KATHOLIEKE UNIVERSITEIT LEUVEN

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

Signal Processing in Hearing Aids: Noise Reduction, Compression and Feedback Cancellation

Kim Ngo

University of Oldenburg, Germany.

Tuesday Jan. 11th, 2011.

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)
- Combined noise reduction and dynamic range compression
- (undesired interaction effect, residual noise amplification)
- Prediction-error-method based adaptive feedback cancellation
- (bias reduction, near-end signal modelling)
- Conclusion
- (current/future work)

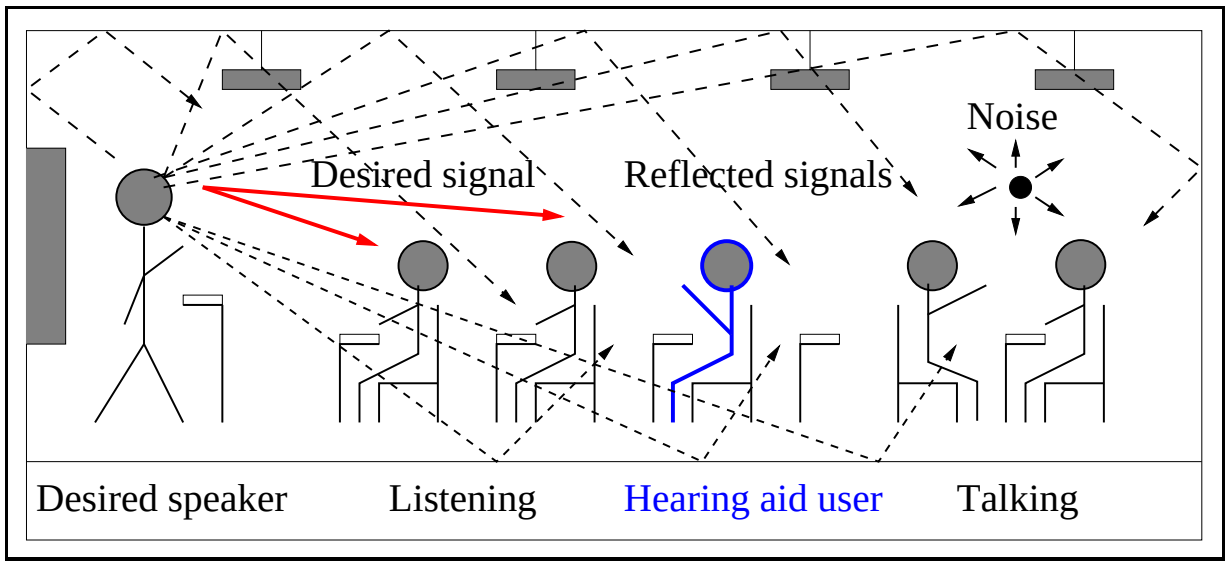
KATHOLIEKE UNIVERSITEIT LEUVEN

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)



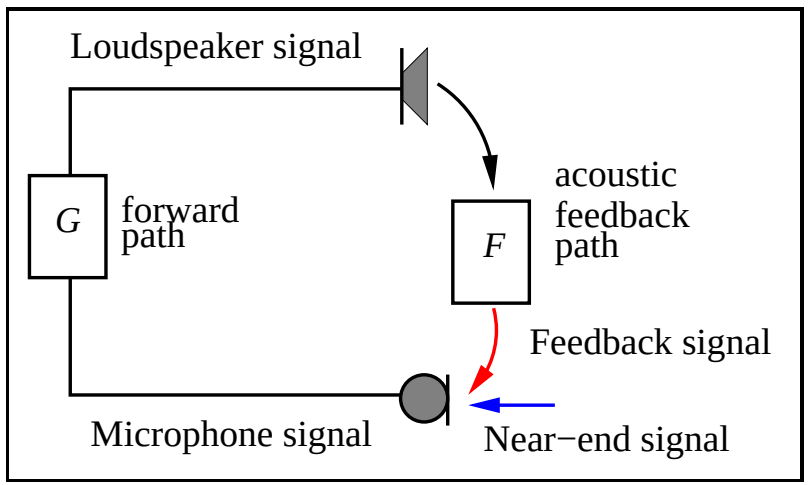
Objective of NR:

- 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:

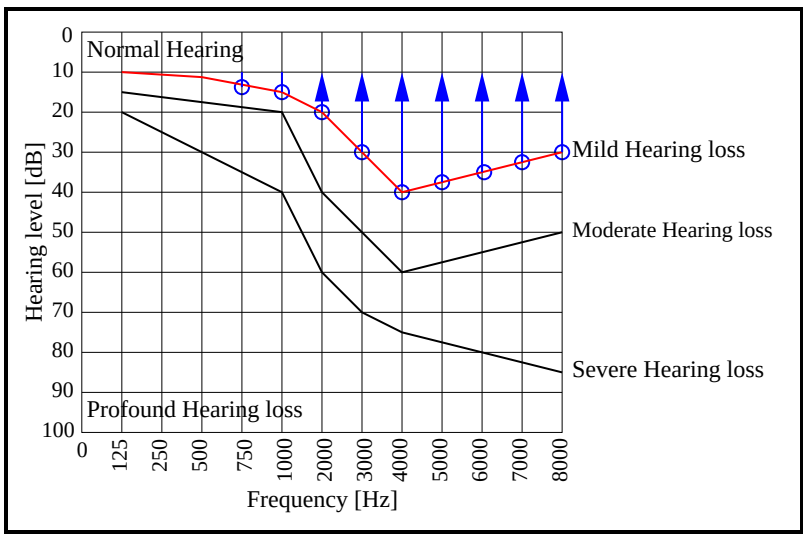
- Increase maximum stable gain (MSG)
- Reduce bias and convergence (misadjustment)
- Minimize speech distortion (sound quality, howling)

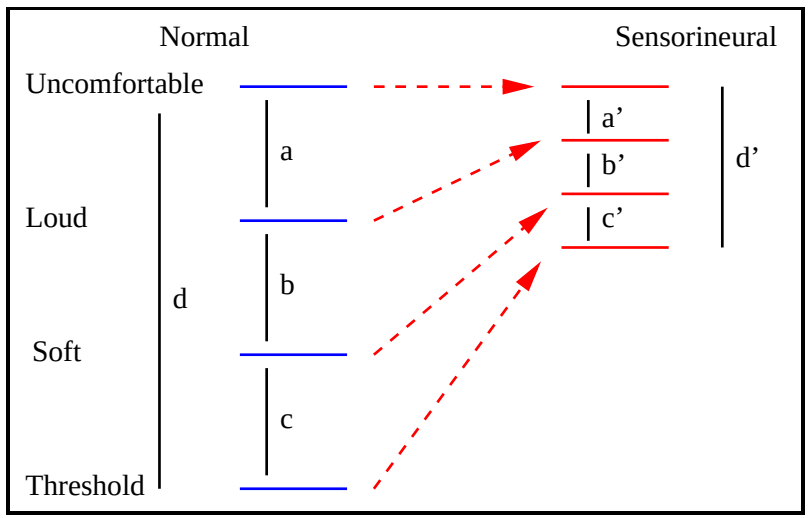
KATHOLIEKE UNIVERSITEIT LEUVEN

Introduction

Reduced dynamic range

- Audibility an important first step in improving the intelligibility of a speech signal.





Objective of DRC:

- Mapping the wide dynamic range of speech into the reduced dynamic range.
- Automatically adjust the gain based on the intensity level.
- 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

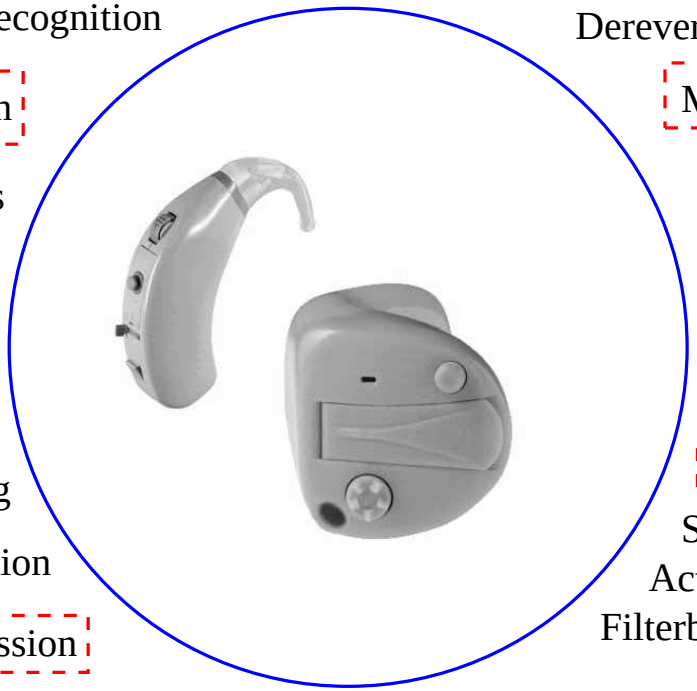
AD/DA Converters

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

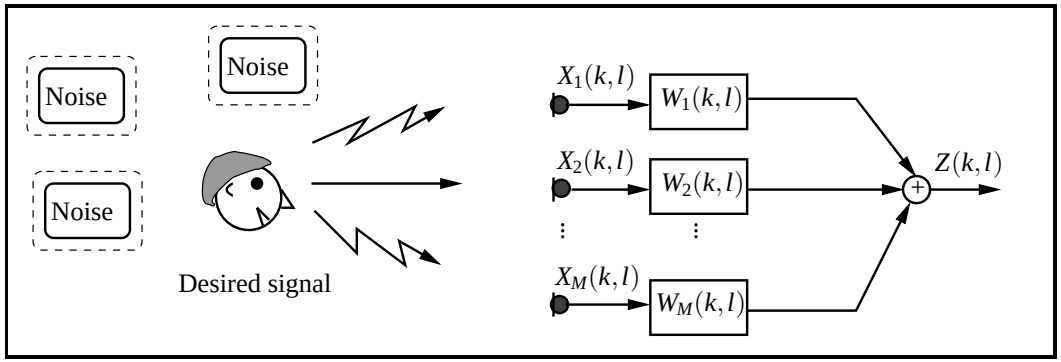
Source Separation

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:

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

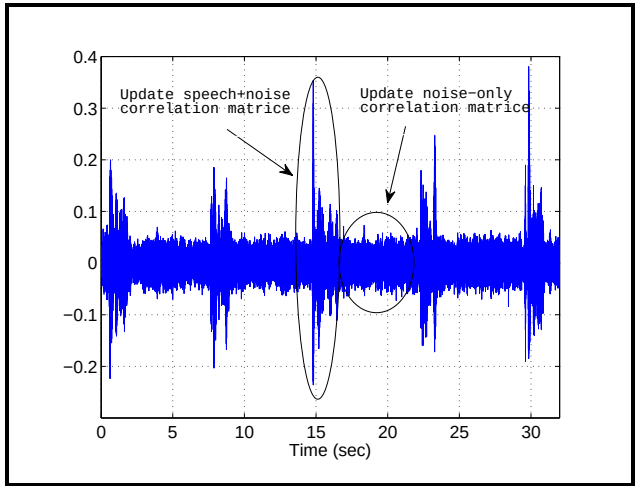
Concept:

- 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

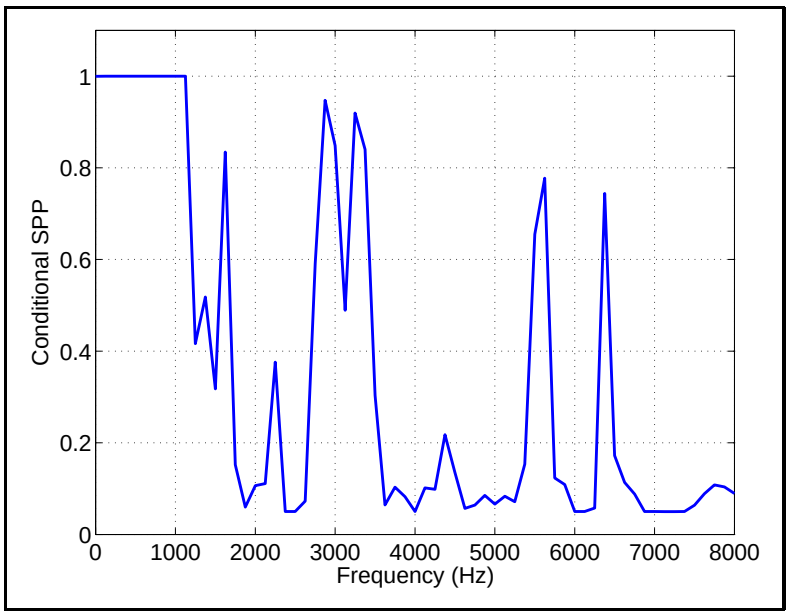
Two-state speech model:

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

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.

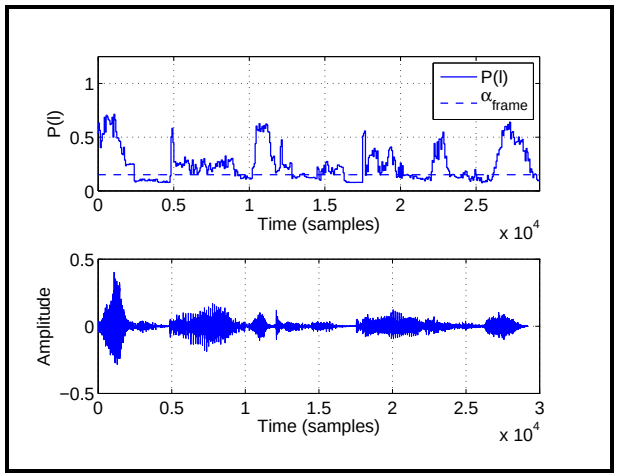
KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

H_0 and H_1 state detection based on SPP

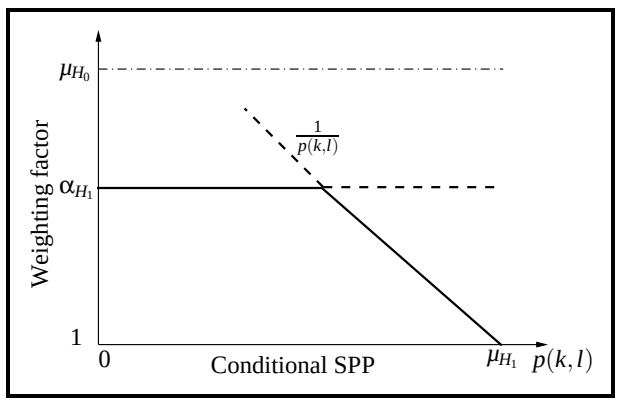
Binary decision:

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



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):

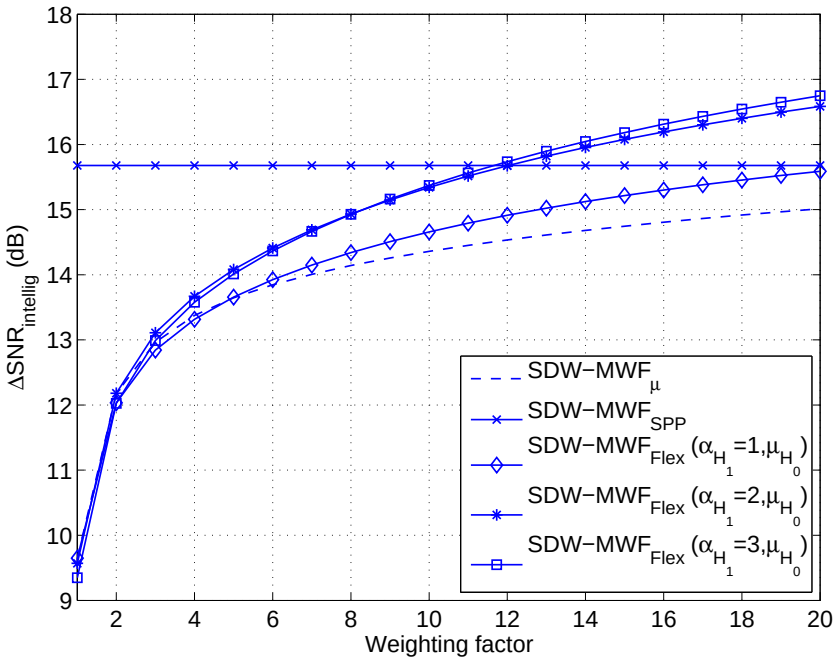
$$\Delta\mathrm{SNR}_{\mathrm{intellig}} = \sum_i I_i \left(\mathrm{SNR}_{i,out} - \mathrm{SNR}_{i,in}\right)$$

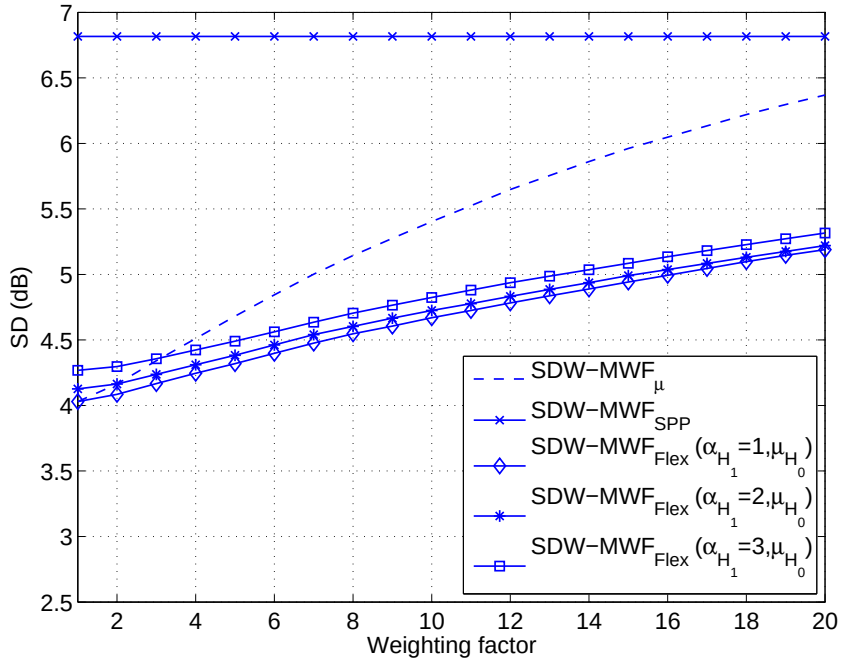
frequency-weighted log-spectral signal distortion (SD)

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

KATHOLIEKE UNIVERSITEIT LEUVEN

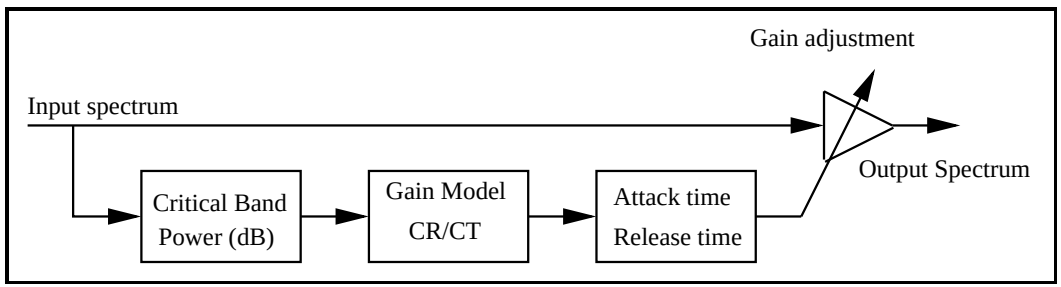
Experimental results MWF based NR

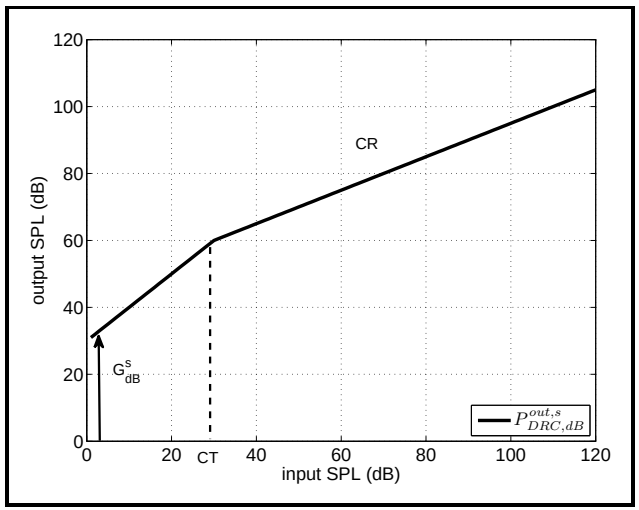




KATHOLIEKE UNIVERSITEIT LEUVEN

Dynamic Range Compression





- Compression Threshold (CT) (point at which the slope change)
- Compression Ratio (CR) (Steepness of the slope)
- Amplification gain G_{dB}

$$P_{\text{lin,dB}}(k,l) = P^{\text{in,s}}_{\text{DRC,dB}}(k,l) + G^s_{\text{dB}}$$

$$P_{\text{cp,dB}}(k,l) = \text{CT} + \frac{1}{\text{CR}} \cdot (P^{\text{in,s}}_{\text{DRC,dB}}(k,l) - \text{CT}) + G^s_{\text{dB}}$$

The output power in dB is then given by

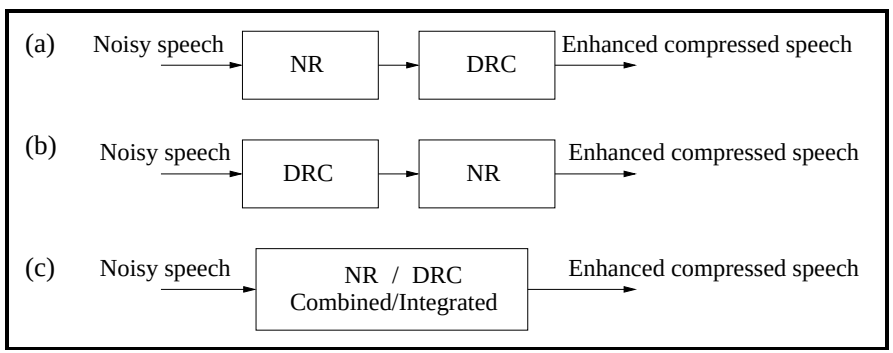
$$P^{\text{out,s}}_{\text{DRC,dB}}(k,l) = \begin{cases} P_{\text{lin,dB}}(k,l) & \text{if } P^{\text{in,s}}_{\text{DRC,dB}}(k,l) < \text{CT} \\ P_{\text{cp,dB}}(k,l) & \text{if } P^{\text{in,s}}_{\text{DRC,dB}}(k,l) \geq \text{CT} \end{cases}$$

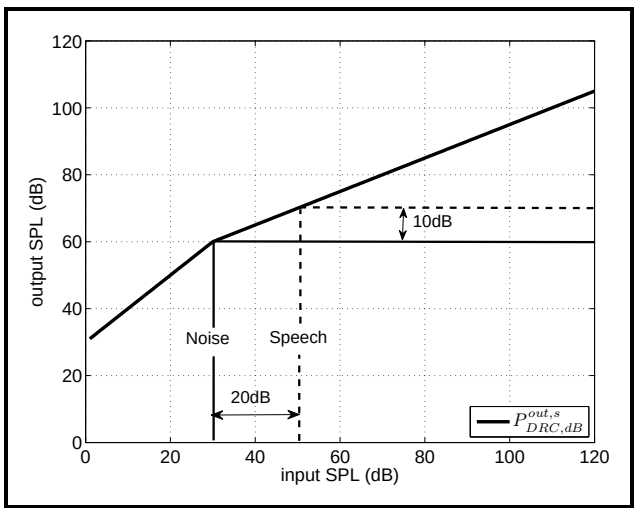
KATHOLIEKE UNIVERSITEIT LEUVEN

Dynamic Range Compression

DRC in presence of noise

- When NR and DRC are serially concatenated, undesired interaction effects occur
- DRC can counteract NR by amplifying the residual noise after NR
- Degrades the SNR and defeats the purpose of using NR
- DRC does not distinction between speech and noise dominant segments
- Low intensity segments are amplified equally (including residual noise)

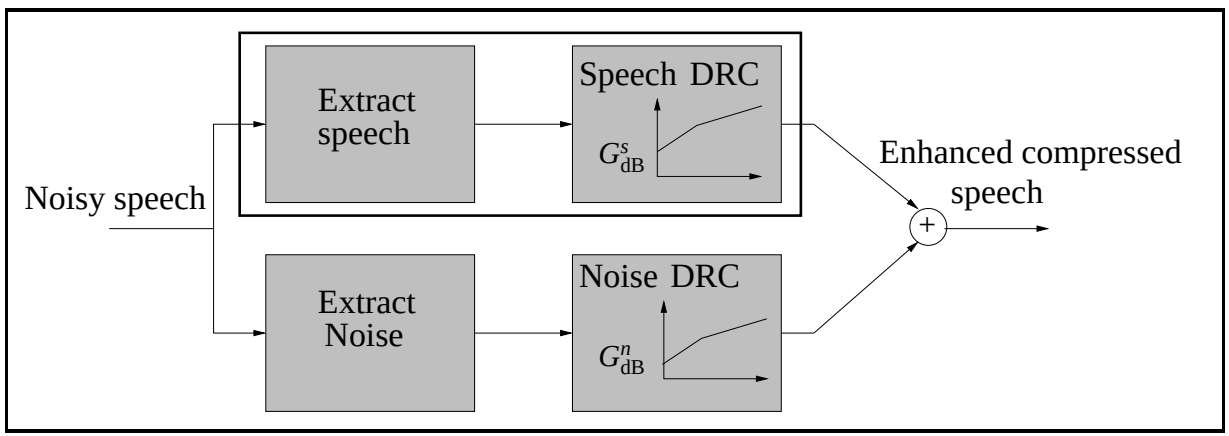




KATHOLIEKE UNIVERSITEIT LEUVEN

Combined MWF and DRC

Ideal system



Gain difference between speech DRC and noise DRC:

$$\Delta G_{\mathrm{dB}} = G^s_{\mathrm{dB}} - G^n_{\mathrm{dB}}$$

KATHOLIEKE UNIVERSITEIT LEUVEN

Combined MWF and DRC

Estimating speech component:

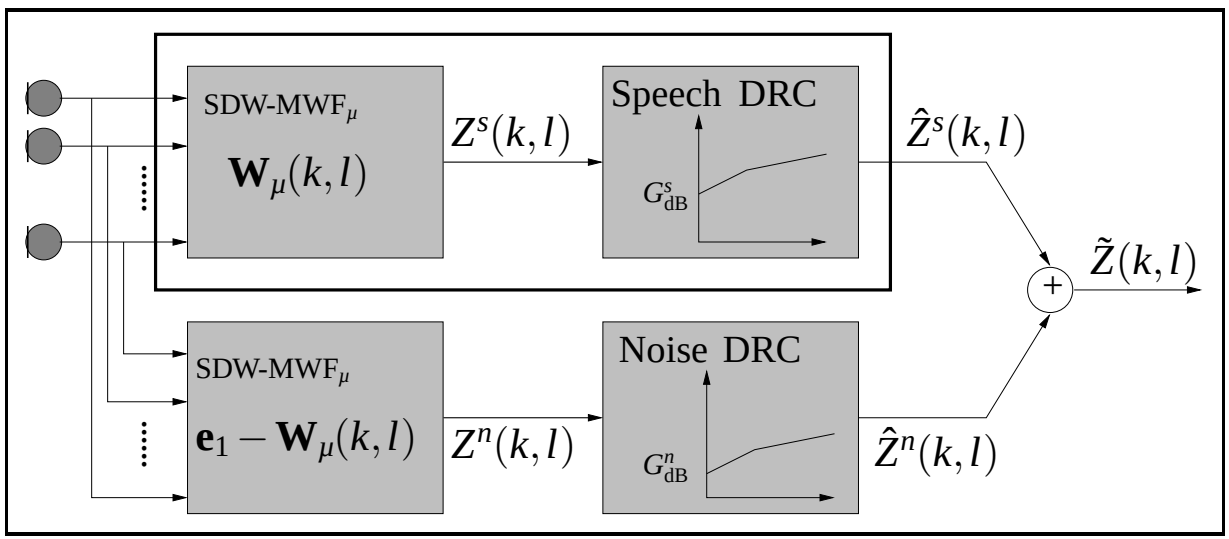
$$\mathbf{W}_\mu(k,l) = \left[\mathbf{R}^s(k,l) + \mu\mathbf{R}^n(k,l)\right]^{-1}\mathbf{R}^s(k,l)\mathbf{e}_1$$

$$Z^s(k,l) = \mathbf{W}_\mu^H(k,l)\mathbf{X}(k,l)$$

Estimating noise component:

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

$$Z^n(k,l) = \mathbf{V}_\mu^H(k,l)\mathbf{X}(k,l)$$



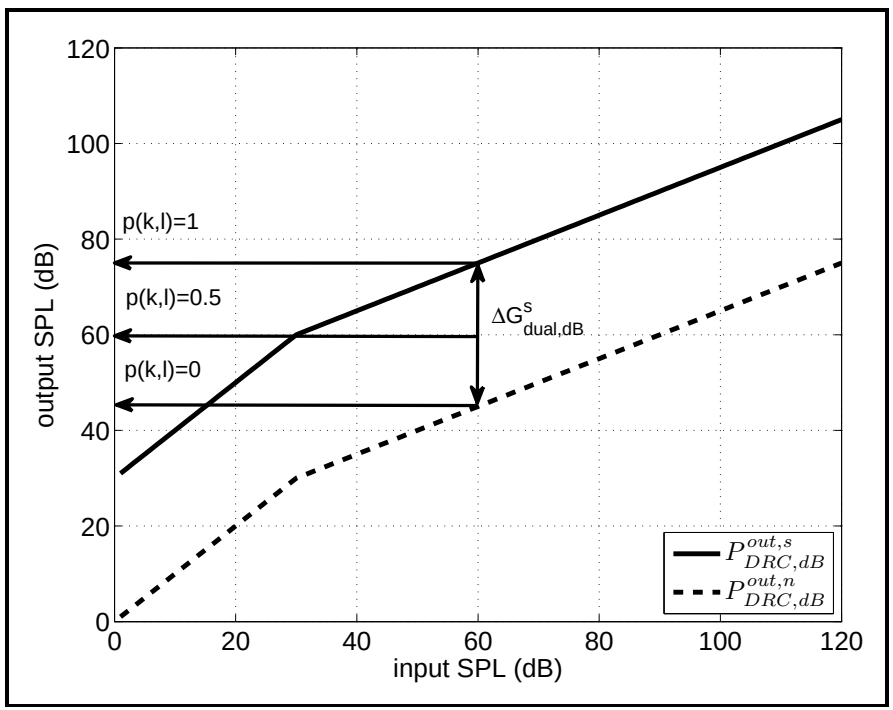
KATHOLIEKE UNIVERSITEIT LEUVEN

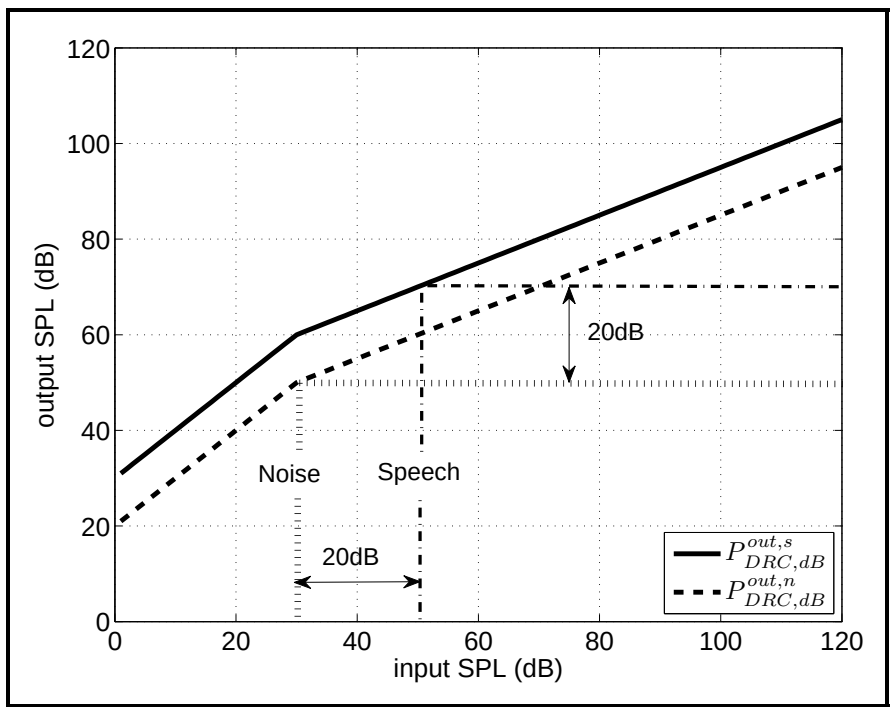
Combined MWF and DRC

Dual-DRC

Concept:

- Reusing the conditional SPP estimated in SDW-MWF$_{SPP}$
- Using a switchable compression characteristic
- If $(p(k,l) = 1)$ the speech DRC applied.
- If $(p(k,l) = 0)$ it is undesirable to amplify and the noise DRC is applied.
- For the in-between cases a weighted sum of the two DRC curves is used





KATHOLIEKE UNIVERSITEIT LEUVEN

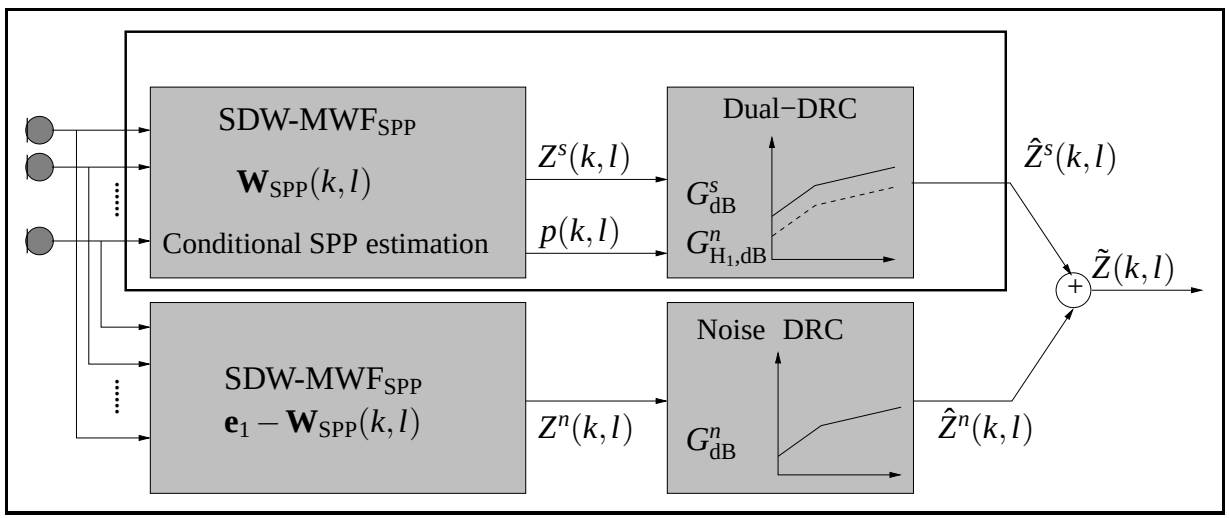
Combined MWF and DRC

Output power in dual-DRC:

$$P^{out}_{\text{dual-DRC,dB}}(k,l) = p(k,l) \cdot P^{out,s}_{\text{DRC,dB}}(k,l) + (1 - p(k,l)) \cdot P^{out,n}_{\text{DRC,dB}}(k,l)$$

Gain difference in dual-DRC:

$$\Delta G_{\text{dual,dB}} = G^{s}_{\text{dB}} - G^{n}_{\text{H}_1,\text{dB}}$$



KATHOLIEKE UNIVERSITEIT LEUVEN

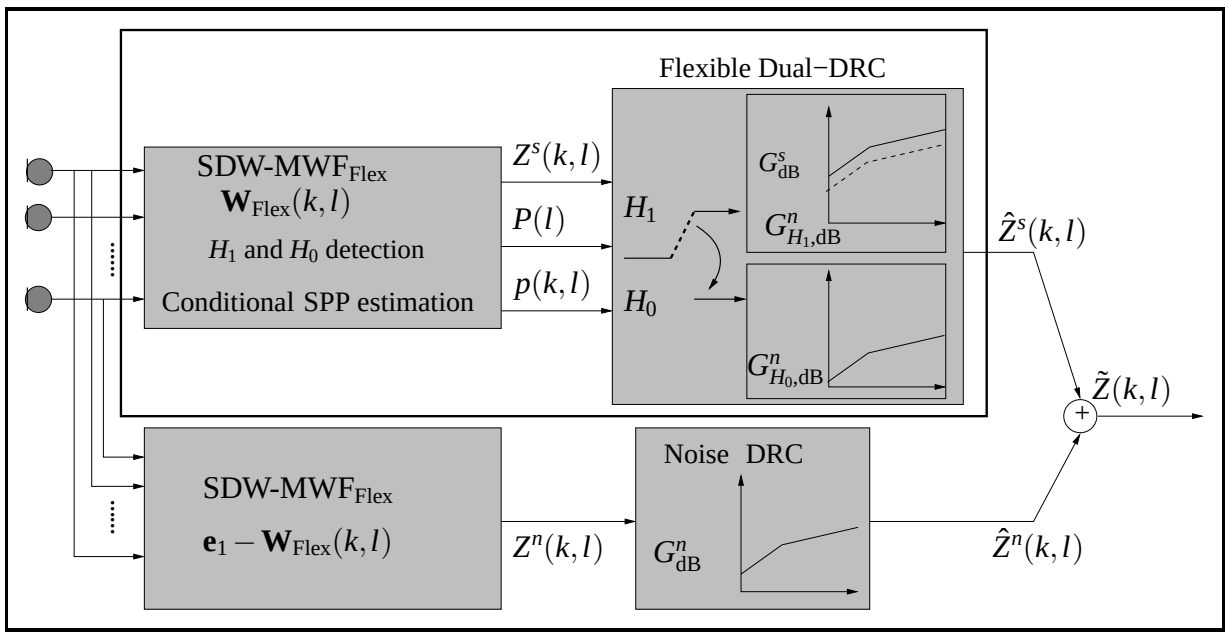
Combined MWF and DRC

Output power in flexible dual-DRC:

$$P^{out}_{\text{flex-DRC,dB}}(k,l) =$$
$$P(l)\left[p(k,l)P^{out,s}_{\text{DRC,dB}}(k,l) + (1-p(k,l))P^{out,n}_{\text{DRC,dB}}(k,l)\right] + (1-P(l))P^{out,n}_{\text{DRC,dB}}(k,l)$$

Gain difference in flexible dual-DRC:

$$\Delta G_{\text{flex,dB}} = P(l)\left[G^{s}_{\text{dB}} - G^{n}_{\text{H}_1\text{dB}}\right] + (1-P(l))G^{n}_{\text{H}_0\text{dB}}$$



KATHOLIEKE UNIVERSITEIT LEUVEN

Experimental set-up Combined MWF and DRC

Simulation parameters for DRC:

- The DRC is implemented based on critical bands
- Input level is set to 65dB SPL at the microphones.
- Attack and release time are set to at=10ms and rt=20ms.
- Compression ratio CR=2.
- Compression threshold is set to CT=30dB.

Simulation parameters for NR:

Method	SDW-MWF$_{\mu}$	SDW-MWF$_{\text{SPP}}$	SDW-MWF$_{\text{Flex}}$
Input SNR	0dB	0dB	0dB
ΔSNR	13.1dB	13.2dB	13.9dB
SD	4.2dB	4.3dB	4.2dB

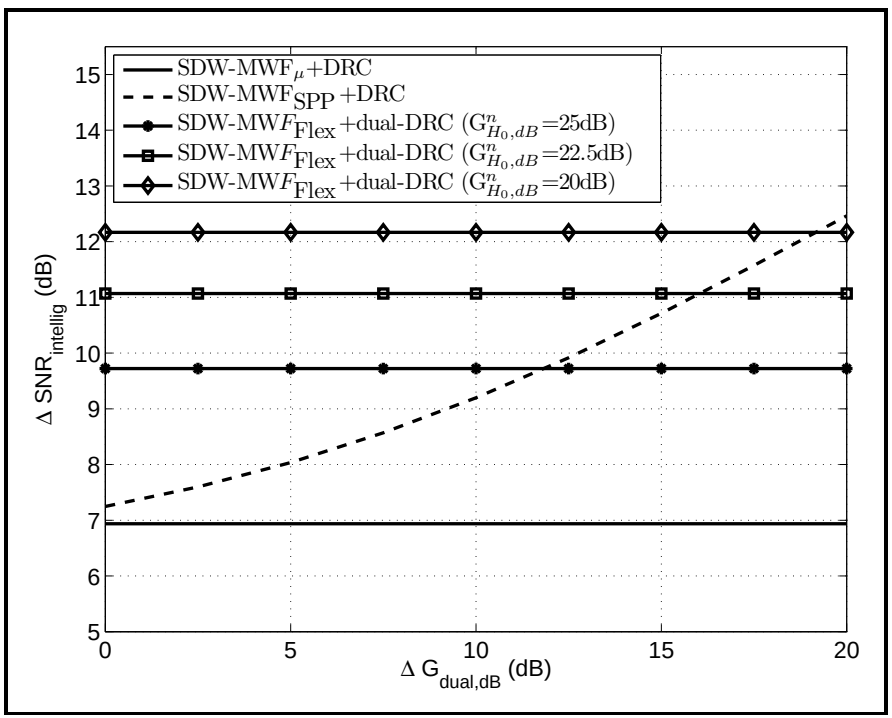
Table: SNR improvement and SD of the different SDW-MWF based NR.

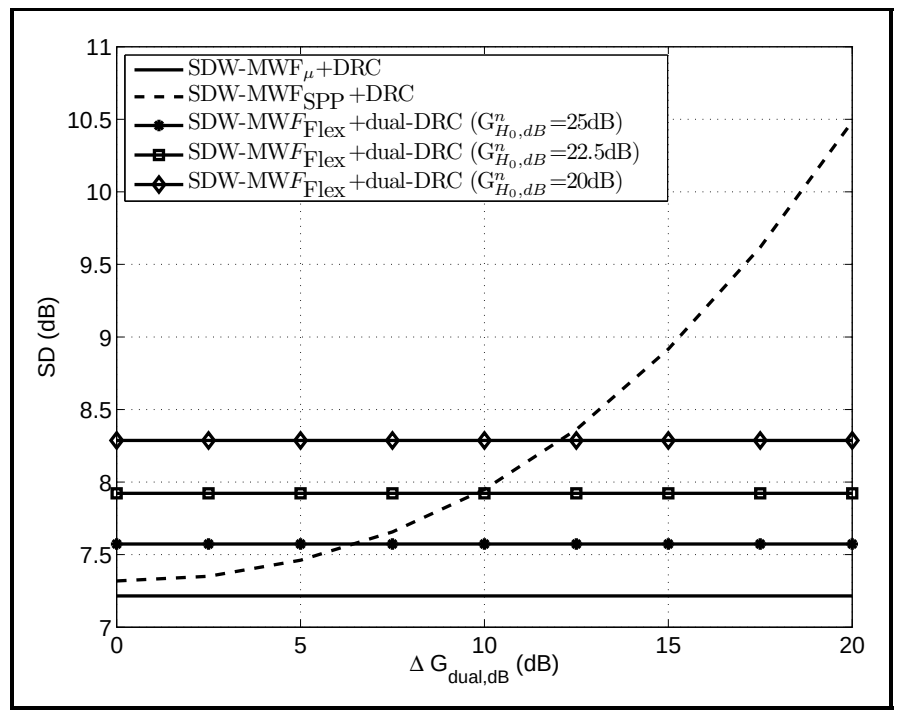
KATHOLIEKE UNIVERSITEIT LEUVEN

Experimental results Combined MWF and DRC

	G^{s}_{dB}	G^{n}_{dB}	$G^{n}_{H_1,\mathrm{dB}}$	$G^{n}_{H_0,\mathrm{dB}}$
SDW-MWF$_{\mu}$	30dB	N/A	N/A	N/A
SDW-MWF$_{\mathrm{SPP}}$	30dB	N/A	10dB-30dB	N/A
SDW-MWF$_{\mathrm{Flex}}$	30dB	N/A	27.5dB	20dB-25dB

Table: Gain settings for first experiment.



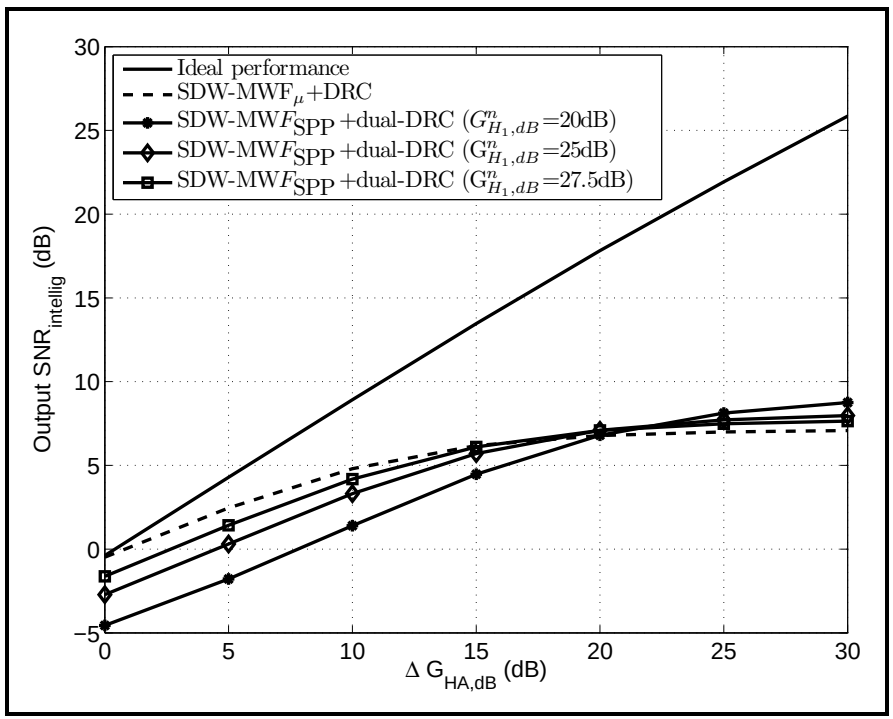


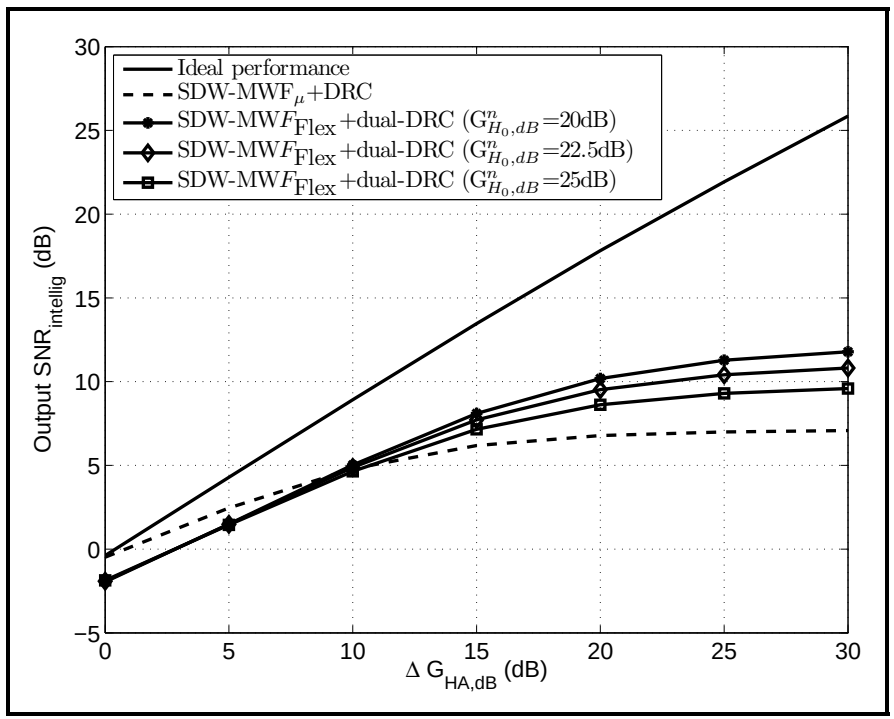
KATHOLIEKE UNIVERSITEIT LEUVEN

Experimental results Combined MWF and DRC

	G^s_{dB}	G^n_{dB}	$G^n_{H_1,\text{dB}}$	$G^n_{H_0,\text{dB}}$
SDW-MWF$_\mu$	30dB	0dB-30dB	N/A	N/A
SDW-MWF$_{\text{SPP}}$	30dB	0dB-30dB	20dB-27.5dB	N/A
SDW-MWF$_{\text{Flex}}$	30dB	0dB-30dB	27.5dB	20dB-25dB

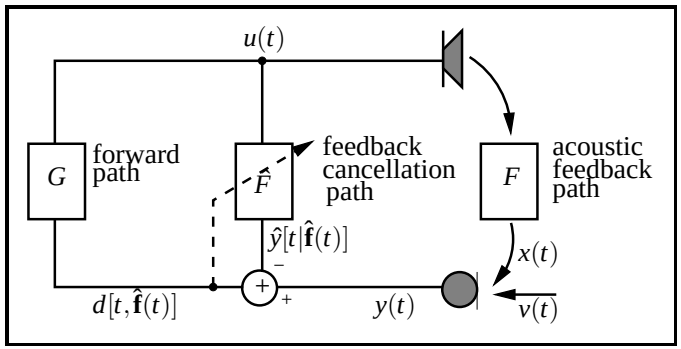
Table: Gain settings for second experiment.





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

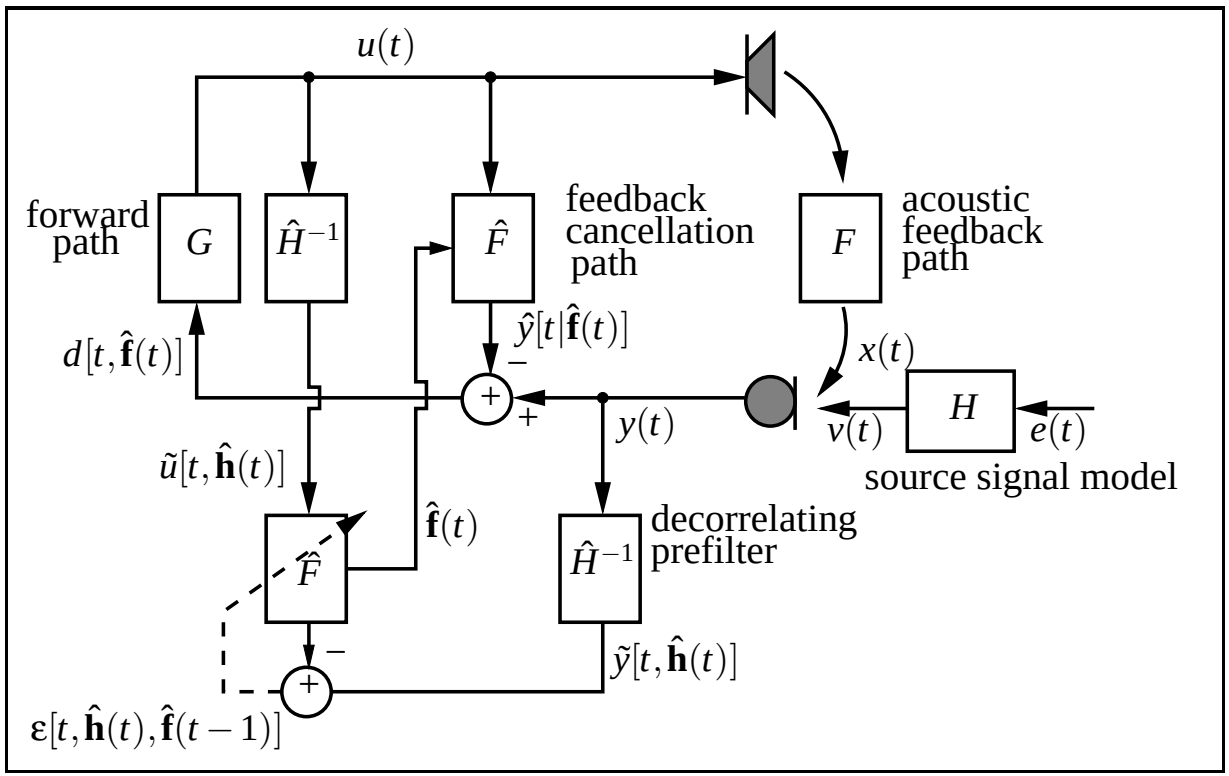
Minimize prediction error:

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

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

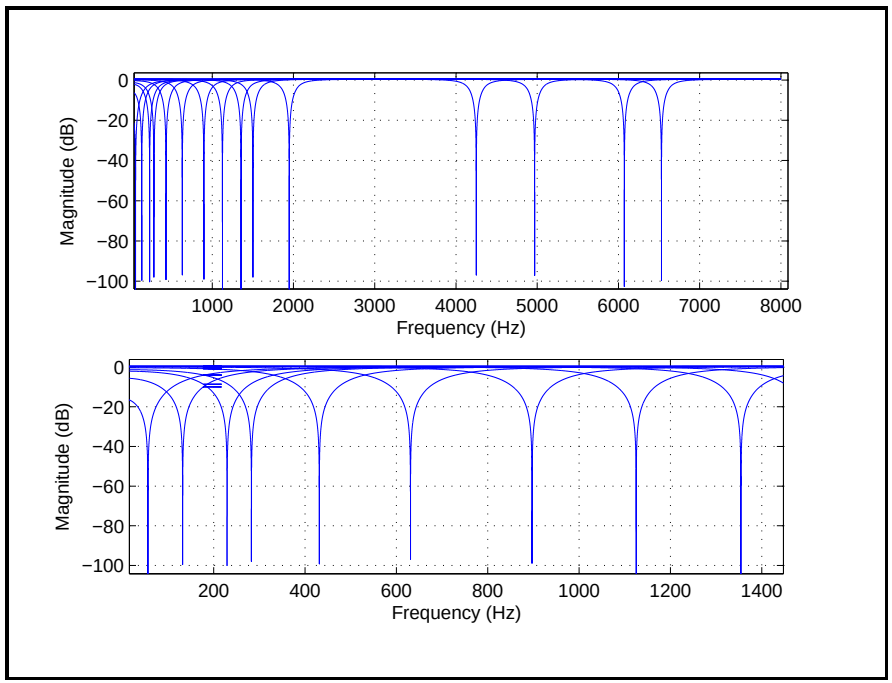
Adaptive feedback cancellation

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

Prediction error using CPZLP+AR model:

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

- CPZLP estimates the sinusoidal frequencies independently
- Applies equal (infinity) suppression for all frequencies



KATHOLIEKE UNIVERSITEIT LEUVEN

Adaptive feedback cancellation

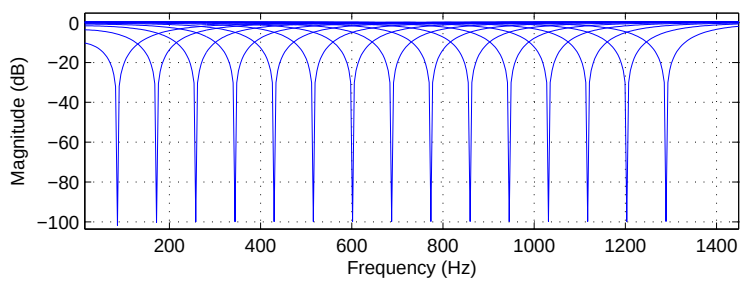
PEM-based AFC (cascaded near-end signal model)

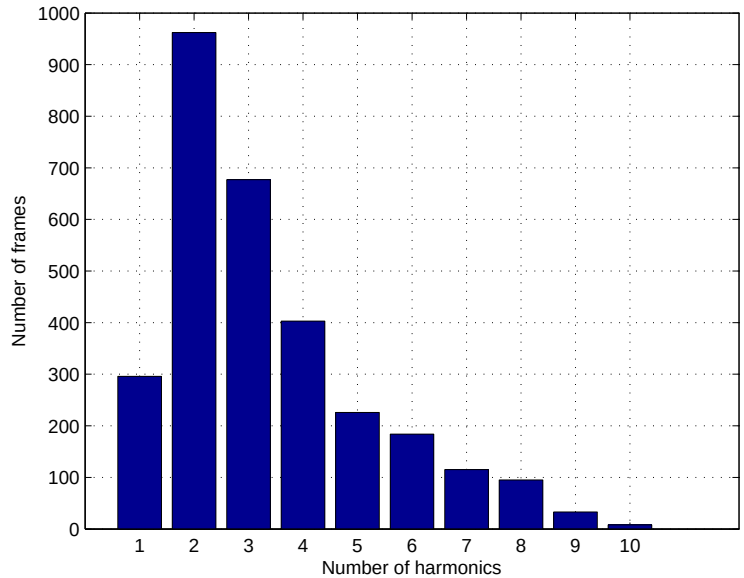
Prediction error using Pitch+AR model:

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

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:

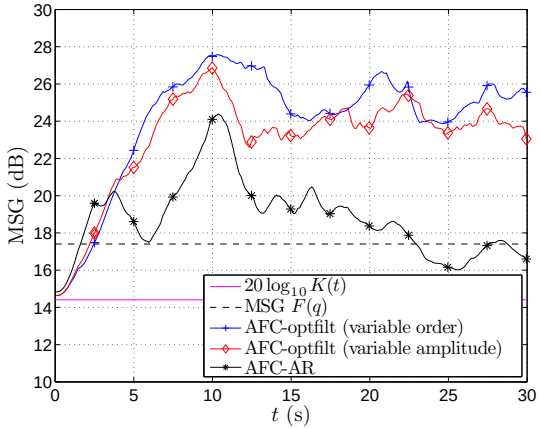
$$\text{MSG}(t) = -20\log_{10}\left[\max_{\omega\in\mathcal{P}} |J(\omega,t)[F(\omega,t)-\hat{F}(\omega,t)]|\right]$$

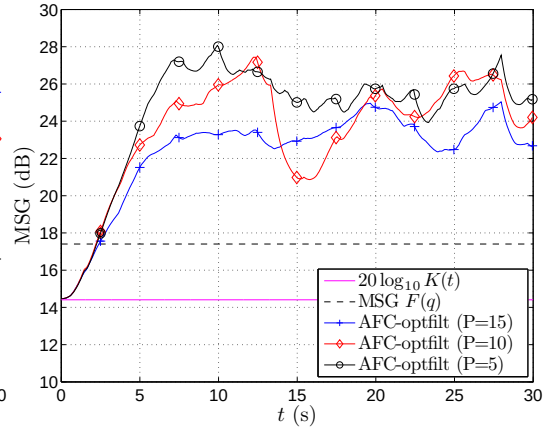
Misadjustment between the estimated feedback path $\hat{\mathbf{f}}(t)$ and the true feedback path $\mathbf{f}$:

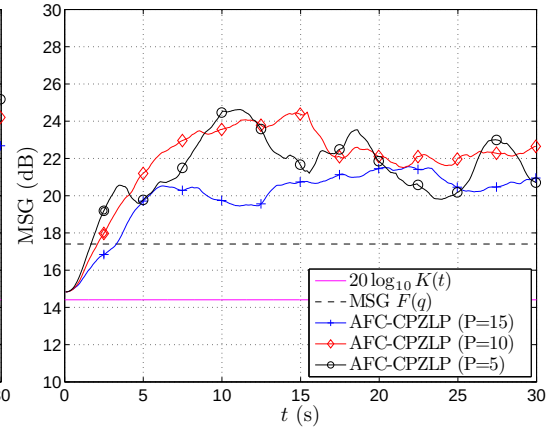
$$\text{MA}_F = 20\log_{10}\frac{||\hat{\mathbf{f}}(t)-\mathbf{f}||_2}{||\mathbf{f}||_2}.$$

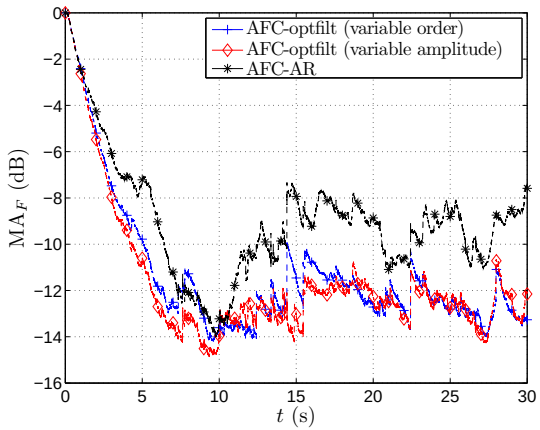
KATHOLIEKE UNIVERSITEIT
LEUVEN

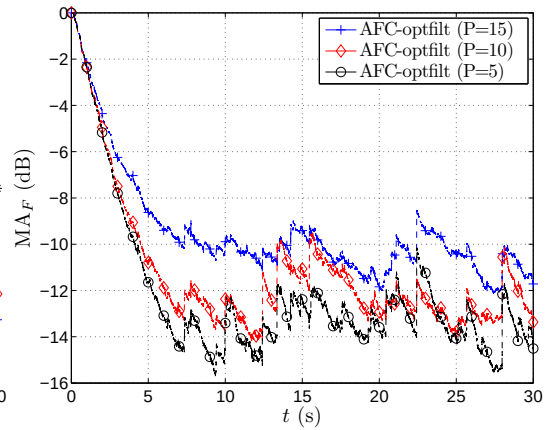
Experimental results PEM-based AFC

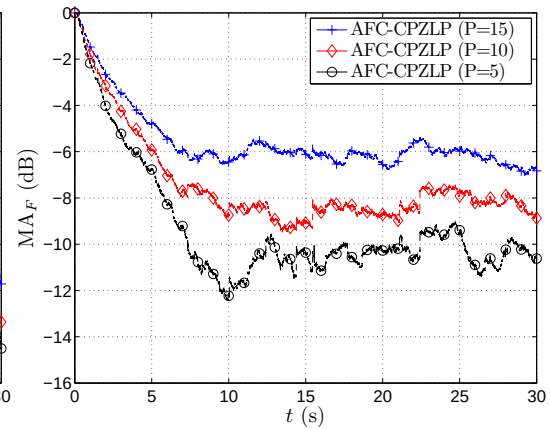












KATHOLIEKE UNIVERSITEIT LEUVEN

Conclusion

Summary:

- Different extension of the SDW-MWF algorithms
- Combined MWF based NR and DRC
- PEM-based AFC using cascaded sinusoidal 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 combined MWF based NR and DRC:

- Fully integrated approach

Current work on PEM-based AFC:

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

Current work on beamformer:

- A robust Capon beamformer based on a predefined uncertainty region

Current work on combined AFC and NR:

- Interaction of single-channel NR on AFC

KATHOLIEKE UNIVERSITEIT
LEUVEN

End of presentation

Thank you!.....Questions?