University of Stuttgart
Germany

Real-Time Capable Robust Noise Reduction

Speech Signal Processing and Speech Enhancement

Maximilian Luz

Summer Semester 2019

IMS

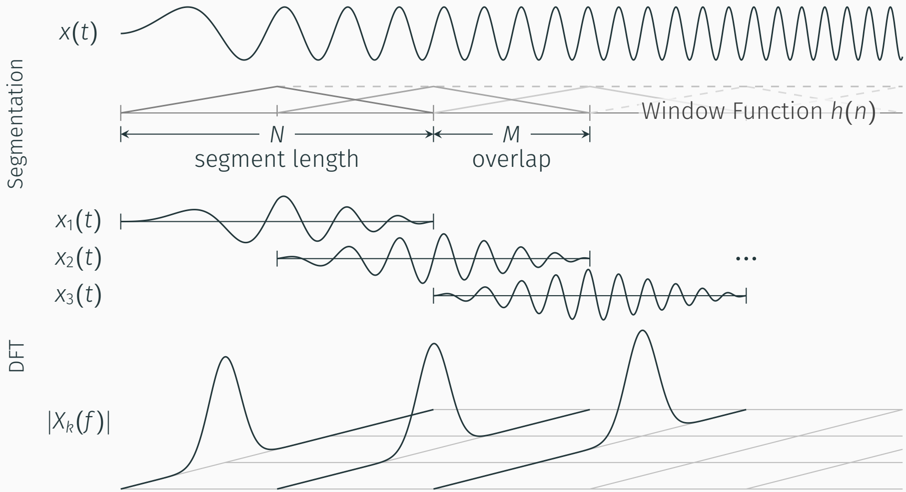
Goals

- Single microphone.
- Real-time capable.
- Adaptive to changes in noise/signal.
- Processing in frequency-domain.
- Unsupervised.

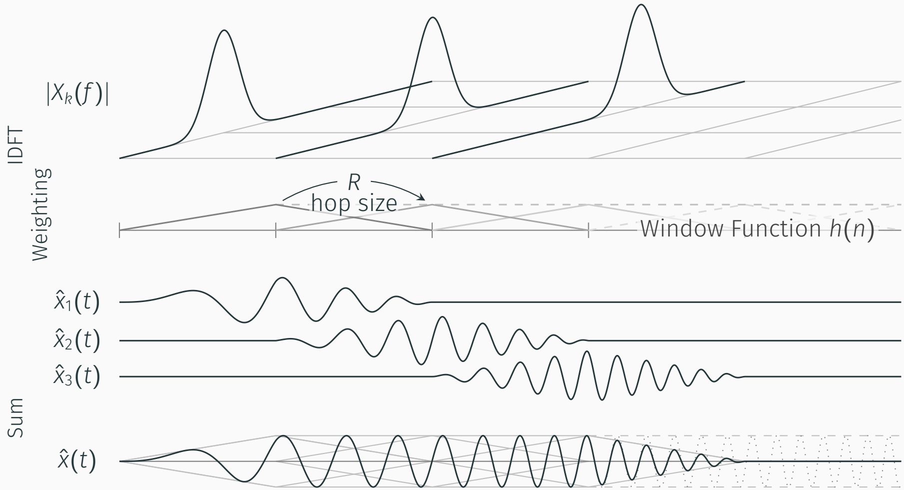
Outline

Basics

Short-Time Fourier Transform



(Weighted) Overlapp and Add

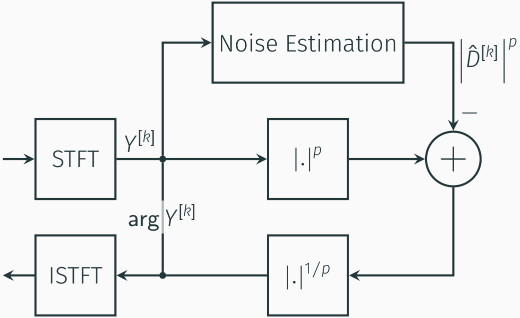


Methods

noisy signal noise

$$y(t) = x(t) + d(t)$$

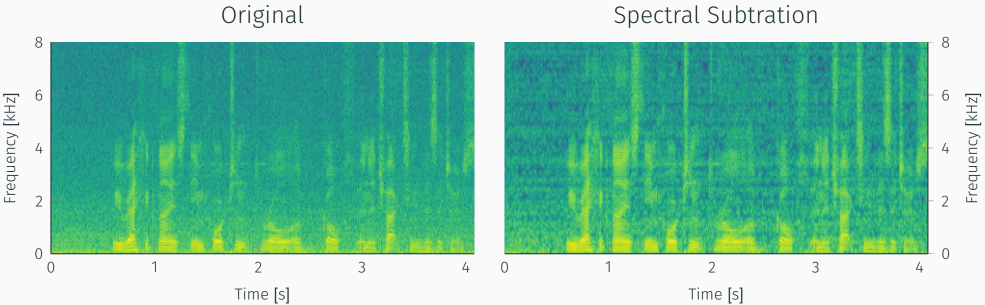
signal



Open Questions:

- How to estimate noise?
- How to handle negative magnitude values after subtraction?

Spectral Subtraction: Results

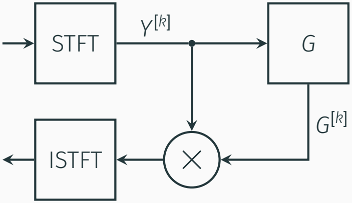


Issues

- Residual (musical) noise.
- Too much subtraction leads to speech distortion.

Gain Function

$$y(t) = x(t) + d(t) \qquad \text{where} \qquad X^{[k]} = A^{[k]} e^{j\alpha^{[k]}}, \quad Y^{[k]} = R^{[k]} e^{j\vartheta^{[k]}},$$



Usually: $G\left(\xi^{[k]}, \gamma^{[k]}\right) \in \mathbb{R}$

spectral signal power

$$\xi^{[k]} := \frac{\lambda_x^{[k]}}{\lambda_d^{[k]}} \qquad (\textit{a priori} \text{ SNR})$$

$$\gamma^{[k]} := \frac{\left(R^{[k]}\right)^2}{\lambda_d^{[k]}} \qquad (\textit{a posteriori} \text{ SNR})$$

spectral noise power

Need to be estimated!

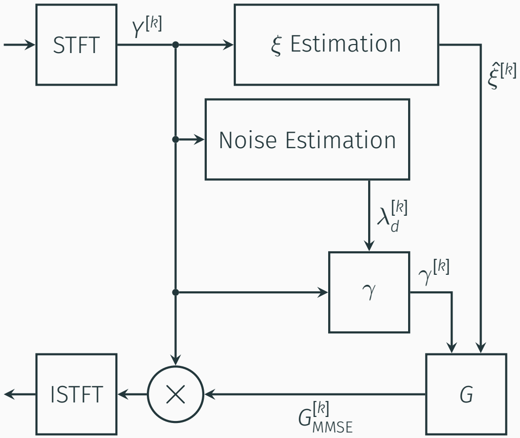
Minimum Mean-Square Error Spectral Amplitude Estimator (MMSE)

Idea: Minimize

$$\mathbb{E}\left\{\left(A^{[k]} - \hat{A}^{[k]}\right)^2\right\}$$

Solution: (Assumes Gaussian distribution)

$$\begin{aligned}\hat{A}^{[k]} &= \mathbb{E}\left\{A^{[k]} \mid Y^{[k]}\right\}\\ &= G_{\mathrm{MMSE}}\left(\xi^{[k]}, \gamma^{[k]}\right) \cdot R^{[k]}\end{aligned}$$



Minimum Mean-Square Error Log-Spectral Amplitude Estimator (log-MMSE)

Idea: Minimize

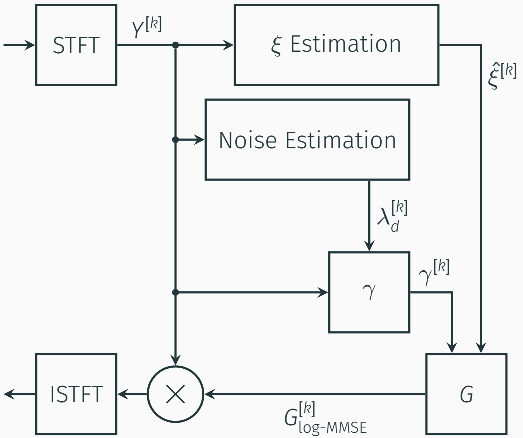
$$\mathbb{E}\left\{\left(\log A^{[k]} - \log \hat{A}^{[k]}\right)^2\right\}$$

Solution: (Assumes Gaussian distribution)

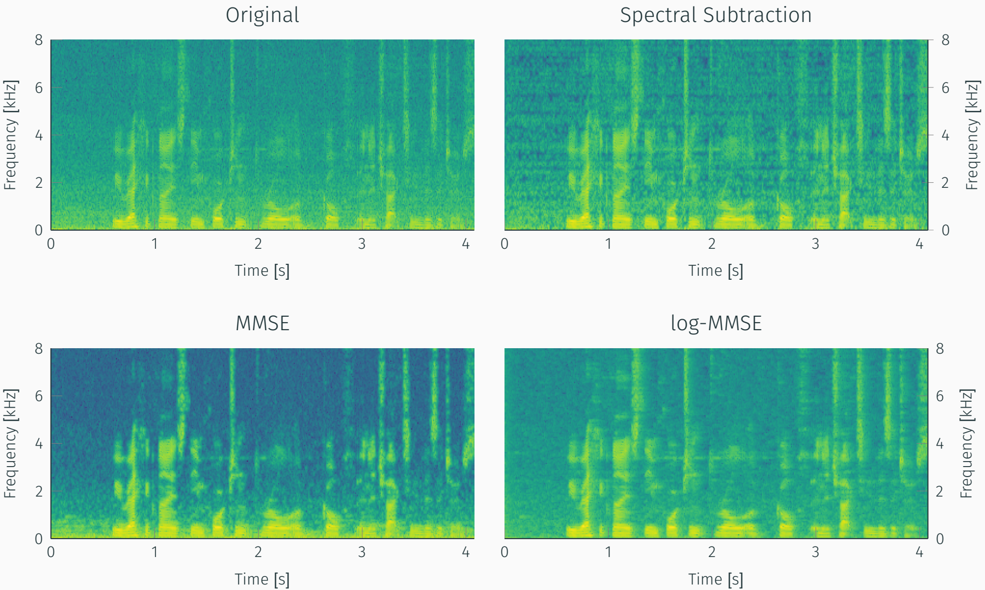
$$\begin{aligned}\hat{A}^{[k]} &= \exp \mathbb{E}\left\{\ln A^{[k]} \mid Y^{[k]}\right\} \\ &= G_{\text{log-MMSE}}\left(\xi^{[k]}, \gamma^{[k]}\right) \cdot R^{[k]}\end{aligned}$$

Notes:

- MMSE with different penalization.
- Better measure for speech [Gray et al. 1980].



MMSE and log-MMSE: Results



Incorporating Signal Presence Uncertainty (OM-LSA) [Cohen and Berdugo 2001]

Idea: Two hypotheses

$$H_0^{[k]} : Y^{[k]} = D^{[k]}$$
$$H_1^{[k]} : Y^{[k]} = X^{[k]} + D^{[k]}$$

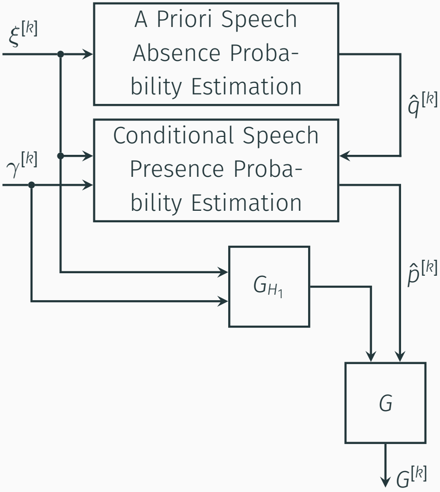
$$p^{[k]} := P\left(H_1^{[k]} \mid Y^{[k]}\right)$$

Solution:

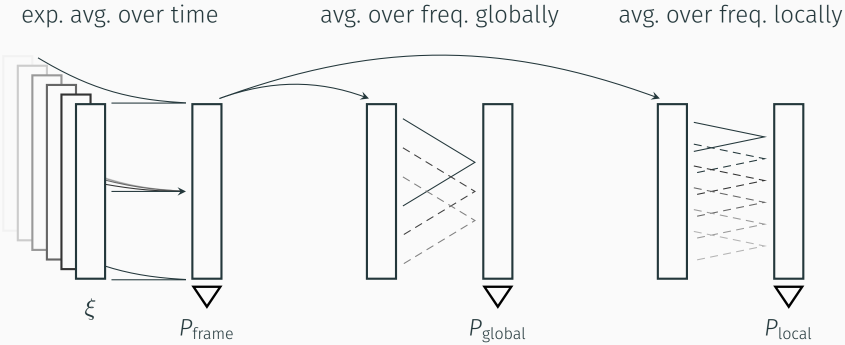
$$G\left(\xi^{[k]}, \gamma^{[k]}\right) = G_{H_1}^{p^{[k]}}\left(\xi^{[k]}, \gamma^{[k]}\right) \cdot G_{\min}^{1-p^{[k]}}$$

Estimate p_k via Gaussian model and

$$q^{[k]} := P\left(H_0^{[k]}\right)$$



Estimating the a priori Speech Absence Probability $q^{[k]}$ [Cohen and Berdugo 2001]



$$\hat{q}^{[k]} = 1 - P_{\text{local}}^{[k]} \cdot P_{\text{global}}^{[k]} \cdot P_{\text{frame}}^{[k]}$$

Maximum Likelihood:

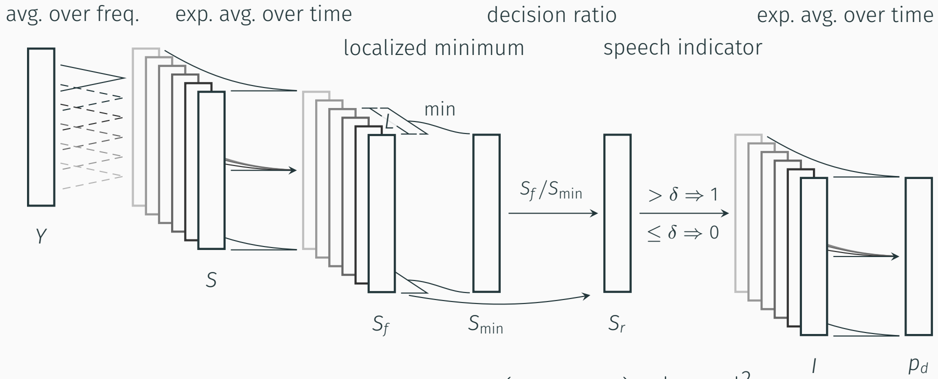
$$\bar{\gamma}^{[k,n]} = \alpha\bar{\gamma}^{[k,n-1]} + (1-\alpha)\frac{\gamma^{[k,n]}}{\beta}, \qquad 0 \leq \alpha \leq 1,\ \beta \geq 1$$
$$\hat{\xi}^{[k,n]} = \max\left\{\bar{\gamma}^{[k,n]} - 1, 0\right\}$$

Decision-Directed:

$$\hat{\xi}^{[k,n]} = \alpha G^2_{H_1}\left(\hat{\xi}^{[k,n-1]}, \gamma^{[k,n-1]}\right) \cdot \gamma^{[k,n-1]} + (1-\alpha)\max\left\{\gamma^{[k,n]} - 1, 0\right\}$$

Decision-directed approach usually has less musical noise.

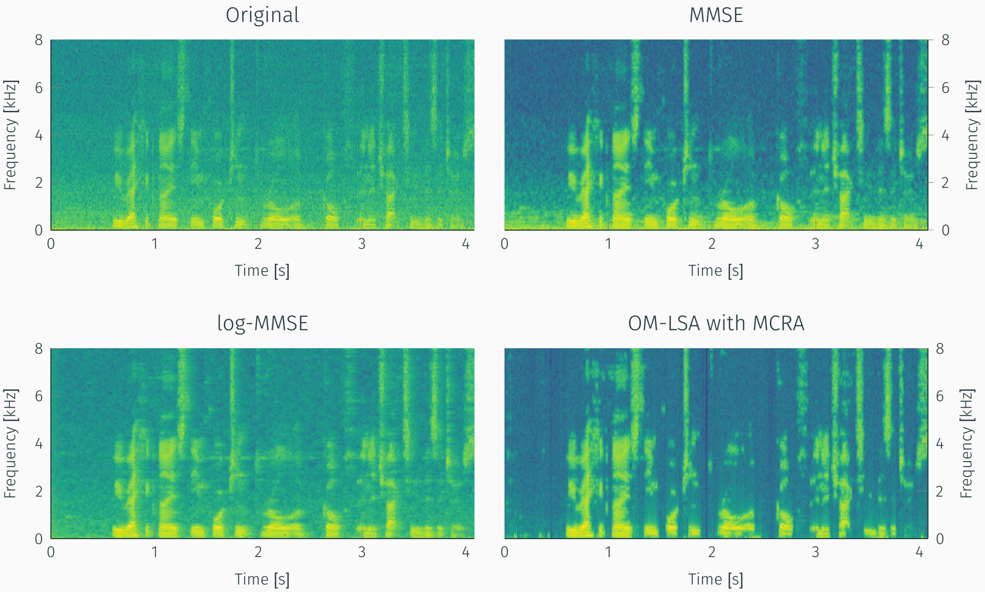
Minima controlled recursive averaging (MCRA):



$$\hat{\lambda}_d^{[k,n+1]} = \tilde{\alpha}_d^{[k,n]} \cdot \hat{\lambda}_d^{[k,n]} + \left(1 - \tilde{\alpha}_d^{[k,n]}\right) \cdot \left|Y^{[k,n]}\right|^2$$

$$\tilde{\alpha}_d^{[k,n]} = \alpha_d + (1 - \alpha_d) p_d^{[k,n]}$$

Results


Original
MMSE
log-MMSE
OM-LSA with MCRA
Frequency [kHz]
Time [s]
0
1
2
3
4
6
8

Demonstration

References

Chen, Jingdong et al. (July 2006). “New insights into the noise reduction Wiener filter”. In: *IEEE Transactions on Audio, Speech and Language Processing* 14.4, pp. 1218–1234.

Cohen, Israel and Baruch Berdugo (Nov. 2001). “Speech enhancement for non-stationary noise environments”. In: *Signal Processing* 81.11, pp. 2403–2418.

Ephraim, Y. and D. Malah (Dec. 1984). “Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 32.6, pp. 1109–1121.

— (Apr. 1985). “Speech enhancement using a minimum mean-square error log-spectral amplitude estimator”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 33.2, pp. 443–445.

Gray, R. et al. (Aug. 1980). “Distortion measures for speech processing”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28.4, pp. 367–376.

Loizou, Philipos C. (Feb. 2013). *Speech Enhancement*. CRC Press.