KATHOLIEKE UNIVERSITEIT
LEUVEN

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

A Flexible Speech Distortion Weighted Multi-Channel Wiener Filter for Noise Reduction in Hearing Aids

Kim Ngo, Marc Moonen, Søren Holdt Jensen and Jan Wouters

Friday May 27th, 2011

The 36th International Conference on Acoustics, Speech and Signal Processing
Prague, Czech Republic

KATHOLIEKE UNIVERSITEIT
LEUVEN

Outline

- Introduction and problem statement
- Speech Distortion Weighted Multi-channel Wiener filter (SDW-MWF)
- Conditional speech presence probability (SPP)
- SDW-MWF incorporating the conditional SPP
- SDW-MWF incorporating a flexible weighting factor
- Conclusion

KATHOLIEKE UNIVERSITEIT LEUVEN

Introduction

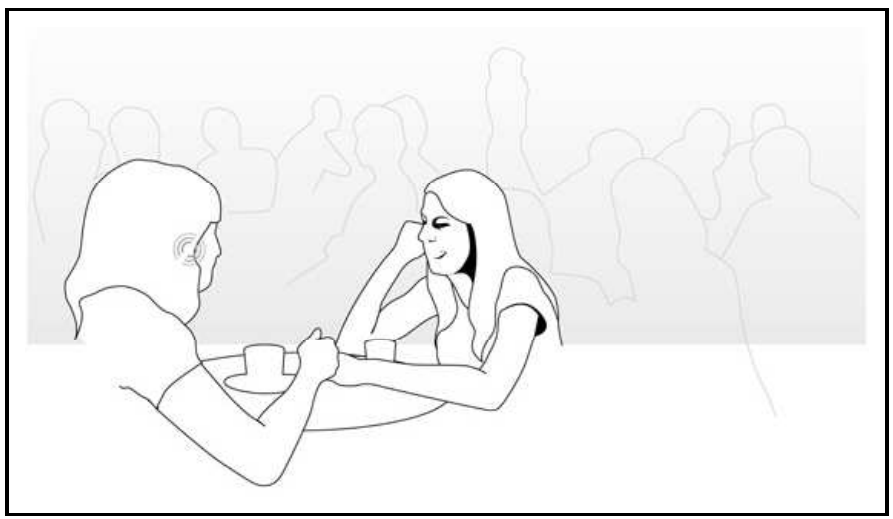
Hearing loss/Background Noise

Common causes of hearing loss:

- Age-related
- Daily exposure to excessive noise in the work environment
- Listening to loud music

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)



- 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.

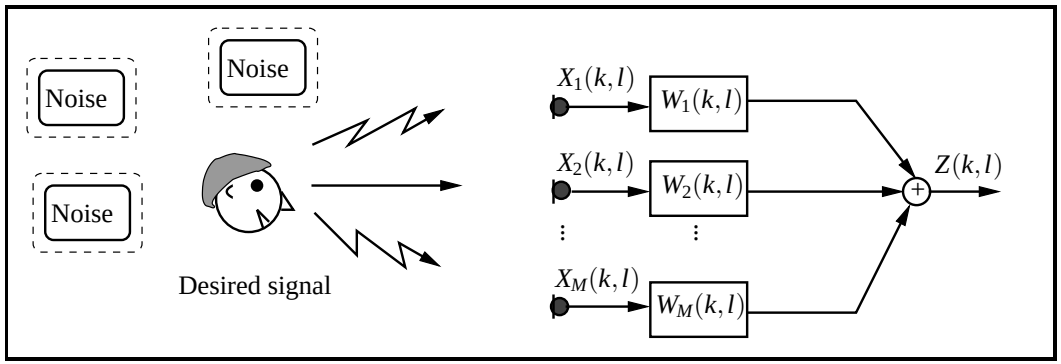
KATHOLIEKE UNIVERSITEIT LEUVEN

Introduction

Multi-microphone noise reduction

Common techniques:

- Directional microphones
- Delay-and-sum beamformers
- GSC, MVDR, LCMV beamformers.
- **Multi-channel Wiener filter**



Objective of NR:

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

KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

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

Frequency-domain microphone signals:

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

- k is the frequency bin index and l is the frame index

MWF MMSE criterion:

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

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:

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

KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

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

Frequency-domain microphone signals:

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

- k is the frequency bin index and l is the frame index

MWF MMSE criterion:

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

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:

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

KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

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

Frequency-domain microphone signals:

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

- k is the frequency bin index and l is the frame index

MWF MMSE criterion:

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

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:

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

KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

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

Frequency-domain microphone signals:

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

- k is the frequency bin index and l is the frame index

MWF MMSE criterion:

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

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:

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

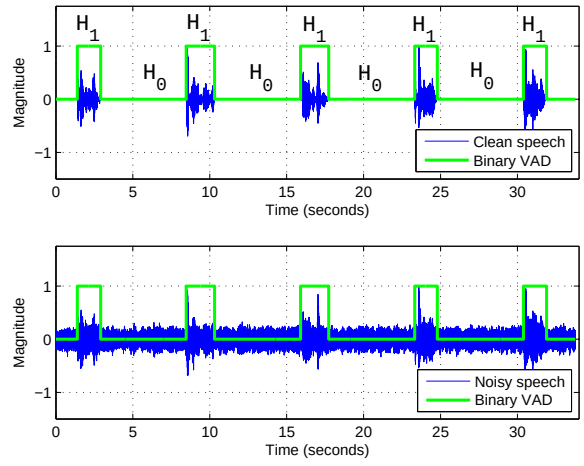
KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

Estimation and update of correlation matrices

- 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.

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

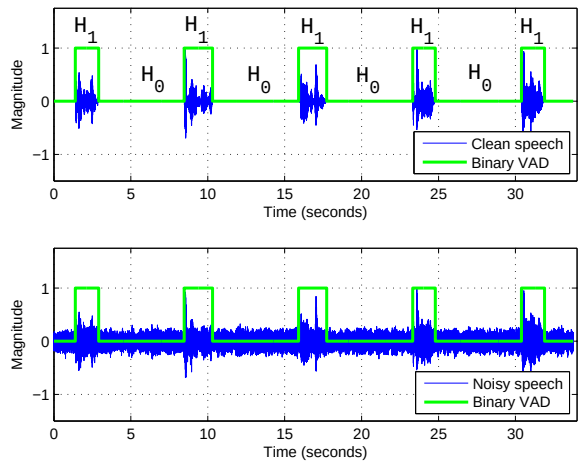
KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

Estimation and update of correlation matrices

- 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.

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

KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

Motivation

Properties of the SDW-MWF$_\mu$:

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

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

Properties not included in the SDW-MWF$_\mu$:

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

Multi-channel Wiener filter

Motivation

Properties of the SDW-MWF$_\mu$:

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

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

Properties not included in the SDW-MWF$_\mu$:

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

KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

Speech presence probability

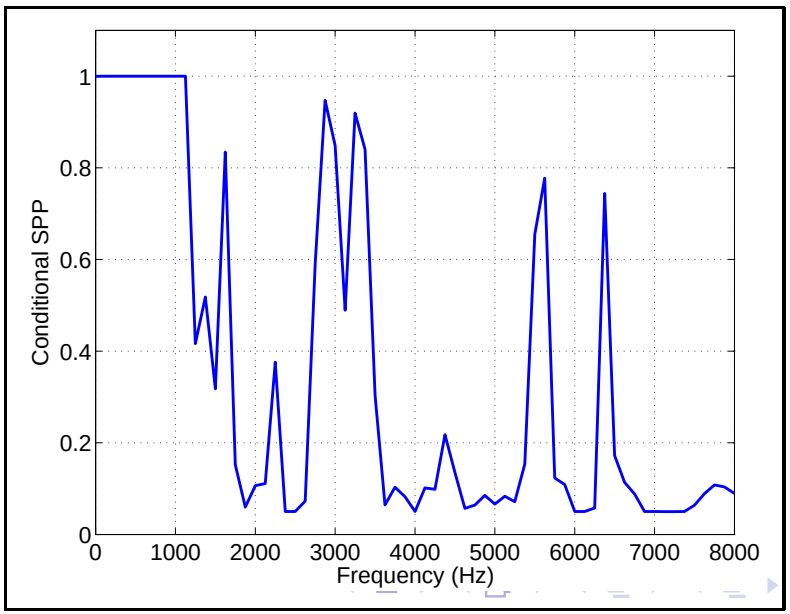
Two-state speech model:

$$H_0(k,l) : \mathbf{X}(k,l) = \mathbf{X}^n(k,l)$$
$$H_1(k,l) : \mathbf{X}(k,l) = \mathbf{X}^s(k,l) + \mathbf{X}^n(k,l).$$

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:

$$\mathbf{W}_{\text{SPP}}(k,l) = \arg\min_{\mathbf{W}} \; p(k,l)\varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}|^2|H_1\} + (1-p(k,l))\varepsilon\{|\mathbf{W}^H\mathbf{X}|^2|H_0\}$$

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

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

- 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

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

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

$$\mathbf{W}_{\text{SPP}}(k,l) = \arg\min_{\mathbf{W}} \; p(k,l)\varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}|^2|H_1\} + (1-p(k,l))\varepsilon\{|\mathbf{W}^H\mathbf{X}|^2|H_0\}$$

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

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

- 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.

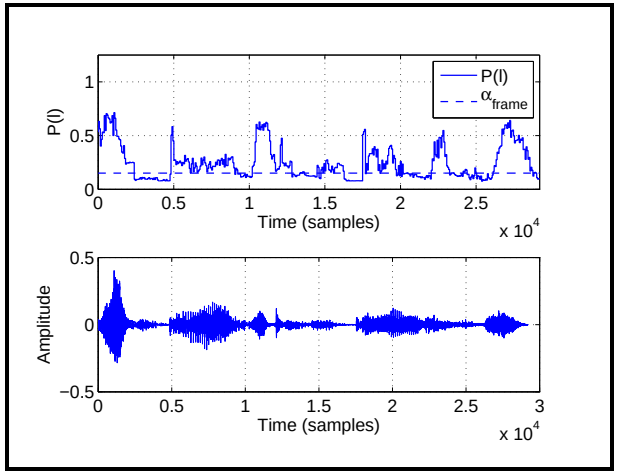
KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

H_0 and H_1 state detection based on SPP

Binary decision:

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



Rationale:

- Reducing the noise in the H_0 state can be related to increasing listening comfort, since speech is not present in the H_0 state, which means that a greater attenuation can be applied.
- Reducing the noise in the H_1 state is a more challenging task since this relates to speech intelligibility and hence the speech distortion weighted concept truly only makes sense in the H_1 state.

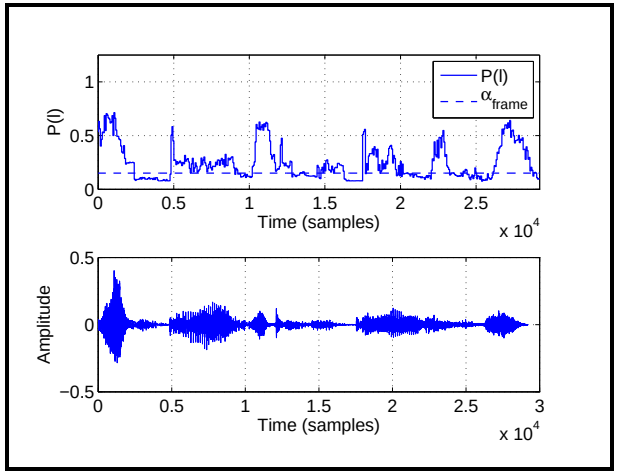
KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

H_0 and H_1 state detection based on SPP

Binary decision:

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



Rationale:

- Reducing the noise in the H_0 state can be related to increasing listening comfort, since speech is not present in the H_0 state, which means that a greater attenuation can be applied.
- Reducing the noise in the H_1 state is a more challenging task since this relates to speech intelligibility and hence the speech distortion weighted concept truly only makes sense in the H_1 state.

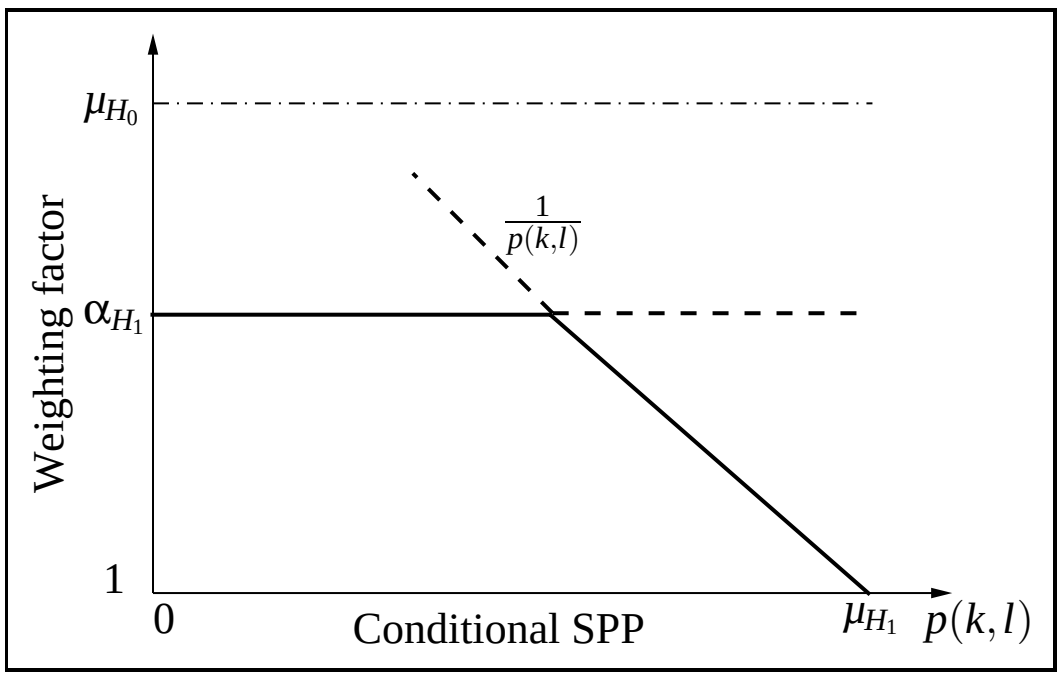
KATHOLIEKE UNIVERSITEIT LEUVEN

Multi-channel Wiener filter

H_0 and H_1 state detection based on SPP

Proposed weighting factor:

- 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:

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

with the weighting factor defined as:

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

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):

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

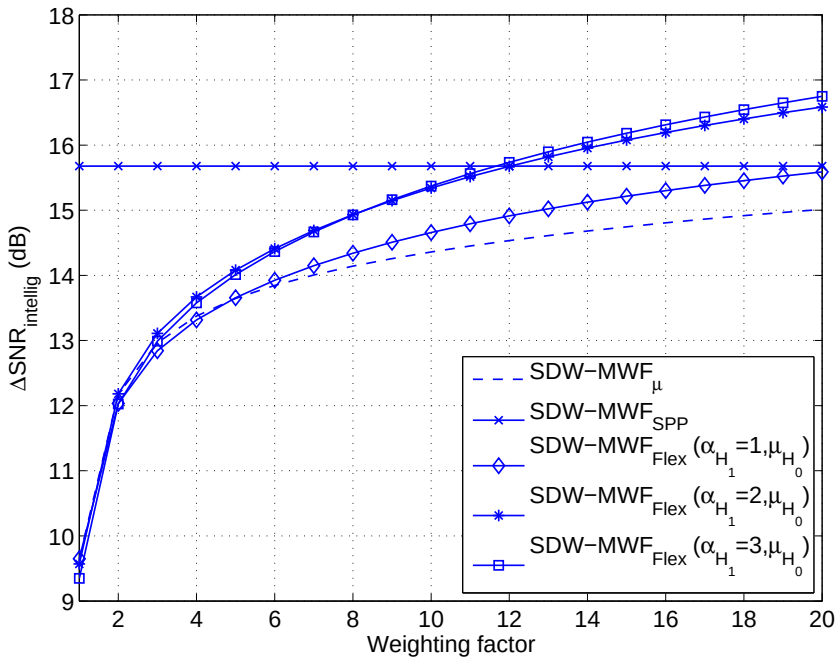
frequency-weighted log-spectral signal distortion (SD)

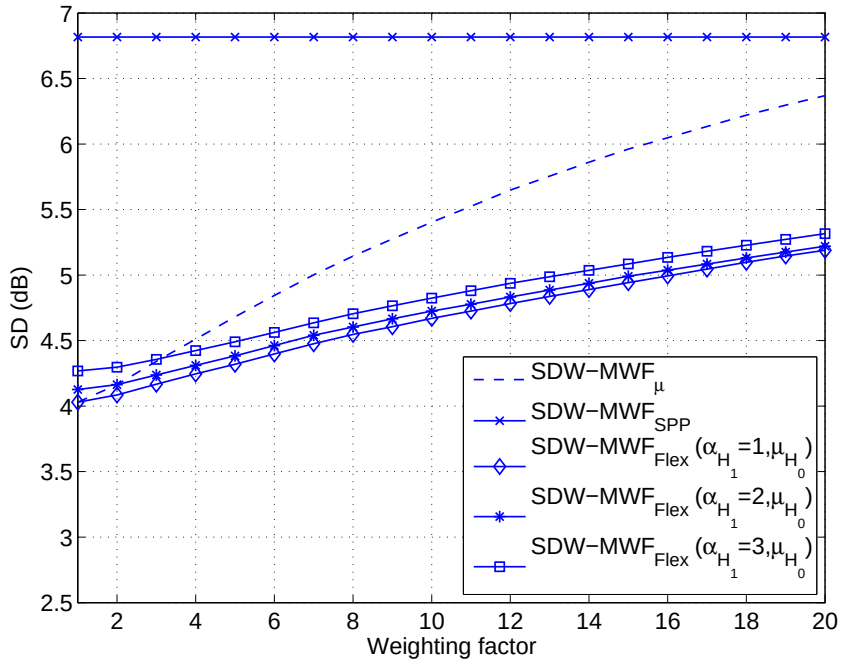
$$\mathrm{SD} = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\int_{f_l}^{f_u} w_{ERB}(f)\left(10log_{10}\frac{P^s_{out,k}(f)}{P^s_{in,k}(f)}\right)^2 df}$$

KATHOLIEKE UNIVERSITEIT LEUVEN

Experimental results MWF based NR

- SDW-MWF$_{\mu}$ - low SNR improvement - high distortion
- SDW-MWF$_{\mathrm{SPP}}$ - high SNR improvement - high distortion
- SDW-MWF$_{\mathrm{Flex}}$ - **high SNR improvement** - **low distortion**





KATHOLIEKE UNIVERSITEIT LEUVEN

Conclusion

Summary:

- Different extensions of the SDW-MWF algorithms
- SDW-MWF$_{\mu}$ fixed weighting for each frame and for each frequency
- SDW-MWF$_{\mathrm{SPP}}$ weighting based on the conditional SPP
- SDW-MWF$_{\mathrm{Flex}}$ weighting based on combined soft and binary detection

Future/current work:

- Perceptual evaluation using hearing impaired listeners
- Using a perceptual motivated weighting factor in the SDW-MWF

KATHOLIEKE UNIVERSITEIT
LEUVEN

End of presentation

Thank you!.....Questions?