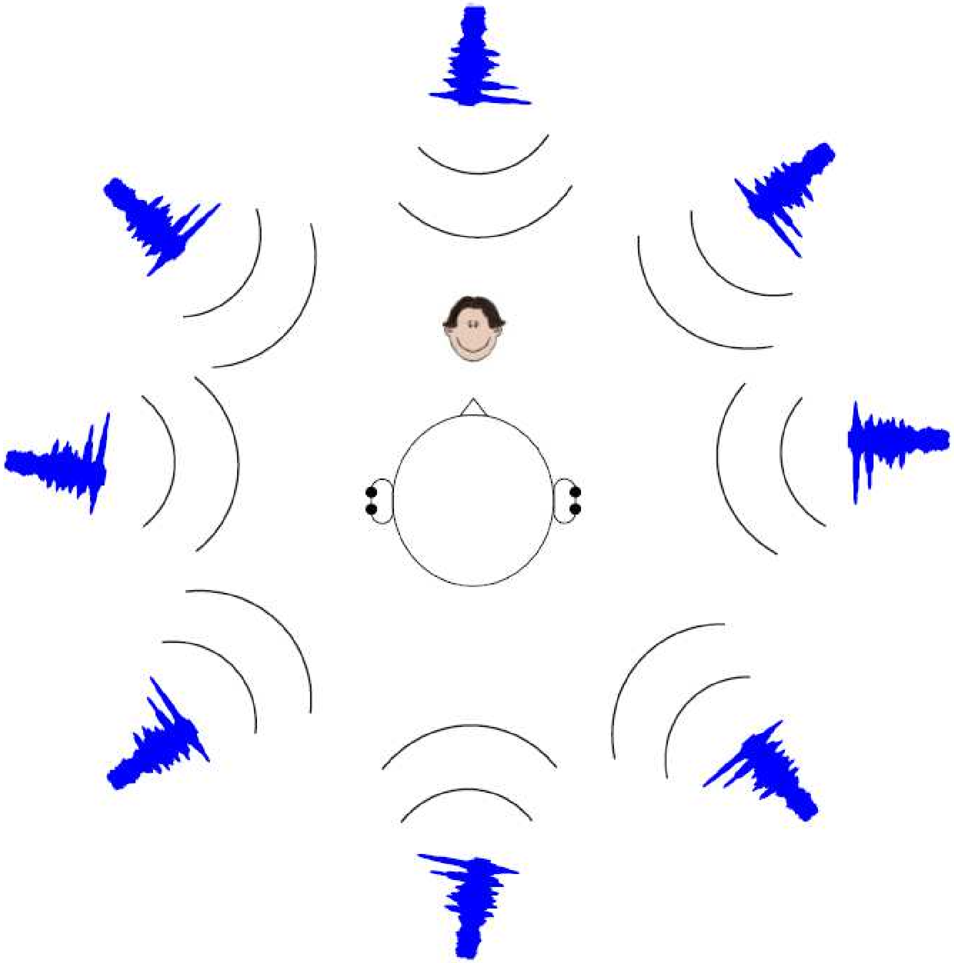
$$J_{\mathrm{MWF}}(\mathbf{W}) = \mathcal{E}\left\{ \underbrace{\left\| \begin{bmatrix} X_0 - \mathbf{W}_0^H \mathbf{X} \\ X_1 - \mathbf{W}_1^H \mathbf{X} \end{bmatrix} \right\|^2}_{\text{speech distortion}} + \mu \underbrace{\left\| \begin{bmatrix} \mathbf{W}_0^H \mathbf{V} \\ \mathbf{W}_1^H \mathbf{V}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

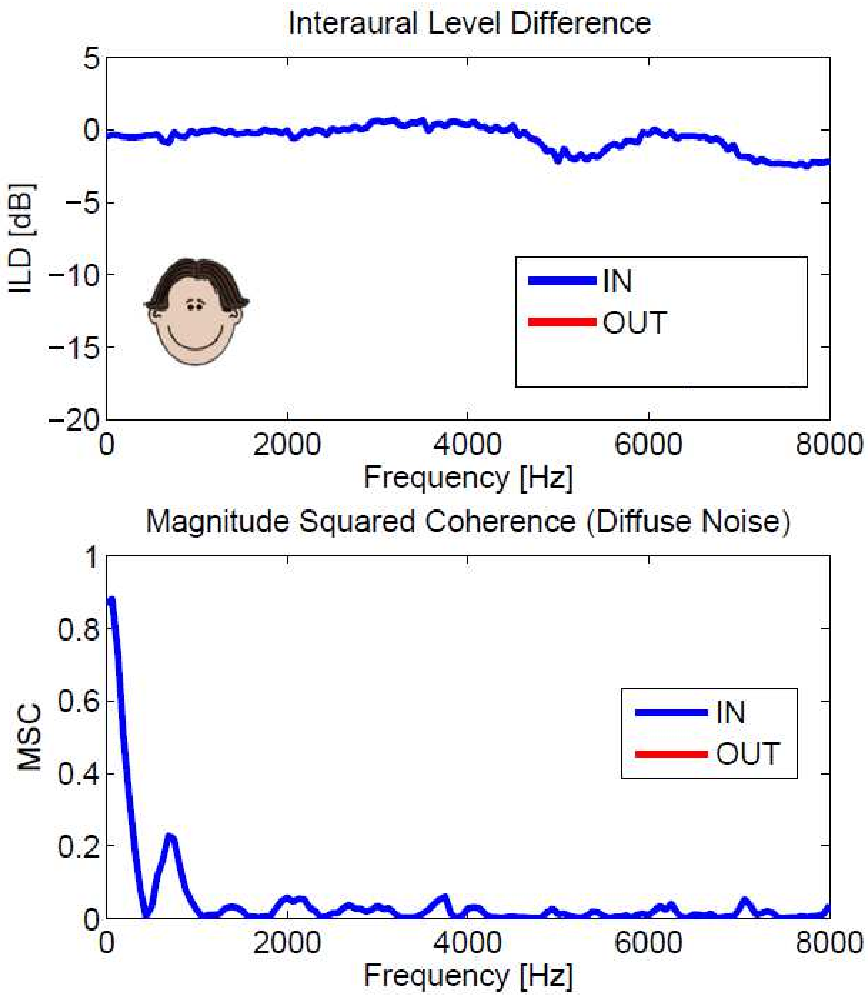
Requires estimate of speech and noise covariance matrices, e.g. based on VAD

Can be decomposed as binaural MVDR beamformer and spectral postfilter

Good noise reduction performance, what about binaural cues ?

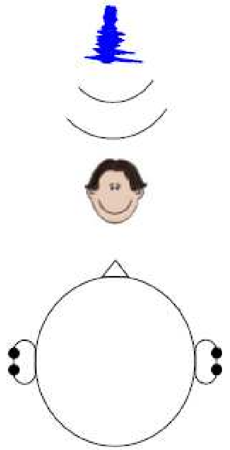
Binaural MVDR and MWF Binaural cues (diffuse noise)

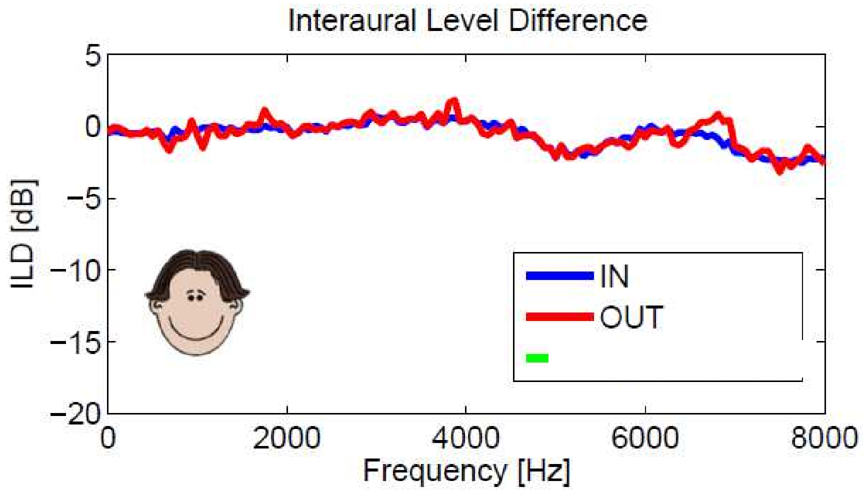


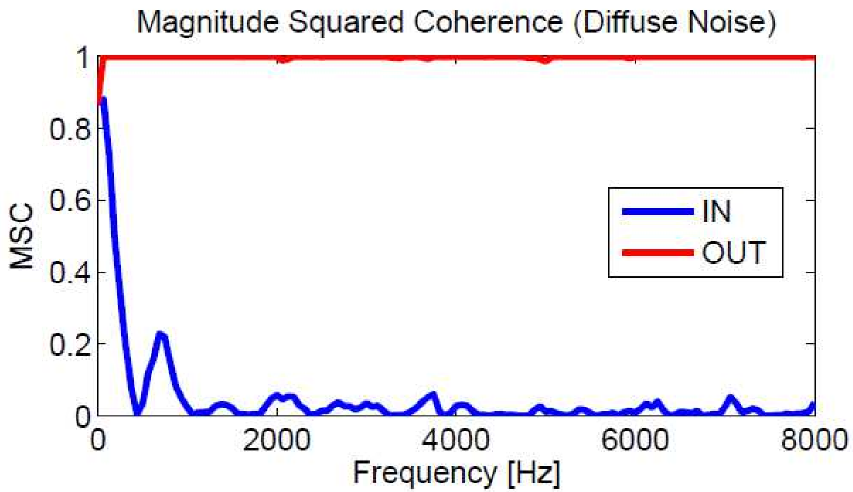


Note: MSC = Magnitude Squared Coherence

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)

$$J_{MWF-IC}(\mathbf{W}) = J_{MWF}(\mathbf{W}) + \lambda \left| \frac{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_1}{\sqrt{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1}} - IC_v^{des} \right|^2$$

⊖ No closed-form solution, iterative optimization procedures required

Partial noise estimation (MWF-N)

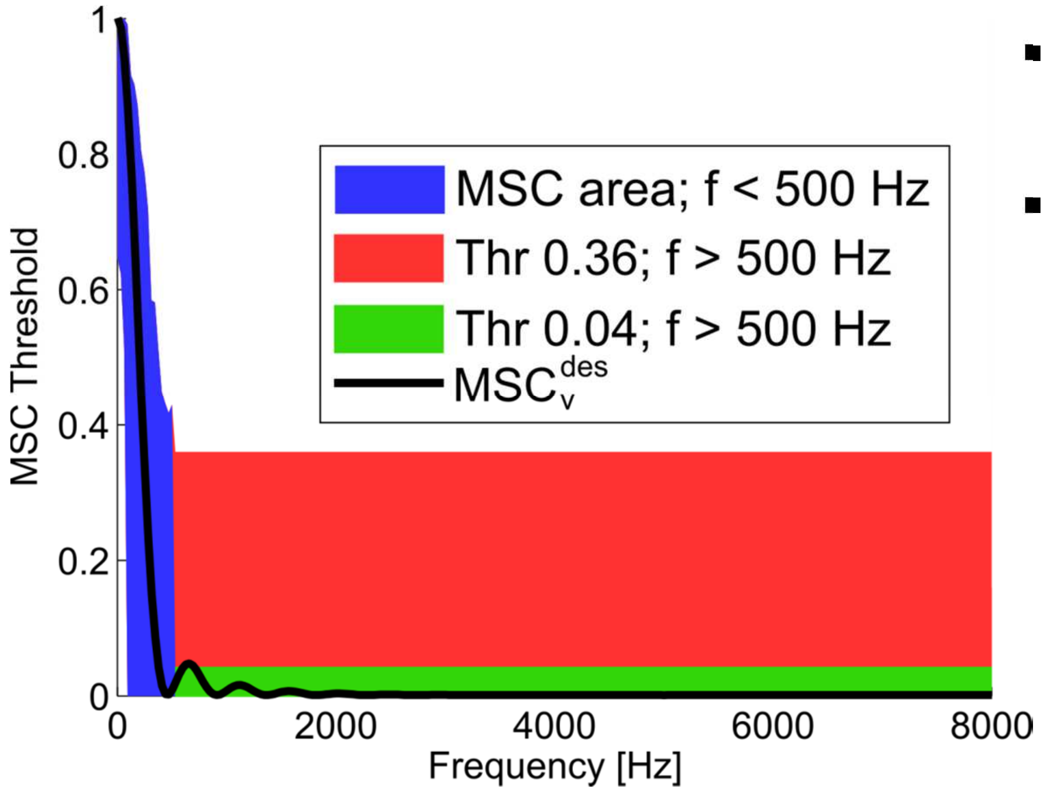
$$J_{\mathrm{MWF-N}}(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_0 - \mathbf{W}_0^H \mathbf{X} \\ X_1 - \mathbf{W}_1^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \eta V_0 - \mathbf{W}_0^H \mathbf{V} \\ \eta V_1 - \mathbf{W}_1^H \mathbf{V} \end{bmatrix} \right\|^2 \right\}$$

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

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

Binaural MWF: Extensions for diffuse noise

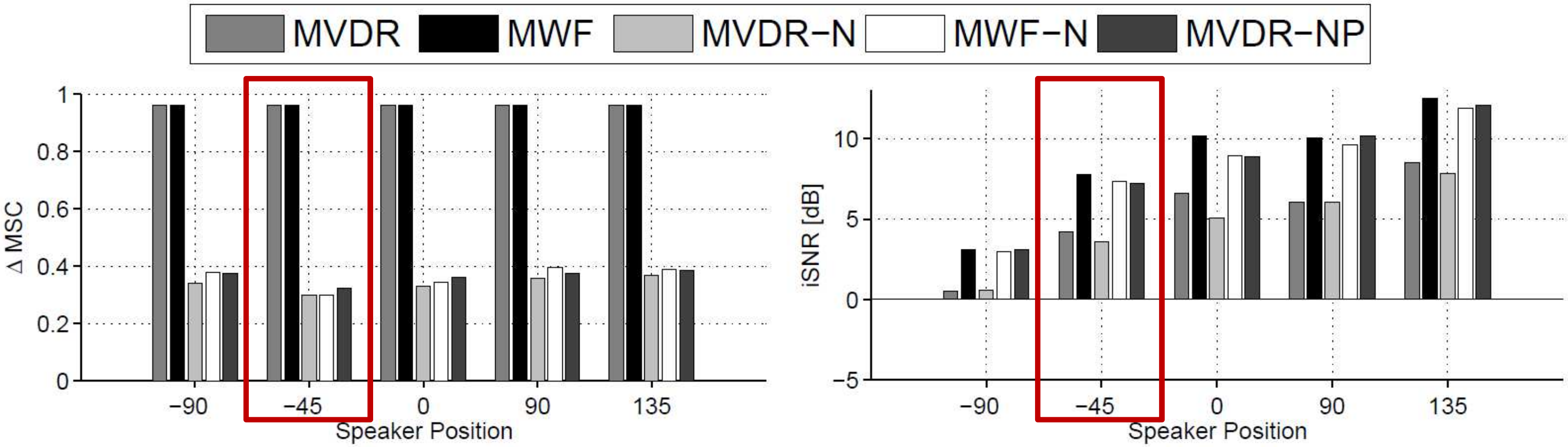
- **Determine (frequency-dependent) trade-off parameters based on psycho-acoustic criteria**
  - Amount of IC preservation based on subjective listening experiments evaluating the IC discrimination abilities of the human auditory system



- 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**)

Binaural MWF: Extensions for diffuse noise

❑ Instrumental evaluation / sound samples



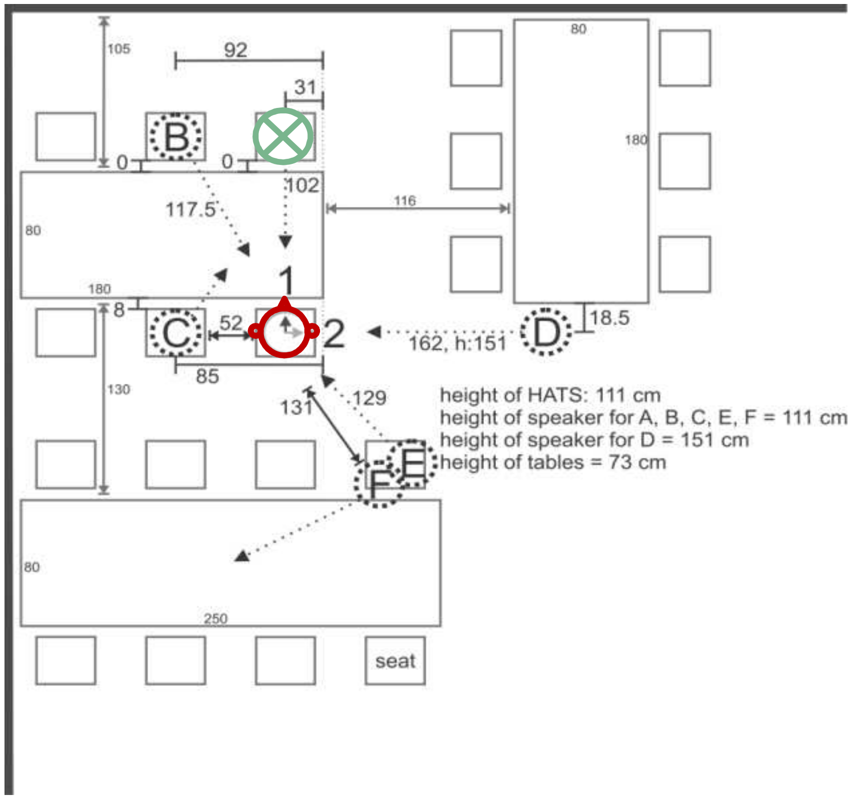
(a) $\Delta\mathrm{MSC_v}$ (b) $\mathrm{iSNR_{be}^{out}}$

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

Office ($T_{60} \approx$ 700ms), M=4 (BRIR), recorded ambient noise, speaker at -45°, 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]

Subjective Evaluation: Test setup

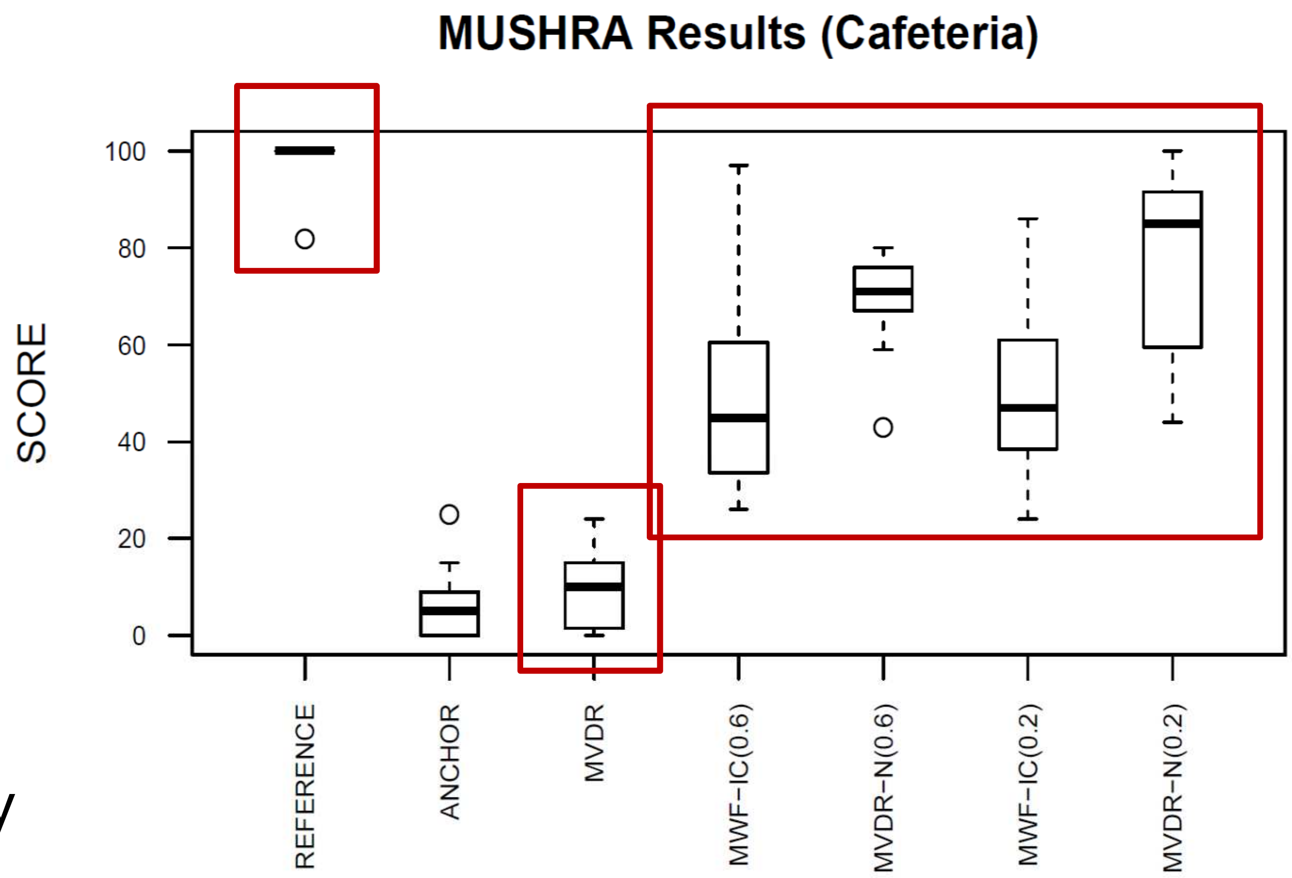


- Binaural hearing aid recordings (M=4 mics) in **cafeteria** (T_{60} ≈ 1250 ms) [Kayser 2009]
- **Noise:** realistic cafeteria ambient noise
- **Algorithms**: binaural MVDR + cue preservation extensions (MWF-IC, MVDR-N) with different MSC boundaries
- **Subjective listening experiments**:
  - 15 normal-hearing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

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

Subjective Evaluation: Spatial quality (MUSHRA)

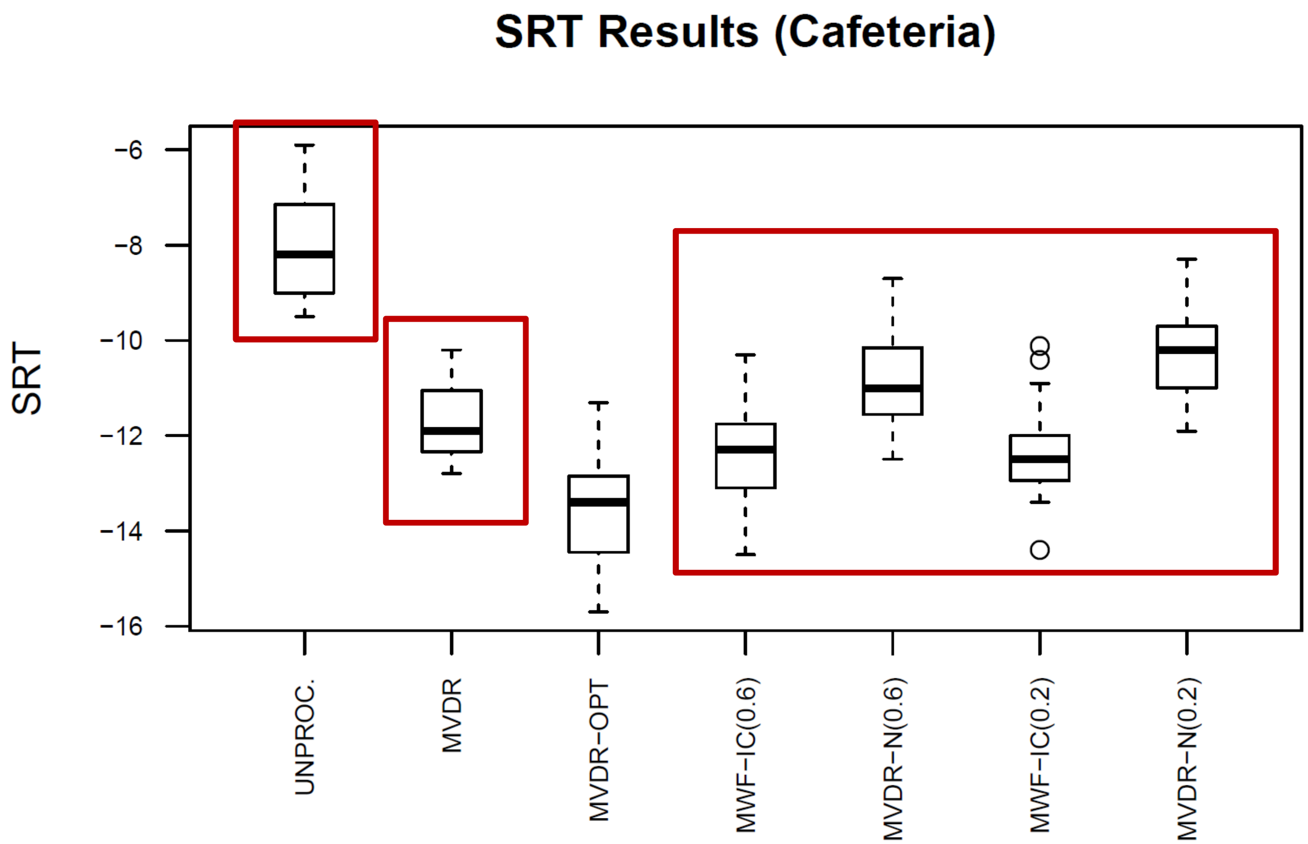
- Evaluate spatial difference between reference and output signal
- **MWF-IC and MVDR-N outperform MVDR**
  - MVDR-N shows better results than MWF-IC
  - Decreasing the MSC threshold slightly improves spatial quality



Binaural cue preservation for diffuse noise improves spatial quality

Subjective 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

Binaural noise reduction

Extensions for interfering sources

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)

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

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

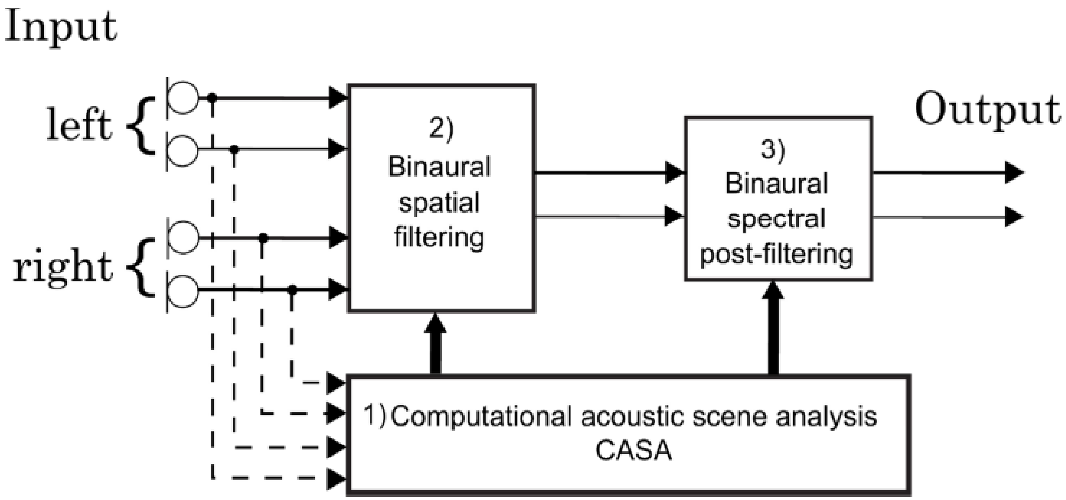
- ⊕ Binaural cues of speech source **and** interfering source preserved
- ⊕ Also binaural MWF-based versions (incl. spectral filtering) can be derived
- ⊖ Background noise: MSC not exactly preserved, possible noise amplification

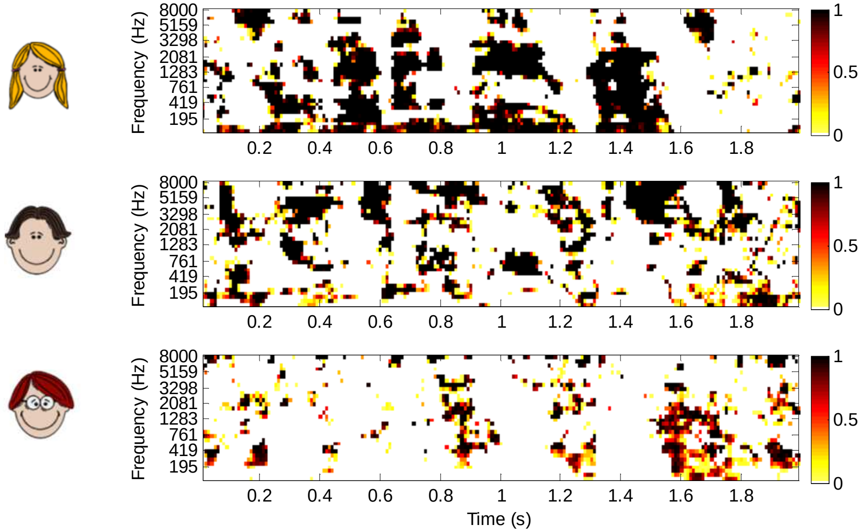
[Hadad 2014/2015/2016, Marquardt 2014/2015]

Current research: Integration with CASA

- For all discussed binaural noise reduction and cue preservation algorithms **several quantities need to be estimated**:
  - Steering vector (RTF/DOA) of desired source (and interfering sources)
  - Correlation matrix of background noise

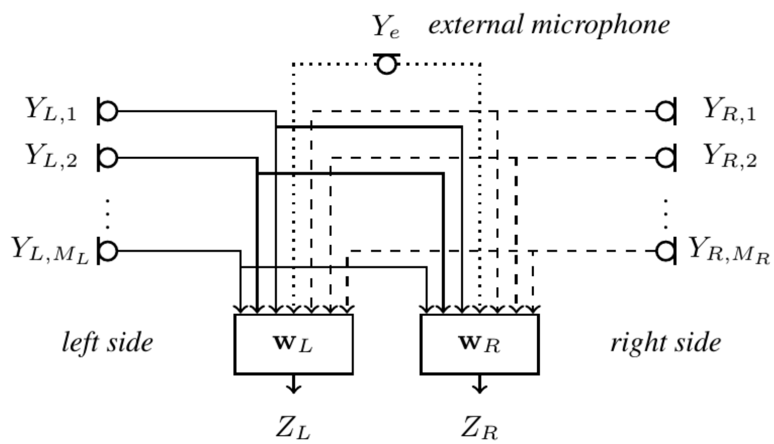
- Non-trivial task for **complex and time-varying acoustic scenarios** → **integration with computational acoustic scene analysis (CASA)** in the control path of speech enhancement algorithms



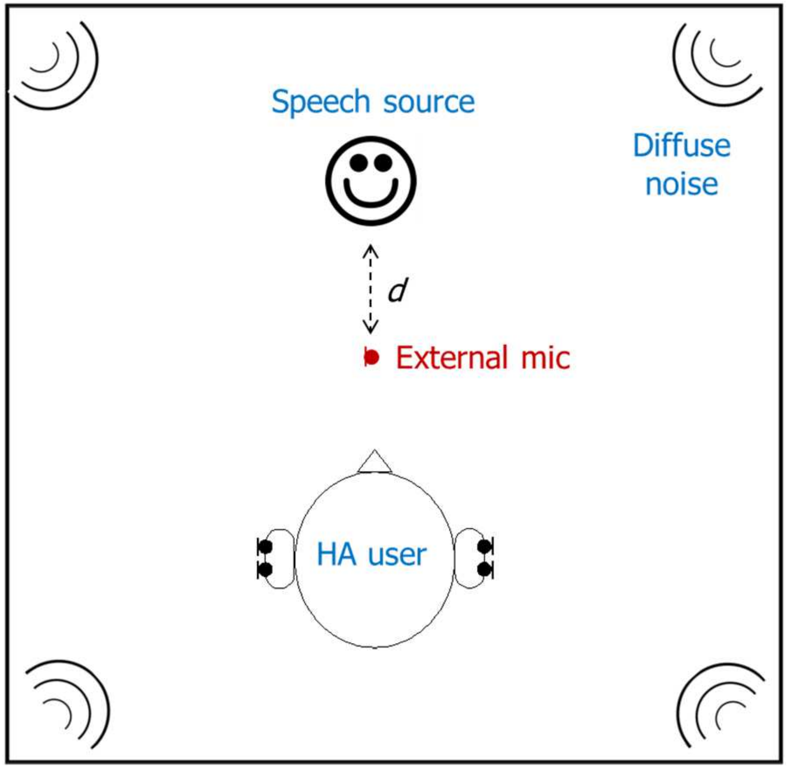


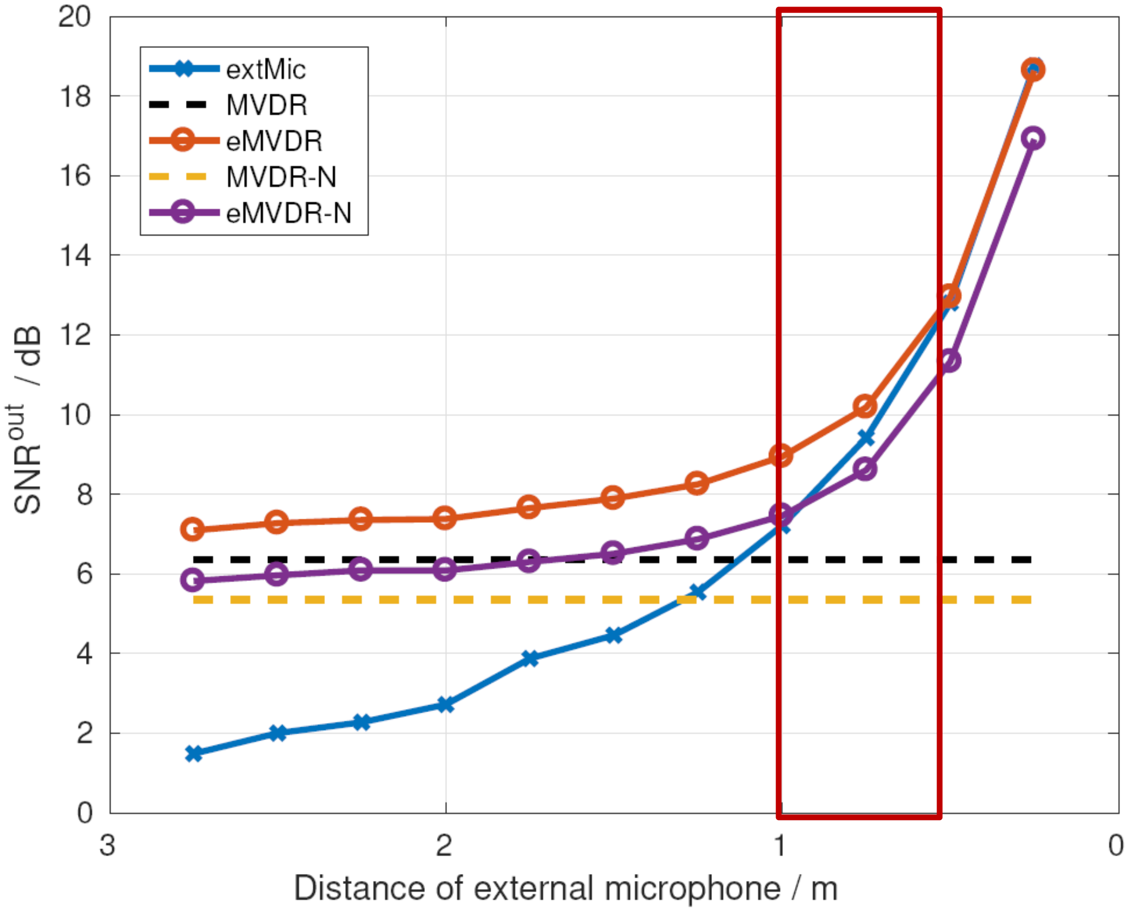
Current research: External microphone(s)

- Exploit the availability of one or more external microphones **(acoustic sensor network)** with hearing aids [Bertrand 2009, Yee 2016]
- **Objective:** improve noise reduction and/or binaural cue preservation performance
- For **binaural MVDR-N beamformer** with external microphone: trade-off between noise reduction performance and binaural cue preservation for
  - Interfering source [Szurley, 2016]
  - Diffuse noise [Gößling, 2017]



Current research: External microphone(s)





- Using external microphone may lead to **significant SNR improvement**
- **eMVDR-N is able to preserve binaural cues of both speech source + residual noise**

[Gößling, HSCMA 2017]

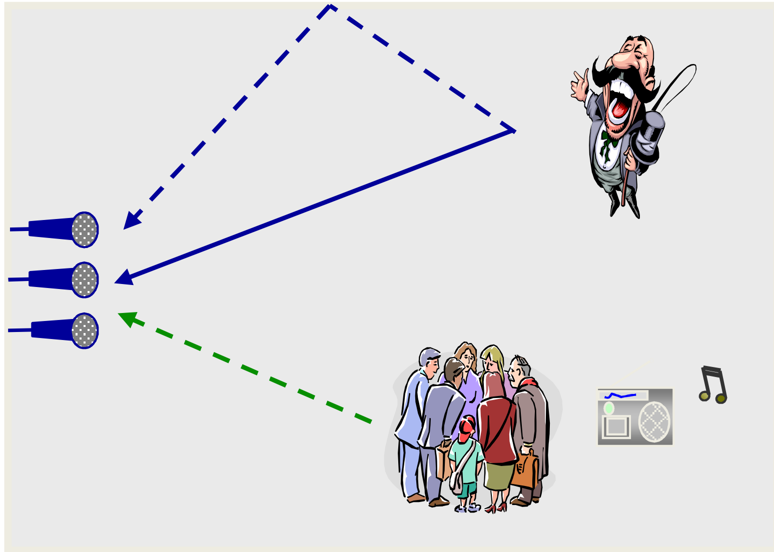
Summary

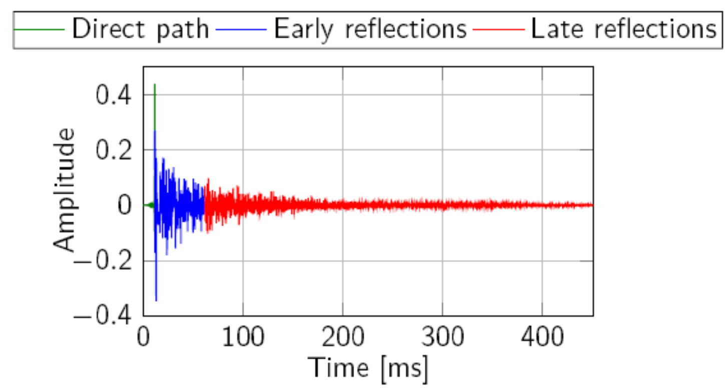
- **Binaural noise reduction algorithms**: 2 main paradigms
  - Spectral post-filtering
  - "True" binaural spatial filtering
- **Extensions of binaural MVDR/MWF** for diffuse noise and interfering speaker, preserving binaural cues of residual noise/interference
- Evaluation of **binaural MVDR extensions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

II. Joint dereverberation and noise reduction

Dereverberation and noise reduction

- **Problem**
  - Noise and reverberation jointly present in typical acoustic environments
  - Speech quality and intelligibility degradation
  - Performance degradation of ASR systems
- **Objectives**
  - Single- and multi-channel joint noise reduction and dereverberation algorithms
  - Exploit **knowledge / statistical models of room acoustics and speech signals**
- **Approaches**
  1. Single- and multi-microphone **spectral enhancement**
  2. **Multi-channel linear prediction**: probabilistic estimation using statistical model of desired signal

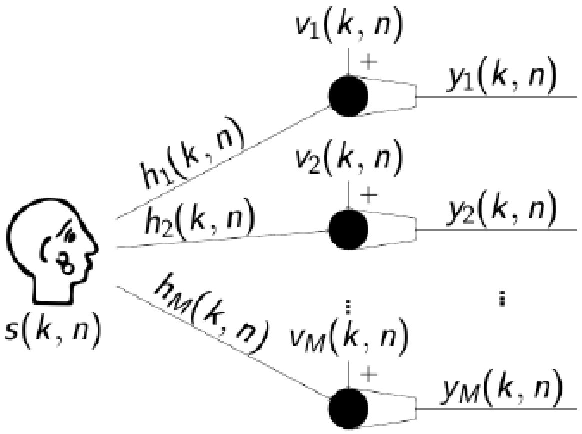


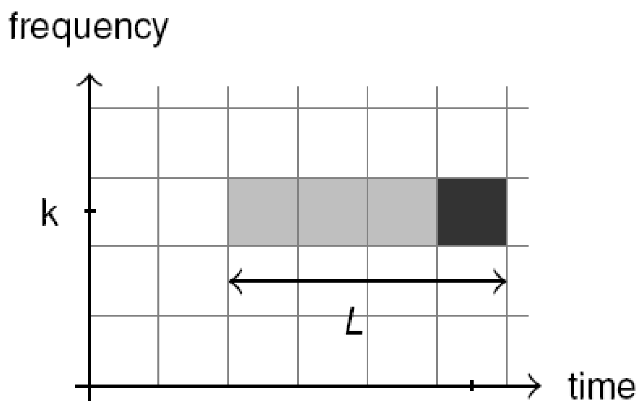


Dereverberation and noise reduction

- **Scenario:** speech source in noisy and reverberant environment, *M* microphones
- **STFT-domain:**
  - approximation of time-domain convolution using convolutive transfer function (CTF)

$$y_m(k,n) = \underbrace{h_m(k,n) * s(k,n)}_{x_m(k,n)} + v_m(k,n)$$



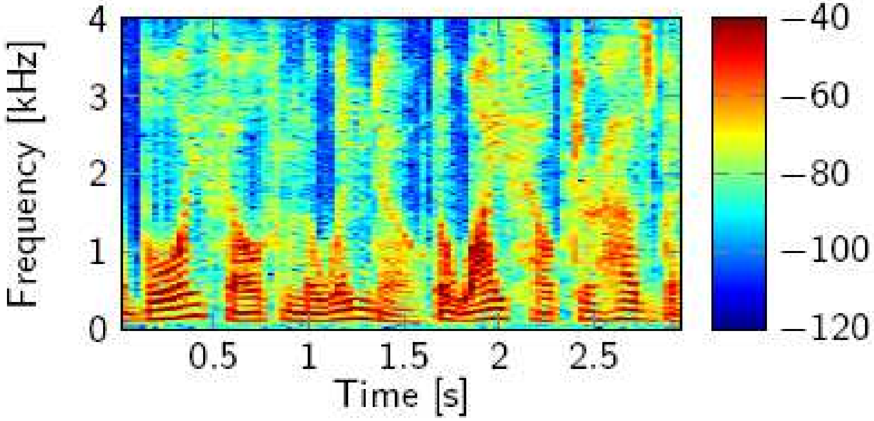


Dereverberation and noise reduction

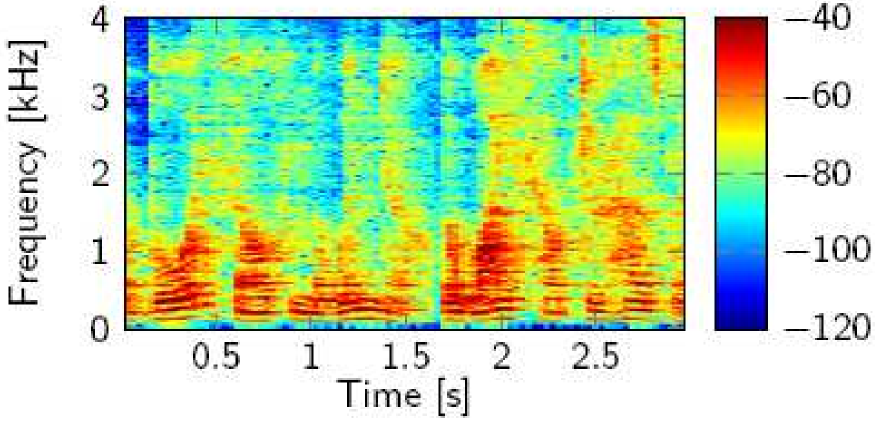
- **Scenario:** speech source in noisy and reverberant environment, *M* microphones
- **STFT-domain:**
  - approximation of time-domain convolution using convolutive transfer function (CTF)

$$y_m(k,n) = \underbrace{h_m(k,n) * s(k,n)}_{x_m(k,n)} + v_m(k,n)$$

  - clean speech is **more sparse** than reverberant speech



Clean



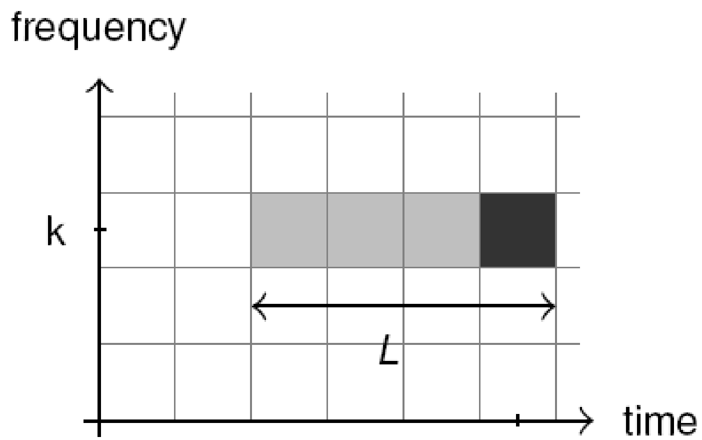
Reverberant

Dereverberation and noise reduction

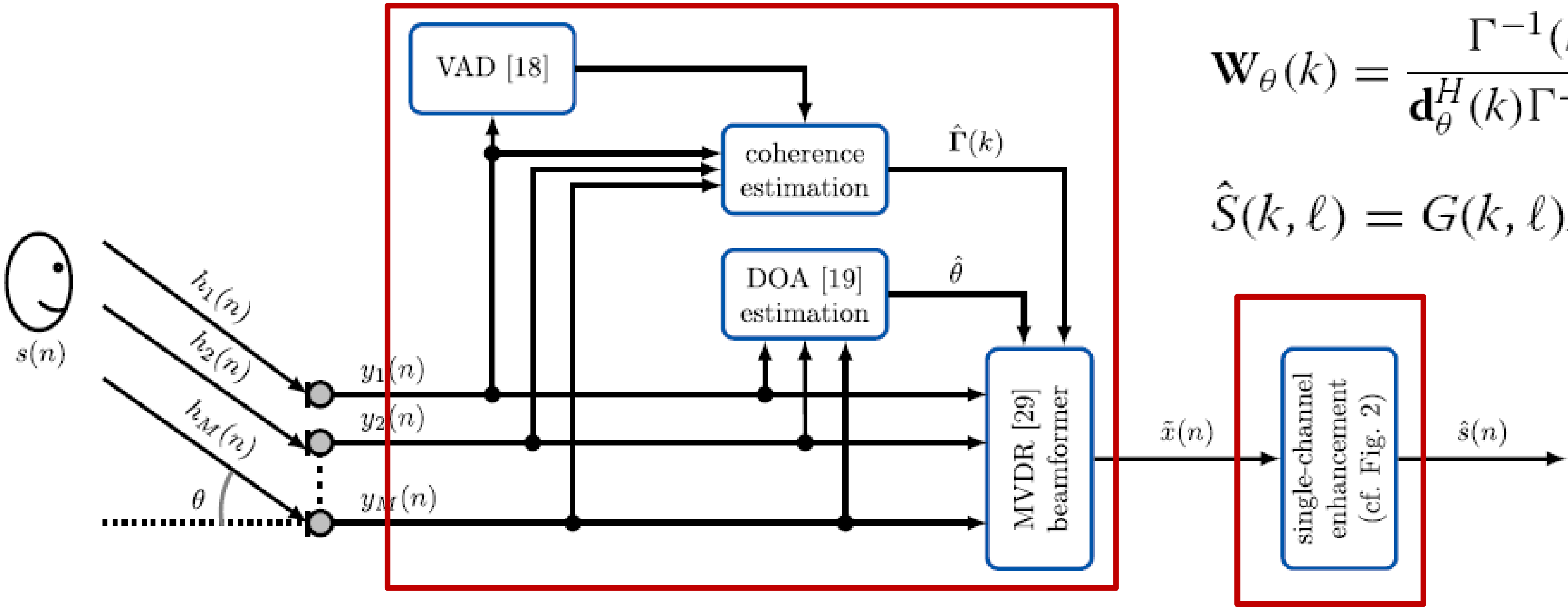
- **Scenario:** speech source in noisy and reverberant environment, *M* microphones
- **STFT-domain:**
  - approximation of time-domain convolution using convolutive transfer function (CTF)

$$y_m(k,n) = \underbrace{h_m(k,n) * s(k,n)}_{x_m(k,n)} + v_m(k,n)$$

  - clean speech is **more sparse** than reverberant speech
- **Dereverberation methods:**
  - **Spatial filtering / beamforming**
  - **Spectral enhancement:** apply real-valued gain to each time-frequency bin
  - **Reverberation suppression:** subtract (complex-valued) estimate of late reverberant component



1. Beamforming + spectral post-filtering



$$\mathbf{W}_\theta(k) = \frac{\Gamma^{-1}(k)\mathbf{d}_\theta(k)}{\mathbf{d}_\theta^H(k)\Gamma^{-1}(k)\mathbf{d}_\theta(k)}$$

$$\hat{S}(k,\ell) = G(k,\ell)\tilde{X}(k,\ell),$$

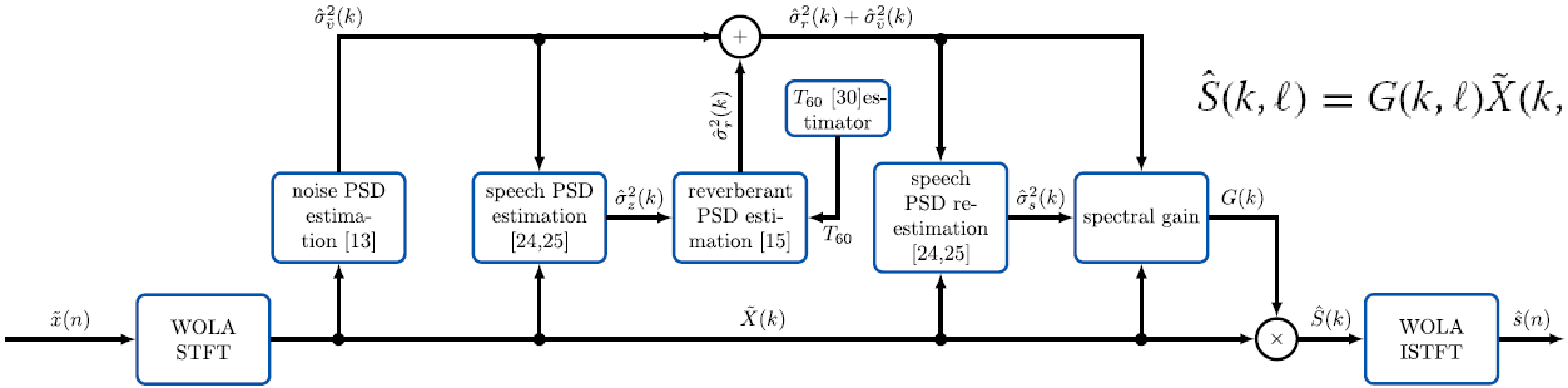
- **MVDR beamformer,** requiring assumption about spatial coherence of late reverberation + direction-of-arrival (DOA) estimate of speech source
- **Spectral post-filter: estimate of late reverberant PSD**
  - **Single-channel estimator,** requiring estimate of reverberation time T_{60}
  - **Multi-channel estimator**, requiring assumption about spatial coherence of late reverberation (+ DOA estimate of speech source)

[Cauchi et al., JASP 2015]

1. Beamforming + spectral post-filtering

- **Spectral post-filter: single-channel estimator**
  1. **Noise PSD**: minimum statistics approach (longer window as usual)
  2. **Reverberant speech PSD**: ML estimate + cepstro-temporal smoothing
  3. **Late reverberant PSD**: assuming exponential decay (requiring T60 estimate)
  4. **Clean speech PSD**: ML estimate + cepstro-temporal smoothing

$$\hat{\sigma}_r^2(k,\ell) = e^{-2\Delta T_d f_s} \hat{\sigma}_z^2(k,\ell - T_d/T_s)$$

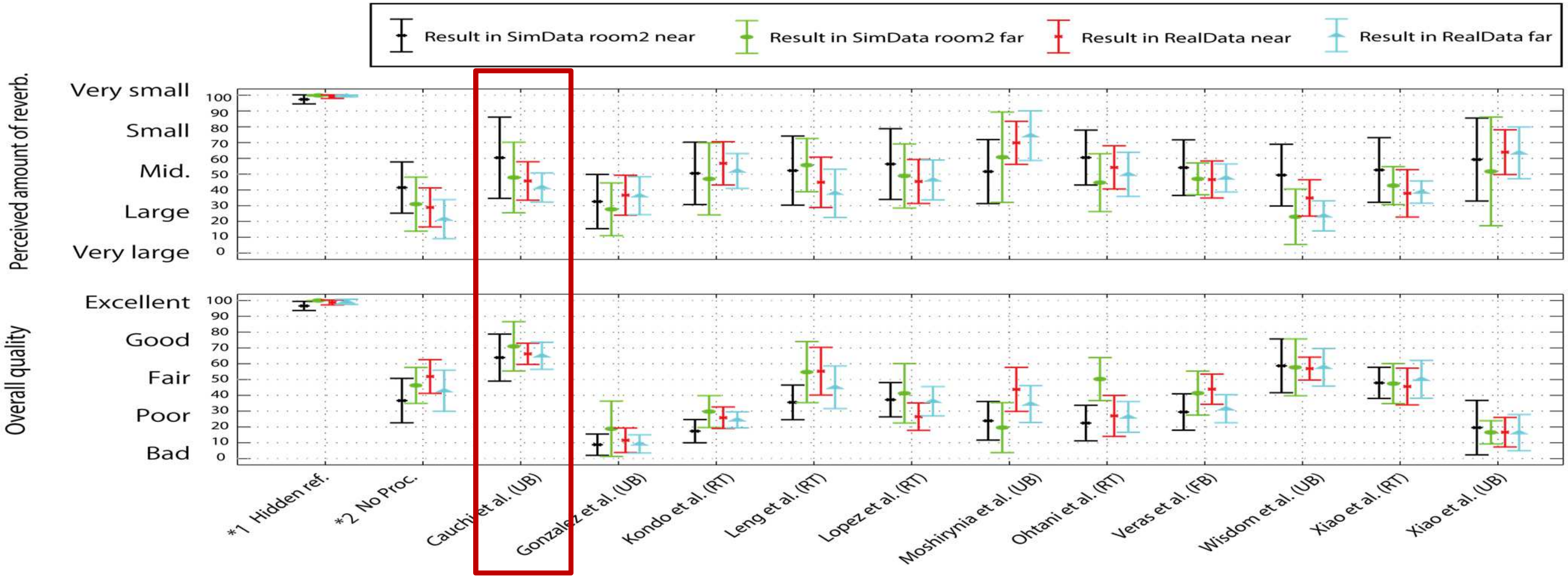


$$\hat{S}(k,\ell) = G(k,\ell)\tilde{X}(k,\ell),$$

1. Beamforming + spectral post-filtering

- **Subjective evaluation (evaluation set of REVERB challenge)**

Anchor | Unprocessed | SE_3 | MVDR | MVDR+SE_3



Circular array (M=8, d = 20 cm), fs = 16 kHz, SNR = 20 dB; S2: T60 = 500 ms (0.5m, 2m), R1: T60 = 700 ms (1m, 2.5m)
STFT: 32 ms, 50% overlap, Hann; MVDR: WNGmax = -10 dB; Postfilter: β=0.5, μ=0.5, Gmin = -10dB, Td = 80 ms, MS window = 3s

[Cauchi et al., JASP 2015] [Cauchi et al., REVERB 2015]

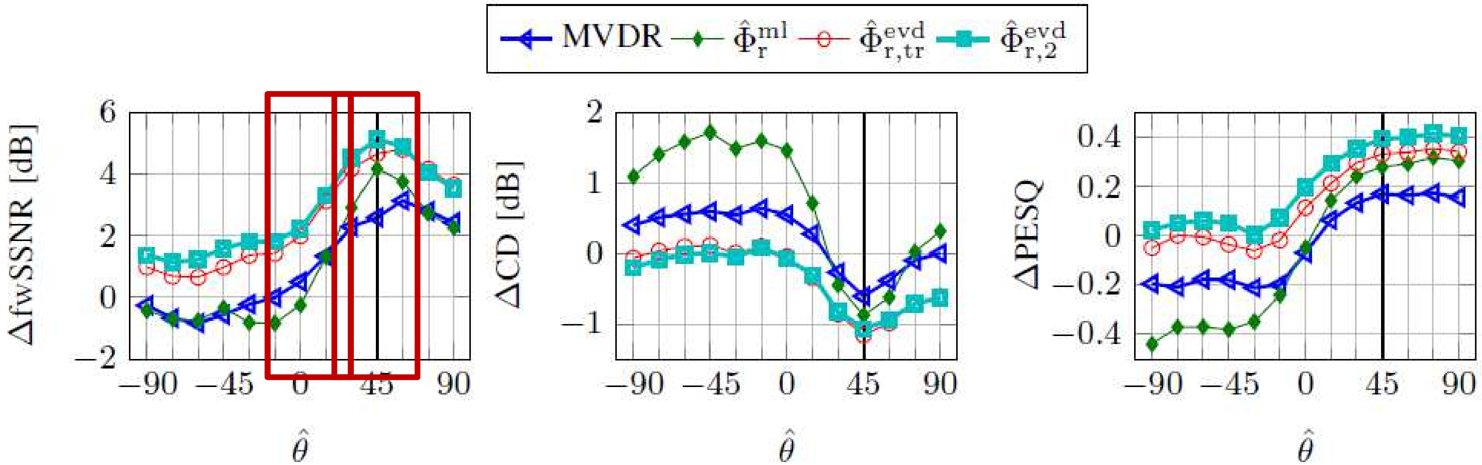
1. Beamforming + spectral post-filtering

Spectral post-filter: multi-channel estimator

- Requires assumption about spatial coherence Γ of late reverberant sound field, e.g. spherically isotropic (diffuse)
- Different estimators have been recently proposed:
  - ML estimator, requiring DOA estimate of speech source [Braun 2013, Kuklasinksi 2016]
  - Estimator based on eigenvalue decomposition, **not** requiring DOA estimate of speech source

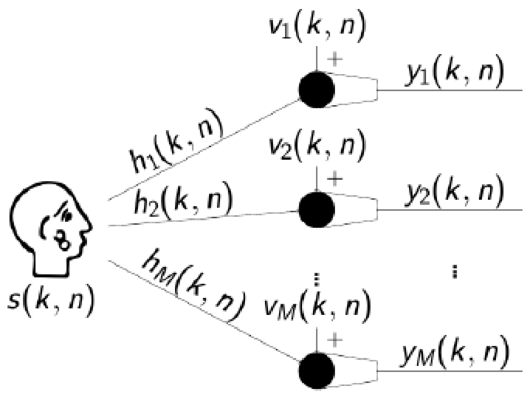
$$\hat{\Phi}_{\mathrm{r}}^{\mathrm{evd}} = \lambda_2\{\mathbf{\Phi}_{\mathrm{x}}\mathbf{\Gamma}^{-1}\} = \cdots = \lambda_M\{\mathbf{\Phi}_{\mathrm{x}}\mathbf{\Gamma}^{-1}\} = \frac{1}{M-1}\left(\mathrm{tr}\left\{\mathbf{\Phi}_{\mathrm{x}}\mathbf{\Gamma}^{-1}\right\} - \lambda_1\left\{\mathbf{\Phi}_{\mathrm{x}}\mathbf{\Gamma}^{-1}\right\}\right)$$

- Robustness against DOA estimation errors (M=4, T_{60}=610 ms, θ=45°)



2. Multi-channel linear prediction

- **Direct STFT-based approach:**
  - directly estimate clean speech STFT coefficients *s(k,n)* from reverberant (and noisy) STFT coefficients $y_m(k,n)$
  - Speech properties (e.g., sparsity) can be modelled naturally in STFT-domain
  - Low computational complexity



$$y_m(k,n) = \underbrace{h_m(k,n) * s(k,n)}_{x_m(k,n)} + v_m(k,n)$$

1. *Using convolutive transfer function (CTF) model*
2. Transform to equivalent AR model → **multi-channel linear prediction (MCLP)**

$$x_1(k,n) = d(k,n) + \sum_{m=1}^{M} \sum_{l=0}^{L_g-1} g_m(k,l) x_m(k, n-\tau-l)$$

clean signal (incl. early reflections) **prediction filters** **delay (early reflections)**

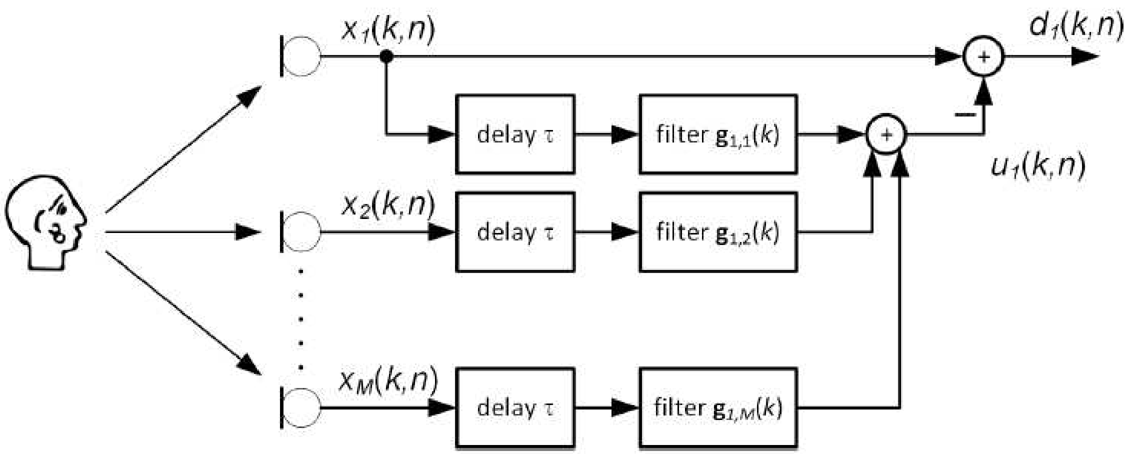
2. Multi-channel linear prediction

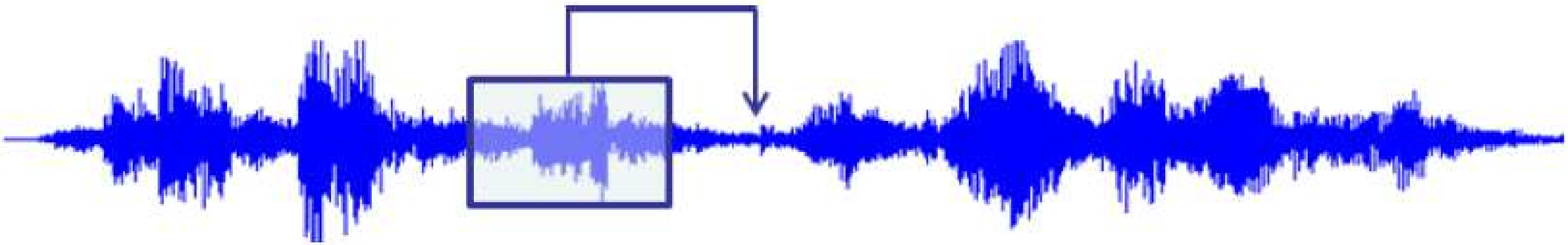
- **AR model of reverberant speech**

$$\mathbf{x}_1(k) = \mathbf{d}(k) + \mathbf{X}_\tau(k)\mathbf{g}(k).$$

$$\hat{\mathbf{d}}(k) = \mathbf{x}_1(k) - \mathbf{X}_\tau(k)\hat{\mathbf{g}}(k)$$

predicted reverberation





How to select suitable cost function for prediction filters ?

2. Multi-channel linear prediction

- **Generalization of original MCLP approach** [Nakatani et al., 2010]
  - STFT coefficients of desired signal are assumed to be independent and modelled using **circular sparse/super-Gaussian prior with time-varying variance** $\lambda(n)$

$$\rho(d(n)) = \max_{\lambda(n)>0} \mathcal{N}_{\mathbb{C}}(d(n); 0, \lambda(n))\psi(\lambda(n))$$

  Scaling function $\psi(.)$ can be interpreted as **hyper-prior on variance**

  - **Maximum-Likelihood Estimation** (batch, per frequency bin)

$$\mathcal{L}(\mathbf{g}) = \prod_{n=1}^{N} \rho(d(n)) \Rightarrow \min_{\boldsymbol{\lambda}>0,\mathbf{g}} \sum_{n=1}^{N} \left(\frac{|d(n)|^2}{\lambda(n)} + \log \pi\lambda(n) - \log \psi(\lambda(n)) \right)$$

  - **Alternating optimization procedure**
    1. Estimate **prediction vector** (assuming fixed variances)

$$\hat{\mathbf{g}}^{(i+1)} = \left(\mathbf{X}_\tau^H \mathcal{D}_{\hat{\boldsymbol{\lambda}}^{(i)}}^{-1} \mathbf{X}_\tau \right)^{-1} \mathbf{X}_\tau^H \mathcal{D}_{\hat{\boldsymbol{\lambda}}^{(i)}}^{-1} \mathbf{x}_1$$

    2. Estimate **variances** (assuming fixed prediction vector)

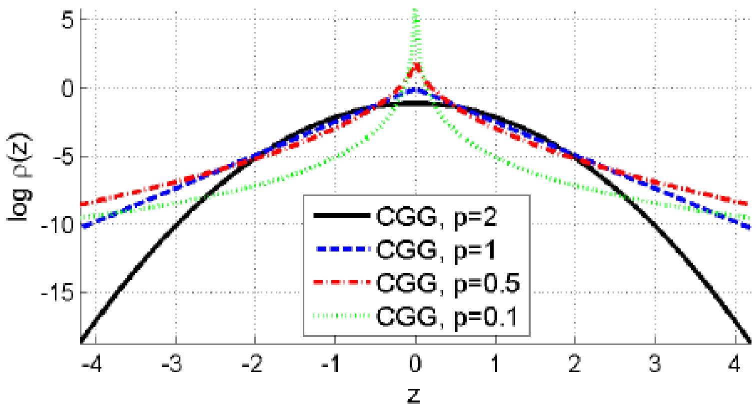
$$\hat{\lambda}^{(i+1)}(n) = \arg\min_{\lambda(n)>0} \frac{|\hat{d}^{(i+1)}(n)|^2}{\lambda(n)} + \log \pi\lambda(n) - \log \psi(\lambda(n))$$

2. Multi-channel linear prediction

- **Example:** complex generalized Gaussian (CGG) prior with shape parameter *p*

$$\rho(z) = \frac{p}{2\pi\gamma\Gamma(2/p)} e^{-\frac{|z|^p}{\gamma^{p/2}}}$$

$$\boxed{\hat{\lambda}^{(i+1)}(n) = |\hat{d}^{(i+1)}(n)|^{2-p},}$$



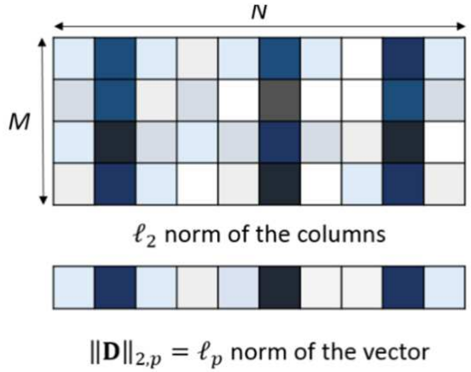
- **Remarks:**
  1. ML estimation using CGG prior is equivalent to **l_p-norm minimization**
     → **promotes sparsity of TF-coefficients across time** (for p < 2)

$$\boxed{\min_{\mathbf{g}} \|\mathbf{d}\|_p^p,}$$

  2. Original approach [Nakatani et al. 2010] corresponds to **p=0:**
     - **Strong sparse prior**, strongly favoring values of desired signal close to zero

2. Multi-channel linear prediction: extensions

1. **Group sparsity for MIMO dereverberation**
   - Maximize sparsity of TF-coefficients across time and simultaneously keep/discard TF-coefficients across microphones → mixed $l_{2,p}$-norm
   - **Multiple outputs** → possibility to apply spatial filtering

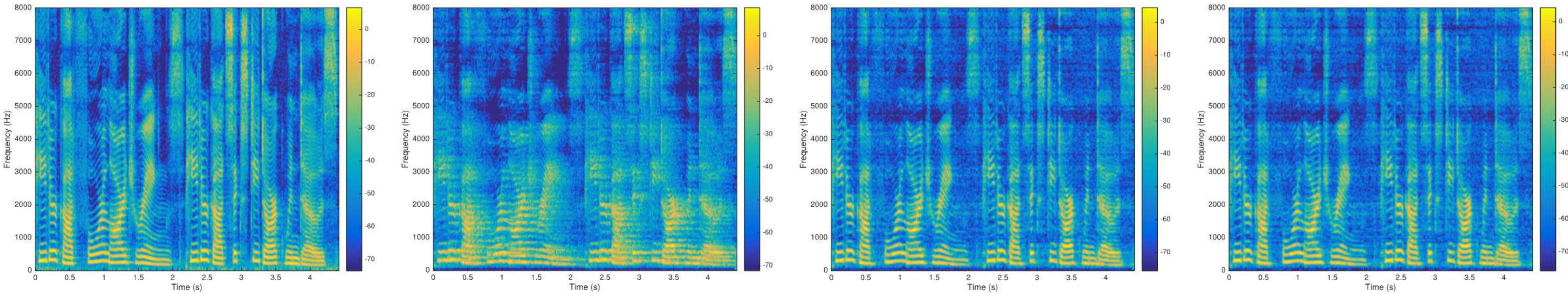


2. Incorporate **low-rank structure** of speech spectrogram
   - Combination with learned/pre-trained spectral dictionaries (NMF)
3. Batch processing → **adaptive processing**
   - Incorporate exponential weighting in cost function
   - **Problem:** overestimation of late reverberation for small forgetting factors γ (dynamic scenarios) → severe distortion in output signal
   - **Solution**: constrain MCLP-based estimate of late reverberation using PSD estimate

$$\check{\mathbf{G}}(n) = \arg\min_{\mathbf{G}(n)} \sum_{t=1}^{n} \gamma^{n-t} w(t) \|\mathbf{d}(t)\|_2^2 \quad \boxed{\text{subject to} \quad |\mathbf{G}^{\mathsf{H}}(n)\tilde{\mathbf{x}}_\tau(n)|^2 \leq \hat{\boldsymbol{\sigma}}_u^2(n)}$$

2. Multi-channel linear prediction: results

- **Instrumental validation (binaural, noiseless, batch)**



Clean | Microphone | MCLP | MCLP+NMF

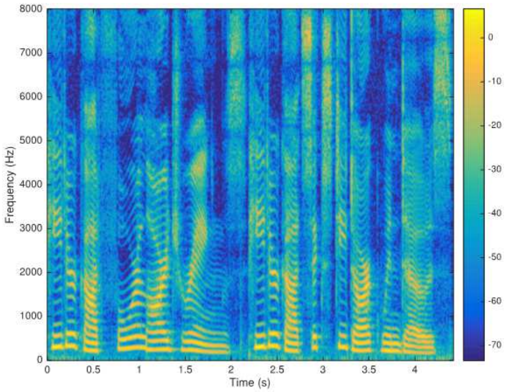
	PESQ	CD	FWSSNR	LLR	SRMR
Microphone	**1.21**	4.27	**3.61**	0.93	2.05
MCLP	2.40	3.15	7.92	0.60	3.83
MCLP+NMF	**2.42**	3.16	**7.84**	0.60	3.88

$T_{60} \approx 700$ms, M=2 (BRIR), distance 4m, fs=16 kHz; STFT: 64ms (overlap 16ms); MCLP: L_g=30, τ=2, p=0

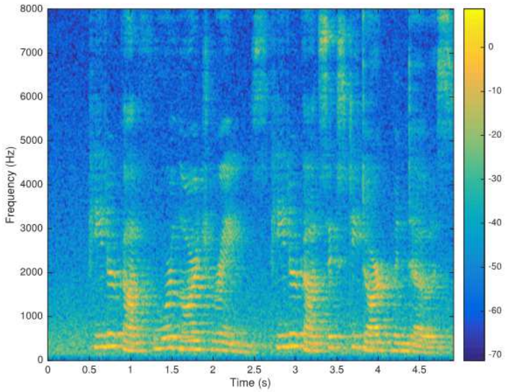
[Jukić et al., ICASSP 2015]

2. Multi-channel linear prediction: results

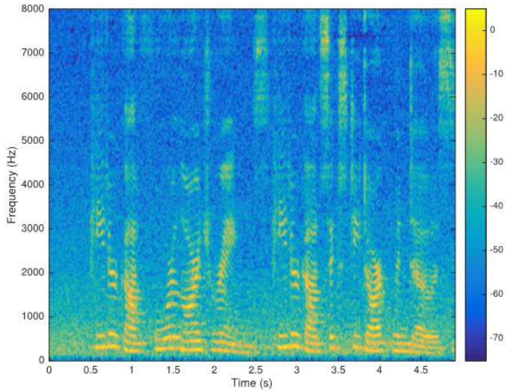
- **Instrumental validation (binaural, noisy 15dB, batch)**



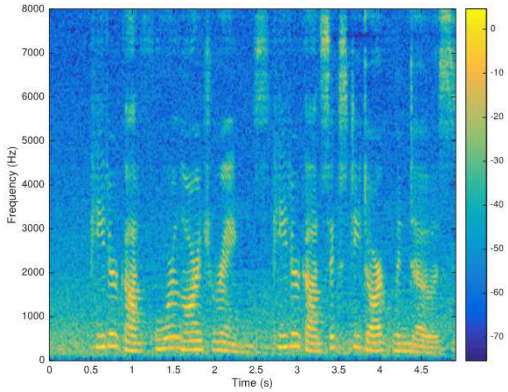
Clean



Microphone



MCLP



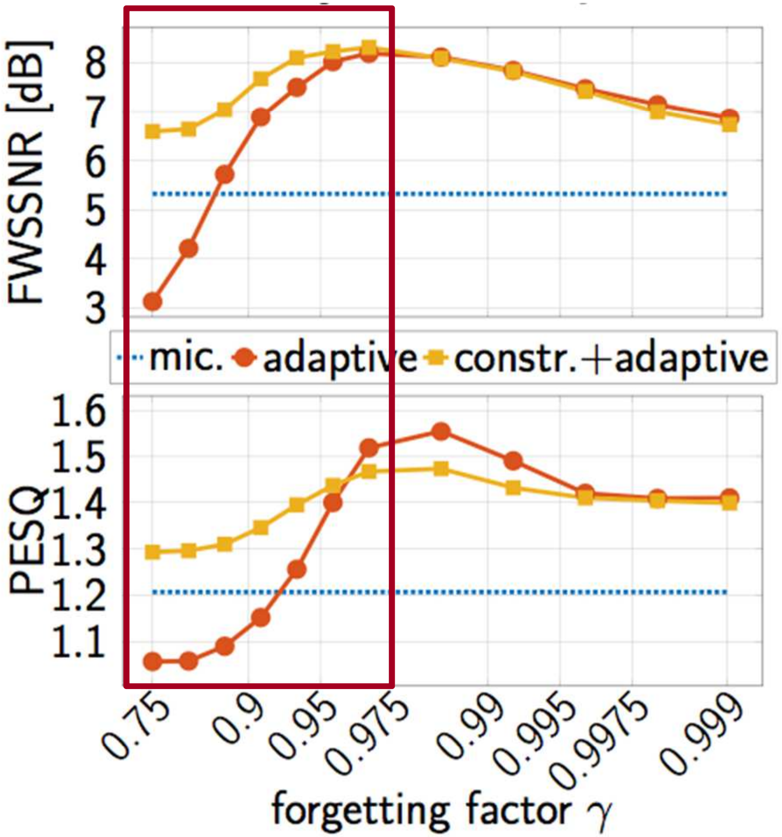
MCLP+NMF

$T_{60} \approx$ 700ms, M=2 (BRIR), distance 4m, fs=16 kHz; STFT: 64ms (overlap 16ms); MCLP: L_g=30, τ=2, p=0

[Jukić et al., ICASSP 2015]

2. Multi-channel linear prediction: results

- **Instrumental validation (noiseless, adaptive)**



clean **microphone**

	ADA	Constr. +ADA
γ=0.98		
γ=0.88		

Constrained MCLP much less sensitive to forgetting factor (especially for small values)

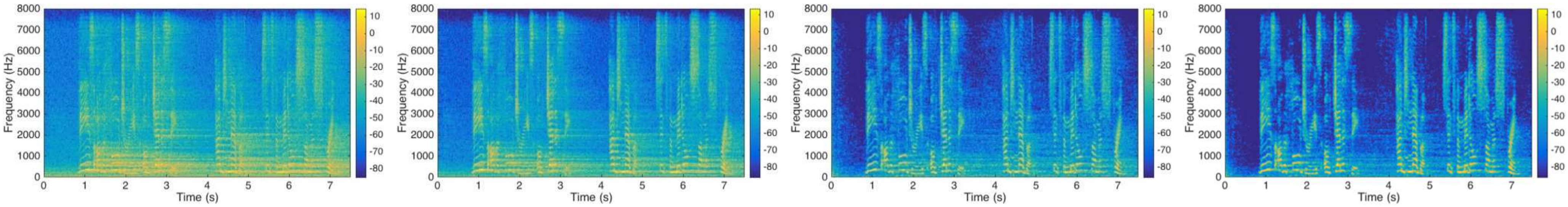
$T_{60} \approx$ 700ms, M=2, distance 2m, **source switching between +45 and -45,** fs=16 kHz; STFT: 64ms (overlap 16ms); L_g=20, τ=2, p=0

[Jukić et al., SPL 2017]

2. Multi-channel linear prediction: results

- **Instrumental validation (high reverberation + noisy, adaptive)**

d ~ 2m



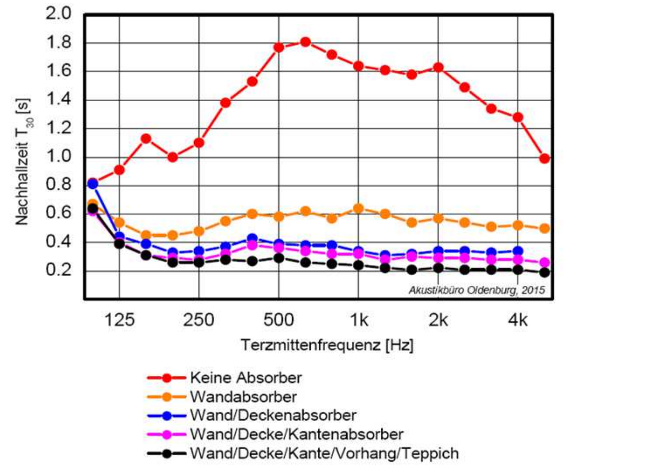
Microphone | 1ch SE [REVERB] | Adaptive MCLP | Adaptive MCLP + SE

T60 ~ 6s (St Alban The Martyr Church, London), M=2 (spacing~1m), fs=16 kHz, **real recordings**
STFT: 64ms (overlap 16ms); MCLP: L_g=30, τ=2, p=0, adaptive (γ=0.96)

Current/future research

- **Combined dereverberation and noise reduction**
  - Extension of multi-channel EVD-based PSD estimator and
  - Extension of blind probabilistic model-based approach
- **Instrumental measures**: prediction of perceived level of reverberation, by optimizing/redesigning SRMR measure (joint project with Prof. Tiago Falk)
- Database in new **varechoic lab**

Abbildung 1: In Raum E10 in den in Tabelle 1 angegebenen Raumzuständen gemessenen Nachhallzeiten in Terzbändern im Vergleich



Summary

- **Blind methods for combined dereverberation and noise reduction**
  - Spectral enhancement by applying real-valued gain to each time-frequency bin (single- and multi-channel PSD estimators)
  - Reverberation suppression by estimating late reverberant component using multi-channel linear prediction
- **Good dereverberation performance possible**, even for moving source and moderate noise
- Application to binaural hearing aids (combination with binaural noise reduction and cue preservation) to be further investigated

Acknowledgments

Dr. Daniel Marquardt | Dr. Ina Kodrasi | Ante Jukić | Nico Gößling | Benjamin Cauchi | Prof. Timo Gerkmann | Prof. Volker Hohmann | Elior Hadad | Prof. Sharon Gannot

- **Funding:**
  - Cluster of Excellence Hearing4All (DFG)
  - Marie-Curie Initial Training Network "Dereverberation and Reverberation of Audio, Music, and Speech" (EU)
  - 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

Questions ?

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

Recent publications

- I. Kodrasi, S. Doclo, *Late reverberant power spectral density estimation based on an eigenvalue decomposition*, in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, USA, Mar. 2017.
- A. Jukić, T. van Waterschoot, S. Doclo, *Adaptive speech dereverberation using constrained sparse multi-channel linear prediction*, *IEEE Signal Processing Letters*, vol. 24, no. 1, pp. 101-105, Jan. 2017.
- A. Jukić, T. van Waterschoot, T. Gerkmann, S. Doclo, *A general framework for incorporating time-frequency domain sparsity in multi-channel speech dereverberation*, *Journal of the Audio Engineering Society*, Jan-Feb 2017.
- I. Kodrasi, B. Cauchi, S. Goetze, S. Doclo, *Instrumental and perceptual evaluation of dereverberation techniques based on robust acoustic multi-channel equalization*, *Journal of the Audio Engineering Society*, Jan-Feb 2017.
- B. Cauchi, J. F. Santos, K. Siedenburg, T. H. Falk, P. A. Naylor, S. Doclo, S. Goetze, *Predicting the quality of processed speech by combining modulation based features and model-trees*, in *Proc. ITG Conference on Speech Communication*, Paderborn, Germany, Oct. 2016, pp. 180-184.
- A. Kuklasinski, S. Doclo, S. H. Jensen, J. Jensen, *Maximum Likelihood PSD Estimation for Speech Enhancement in Reverberation and Noise*, *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 24, pp. 1595-1608, Sep. 2016.
- I. Kodrasi, S. Doclo, *Joint Dereverberation and Noise Reduction Based on Acoustic Multichannel Equalization*, *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 24, no. 4, pp. 680-693, Apr. 2016.
- A. Jukić, T. van Waterschoot, T. Gerkmann, S. Doclo, *Group sparsity for MIMO speech dereverberation*, in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, USA, Oct. 2015, pp. 1-5.
- A. Jukić, T. van Waterschoot, T. Gerkmann, S. Doclo, *Multi-channel linear prediction-based speech dereverberation with sparse priors*, *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 23, no. 9, pp. 1509-1520, Sep. 2015.
- B. Cauchi, I. Kodrasi, R. Rehr, S. Gerlach, A. Jukić, T. Gerkmann, S. Doclo, S. Goetze, *Combination of MVDR beamforming and single-channel spectral processing for enhancing noisy and reverberant speech*, *EURASIP Journal on Advances in Signal Processing*, 2015:61, pp. 1-12.
- I. Kodrasi, S. Goetze, S. Doclo, *Regularization for Partial Multichannel Equalization for Speech Dereverberation*, *IEEE Trans. Audio, Speech and Language Processing*, vol. 21, no. 9, pp. 1879-1890, Sep. 2013.

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