Speech Dereverberation for Hearing Aids with a Binaural Data Link

Heiner Löllmann, Marco Jeub and Peter Vary

Diskussionssitzung Binaurales Hören mit Hörgeräten und Cochleaimplantaten

September 28, 2011

Institute of Communication Systems and Data Processing
Prof. Dr.-Ing. P. Vary

RWTH AACHEN

Overview

- Introduction
- Influence of Bilateral Dereverberation
- Two-Stage Binaural Speech Enhancement System
  - Model Assumptions
  - Stage I: Reduction of late reverberation
  - Stage II: Reduction of early reverberation
  - Low Delay Processing
  - Simulation Results & Audio Example
- Conclusions & Outlook

Motivation

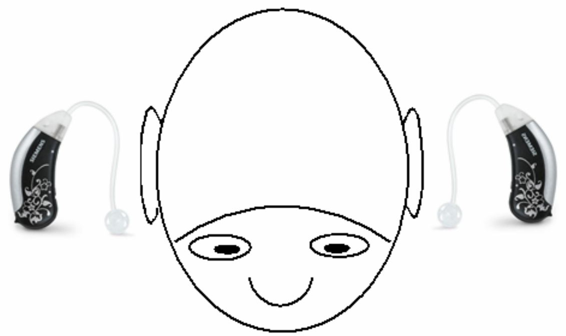
- Speech Enhancement in Hearing Aids
  - essential for listening comfort and speech intelligibility
- Current Systems
  - perform mostly noise reduction
  - speech dereverberation less common
- Problem
  - most published speech dereverberation algorithms do not fulfill the demanding requirements for *hearing aids*
    - low computational complexity
    - very low signal delay
    - robustness w.r.t. noise

    …

→ Motivates New Approach (and Further Research!)

Motivation

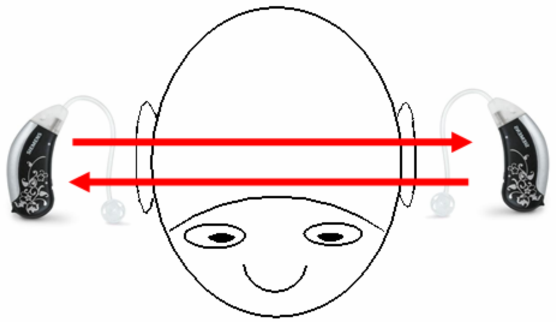
Common Technology



Bilateral Processing

- two independent devices
- no data-link

New Technology



Binaural Processing

- two linked devices
- (full) data-link

Binaural cues (interaural time & level difference) can be severely degraded by independent *bilateral* processing!

Effect of Bilateral Processing on Binaural Cues

- Interaural Time Difference (ITD)

$$\boxed{\Delta t = \frac{3r}{c} \sin\,\theta}$$

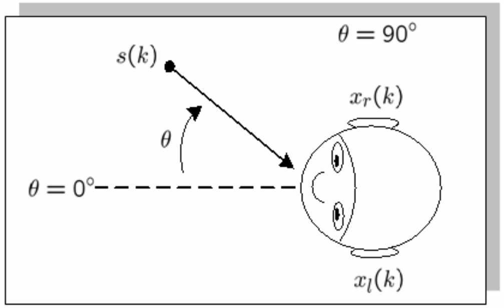
r: Radius of the head (0.085m)

c: Speed of sound (340m/s)

- Interaural Level Difference (ILD)

$$\boxed{\Delta E = 10 \cdot \log_{10}\left(\frac{E_l}{E_r}\right)}$$

$E_{l|r}$: energy of right or left signal

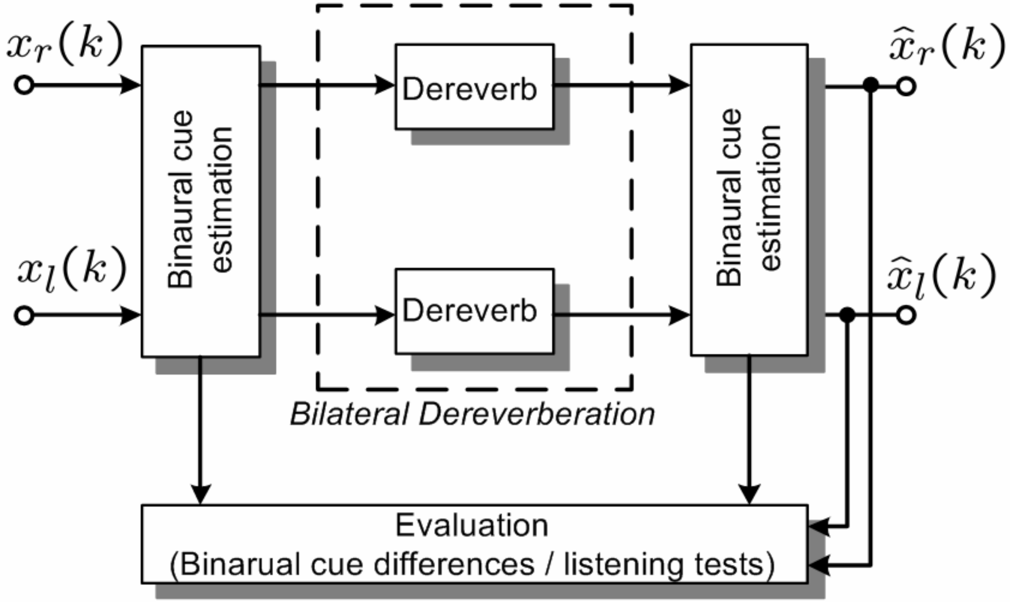


$\theta = -90^\circ$

- Rule of Thumb
  - ITD relevant for frequencies below 1.5kHz
  - ILD relevant for frequencies above 1.5kHz

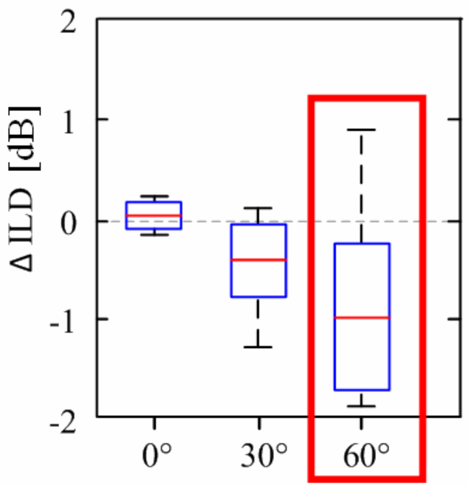
Effect of Bilateral Processing on Binaural Cues

- Investigations of bilateral noise reduction [v.d.Bogeat, JASA 2005]
- Here: Investigation of *bilateral dereverberation*



- octave subbands
- only cues with high interaural coherence (precedence effect)
- evaluate frame-by-frame fluctuations

Results for Bilateral Dereverberation on ILD

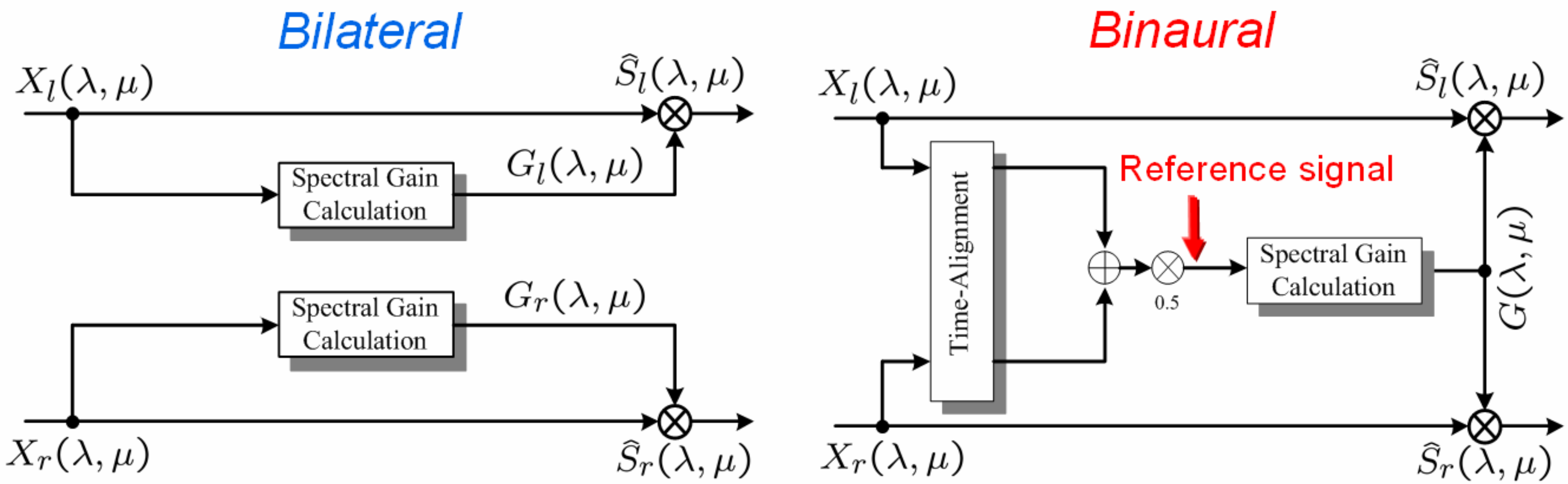


- source from different angles (RT=0.8s)
- frame-by-frame fluctuations
- octave subbands
- exemplarily for dereverberation proposed in [Lebart, PhD 1999]

► Minimum audible ILD difference: 0.5dB

> ➢ Binaural cues can be severely degraded by independent *bilateral dereverberation*

Influence of Bilateral Dereverberation



- Results of Listening Test (17 persons)

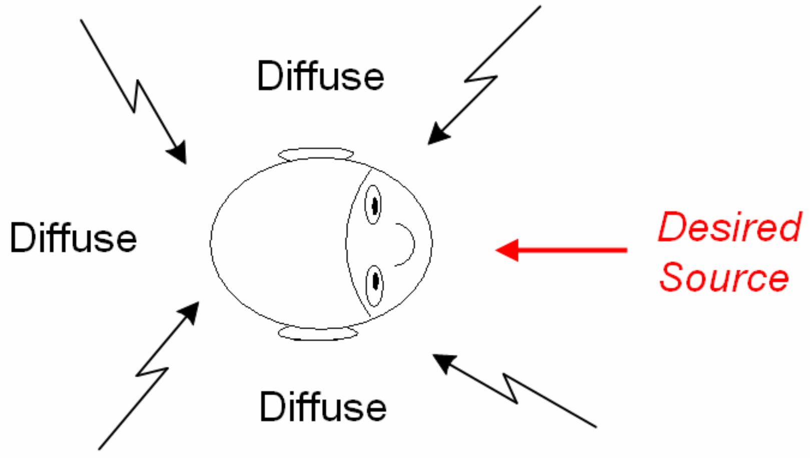
λ frame index
μ frequency index

Simulation Setup	No Preference	Binaural Dereverberation	Bilateral Dereverberation
40° ,d=2m	5.9%	82.4%	11.7%
0° ,d=1m	11.7%	64.8%	23.5%
Average	**8.8%**	**73.6%**	**17.6%**

How to Preserve the Binaural Cues?

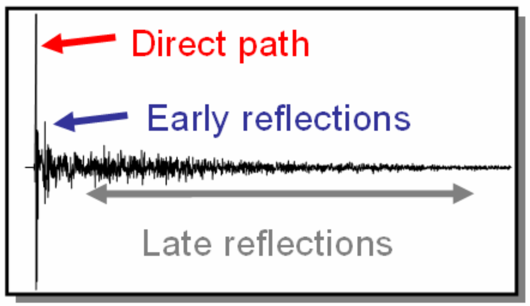
- Reconstruction after bilateral processing
  - Blind Source Separation (BSS) [Wehr, ITG 2008]
  - Binaural Artificial Bandwidth Extension (ABE)
    [Laaksonen, ICASSP 2009]
    Modify (estimated) high band cues based on low band cues
- Modify speech enhancement algorithm
  - Special constraints for a multichannel Wiener filter
    [Doclo, TASLP 2010], [v.d. Bogeart, PhD 2008]
  - Spectral subtraction: apply same gains to both channels
    [Lotter & Vary, Journal on Appl. SP 2006], [Peissig, PhD 1992]

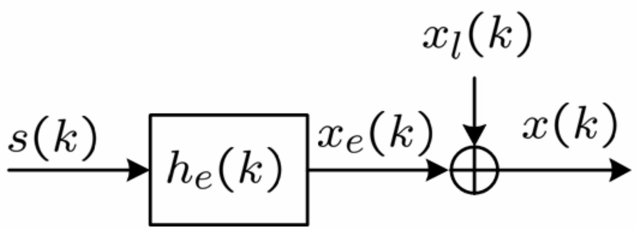
Assumptions for Binaural Dereverberation



- high direct path energy of desired source (*high coherence*)
- diffuse sound field (*low coherence*)

► *Late Reverberation* modeled as Additive Noise [Lebart, PhD 1999]

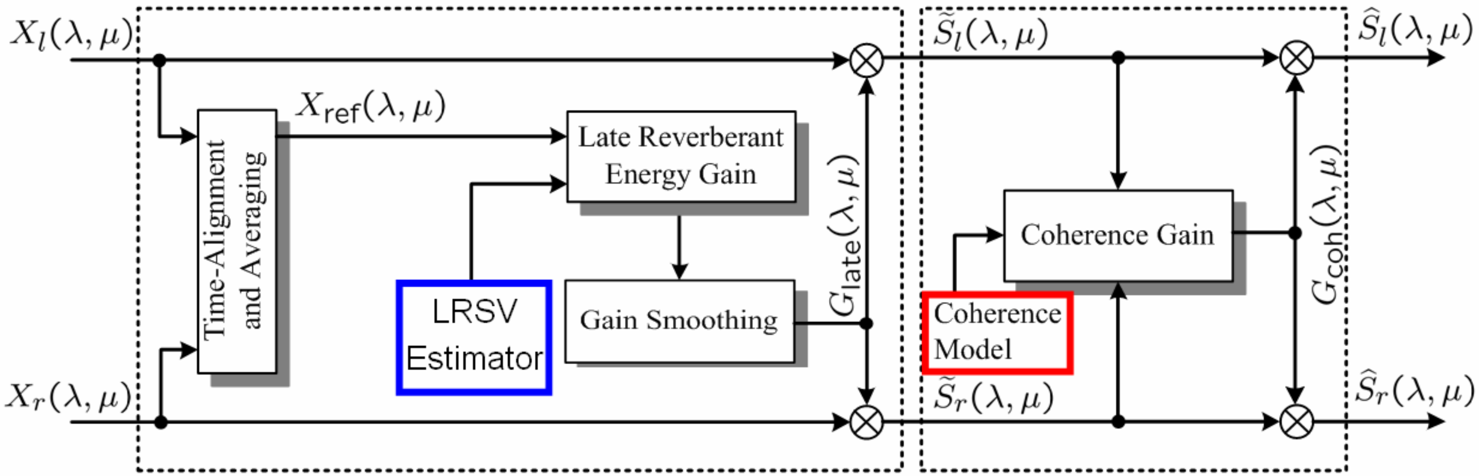




Binaural Speech Enhancement System

- System Overview (in frequency-domain)

λ frame index
μ frequency index

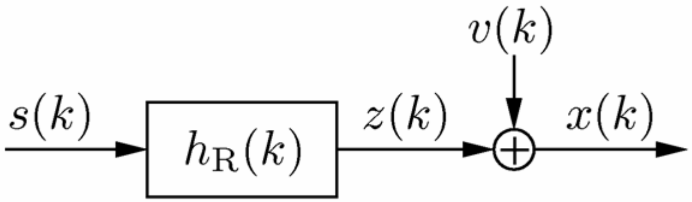


- Stage I reduces *late reverberation*
  - relies on Late Reverberant Spectral Variance (LRSV) estimator
- Stage II reduces *non-coherent* components
  - early and residual late reverberation
  - relies on a coherence model

Principle of Stage I

- Model of Reverberant and Noisy Speech

$$x(k) = \sum_{n=0}^{L_R-1} s(k-n) \cdot h_{\mathrm{R}}(n) + v(k)$$

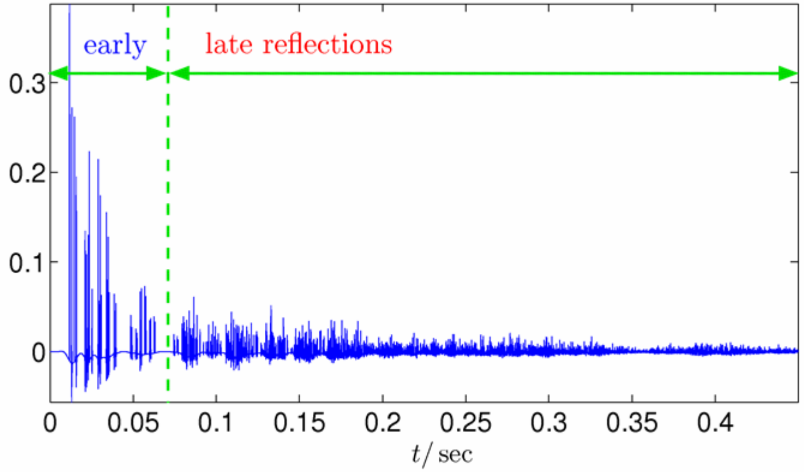


Principle of Stage I

- Model of Reverberant and Noisy Speech

$$x(k) = \sum_{n=0}^{L_R-1} s(k-n) \cdot h_{\mathrm{R}}(n) + v(k)$$

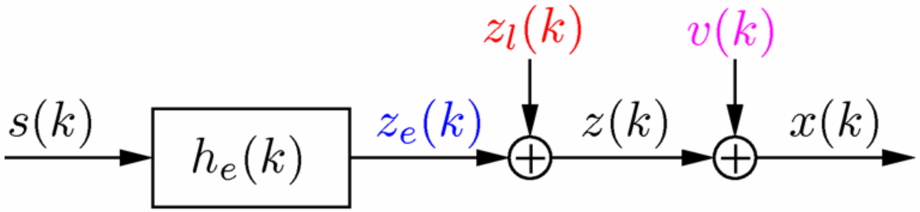
$$= \underbrace{\sum_{n=0}^{L_e-1} s(k-n) \cdot h_e(n)}_{z_e(k):\ \text{early reverberant speech}} + \underbrace{\sum_{n=L_e}^{L_R-1} s(k-n) \cdot h_l(n)}_{z_l(k):\ \text{late reverberant speech}} + \underbrace{v(k)}_{\text{noise}}$$



Principle of Stage I

- Model of Reverberant and Noisy Speech

$$
\begin{aligned}
x(k) &= \sum_{n=0}^{L_R-1} s(k-n)\cdot h_{\mathrm{R}}(n) + v(k) \\
&= \underbrace{\sum_{n=0}^{L_e-1} s(k-n)\cdot h_e(n)}_{z_e(k):\ \text{early reverberant speech}} + \underbrace{\sum_{n=L_e}^{L_R-1} s(k-n)\cdot h_l(n)}_{z_l(k):\ \text{late reverberant speech}} + \underbrace{v(k)}_{\text{noise}}
\end{aligned}
$$



- Benefit
  - *additive* distortions due to late reverberant speech and noise can be suppressed by *common* spectral subtraction

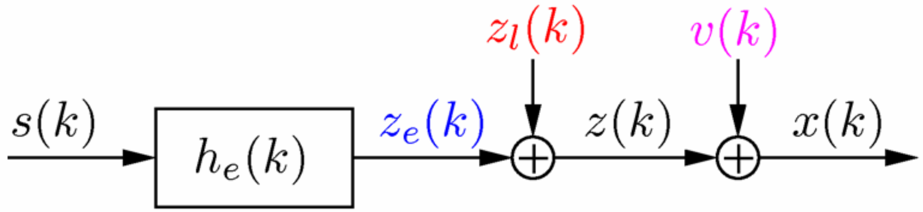
Principle of Stage I

▶ Weights calculated by Signal-to-Interference Ratio (SIR)

$$G'_{\text{late}}(\lambda, \mu) = 1 - \frac{1}{\sqrt{\gamma(\lambda, \mu)}}$$

with *a posteriori* SIR

$$\gamma(\lambda, \mu) = \frac{\mathrm{E}\{|X(\lambda, \mu)|^2}{\sigma_{z_l}^2(\lambda, \mu) + \sigma_v^2(\lambda, \mu)}$$



▶ Benefit

- *additive* distortions due to late reverberant speech and noise can be suppressed by *common* spectral subtraction

Principle of Stage I

- Weights calculated by Signal-to-Interference Ratio (SIR)

$$G'_{\text{late}}(\lambda,\mu) = 1 - \frac{1}{\sqrt{\gamma(\lambda,\mu)}}$$

with *a posteriori* SIR

$$\gamma(\lambda,\mu) = \frac{\mathrm{E}\{|X(\lambda,\mu)|^2}{\sigma^2_{z_l}(\lambda,\mu) + \sigma^2_v(\lambda,\mu)}$$

- Estimation of Noise PSD $\sigma^2_v(\lambda,\mu)$
  - can be done by well-known algorithms (e.g. minimum statistics)
- Estimation of LRSV $\sigma^2_{z_l}(\lambda,\mu)$
  - less explored
  - crucial for the achieved speech quality

LRSV Estimation for Stage I

- Model-based approach [Lebart, PhD 1999]
  - LRSV estimator derived by simple statistical model for RIR

$$\sigma^2_{\text{late}}(\lambda, \mu) = e^{-6 \ln 10\, T_l / T_{60}} \cdot \sigma_z^2(\lambda - L_e, \mu)$$

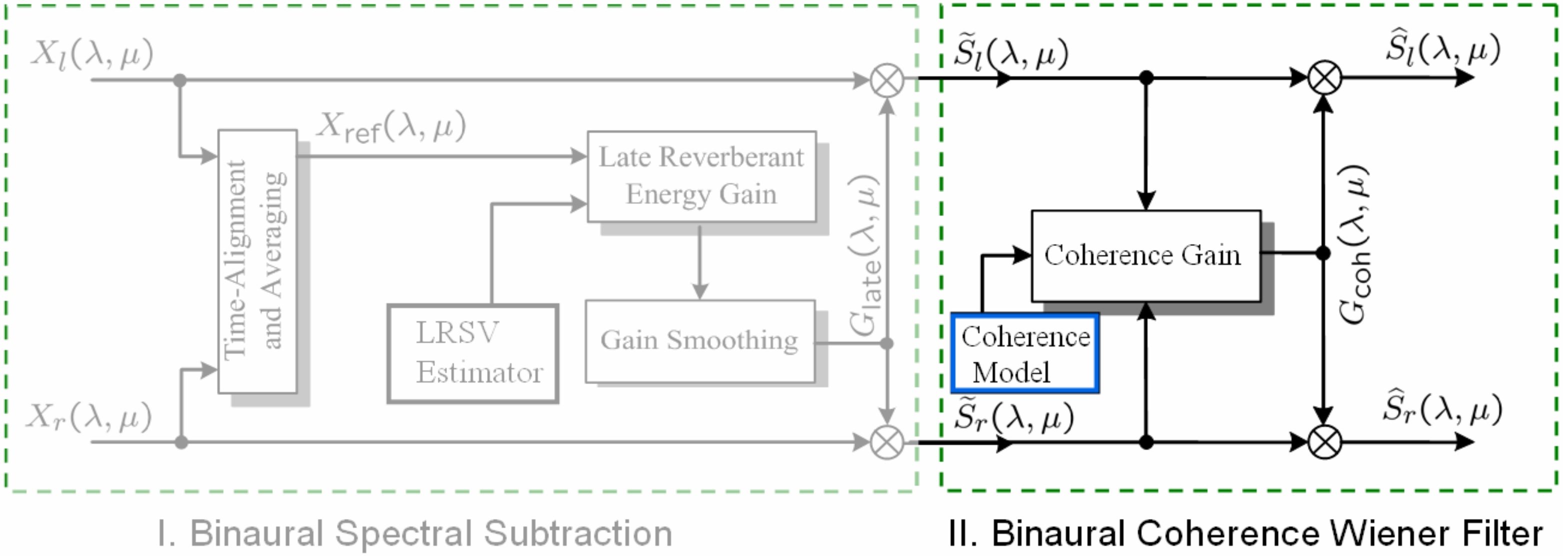
  - parameters $T_l,\ L_e$ fixed
  - variance of reverberant speech $\sigma_z^2(\lambda, \mu)$ given (after denoising)

- Reverberation Time T_{60}
  - *blind* estimation with low complexity [Löllmann & Vary, IWAENC 2008, 2010]

Binaural Speech Enhancement System

- System Overview (in frequency-domain)

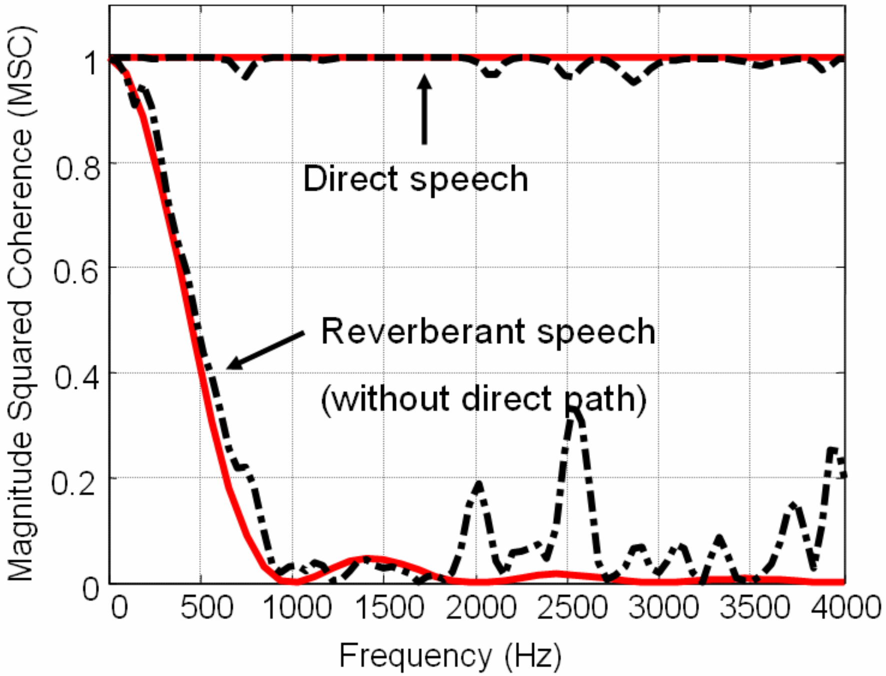
λ frame index
μ frequency index



- Stage II
  - reduction of (residual) non-coherent speech components and noise
  - relies on a suitable coherence model

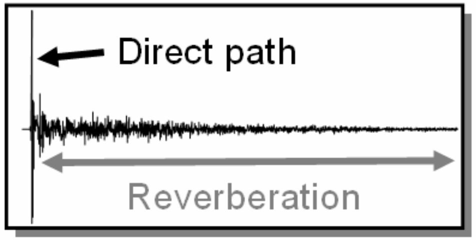
Stage II: Coherence-Based Supression

- ▶ Coherence of reverberant sound fields (without head)



Theoretical coherence of ideal diffuse sound field:

$$\Gamma^{(\text{diff})}_{x_l x_r}(f) = \text{sinc}\left(\frac{2\pi f d_{mic}}{c}\right)$$



➢ Reverberation results in diffuse sound field

Binaural Speech Enhancement System

▶ Stage II - Coherence-based dereverberation

- Calculation of (Wiener filter) gains [McCowan, TASLP 2003]

$$G_{\mathsf{coh}}(\lambda,\mu) = \frac{\hat{\Phi}_{ss}(\lambda,\mu)}{\frac{1}{2}\cdot\left(\hat{\Phi}_{\tilde{s}_l\tilde{s}_l}(\lambda,\mu)+\hat{\Phi}_{\tilde{s}_r\tilde{s}_r}(\lambda,\mu)\right)}$$

- Requires estimate of (direct) speech auto-PSD

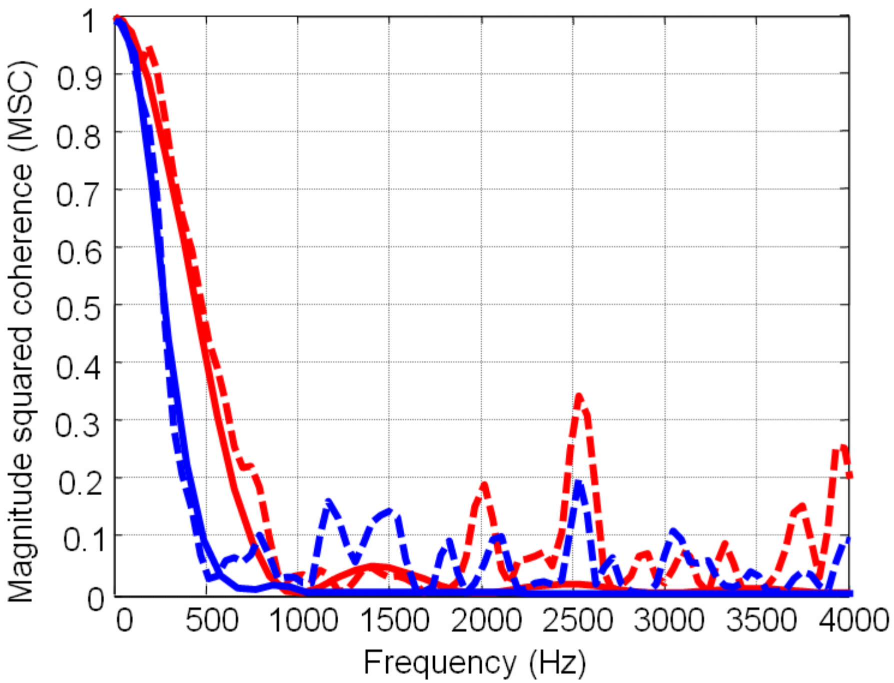
$$\hat{\Phi}_{ss}(\lambda,\mu) = \frac{\mathrm{Re}\{\hat{\Phi}_{\tilde{s}_l\tilde{s}_r}(\lambda,\mu)\} - \frac{1}{2}\mathrm{Re}\{\Gamma_{\tilde{s}_l\tilde{s}_r}(\Omega)\}\left(\hat{\Phi}_{\tilde{s}_l\tilde{s}_l}(\lambda,\mu)+\hat{\Phi}_{\tilde{s}_r\tilde{s}_r}(\lambda,\mu)\right)}{1-\mathrm{Re}\{\Gamma_{\tilde{s}_l\tilde{s}_r}(\Omega)\}}$$

(fixed) coherence function

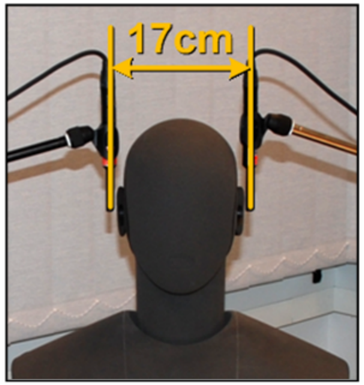
➢ Suitable coherence model ”only” needed!

Stage II: Coherence Model

- Influence of head shadowing on coherence



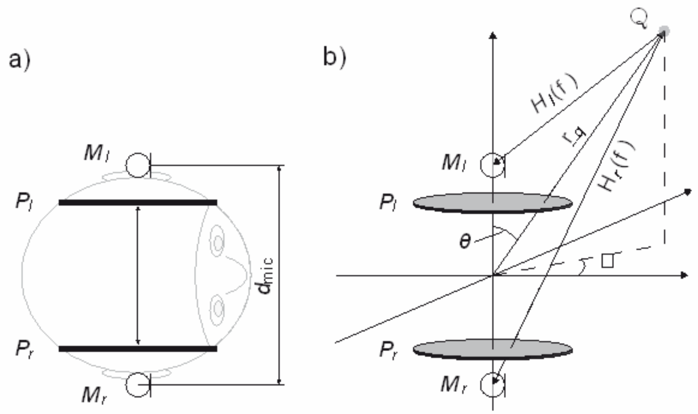
[Dörbecker, PhD 1998]



> ➢ Head shadowing must be taken into account

Stage II: Coherence Model

- Head modelled by two circular plates [Dörbecker, PhD 1998, Jeub, SP Letters 2010]



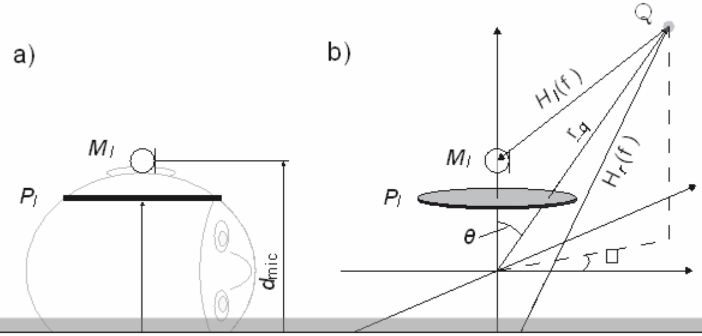
- Integration over all possible directions of incident

$$\Gamma_{x_l x_r}(\Omega) = \frac{\int_0^{\pi/2} \left(H_l(\Omega,\theta)\,H_r^*(\Omega,\theta) + H_r(\Omega,\theta)\,H_l^*(\Omega,\theta)\right)\sin\theta\,\mathrm{d}\theta}{\sqrt{\int_0^{\pi/2}\left(|H_l(\Omega,\theta)|^2 + |H_r(\Omega,\theta)|^2\right)\sin\theta\,\mathrm{d}\theta}}$$

Stage II: Coherence Model

- Head modelled by two circular plates

[Dörbecker, PhD 1998, Jeub, SP Letters 2010]



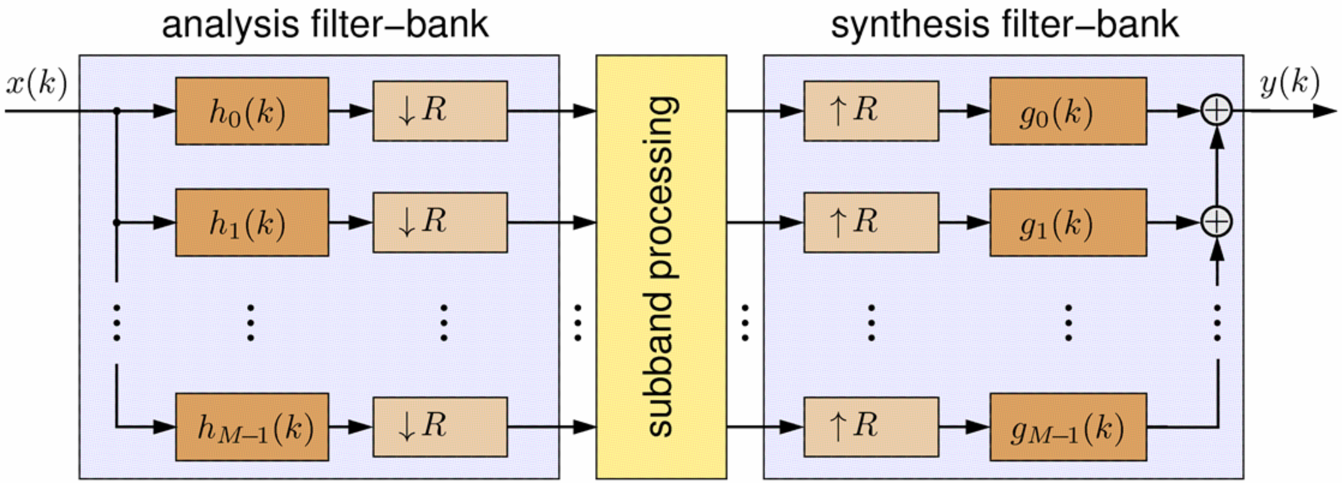
For digital hearing aids:
- Directions of incident are not uniformly distributed
- 2D (cylindrically isotropic) sound field is more suitable

$$\Gamma_{x_l x_r}(\Omega) = \frac{\int_0^{\pi/2} \left(H_l(\Omega,\theta)\, H_r^*(\Omega,\theta) + H_r(\Omega,\theta)\, H_l^*(\Omega,\theta) \right) \mathrm{d}\theta}{\sqrt{\int_0^{\pi/2} \left(|H_l(\Omega,\theta)|^2 + |H_r(\Omega,\theta)|^2 \right) \mathrm{d}\theta}}$$

Low Delay Processing

- Problem
  - proposed algorithm relies on *spectral weighting*
  - performed by an analysis-synthesis filter-bank (AS FB)

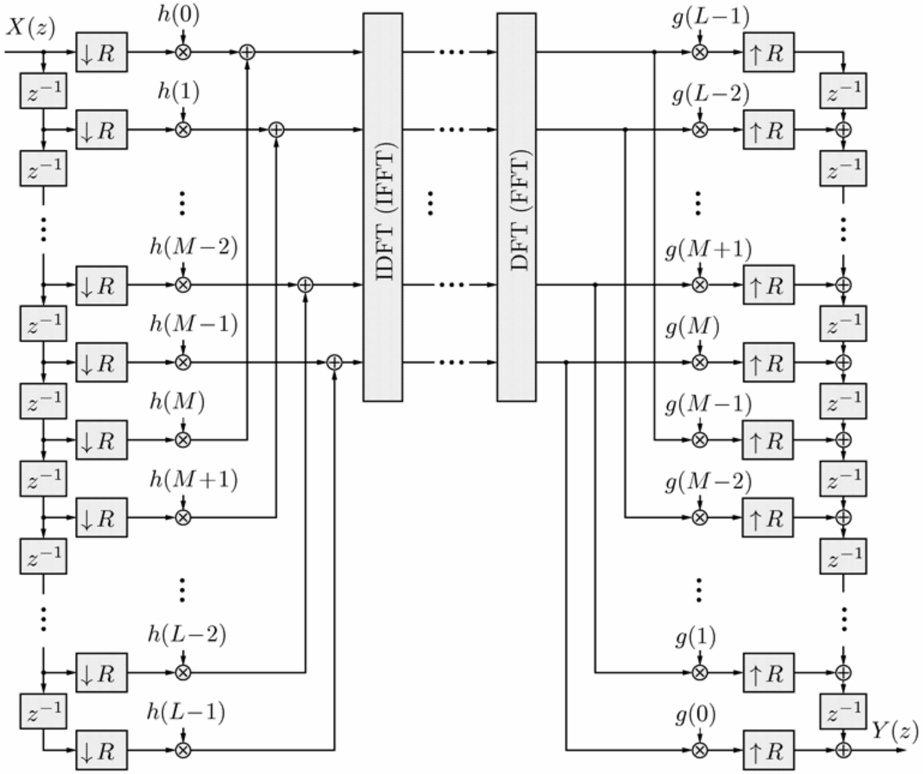
  ➔ additional signal delay



➢ Low delay filter-bank needed to avoid comb-filter effect!

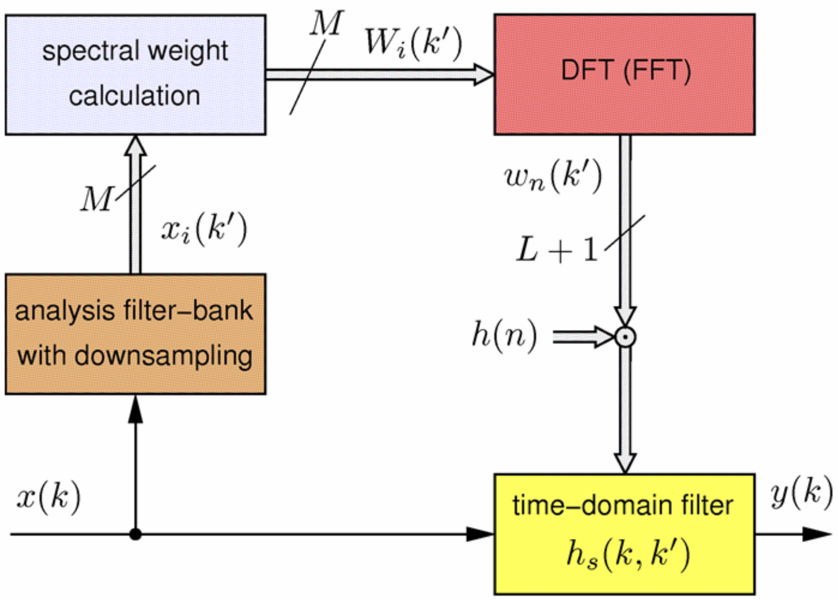
Low Delay Processing

- Uniform DFT AS Filter-Bank [Bäuml & Sörgel, EUSIPCO 2008]
  - low delay achieved by short prototype lowpass filter (latency ≈ 6ms)
  - efficient realization by FFT and polyphase network



Low Delay Processing

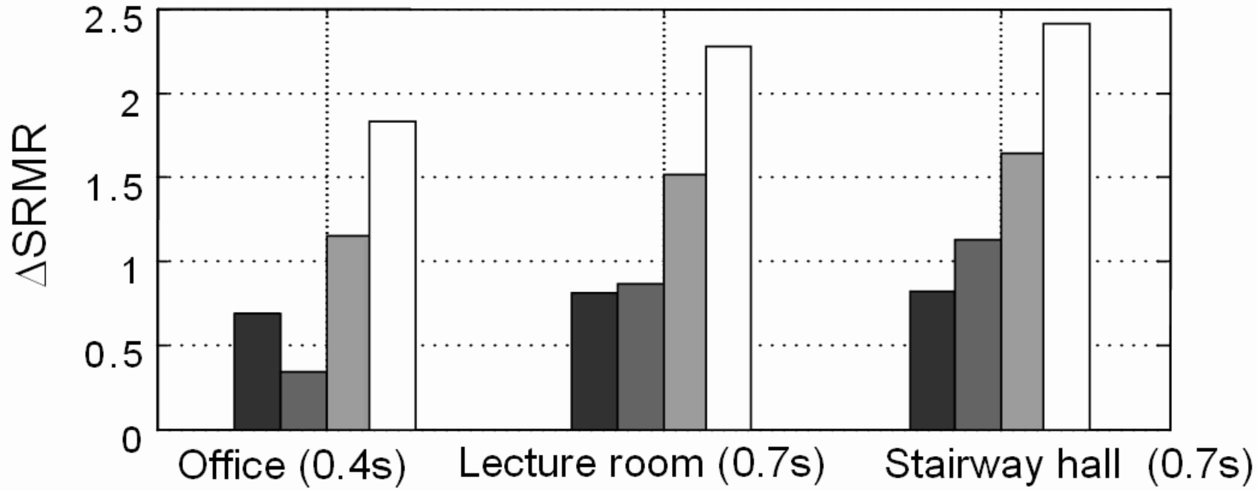
- Warped Filter-Bank Equalizer [Löllmann, Elsevier Journal 2007]
  - low delay achieved by time-domain filtering
  - efficient realization by FFT and polyphase network (not shown)
  - only 32 non-uniform (auditory) frequency bands



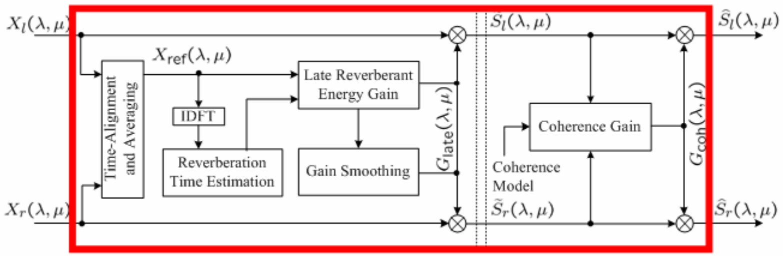
Evaluation

- **Quality Measure**

  Speech to Reverberation Modulation energy Ratio (SRMR)



algorithm of Allen (adapted to binaural)
Stage II using diffuse coherence model
Stage II using binaural coherence model
Two-stage system



Audio Example

- Setup
  - IND lecture room
  - $T_{60} \approx 0.7\,\mathrm{s}$
  - distance of 8m
  - recordings with HA dummies

- Two-Stage Binaural Dereverberation
  - clean speech
  - reverberant speech
  - enhanced speech

Conclusions & Outlook

- Two-Stage Binaural Speech Dereverberation System presented
  - *binaural cues* (ILD and ITD) are preserved
  - joint speech dereverberation *and* noise suppression
  - reduction of early *and* late reverberation with *low delay*
  - all quantities for Stage I estimated blindly (no a priori knowledge)
  - Stage II based on an improved coherence model
  - better speech dereverberation in comparison to related state-of-the-art approaches
  - approach also suitable for hands-free mode of *mobile phones*

- Further & Ongoing Work
  - algorithm taking the direct-to-reverberation ratio (DRR) into account
  - psychoacoustical weight calculation to reduce musical tones
  - required data rate for binaural processing