

Noise Reduction for AMR-WB

**Christophe Beaugeant (1), Imre Varga (1),
Peter Jax (2), Peter Vary (2), Thomas Lotter (2) (3)**

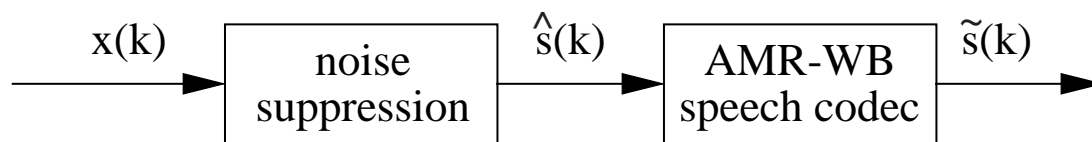
(1) Siemens AG, COM, Munich, Germany

(2) RWTH Aachen, IND

(3) Now with Siemens Audiologische Technik, Erlangen

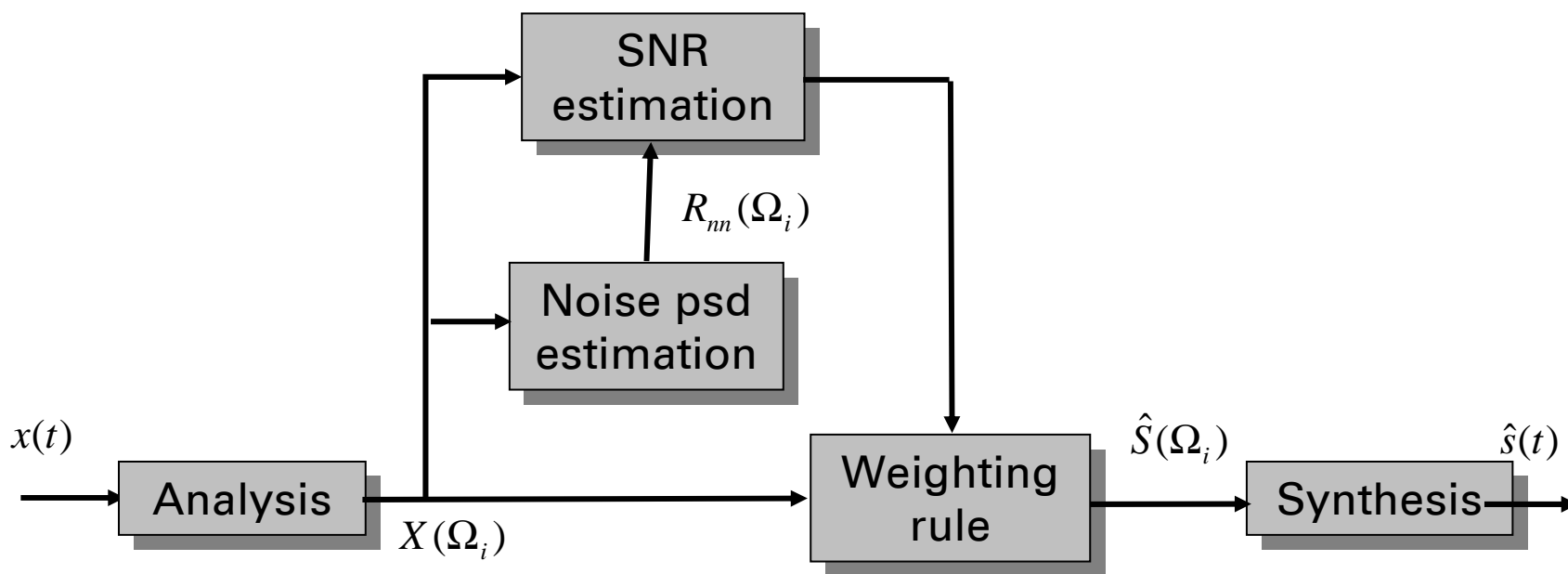
Introduction

- ❑ Development, implementation and optimization of a single-channel noise suppression algorithm as a preprocessing module for the Adaptive Multi-Rate Wideband (AMR-WB) speech codec .
- ❑ Performance requirements for the noise suppression :
 - Significant reduction of acoustic background noise,
 - No annoying artifacts in the residual noise
 - Good speech quality, especially no “clipping” and no degradations for clean input speech.



General description

- ❑ Processing in the frequency domain
- ❑ Noise estimation based on minimum statistics
- ❑ A priori SNR estimation
- ❑ MMSE_LSA weighting rule



Frequency analysis

Analysis:

- Pre-emphasis
- 20 ms frame length
- Windowing
- Zero-padding
- FFT

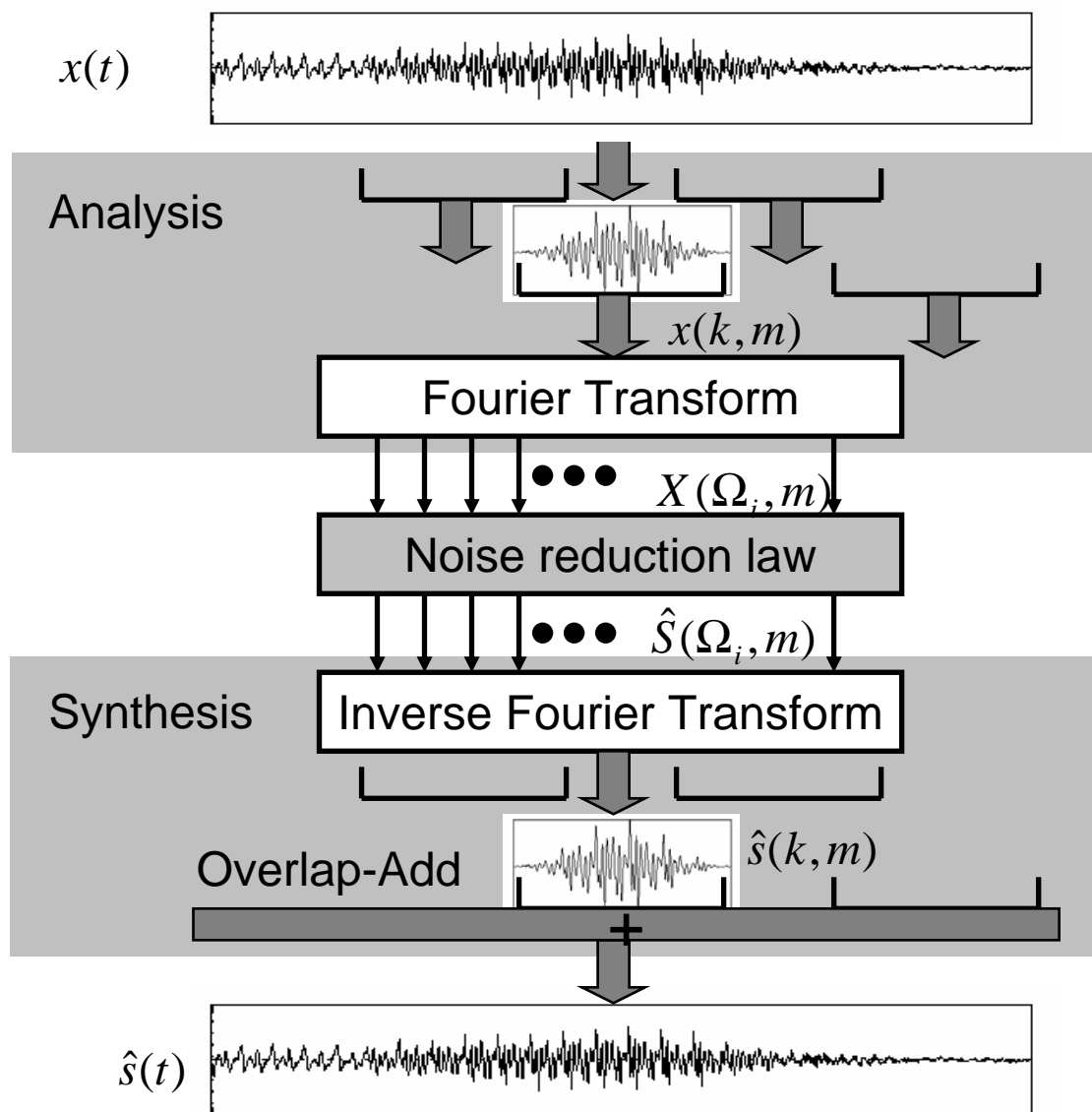
Synthesis

- IFFT
- Overlap-add
- De-emphasis

m : frame index

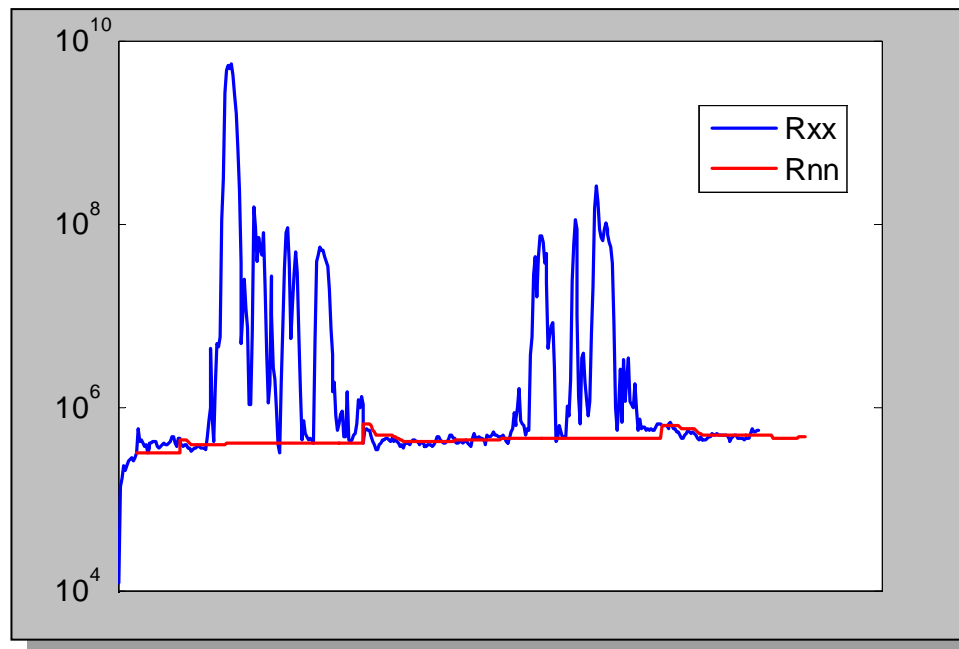
k : time index

Ω_i : frequency



Noise estimation (1)

- ❑ Noise estimation based on minimum statistic analysis [1]
- ❑ Principle:
 - Speech power spectra show peaks, considered as speech periods and valleys, considered as speech absence.
 - Noise power is determined during speech inactivity



[1] R. Martin, " [Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics](#) ", IEEE Transactions on Speech and Audio Processing, Vol. 9, No. 5, July 2001, pp. 504-512

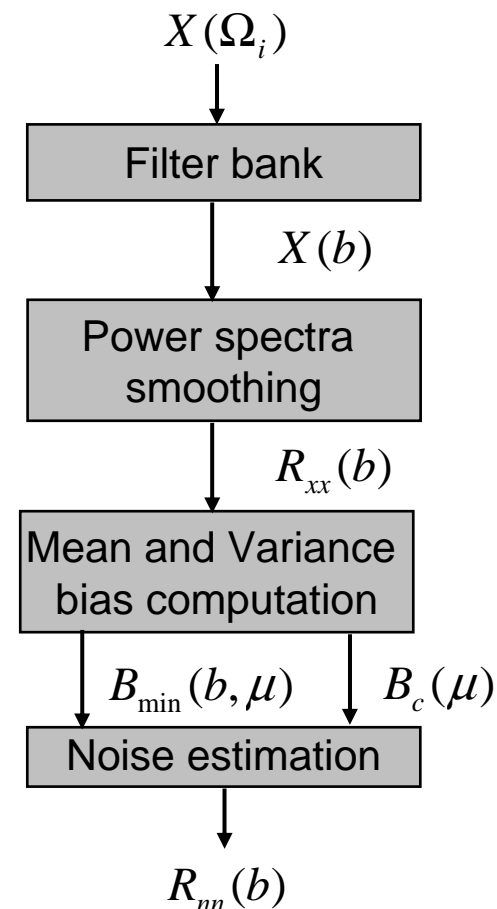
Noise estimation (2)

- ❑ Frequency bins grouped in bands
- ❑ Computation of noisy power spectrum

$$R_{xx}(b, m) = \alpha R_{xx}(b, m-1) + (1-\alpha) |\hat{X}(b, m)|^2$$

where $|\hat{X}(b, m)|^2 = \frac{1}{N_b} \sum_{i \in \text{band } b} |\hat{X}(\Omega_i)|^2$

Smoothing factor α varies as a function of *a posteriori* SNR.



Noise estimation (3)

- Minimum value of the L_{MS} length windowed power spectra (valleys) is benefited as the noise estimate

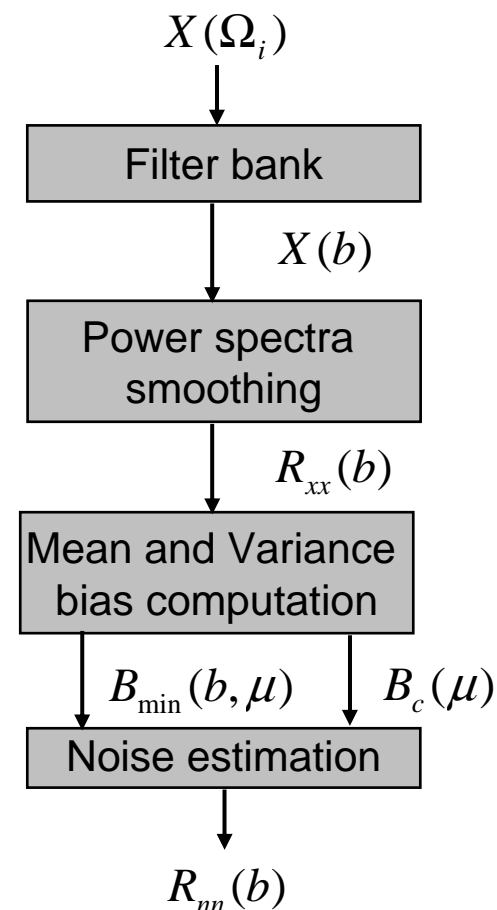
$$R_{nn}(b, m) = \min_{\mu=m-L_{MS}+1, m} \{B_{\min}(b, \mu)B_c(\mu)R_{xx}(b, \mu)\}$$

- $\{B_{\min}(m, k); B_c(m)\}$ to compensate mean bias and variance bias
- Overestimation of noise power

$$R_{nn}(b) = \gamma_{ov} \cdot R_{nn}(b)$$

- Noise estimate computed per frequency band

$$\{R_{nn}(\Omega_i)\}_{\Omega_i \in b} = R_{nn}(b)$$



SNR estimation (1)

- *A priori* SNR estimation with *Decision-Directed* approach [2]
 - To estimate the *a priori* SNR recursively

$$SNR_{prio}(\Omega_i, m) = \varepsilon \frac{\hat{S}^2(\Omega_i, m-1)}{R_{nn}(\Omega_i)} + (1 - \varepsilon) \cdot P[SNR_{post}(\Omega_i) - 1]$$

$$\varepsilon = 0.97 \dots 0.99$$

with $SNR_{post}(\Omega_i) = \frac{|X(\Omega_i)|^2}{R_{nn}(\Omega_i)}$
 $P[.]$: half - wave rectifier

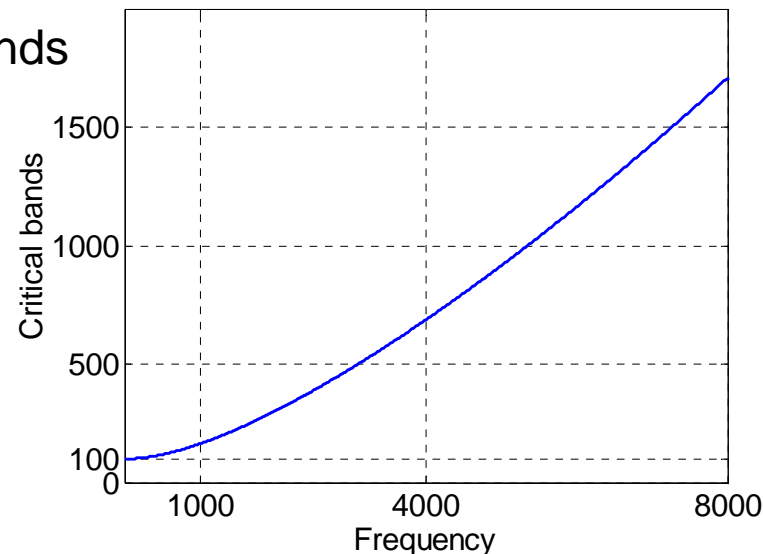
- *A priori* and *a posteriori* SNRs limitation to reduce the speech distortion

$$SNR(\Omega_i) = \max(SNR(\Omega_i), SNR_{global})$$

[2] Y. Ephraim und D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator". IEEE Trans. On ASSP, ASSP-32, Nr. 6, Dec. 1984

SNR estimation (2)

- ❑ Issue: Resolution of the FFT is not in accordance with human hearing system (bark scale)
- ❑ Higher mismatch in high frequency bands because of wider bark domains
- ❑ Consequence: Important difference between adjacent *a priori* SNRs is problematic principally around 6-7kHz
- ❑ Solution: Moving average smoothing



$$\hat{SNR}(\Omega_i) = \frac{1}{sw(i)} \sum_{j=0}^{sw(i)-1} SNR(\Omega_i - (sw(i)/2 + j) \cdot \Delta\Omega)$$

$sw(i)$ size of the moving average window
(depends on bark domain)

$$\Delta\Omega = \frac{2\pi \cdot f_s}{FFT_SIZE}$$

Weighting Rule [3-4]

□ MMSA LSA Weighting Rule

$$\hat{S}(\Omega_i) = \exp \left[E \left\{ \log [S(\Omega_i)] \mid X(\Omega_i), H_1^i \right\} \cdot P(H_1^i \mid Y_i) \right]$$

$$\hat{S}(\Omega_i) = H_{LSA}(\Omega_i) \cdot H_{soft}(\Omega_i) X(\Omega_i) \quad H_1^i \text{ State of speech presence}$$

□ Soft Decision weighting

$$H^{LSA}(\Omega_i) = \frac{SNR_{prio}(\Omega_i)}{1 + SNR_{prio}(\Omega_i)} \exp \left\{ \frac{1}{2} \int_{\psi(\Omega_i)}^{\infty} \frac{e^{-\tau}}{\tau} d\tau \right\} \quad \psi(\Omega_i) = \frac{SNR_{prio}(\Omega_i)}{1 - q(\Omega_i) + SNR_{prio}(\Omega_i)} \cdot SNR_{pos}$$

$$H_{soft}(\Omega_i) = \frac{\Lambda(\Omega_i)}{1 + \Lambda(\Omega_i)} \quad \Lambda(\Omega_i) = \frac{1 - q(\Omega_i)}{q(\Omega_i)} \cdot \frac{e^{\psi(\Omega_i)}}{1 + \frac{SNR_{post}(\Omega_i)}{1 - q(\Omega_i)}}$$

$q(\Omega_i)$ Probability of speech presence

[3] Y. Ephraim und D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator". IEEE Trans. on ASSP, ASSP-33, Nr. 2, April 1985

[4] D. Malah, R.V. Cox und A. Accardi, "Tracking Speech Presence Uncertainty to Improve Speech Enhancement in Non-Stationary Noise Environments, Proc.ICASSP, 1999

Integration

- ❑ Reduction of complexity without decrease of quality (Use of hearing properties)
 - Use of barkscale
 - Code optimization
- ❑ Low cost pre-processing
- ❑ Memory increase compared to narrow-band case is significant but acceptable

Complexity comparison	Wideband	Narrowband
WMOPS	6.33	3.16
SRAM	1466 Words	716 Words
DROM	1428 Words	840 Words
PROM	816 Ops	772 Ops
Scratch RAM	3564 Words	2226 Words

Conclusion

- ❑ Noise reduction system as pre-processing for AMR-WB
- ❑ Transcription of already used solution in narrow-band case
 - Proven quality
 - Proven complexity
- ❑ Use of human hear properties to enhance quality and to reduce complexity
- ❑ First step to exploit specific characteristics of wideband signal: limited resolution of human hear in high frequency range
- ❑ Perspective: advanced exploitation of the specific characteristics of wideband signals
 - Noise and speech properties above 4 kHz
 - Psychoacoustic