

Improved Signal Processing in Hearing Aids: A system Approach

Kim Ngo

ESAT-SCD, Katholieke Universiteit Leuven, Belgium

EST-SIGNAL Meeting September 2009.

KATHOLIEKE UNIVERSITEIT LEUVEN



Outline

- Introduction (Hearing aids, hearing loss, acoustic feedback, background noise).
- Problem statement and motivation.
- Multi-Channel Wiener Filter (MWF) based Noise Reduction (NR).
- SDW-MWF based aprroach to integrate NR and DRC.
- Adaptive Feedback Cancellation (AFC).
- Conclusion.
- Publications.
- Timeline
- Current and future work.



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

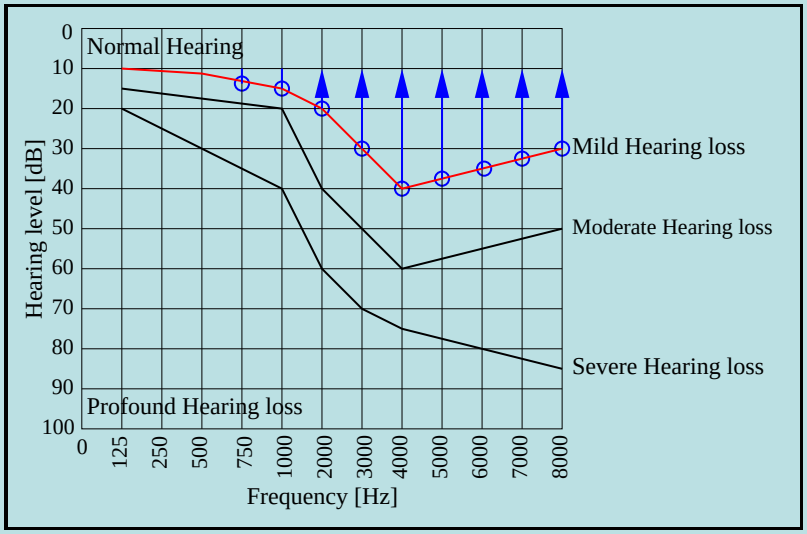
Active Noise Control

Filterbank Design

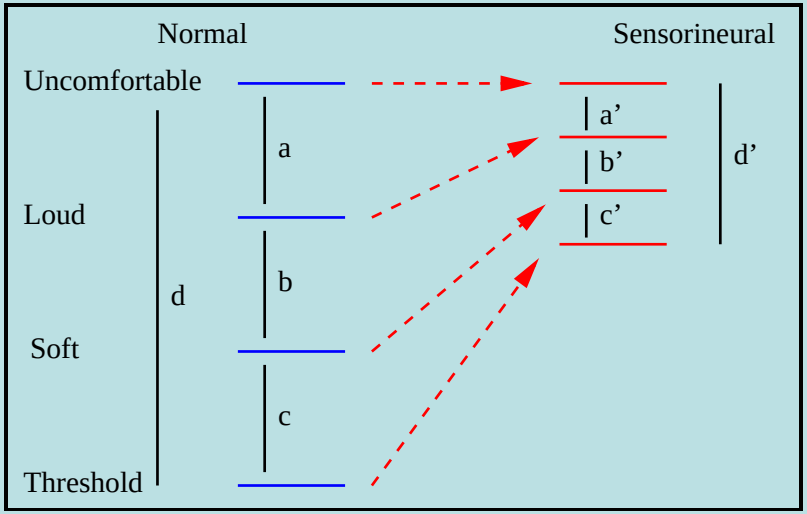


Introduction

Sensorineural hearing loss



Dynamic Range

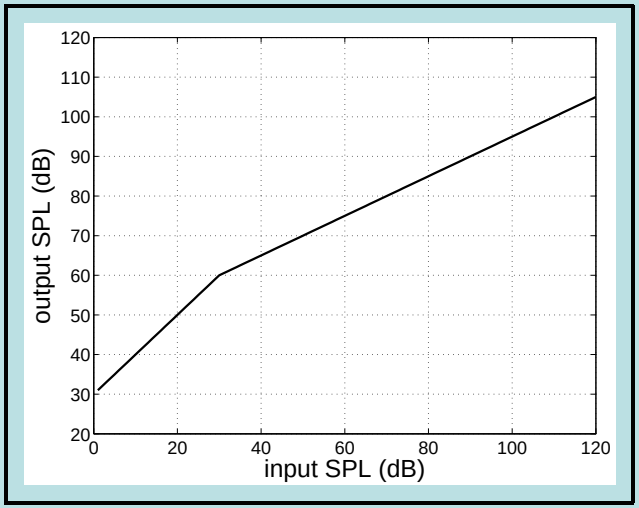


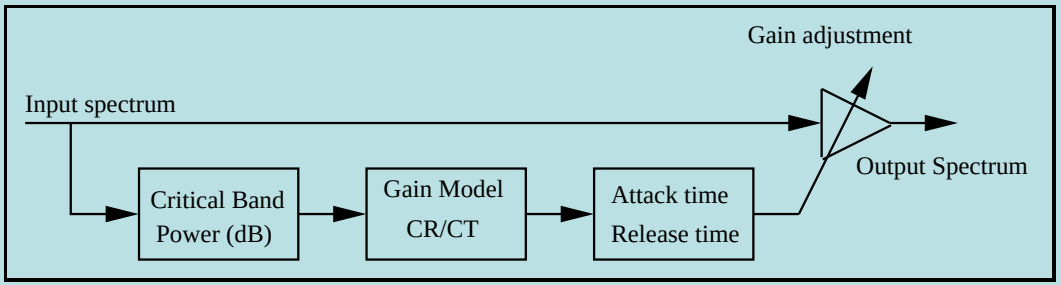
- Hearing threhold increases with increasing frequency.
- Threshold of hearing raised as a result of the hearing loss.
- Threshold of loudness discomfort remains the same.
- Reduced dynamic range (threshold to discomfort).



Dynamic Range Compression

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





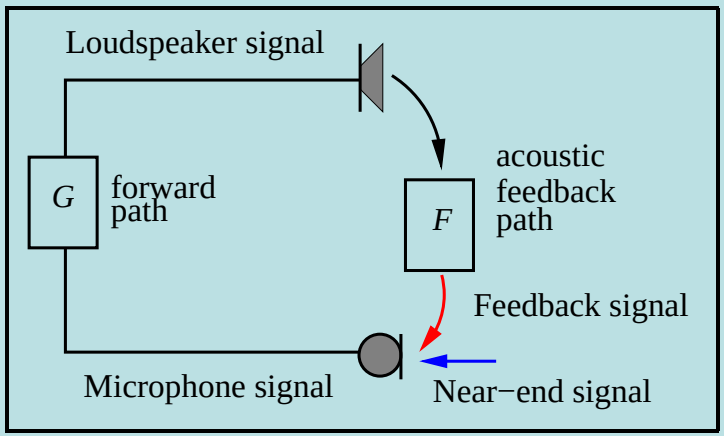
- Automatically adjust the gain based on the intensity level.
- High intensity attenuated - low intensity amplified.
- Compression Threshold (CT) (point at which the slope change)
- Compression Ratio (CR) (Steepness of the slope)
- Amplification gain G_{dB}

Objective:

- Mapping the wide dynamic range of speech into the reduced dynamic range.
- Weak sounds audible - loud sounds not uncomfortably loud.



Acoustic Feedback



- 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.
- Correlation between near-end signal and loudspeaker signal.
- Standard adaptive filtering converge to a biased solution.

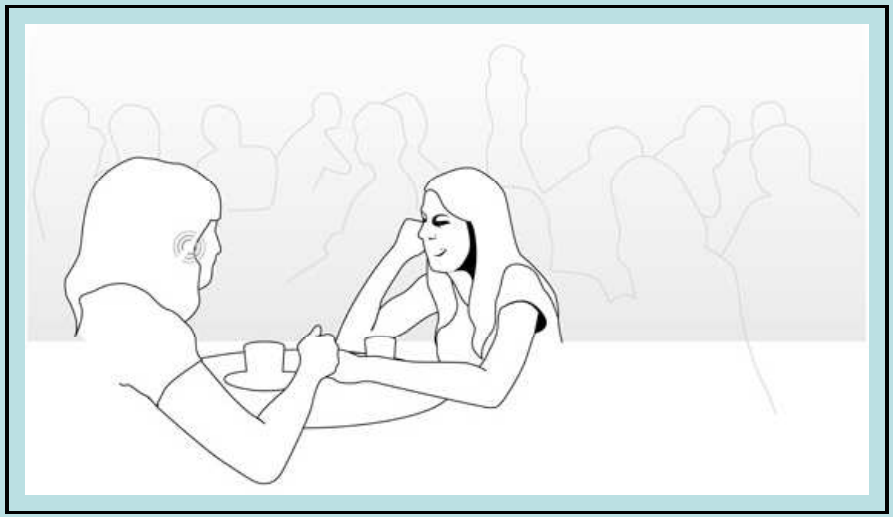
Objective:

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



Background Noise

- Reduced frequency resolution (separating sounds of different frequencies)
- Reduced temporal resolution (intense sounds mask weaker sounds)



Hearing aid users

- Understanding speech in noise is a major problem
- multiple speakers, fans, traffic etc.
- Reduces the intelligibility of speech.
- More sensitive to the noise level.
- Need higher SNR to communicate.

Objective:

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



Problem Statement and Motivation

- Compensation of sensorineural hearing loss requires NR, DRC and AFC.
- The general problem of NR, DRC and AFC is not new.
- Each of these areas are usually developed and evaluated independently.

Existing mehtod

- Hearing aids typically use a serial concatenation of NR and DRC and AFC
- Counteract and limit functionality of other algorithms

Short-term objective:

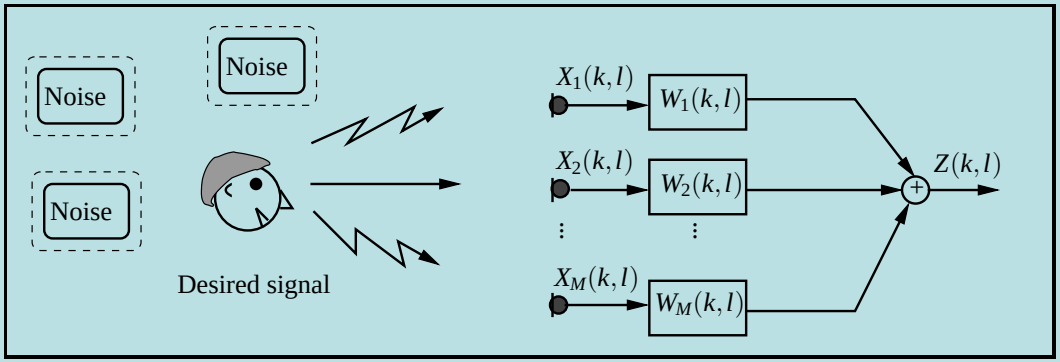
- Development on Multi-channel NR (SDW-MWF).
- Integration of SDW-MWF and DRC.
- Development on Adaptive feedback Cancellation (PEM-based AFC)
- Analyse any undesired effects in the integration process.

Long-term objective:

- Integration of NR, DRC and AFC into one signal processing scheme.
- SNR improvement vs. Audibility vs. speech distortion vs. Increase MSG



Speech Distortion Weighted Multi-channel Wiener Filter (SDW-MWF)



Frequency-domain microphone signals,

$$\mathbf{X}(k,l) = \mathbf{X}^s(k,l) + \mathbf{X}^n(k,l) \tag{1}$$

MWF MMSE criterion,

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

SDW-MWF MMSE criterion,

$$\mathbf{W}^*(k,l) = \arg\min_{\mathbf{W}} \ \varepsilon\{|X_1^s(k,l) - \mathbf{W}^H\mathbf{X}^s(k,l)|^2\} + \mu\varepsilon\{|\mathbf{W}^H\mathbf{X}^n(k,l)|^2\} \tag{3}$$

Optimal SDW-MWF

$$\mathbf{W}^*(k,l) = \left(\mathbf{R}^s(k,l) + \mu\mathbf{R}^n(k,l)\right)^{-1}\mathbf{R}^s(k,l)\mathbf{e}_1 \tag{4}$$

Output of the SDW-MWF can be written as

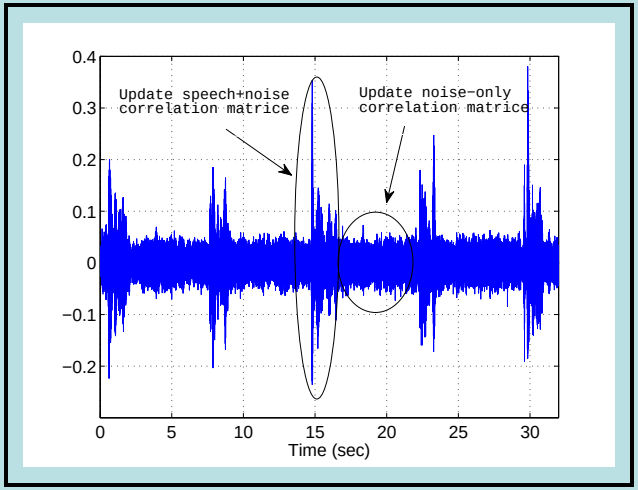
$$Z(k,l) = \mathbf{W}^{*,H}(k,l)\mathbf{X}(k,l). \tag{5}$$



Concept SDW-MWFμ

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

$$\mathbf{R}^s(k,l) = \mathbf{R}^x(k,l) - \mathbf{R}^n(k,l) \tag{6}$$



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

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



Speech Presence Probability

Two-state speech model

$$\begin{aligned} H_0(k,l) : X_i(k,l) &= X_i^n(k,l) \\ H_1(k,l) : X_i(k,l) &= X_i^s(k,l) + X_i^n(k,l) \end{aligned} \tag{7}$$

Conditional Probability Density Functions of the observed signals

$$\begin{aligned} p(X_i(k,l)|H_0(k,l)) &= \frac{1}{\pi\lambda_i^n(k,l)} \exp\left\{-\frac{|X_i(k,l)|^2}{\lambda_i^n(k,l)}\right\} \\ p(X_i(k,l)|H_1(k,l)) &= \frac{1}{\pi(\lambda_i^s(k,l)+\lambda_i^n(k,l))} \exp\left\{-\frac{|X_i(k,l)|^2}{\lambda_i^s(k,l)+\lambda_i^n(k,l)}\right\} \end{aligned} \tag{8}$$

Speech Presence Probability

$$p(k,l) = \left\{1 + \frac{q(k,l)}{1-q(k,l)}(1+\xi(k,l))\exp(-\upsilon(k,l))\right\}^{-1} \tag{9}$$

- Conditional SPP is estimated for each frequency bin and each frame



Extension of SDW-MWF$_\mu$ into SDW-MWF$_{SPP}$

Incorporating the conditional Speech Presence probability in SDW-MWF

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} \; p\varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}|^2|H_1\} + (1-p)\varepsilon\{|\mathbf{W}^H\mathbf{X}^n|^2\} \tag{10}$$

The SDW-MWF incorporating the conditional SPP can then be written as

$$\mathbf{W}^*_{SPP} = \left(\mathbf{R}^s + \left(\tfrac{1}{p}\right)\mathbf{R}^n\right)^{-1}\mathbf{R}^s\mathbf{e}_1. \tag{11}$$

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

The combined solution can then be written as

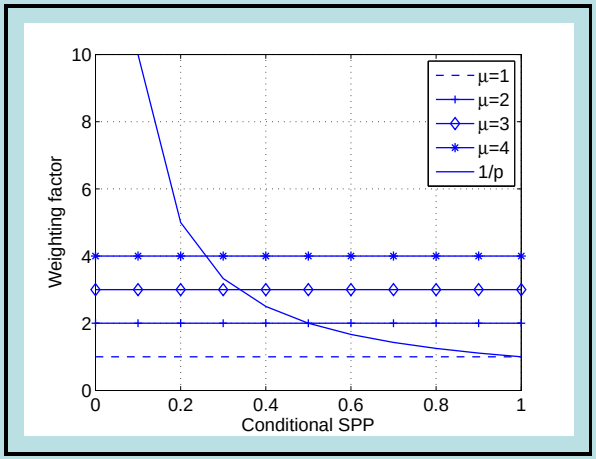
$$\mathbf{W}^*_{SPP} = \left(\mathbf{R}^s + \left(\frac{1}{\alpha(\frac{1}{\mu})+(1-\alpha)p}\right)\mathbf{R}^n\right)^{-1}\mathbf{R}^s\mathbf{e}_1 \tag{12}$$

- Avoids undesired noise modelling

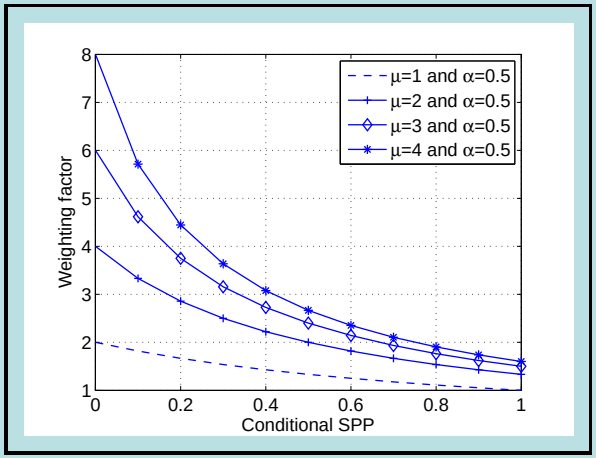


Extension of SDW-MWF$_\mu$ into SDW-MWF$_{SPP}$

- Example of SPP for a frame



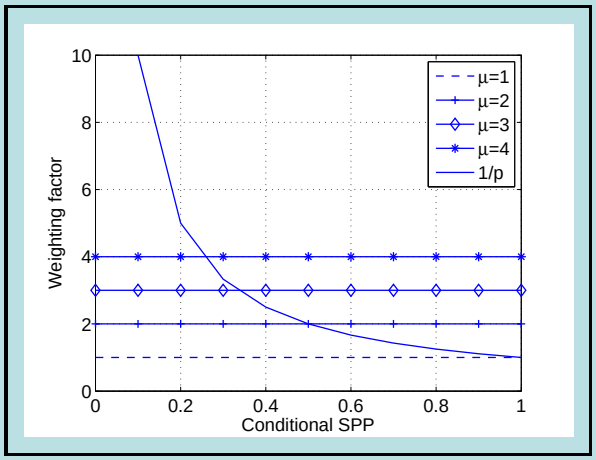
- SDW-based on SPP



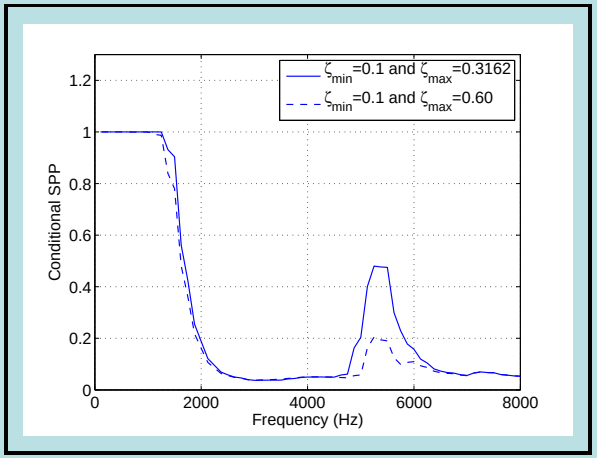


Extension of SDW-MWF$_\mu$ into SDW-MWF$_{SPP}$

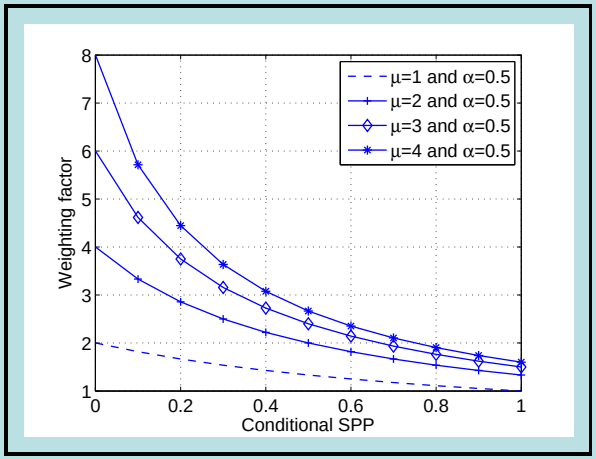
- Example of SPP for a frame



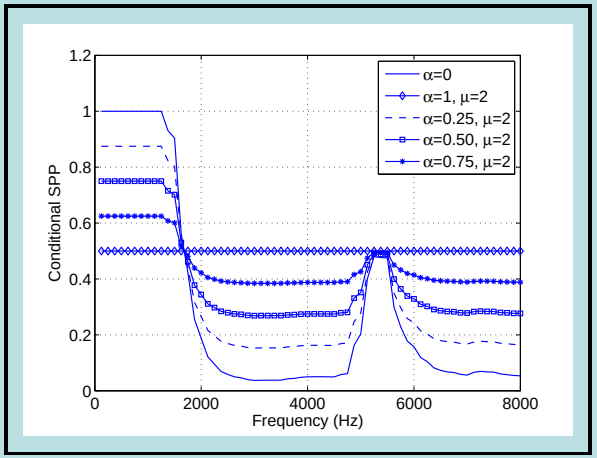
- Example of SPP for a frame (modified)



- SDW-based on SPP

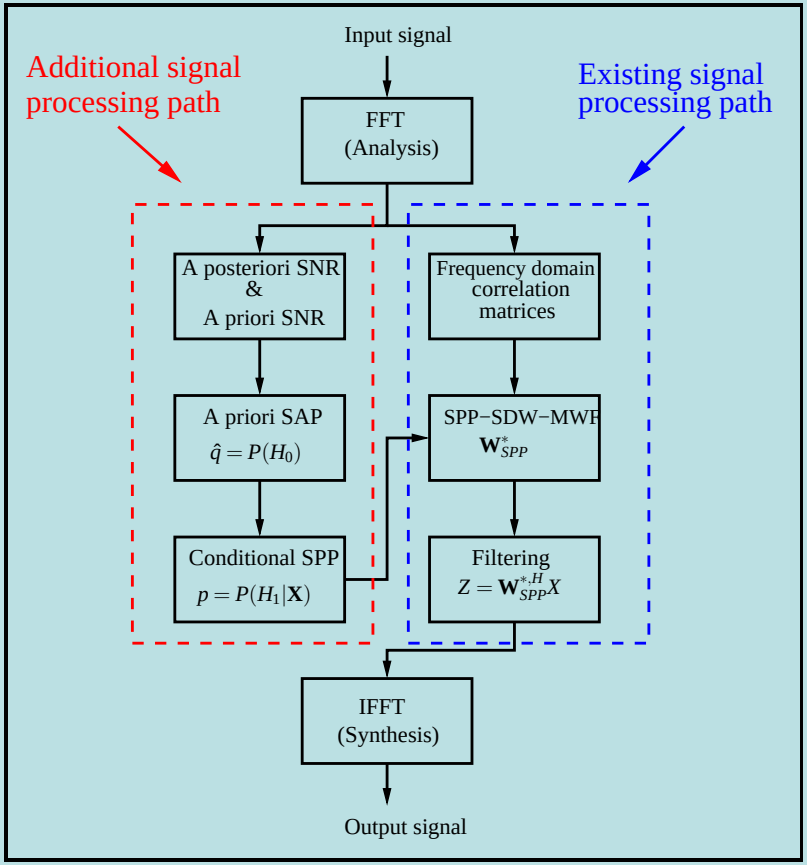


- SDW-based on SPP (modified)





Extension of SDW-MWF$_\mu$ into SDW-MWF$_{SPP}$



Challenges on additional signal processing path

- Incorporating a psychoacoustical Model
- Masking properties of the human auditory system
- Auditory properties of speech perception
- Define perceptual relevant criteria
- Make residual noise perceptually inaudible

Challenges on existing signal processing path

- Limited opportunities
- Continously updating correlation matrices in speech+noise periods.
- development on Voice Activity Detection



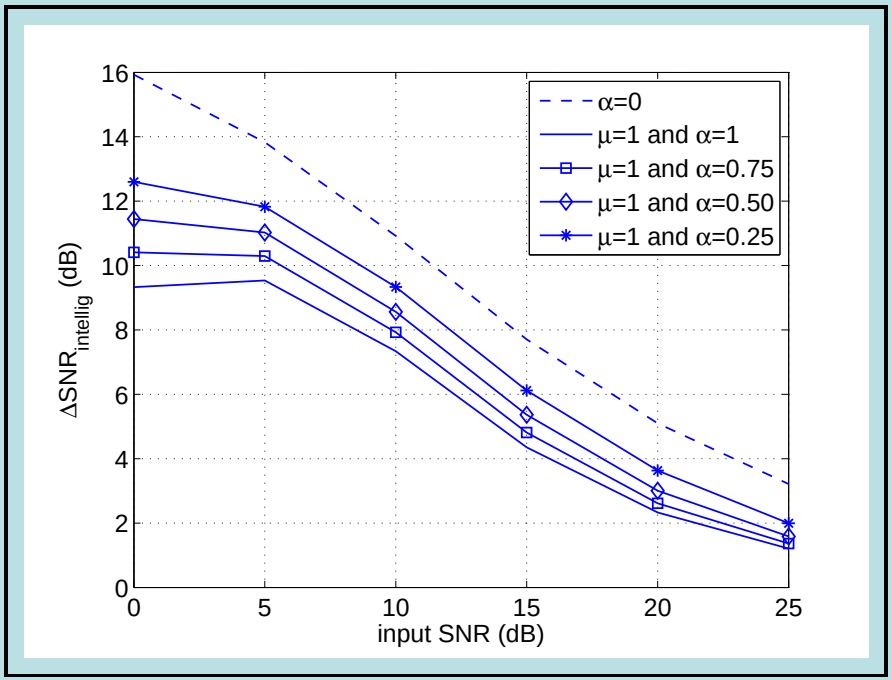
Experimental Set-up for SDW-MWF$_\mu$ and SDW-MWF$_{SPP}$

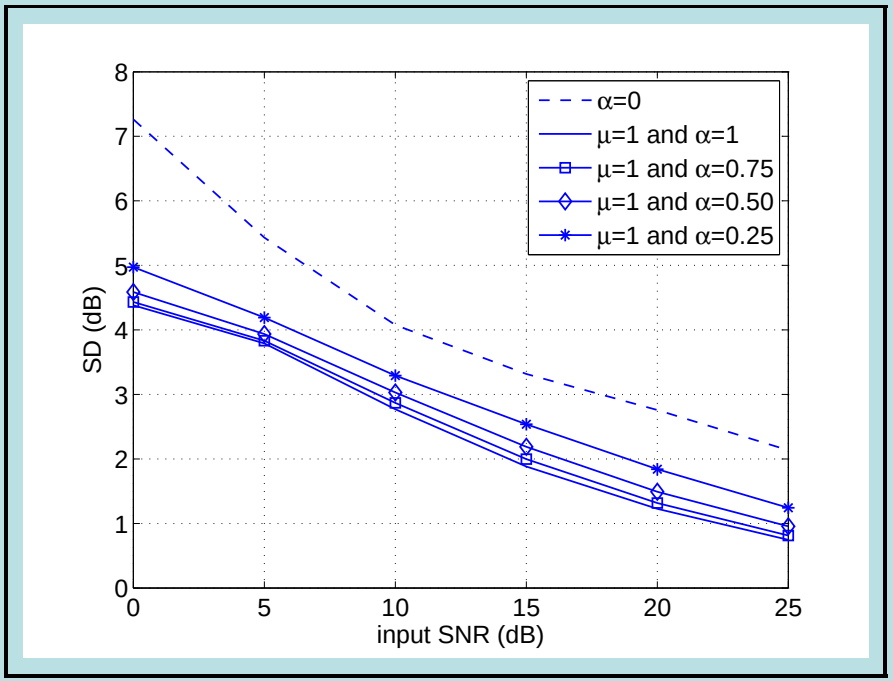
Simulations have been performed with

- A 2-microphone behind-the-ear hearing aid mounted on a CORTEX MK2 manikin.
- The loudspeakers (FOSTEX 6301B) are positioned at 1 meter from the center of the head.
- The reverberation time $T_{60} = 0.21s$.
- The speech is located at 0° and the two multi-talker babble noise sources are located 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.



Experimental results for SDW-MWF$_\mu$ and SDW-MWF$_{SPP}$



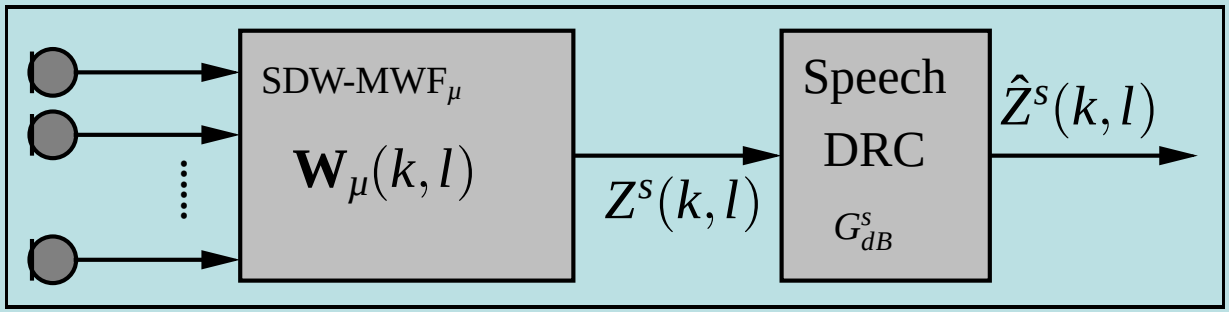




Integration of MWF based NR and DRC

Motivation

- 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



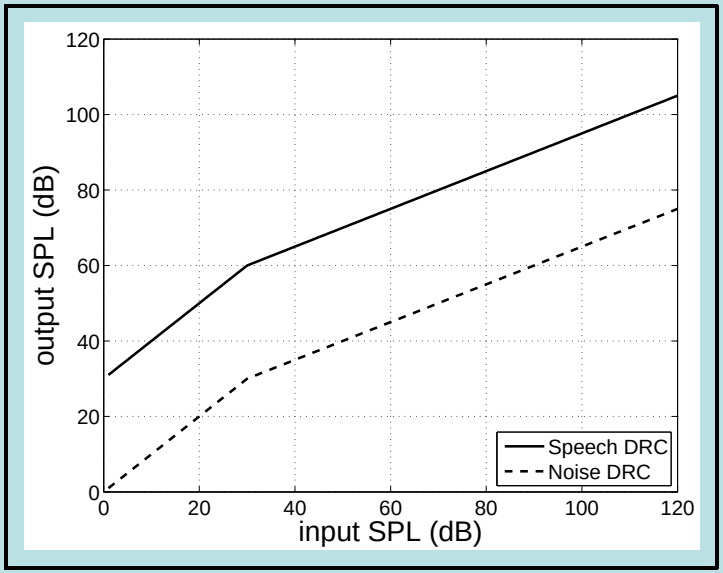
- DRC does not distinction between speech and noise dominant segments
- Low intensity segments are amplified equally (including residual noise)

$$Z(k,l) = \mathbf{W}^{*,H}(k,l)\mathbf{X}(k,l) \quad (13)$$

$$Z(k,l) = \hat{X}_1^s(k,l) + \varepsilon^n(k,l)(\text{residual noise}). \quad (14)$$



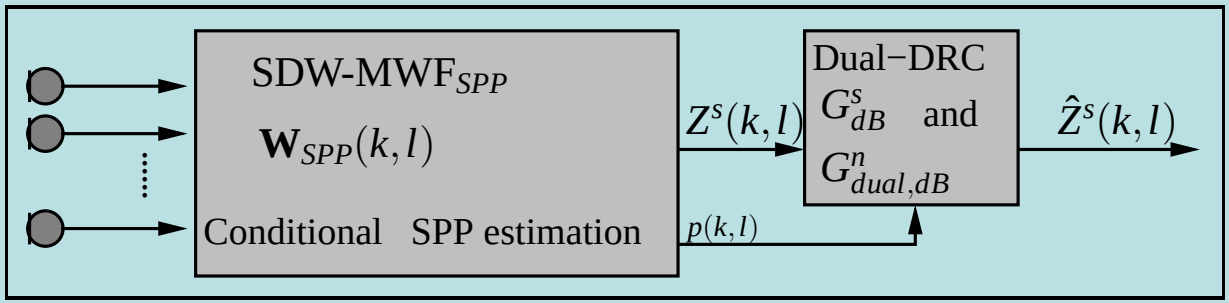
Extension of DRC into Dual-DRC



- Reusing the conditional SPP estimated in SDW-MWF$_{SPP}$
- A dual-DRC concept can be introduced
- Using switchable compression characteristic

Dual-DRC conept

- 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





Experimental Set-up for SDW-MWF based NR and DRC

Simulations have been performed

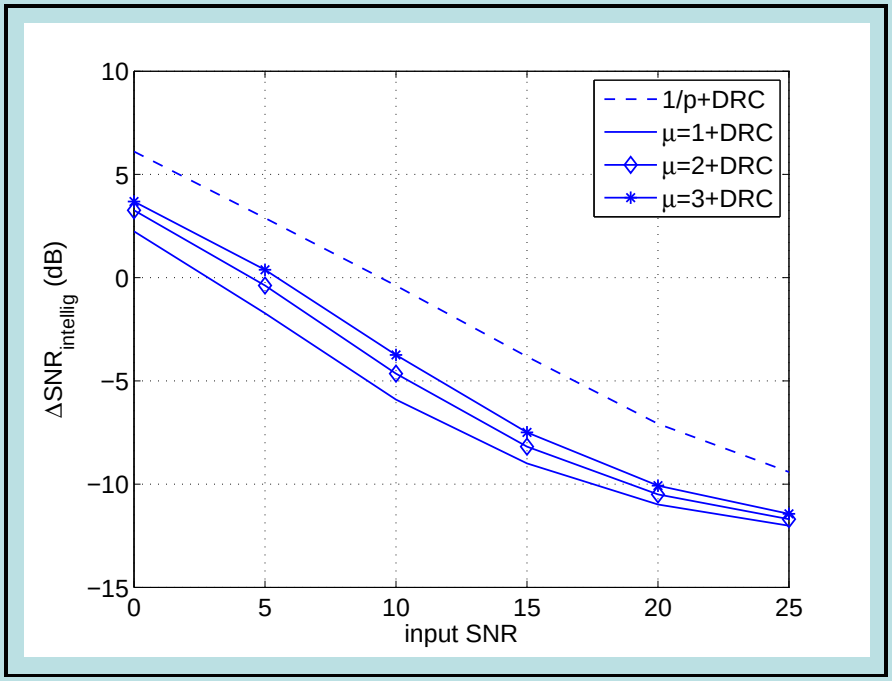
- A 2-microphone behind-the-ear hearing aid mounted on a CORTEX MK2 manikin.
- The loudspeakers (FOSTEX 6301B) are positioned at 1 meter from the center of the head.
- The reverberation time $T_{60} = 0.21s$.
- The speech is located at 0° and the two multi-talker babble noise sources are located at 120° and 180°.
- The speech signals consist of male sentences from the HINT-database
- The noise signals consist of a multi-talker babble from Auditec
- The signals are sampled at 16kHz.
- an FFT length of 128 with half overlapping frames.
- The DRC is implemented based on 20 critical bands.

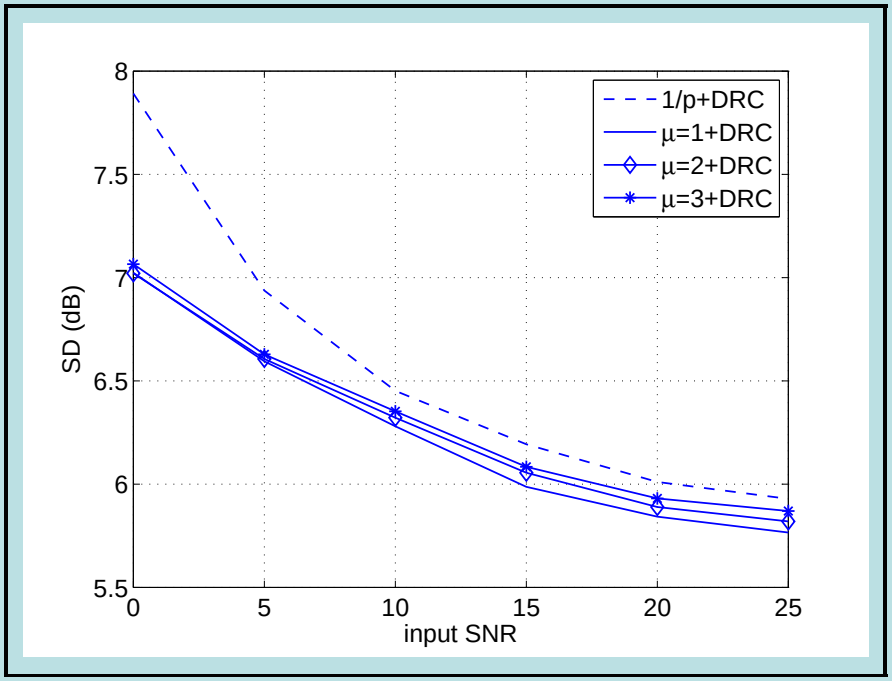
The following parameters are fixed during all simulations:

- The input level is set to 65dB SPL at the hearing aid microphones.
- The DRC attack and release time are set to at=10ms and rt=150ms.
- The compression threshold is set to CT=30dB.
- The hearing aid gain is set to G_{dB}=30dB.



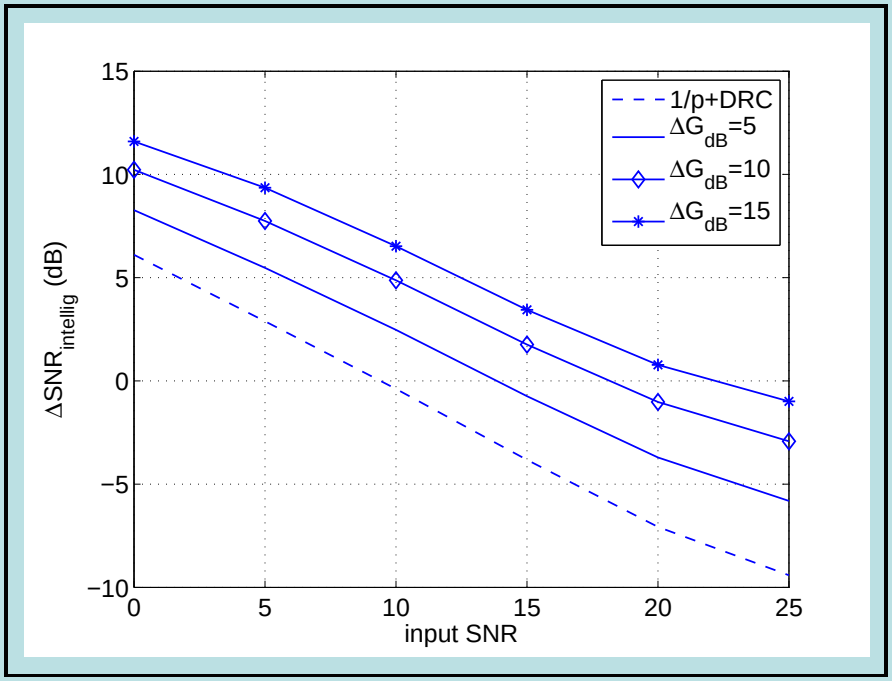
Experimental results for SDW-MWF$_\mu$-based NR and DRC

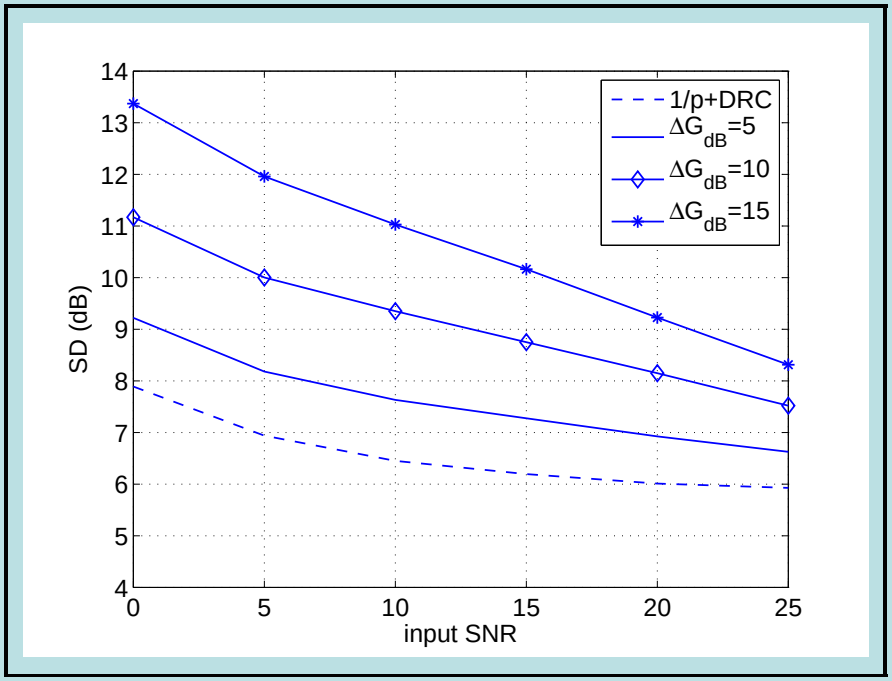






Experimental results for SDW-MWF$_{SPP}$-based NR and dual-DRC



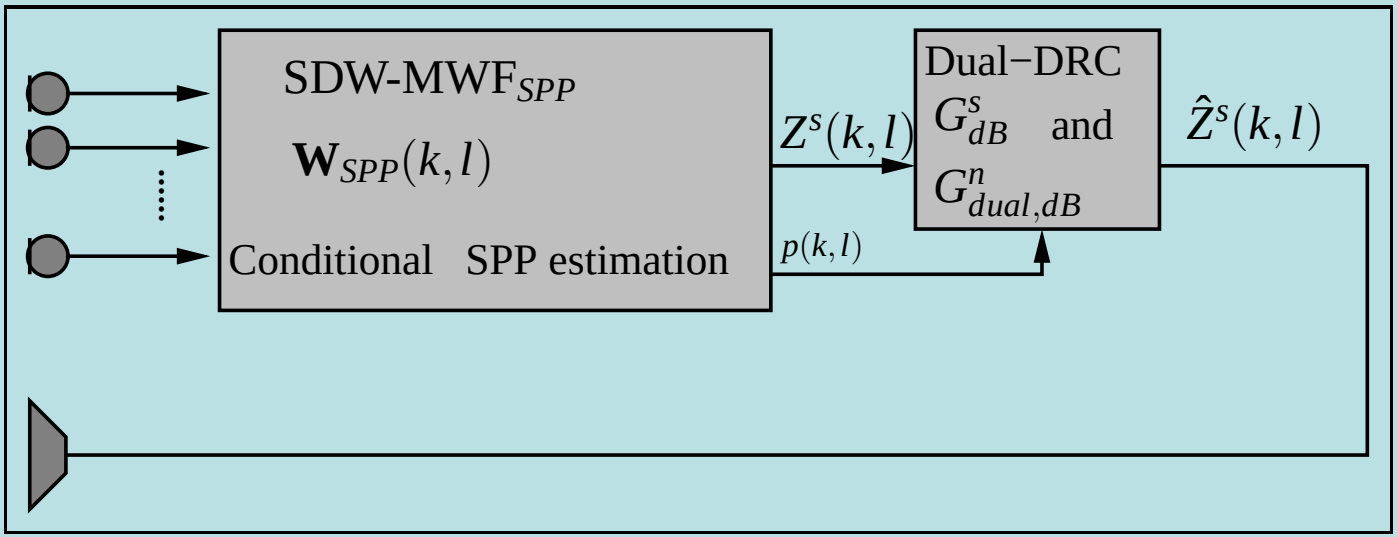




Combined NR and DRC and AFC

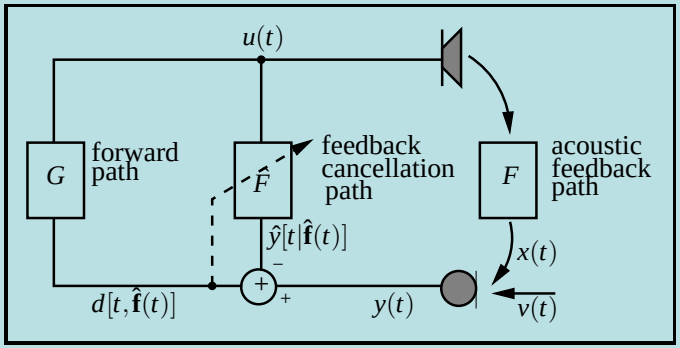
Long-term objective:

- Integration of NR, DRC and AFC into one signal processing scheme.





Adaptive Feedback Cancellation



The microphone signal

$$y(t) = v(t) + x(t) = v(t) + F(q,t)u(t) \quad (15)$$

The feedback-compensated

$$d(t) = v(t) + [F(q,t) - \hat{F}(q,t)]u(t). \quad (16)$$

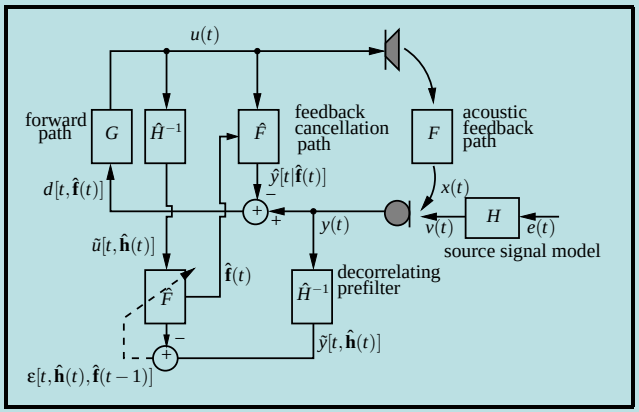
- Adaptively model the feedback path
- Estimate the feedback signal
- correlation between the near-end signal and the loudspeaker signal
- Caused by the closed signal loop

Main Challenge

- Reduce the correlation between the near-end signal and the loudspeaker signal
- Prediction error method-based AFC (PEM-based AFC)



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



Microphone signal

- $y(t) = F(q,t)u(t) + H(q,t)e(t)$
- Prefiltering of loudspeaker and microphone signals.
- Inverse near-end signal model.

The all-pole model can be written as

$$H(q,t) = \frac{1}{C(q,t)} = \frac{1}{1 + c_1(t)q^{-1} + \ldots + c_{n_c}(t)q^{-n_c}} \tag{17}$$

Prediction error

$$\varepsilon[t,\boldsymbol{\xi}(t)] = H^{-1}(q,t)[y(t) - F(q,t)u(t)] \tag{18}$$

Minimize prediction error

$$\min_{\boldsymbol{\xi}(t)} = \frac{1}{2N}\sum_{k=1}^{t}\varepsilon^2[k,\boldsymbol{\xi}(t)] \tag{19}$$

- Single all-pole model (Short-term predictor) fails to remove the periodicity



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

- A cascade of near-end signal models removes the coloring and periodicity

Sinusoidal model

$$d(t) = \sum_{n=1}^{P} A_n \cos(\omega_n t + \phi_n) + r(t), \quad t = 1, ..., M \tag{20}$$

Cascaded near-end signal model

$$y(t) = F(q,t)u(t) + H_1(q,t)H_2(q,t)e(t) \tag{21}$$

The CPZLP model can be written as

$$d(t) = \left(\prod_{n=1}^{P} \frac{1 - 2\rho \cos \omega_n z^{-1} + \rho^2 z^{-2}}{1 - 2 \cos \omega_n z^{-1} + z^{-2}} \right) e(t) \tag{22}$$

the output from the prediction error filter

$$e(t, \boldsymbol{\omega}) = \left(\prod_{n=1}^{P} \frac{1 - 2 \cos \omega_n z^{-1} + z^{-2}}{1 - 2\rho \cos \omega_n z^{-1} + \rho^2 z^{-2}} \right) d(t) \tag{23}$$



Incorporating pitch estimation in PEM-based AFC

- Speech signals are usually considered as voiced or unvoiced
- Voiced sounds consist of fundamental frequency ω_0 and its harmonic components.
- CPZLP estimates all frequencies independently
- Does not exploits the harmonicity of speech.

Fundamental frequency estimation (pitch estimation)

- Sinusoids are having frequencies that are ω_0, i.e., $\omega_n = \omega_0 n$
- This follows naturally from voiced speech being quasi-periodic.

Applying pitch estimation in PEM-based AFC

$$e(t, \boldsymbol{\omega}) = \left(\prod_{n=1}^{P} \frac{1 - 2\cos\omega_n z^{-1} + z^{-2}}{1 - 2\rho\cos\omega_n z^{-1} + \rho^2 z^{-2}} \right) d(t) \tag{24}$$

Pitch estimation considered are

- Subspace-orthogonality-based pitch estimation.
- Subspace-shiftinvariance-based pitch estimation
- Optimal-filtering based pitch estimation.



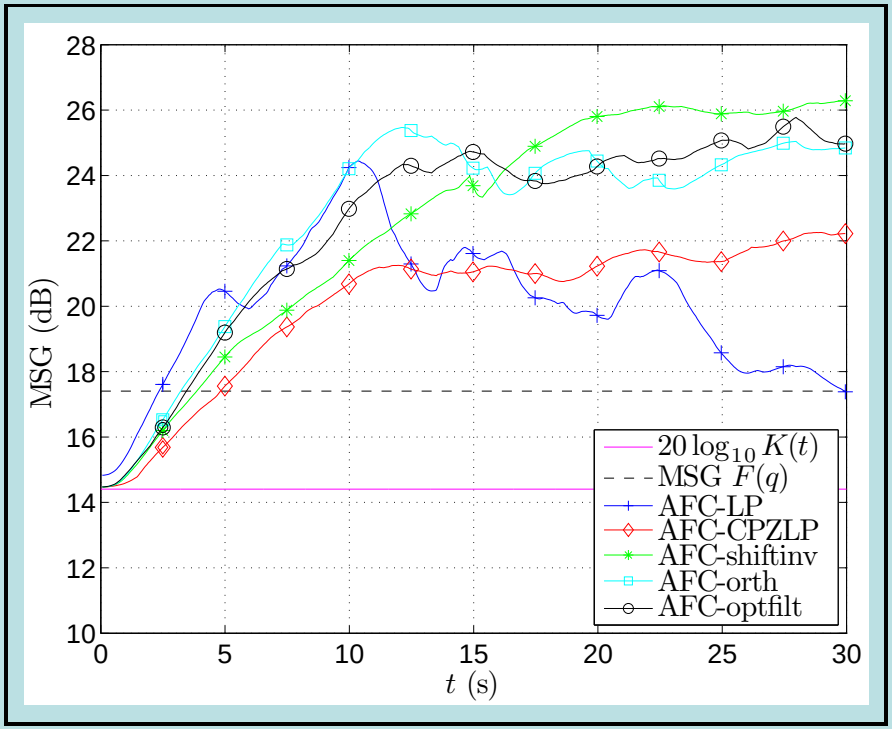
Experimental Set-up for PEM-based AFC

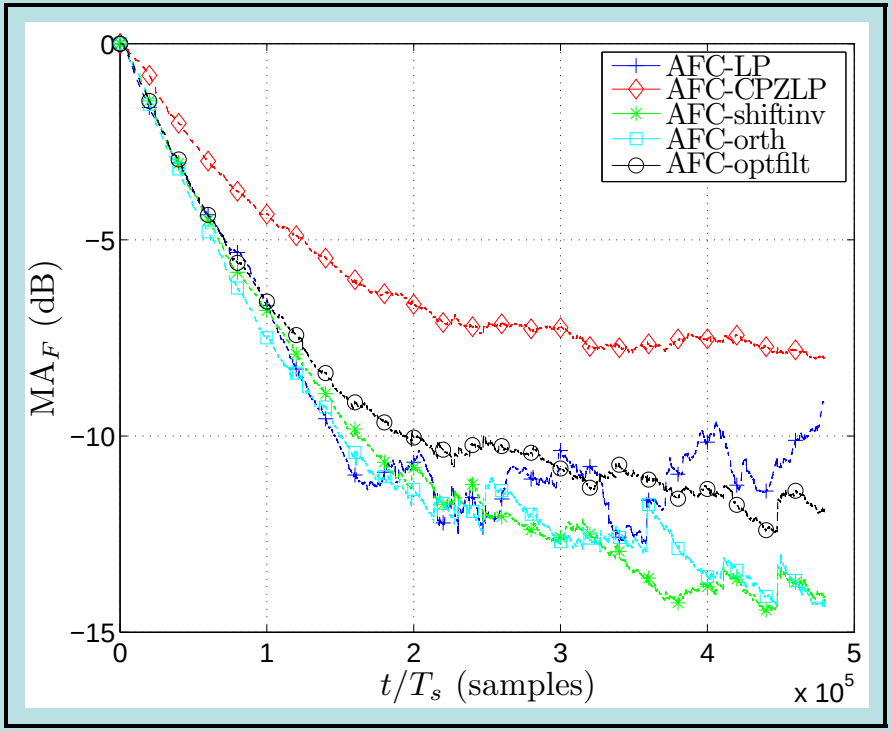
Simulations have been performed

- The near-end sinusoidal model order is set to $P = 15$
- 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.



Experimental results for PEM-based AFC







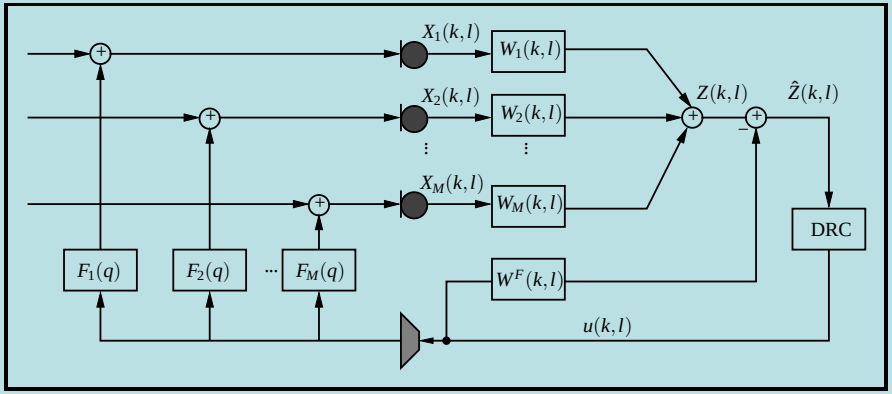
Conclusion

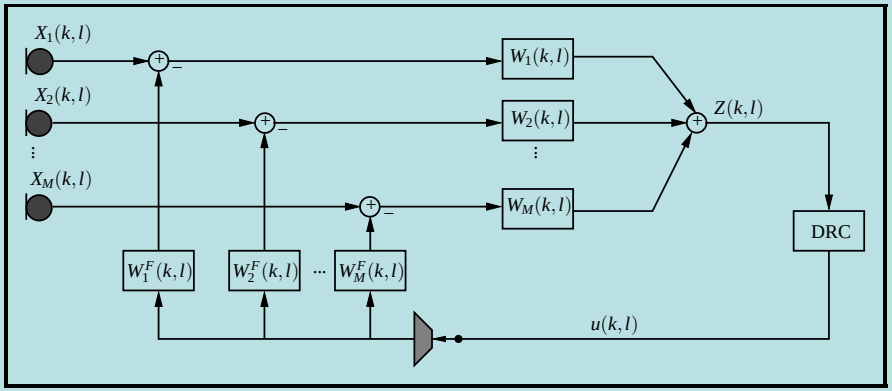
- SDW-MWF$_{\mu}$ can be further improved by using SDW-MWF$_{SPP}$.
- SDW-MWF$_{SPP}$ can be further extended with more advance SDW criteria
- An MWF-based NR and Dynamic Range Compression have been proposed.
- Dual-DRC concept using switchable compression characteristic
- PEM-based AFC using cascaded near-end signal model (using pitch estimation)



Future and current work

Combined MWF and AFC



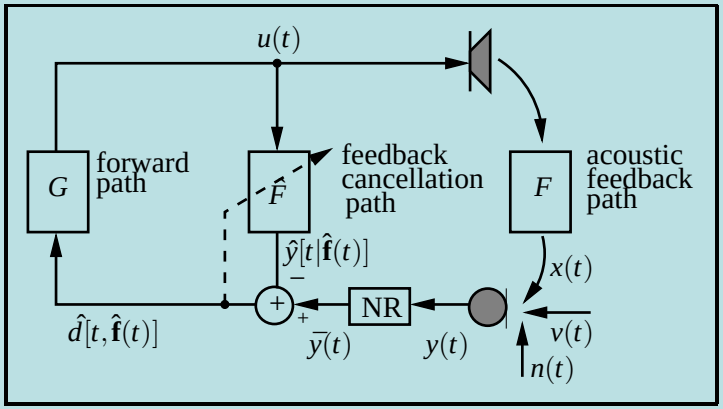


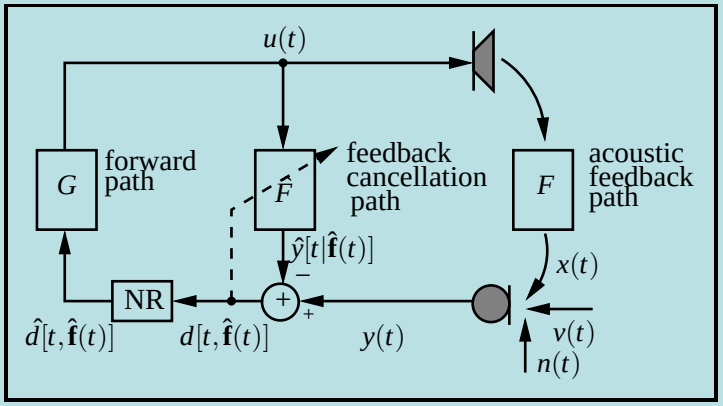
- Applying NR before AFC...
- Applying AFC before NR ...



Future and current work

Combined Single-channel NR and AFC





- Applying NR before AFC...
- Applying AFC before NR ...



Future and current work

further exploit PEM-based AFC using pitch estimation

- Using pitch estimation amplitudes can be estimated
- Possibility to design more accurat CPZLP model (better decorrelation)

Remove the CPZLP model

- by re-using the optimal filter for both fundamental frequency estimation and filtering.
- Decorrelation by sinusoidal subtraction
- Further investigate the use of subspace methods.

Combined Speech Coding and AFC

