KATHOLIEKE UNIVERSITEIT
LEUVEN

Arenberg Doctoral School of Science, Engineering & Technology
Faculty of Engineering
Department of Electrical Engineering

Signal Processing in Hearing Aids: Noise Reduction and Adaptive Feedback Cancellation

Kim Ngo

Starkey, Eden Prairie, MN, USA.

Thursday Nov. 4th, 2010.











KATHOLIEKE UNIVERSITEIT
LEUVEN

Outline

- Introduction and problem statement
- (background noise, acoustic feedback, reduced dynamic range)
- Multi-channel Wiener filter based noise reduction
- (tracking: speech and noise nonstationarities)
- Capon beamformer
- (robust steering vector estimation)
- Prediction-error-method based adaptive feedback cancellation
- (bias reduction, near-end signal modelling)
- Conclusion

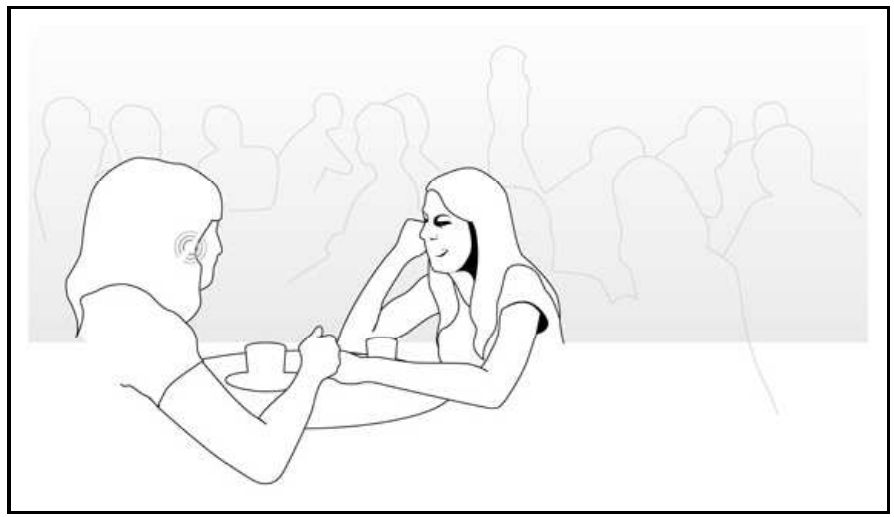
KATHOLIEKE UNIVERSITEIT LEUVEN

Introduction

Background Noise

Hearing impaired people:

- Reduced frequency resolution (separating sounds of different frequencies)
- Reduced temporal resolution (intense sounds mask weaker sounds)
- Reduced spatial cues (spatially separating desired signal from noise)



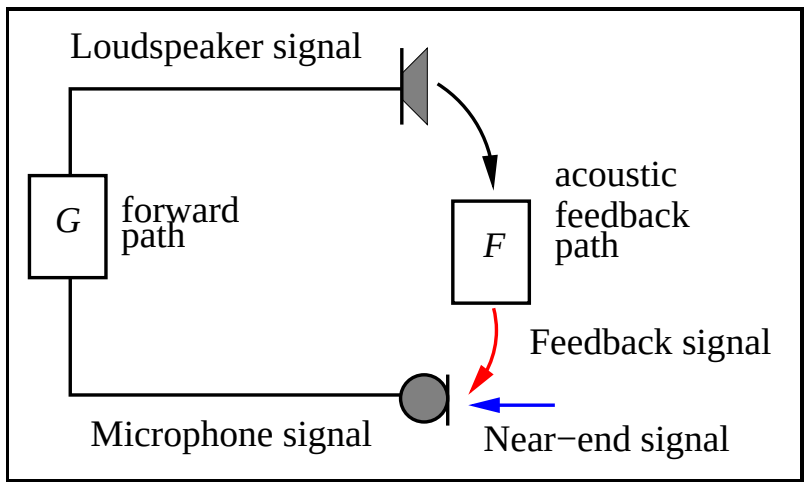
Objective of NR:

- Maximally reduce the noise (SNR improvement)
- Minimize speech distortion (sound quality)
- Improve intelligibility of speech (perceptual quality)

Acoustic Feedback

Design of hearing aid:

- hearing aids become smaller and smaller, open fitting etc.
- Undesired acoustic coupling between loudspeaker and microphone.
- Limits the maximum amplification.
- Feedback are most severe at high frequencies.
- Instability results in a high frequency tone.



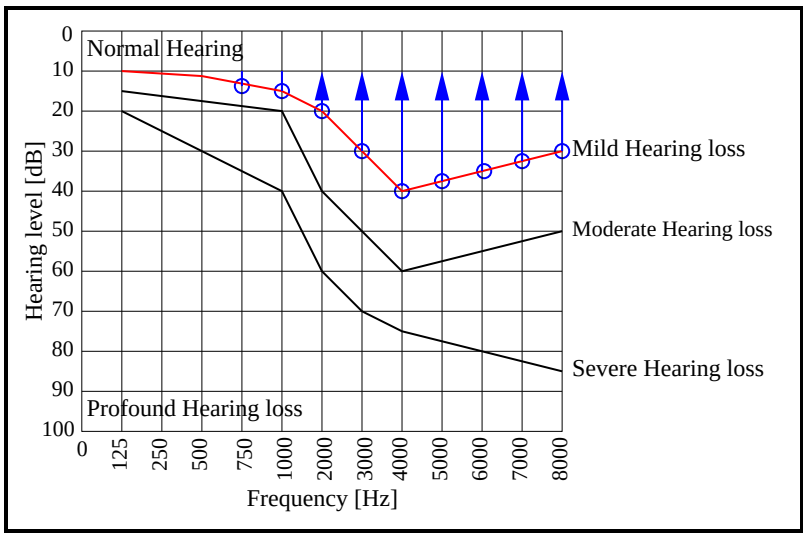
Objective of AFC:

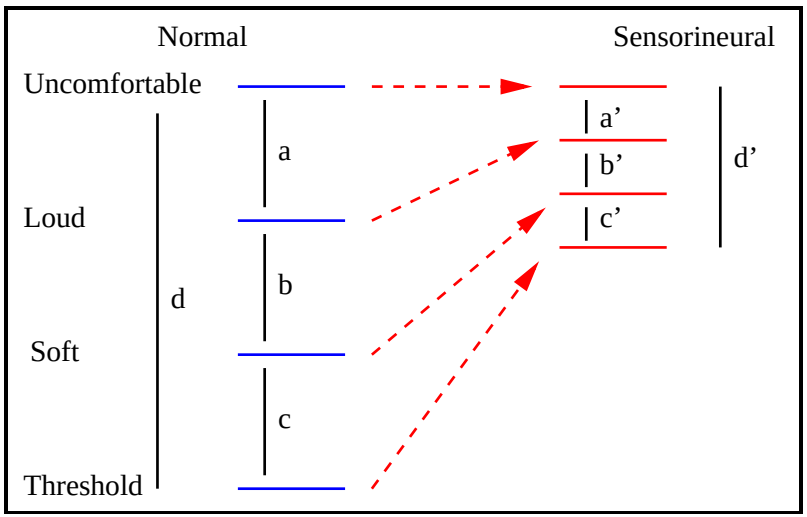
- Increase maximum stable gain (MSG)
- Reduce bias and convergence (misadjustment)
- Minimize speech distortion (sound quality, howling)

KATHOLIEKE UNIVERSITEIT LEUVEN

Introduction

Reduced dynamic range





Objective of DRC:

- Mapping the wide dynamic range of speech into the reduced dynamic range.
- Weak sounds audible - loud sounds not uncomfortably loud.
- High intensity attenuated - low intensity amplified.

KATHOLIEKE UNIVERSITEIT LEUVEN

Introduction

Research areas in Hearing Aids

Automatic Speech Recognition

Feedback Cancellation

Wireless Communications

Binaural Processing

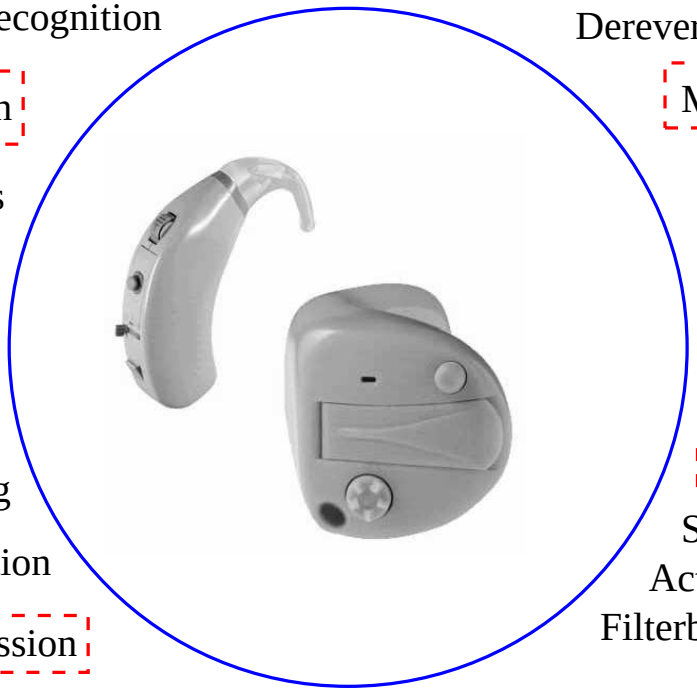
AD/DA Converters

Digital Signal Processing

Analog Signal Processing

Automatic Sound Classification

Dynamic Range Compression



Dereverberation

Multi−channel Noise Reduction

Single−channel Noise Reduction

Directional Microphones

Loudspeaker

Speech/Audio Coding

Source Localisation

Beamforming

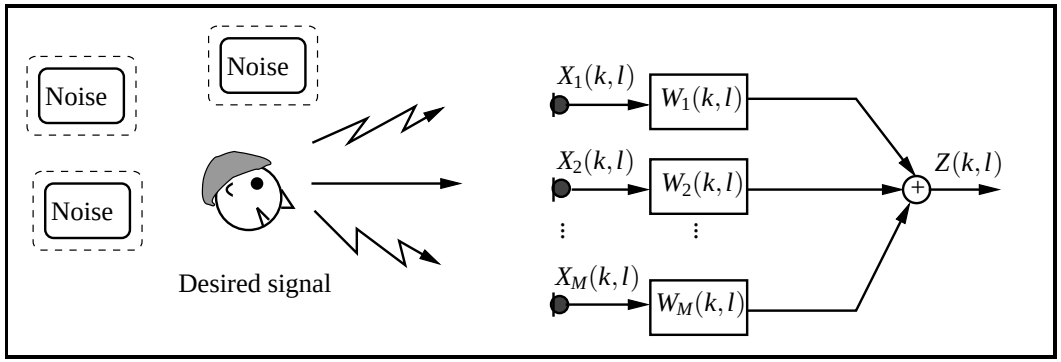
Source Separation

Active Noise Control

Filterbank Design

KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter



Frequency-domain microphone signals:

$$\mathbf{X}(k,l) = \mathbf{X}^s(k,l) + \mathbf{X}^n(k,l)$$

MWF MMSE criterion:

$$\mathbf{W}_{\text{MMSE}}(k,l) = \arg\min_{\mathbf{W}} \ \varepsilon\{|X_1^s(k,l) - \mathbf{W}^H\mathbf{X}(k,l)|^2\}$$

Optimal MWF solution:

$$\boxed{\mathbf{W}_{\text{MMSE}}(k,l) = \left(\mathbf{R}^s(k,l) + \mathbf{R}^n(k,l)\right)^{-1}\mathbf{R}^s(k,l)\mathbf{e}_1}$$

Output of the MWF:

$$Z(k,l) = \mathbf{W}_{\text{MMSE}}^H(k,l)\mathbf{X}(k,l).$$

KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

Speech distortion weighted multi-channel Wiener filter (SDW-MWF$_\mu$)

SDW-MWF$_\mu$ MMSE criterion:

$$\mathbf{W}_\mu(k,l) = \arg\min_{\mathbf{W}} \ \varepsilon\{|X_1^s(k,l) - \mathbf{W}^H\mathbf{X}^s(k,l)|^2\} + \mu\varepsilon\{|\mathbf{W}^H\mathbf{X}^n(k,l)|^2\}$$

Optimal SDW-MWF$_\mu$ solution:

$$\boxed{\mathbf{W}_\mu(k,l) = \left(\mathbf{R}^s(k,l) + \mu\mathbf{R}^n(k,l)\right)^{-1}\mathbf{R}^s(k,l)\mathbf{e}_1}$$

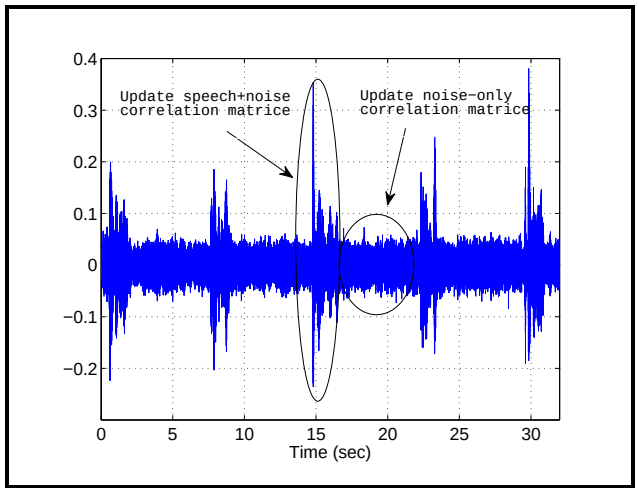
Concept:

- Second-order statistics of the noise are assumed to be stationary

$$\mathbf{R}^s(k,l) = \mathbf{R}^x(k,l) - \mathbf{R}^n(k,l)$$

- Estimation of $\mathbf{R}^x(k,l)$ and $\mathbf{R}^n(k,l)$ an averaging time window of 2-3 seconds is typically used.

$$\boxed{\begin{aligned}\mathbf{R}_n(k,l+1) &= \alpha_n\mathbf{R}_n(k,l) + (1-\alpha_n)\mathbf{X}(k,l)\mathbf{X}^H(k,l)\\ \mathbf{R}_x(k,l+1) &= \alpha_x\mathbf{R}_x(k,l) + (1-\alpha_x)\mathbf{X}(k,l)\mathbf{X}^H(k,l)\end{aligned}}$$



KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

Design concerns

Properties of SDW-MWF:

- SDW-MWF depends on long-term average of spectral and spatial characteristics.
- Eliminates short-time effects, such as musical noise
- SDW-parameter μ is a fixed value for all frequencies

Properties not included in SDW-MWF:

- Speech and noise can be non-stationary spectrally and temporally.
- speech contains many pauses while noise can be continously present.
- Different weight to speech dominant segments and to noise dominant segments

KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

Speech Presence Probability

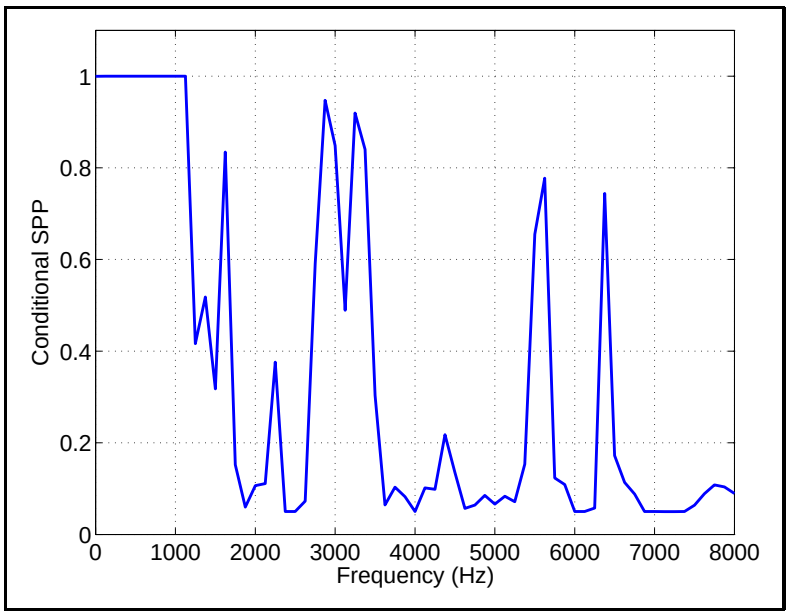
Two-state speech model:

$$\begin{aligned} H_0(k,l) &: X_i(k,l) = X_i^n(k,l) \\ H_1(k,l) &: X_i(k,l) = X_i^s(k,l) + X_i^n(k,l) \end{aligned}$$

conditional speech presence probability $p(k,l) \triangleq P(H_1(k,l)|X_i(k,l))$:

$$p(k,l) = \left\{1 + \frac{q(k,l)}{1-q(k,l)}(1+\xi(k,l))\exp(-\upsilon(k,l))\right\}^{-1}$$

- $q(k,l) \triangleq P(H_0(k,l))$ is the a priori speech absence probability (SAP)
- $\xi(k,l)$ is the a priori SNR
- $\gamma(k,l)$ is the a posteriori SNR
- $p(k,l)$ is estimated for each frequency bin and each frame



KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

SDW-MWF incorporating the conditional SPP (SDW-MWF$_{\text{SPP}}$)

SDW-MWF$_{\text{SPP}}$ MMSE criterion:

$$\mathbf{W}_{\text{SPP}} = \arg\min_{\mathbf{W}} \; p\varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}|^2|H_1\} + (1-p)\varepsilon\{|\mathbf{W}^H\mathbf{X}^n|^2\}$$

Optimal SDW-MWF$_{\text{SPP}}$ solution:

$$\boxed{\mathbf{W}_{\text{SPP}} = \left(\mathbf{R}^s + \left(\tfrac{1}{p}\right)\mathbf{R}^n\right)^{-1}\mathbf{R}^s\mathbf{e}_1.}$$

concept:

- If $p = 0$, the SDW-MWF$_{\text{SPP}}$ attenuates the noise by applying $\mathbf{W}_{\text{SPP}} \leftarrow 0$.
- If $p = 1$, the SDW-MWF$_{\text{SPP}}$ solution corresponds to the MWF solution (μ=1).
- If $0 < p < 1$ there is a trade-off between noise reduction and speech distortion.

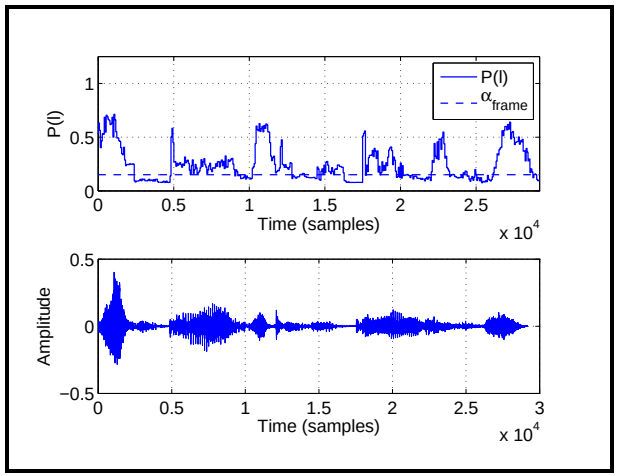
KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

H_0 and H_1 state detection based on SPP

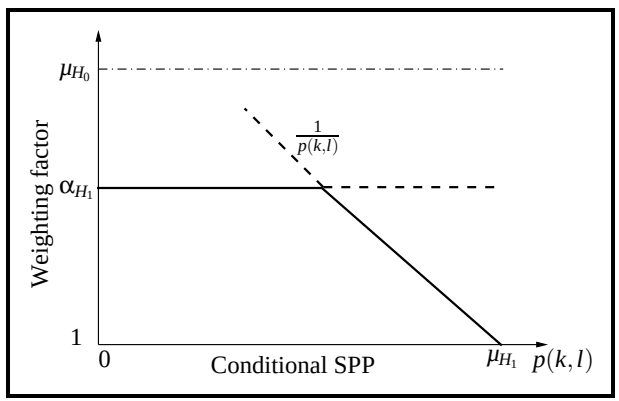
Binary decision:

$$P(l) = \begin{cases} H_1 : 1 & \text{if} \quad \frac{1}{K}\sum_{k=1}^{K} p(k,l) \geq \alpha_{\text{frame}} \\ H_0 : 0 & \text{otherwise} \end{cases}$$



Concept:

- A weighting factor μ_{H_1} is introduced, which is a function of $p(k,l)$, that defines the amount of noise reduction that can be applied in the H_1 state.
- A weighting factor μ_{H_0} is introduced, which is a constant weighting factor, that defines the amount of noise reduction that can be applied in the H_0 state.



KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

SDW-MWF incorporating a flexible weighting factor (SDW-MWF$_{\text{flex}}$)

SDW-MWF$_{\text{flex}}$ MMSE criterion:

$$\mathbf{W}_{\text{Flex}}(k,l) = \arg\min_{\mathbf{W}} \begin{cases} P(l)\left[\mu_{H_1}\varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}|^2|H_1\} + (1-\mu_{H_1})\varepsilon\{|\mathbf{W}^H\mathbf{X}|^2|H_0\}\right] \\ (1-P(l))\left[\frac{1}{\mu_{H_0}}\varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}^s|^2\} + \varepsilon\{|\mathbf{W}^H\mathbf{X}^n|^2\}\right] \end{cases}$$

where $\mu_{H_1} = \max(p(k,l), \frac{1}{\alpha_{H_1}})$

Optimal SDW-MWF$_{\text{flex}}$ solution:

$$\boxed{\mathbf{W}_{\text{Flex}}(k,l) = \left[\mathbf{R}^s + \gamma(k,l)\mathbf{R}^n\right]^{-1}\mathbf{R}^s\mathbf{e}_1}$$

with the weighting factor defined as:

$$\gamma(k,l) = \left[P(l)\max(p(k,l), \frac{1}{\alpha_{H_1}}) + (1-P(l))\frac{1}{\mu_{H_0}}\right]^{-1}$$

KATHOLIEKE UNIVERSITEIT LEUVEN

Experimental set-up MWF based NR

Simulation parameters:

- A 2-microphone behind-the-ear hearing aid mounted on a CORTEX MK2 manikin.
- The loudspeakers are positioned at 1 meter from the center of the head.
- The reverberation time $T_{60} = 0.21s$.
- Speech is located at $0°$, two multi-talker babble noise sources at $120°$ and $180°$.
- The speech signals consist of male sentences from the HINT-database
- and the noise signal consist of a multi-talker babble from Auditec
- The speech signals are sampled at 16kHz
- An FFT length of 128 with 50% overlap was used.

Intelligibility-weighted signal-to-noise ratio (SNR):

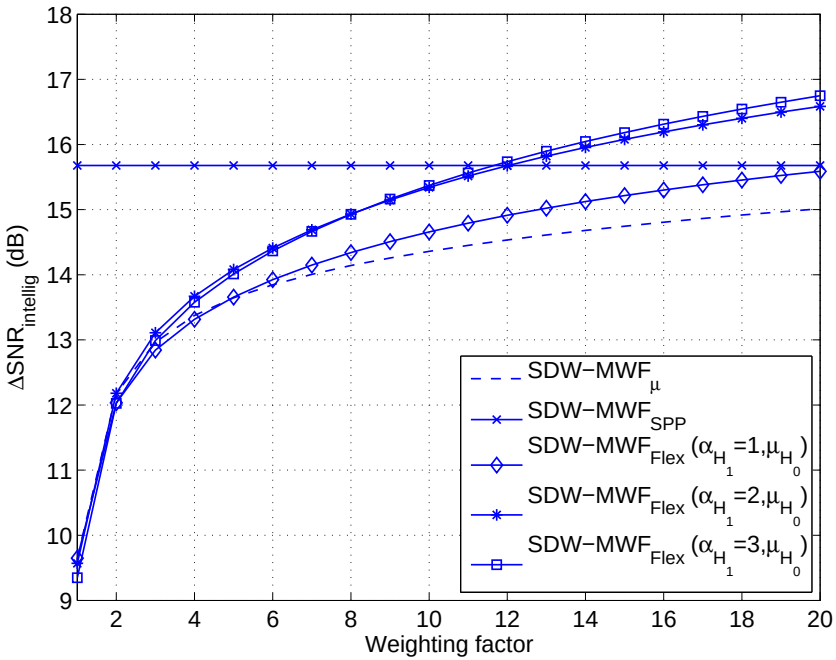
$$\Delta\mathrm{SNR}_{\mathrm{intellig}} = \sum_i I_i \left(\mathrm{SNR}_{i,out} - \mathrm{SNR}_{i,in}\right)$$

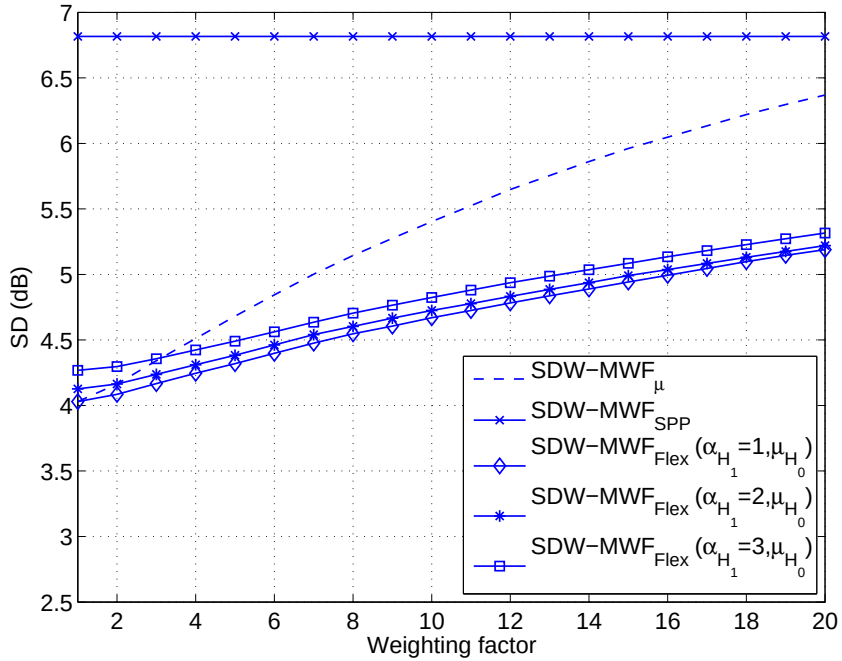
frequency-weighted log-spectral signal distortion (SD)

$$\mathrm{SD} = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\int_{f_l}^{f_u} w_{ERB}(f)\left(10log_{10}\frac{P^s_{out,k}(f)}{P^s_{in,k}(f)}\right)^2 df}$$

KATHOLIEKE UNIVERSITEIT LEUVEN

Experimental results MWF based NR





KATHOLIEKE UNIVERSITEIT LEUVEN

Capon beamformer

Design concerns

- case where the number of interferences exceeds the number of microphones
- Frequency-domain vs. time-domain implementation
- small arrays (degrees of freedom)
- Robustness against interferences (competing speaker)
- Low complexity and simple implementation
- closed-form update expression

KATHOLIEKE UNIVERSITEIT LEUVEN

Capon beamformer

Standard Capon beamformer (SCB)

SCB formulation:

$$\mathbf{W}_{\text{SCB}} = \arg\min_{\mathbf{W}} \mathbf{W}^H \mathbf{R}\mathbf{W}, \quad \text{s. t.} \quad \mathbf{e}^H\mathbf{W} = 1$$

Closed-form solution:

$$\mathbf{W}_{\text{SCB}} = \frac{\mathbf{R}^{-1}\mathbf{e}}{\mathbf{e}^H\mathbf{R}^{-1}\mathbf{e}},$$

Mismatch between presumed and actual steering vector

$$\bar{\mathbf{e}} = \tilde{\mathbf{e}} + \Delta \neq \tilde{\mathbf{e}}$$

- can lead to substantial performance degradation
- attempts to minimize the outpower by suppressing signal+inteference.

KATHOLIEKE UNIVERSITEIT LEUVEN

Capon beamformer

Review beamformer

LCMV beamformer:

$$\min_{\mathbf{W}} \mathbf{W}^H \mathbf{R} \mathbf{W}, \quad \text{s. t.} \quad \mathbf{C}^H \mathbf{W} = \mathbf{f}$$

- For each linear constraint the beamformer loses one degree of freedom.

Diagonal-loading beamformer:

$$\min_{\mathbf{W}} \mathbf{W}^H (\mathbf{R} + \gamma \mathbf{I}) \mathbf{W}, \quad \text{s. t.} \quad \mathbf{e}^H \mathbf{W} = 1$$

- Difficulty specifying the optimal value of γ.

Uncertainty-based beamformer:

$$\min_{\mathbf{W}} \mathbf{W}^H \mathbf{R} \mathbf{W}, \quad \text{s. t.} \quad |\mathbf{W}^H (\mathbf{e} + \Delta)| \geq 1,$$
$$\text{for all} \quad \|\Delta\| \leq \epsilon$$

- Infinite number of non-convex constraints.

KATHOLIEKE UNIVERSITEIT LEUVEN

Capon beamformer

Review beamformer

Max-min optimization:

$$\max_{\mathbf{e}} \min_{\mathbf{W}} \mathbf{W}^H \mathbf{R} \mathbf{W}, \quad \text{s. t.} \quad \mathbf{e}^H \mathbf{W} = 1, \|\mathbf{e}\|^2 = 1$$

expression above can be simplified to:

$$\max_{\mathbf{e}} \frac{1}{\mathbf{e}^H \mathbf{R}^{-1} \mathbf{e}}, \quad \text{s. t.} \quad \|\mathbf{e}\|^2 = 1,$$

and is equivalent to:

$$\min_{\mathbf{e}} \mathbf{e}^H \mathbf{R}^{-1} \mathbf{e}, \quad \text{s. t.} \quad \|\mathbf{e}\|^2 = 1$$

- which is a principal eigenvector problem.

If the target signal is the dominant one:

- the largest eigenvalue of $\mathbf{R}$ results in the optimal steering vector $\mathbf{e}$

If the target signal is not the dominant one:

- a subset of eigenvalues (signal plus interference) can be choosen
- the cost is reduced resolution

KATHOLIEKE UNIVERSITEIT LEUVEN

Capon beamformer

Proposed robust Capon beamformer (RCB)

Prior knowledge:

- we do not know the exact steering vector, only a region in which it should lie

proposed RCB formulation:

$$\max_{\mathbf{e}} \min_{\mathbf{W}} \mathbf{W}^H \mathbf{R} \mathbf{W}, \quad \text{s. t.} \quad \mathbf{e}^H \mathbf{W} = 1, \quad \|\mathbf{e} - \bar{\mathbf{e}}\|^2 \leq \|\beta \mathbf{e}\|^2, \quad \|\mathbf{e}\|^2 = \|\bar{\mathbf{e}}\|^2$$

expression above can be simplified to:

$$\max_{\mathbf{e}} \frac{1}{\mathbf{e}^H \mathbf{R}^{-1} \mathbf{e}}, \quad \text{s. t.} \quad \|\mathbf{e} - \bar{\mathbf{e}}\|^2 \leq \|\beta \mathbf{e}\|^2, \quad \|\mathbf{e}\|^2 = \|\bar{\mathbf{e}}\|^2$$

Concept:

- estimate the optimal steering vector within the uncertainty region
- instead of having multiple constraints we now have a single adapted constraint
- no infinite number of non-convex constraint
- gradient update of the steering vector proposed

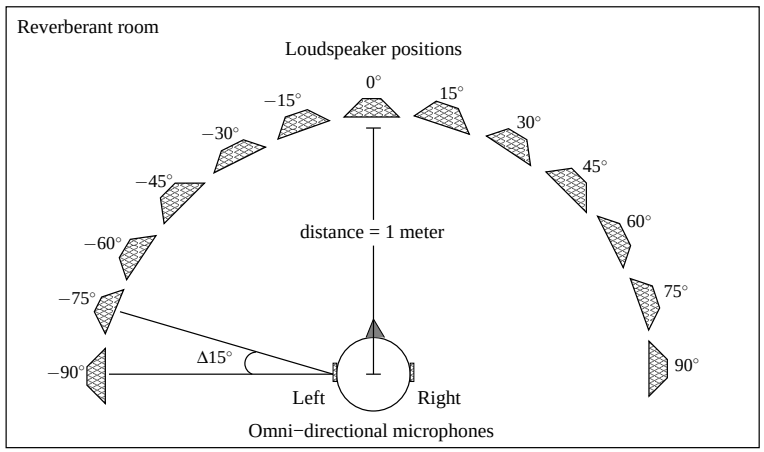
KATHOLIEKE UNIVERSITEIT LEUVEN

Experimental set-up Capon beamformer

Simulation parameters:

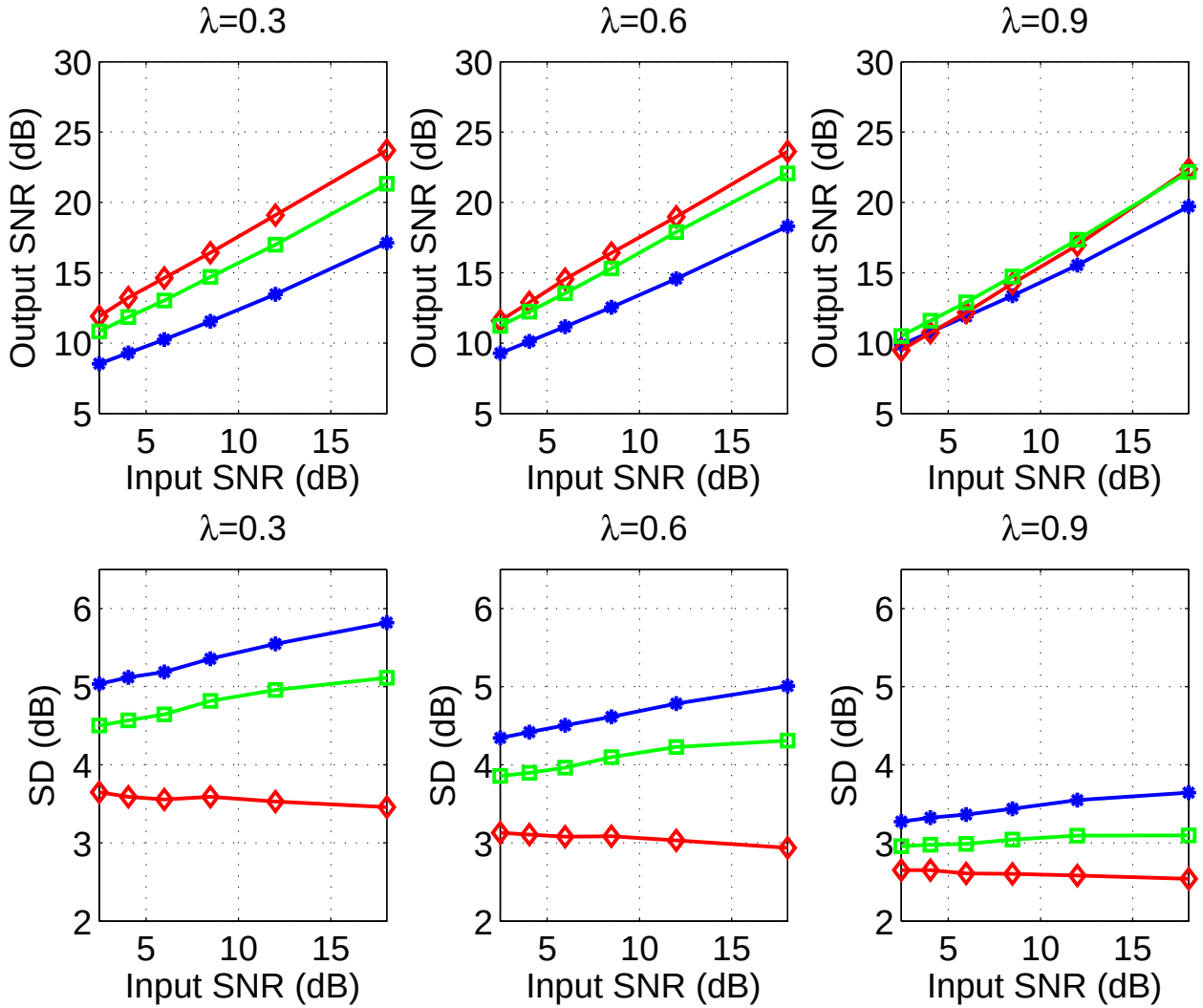
- Reverberant rectangular conference room with dimensions 9m x 4m x 2.75m
- Reverberation time T_{60}=0.37s
- Two omnidirectional microphones in free-field separated by 15cm
- The speech signal consists of male and female sentences each 2.5s. in duration
- The signals are sampled at 22.05 kHz
- An FFT length of 1024 with 50% overlap was used in all simulations.

Spatial scenarios	
Case	Azimuth angle of source
3a	Target at 0°, single-talker at -90°, -45°, 45° and 90°
3b	Target at 0°, single-talker at -90°, -30°, 30° and 90°



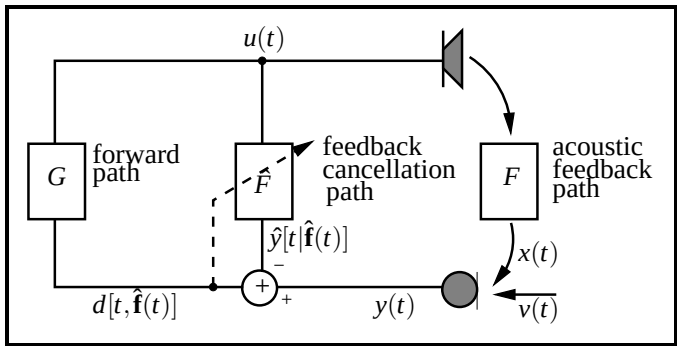
KATHOLIEKE UNIVERSITEIT
LEUVEN

Experimental results
Capon beamformer



KATHOLIEKE UNIVERSITEIT LEUVEN

Adaptive feedback cancellation



The microphone signal:

$$y(t) = v(t) + x(t) = v(t) + F(q,t)u(t)$$

The feedback-compensated:

$$d(t) = v(t) + [F(q,t) - \hat{F}(q,t)]u(t)$$

Concept:

- Adaptively model the feedback path and estimate the feedback signal
- subtract feedback signal from the microphone signal

Problem:

- correlation between the near-end signal and the loudspeaker signal
- Standard adaptive filtering converge to a biased solution.
- distortion: part of the near-end signal cancelled

Main Challenge:

- Reduce the correlation between the near-end signal and the loudspeaker signal
- Prediction error method-based AFC
- PEM delivers an unbiased estimate of the feedback path

KATHOLIEKE UNIVERSITEIT LEUVEN

Adaptive feedback cancellation

Prediction-error-method

Microphone signal:

$$y(t) = H(q,t)r(t) + F(q,t)u(t)$$

- $r(t)$ is a white noise signal.
- $H(q,t)$ is a linear model for the source signal $v(t)$

Prediction error:

$$\varepsilon(t) = H^{-1}(q,t)\left[y(t) - F(q,t)u(t)\right]$$

- Prefiltering of loudspeaker and microphone signals
- Inverse near-end signal model

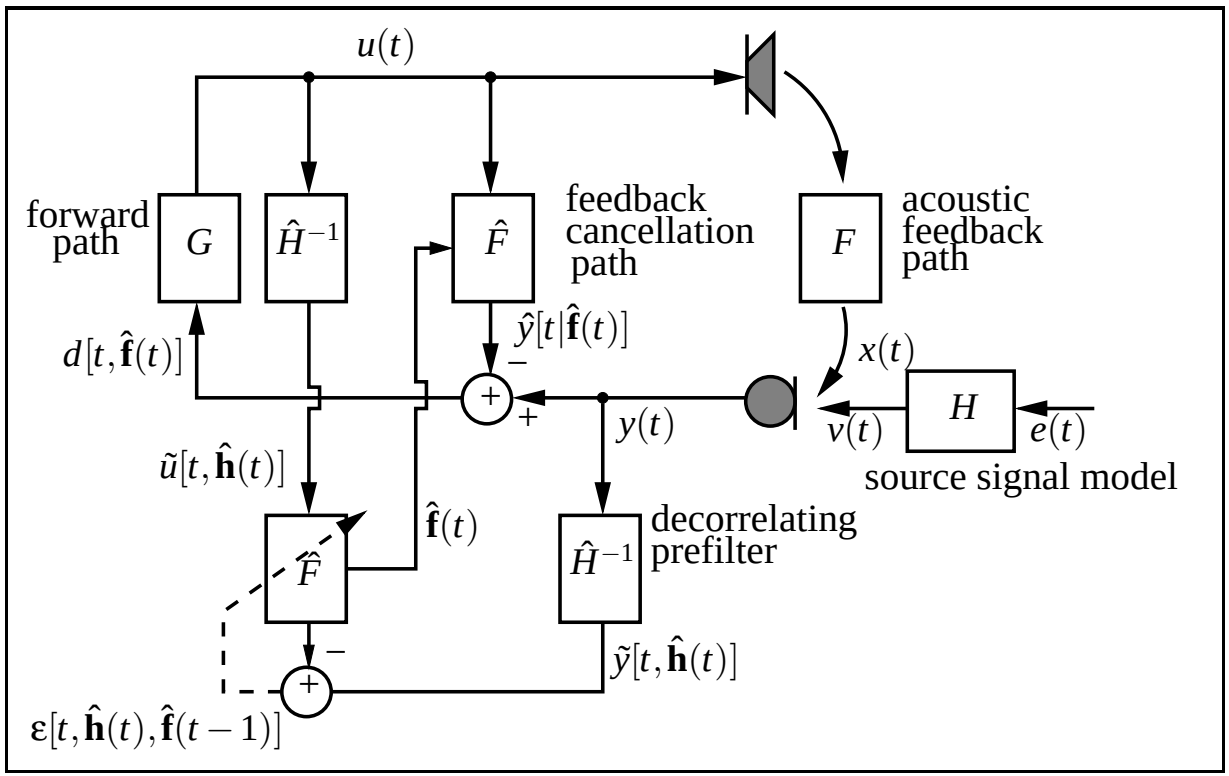
Minimize prediction error:

$$\min_{\mathbf{f}(t)} \sum_{k=1}^{t} \varepsilon^2(k)$$

KATHOLIEKE UNIVERSITEIT LEUVEN

Adaptive feedback cancellation

PEM-based AFC



KATHOLIEKE UNIVERSITEIT LEUVEN

Adaptive feedback cancellation

PEM-based AFC (single near-end signal model)

The AR model can be written as:

$$H_{\mathrm{AR}}(q,t) = \frac{1}{C(q,t)} = \frac{1}{1+\sum_{i=1}^{n_C} c^{(i)}(t) q^{-i}}$$

Near-end signal using AR model:

$$v(t) = \frac{1}{1+\sum_{i=1}^{n_C} c^{(i)}(t) q^{-i}} r(t).$$

Prediction error using AR model:

$$\boxed{\varepsilon(t) = H_{\mathrm{AR}}^{-1}(q,t)[y(t) - F(q,t)u(t)]}$$

- single AR model (Short-term predictor) fails to remove the periodicity
- white noise assumption is not valid for period signals such as voiced signals

KATHOLIEKE UNIVERSITEIT LEUVEN

Adaptive feedback cancellation

PEM-based AFC (cascaded near-end signal model)

Sinusoidal near-end signal model:

$$d(t) = \sum_{n=1}^{P} A_n \cos(\omega_n t + \phi_n) + r(t), \quad t = 1, ..., M$$

Cascaded near-end signal model:

$$y(t) = F(q,t)u(t) + H_1(q,t)H_2(q,t)r(t)$$

- $H_1(q,t)$ predicts the sinusoidal components in the near-end signal
- $H_2(q,t)$ predicts the noise components in the near-end signal
- A cascade of near-end signal models removes the coloring and periodicity

KATHOLIEKE UNIVERSITEIT LEUVEN

Adaptive feedback cancellation

PEM-based AFC (cascaded near-end signal model)

Prediction error using CPZLP+AR model:

$$\varepsilon(t) = H_{\text{CPZLP}}^{-1}(q,t)H_{\text{AR}}^{-1}(q,t)\left[y(t) - F(q,t)u(t)\right]$$

- CPZLP estimates the sinusoidal frequencies independently

Prediction error using Pitch+AR model:

$$\varepsilon(t) = H_{\text{Pitch}}^{-1}(q,t)H_{\text{AR}}^{-1}(q,t)\left[y(t) - F(q,t)u(t)\right]$$

Pitch estimation:

- motivated by voiced speech being quasi-periodic
- offers amplitude estimation and order estimation (number of harmonics)

KATHOLIEKE UNIVERSITEIT LEUVEN

Experimental set-up PEM-based AFC

Simulation parameters:

- The near-end sinusoidal model order is set to P = 15,10 and 5
- The near-end noise model order is set to 30.
- Both near-end signal models are estimated using 50% overlapping data windows of length M = 320 samples.
- The NLMS adaptive filter length is set to n_F = 200.
- The near-end signal is a 30 s speech signal at f_s= 16 kHz.
- The forward path gain $K(t)$ is set 3 dB below the maximum stable gain (MSG) without feedback cancellation.

Achievable amplification before instability occurs is measured by the MSG:

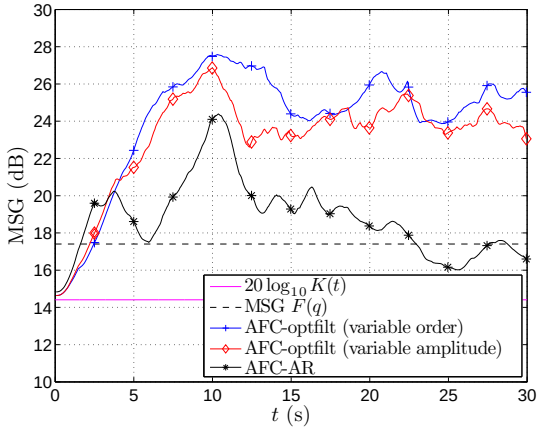
$$\text{MSG}(t) = -20\log_{10}\left[\max_{\omega\in\mathcal{P}} |J(\omega,t)[F(\omega,t)-\hat{F}(\omega,t)]|\right]$$

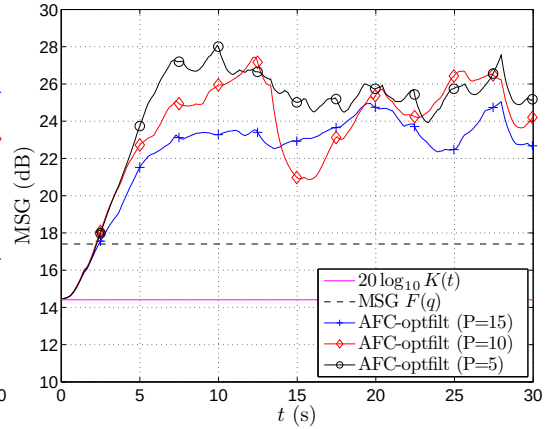
Misadjustment between the estimated feedback path $\hat{\mathbf{f}}(t)$ and the true feedback path $\mathbf{f}$:

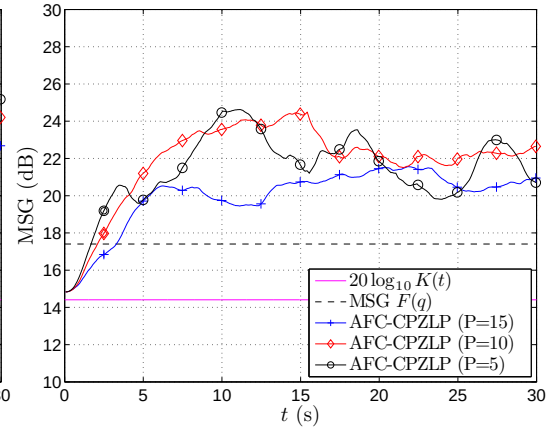
$$\text{MA}_F = 20\log_{10}\frac{||\hat{\mathbf{f}}(t)-\mathbf{f}||_2}{||\mathbf{f}||_2}.$$

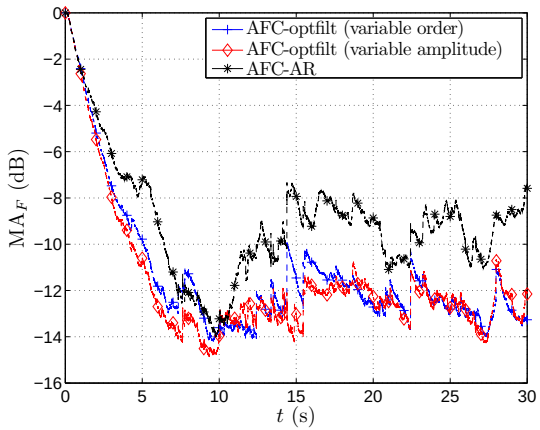
KATHOLIEKE UNIVERSITEIT
LEUVEN

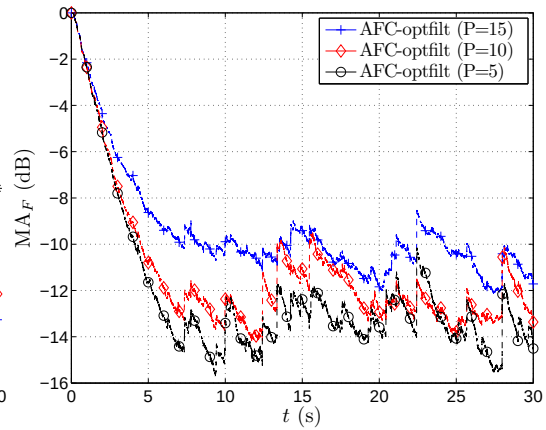
Experimental results PEM-based AFC

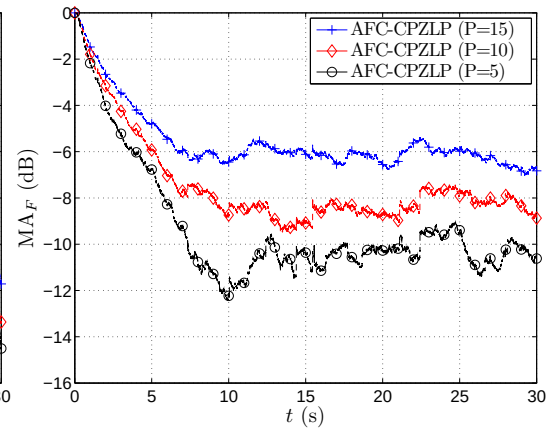












KATHOLIEKE UNIVERSITEIT LEUVEN

Conclusion

Summary:

- Different extension of the SDW-MWF algorithms
- A robust Capon beamformer based on a predefined uncertainty region
- PEM-based AFC using cascaded near-end signal model

Current work on MWF based NR:

- Tracking: continously update of correlation matrices
- Robustness: use of prior knowledge for updating correlation matrices

Current work on PEM-based AFC:

- Near-end signal model: using voice-unvoice detector
- Prediction error filter: sinusoidal subtraction

Current work on combined MWF based NR and DRC:

- Avoid residual noise amplification
- Reuse the conditional SPP information

KATHOLIEKE UNIVERSITEIT
LEUVEN

End of presentation

Thank you!.....Questions?