World Class Standards

# Single channel noise reduction

## Basics and processing used for ETSI STF 294

ETSI Workshop on Speech and Noise in Wideband Communication

Claude Marro
France Telecom

# Outline

- Scope of the presentation
- Classical speech enhancement techniques
- Tuning for real-world communications
- Processing used for ETSI STF 294

# Scope of the presentation

- **Single microphone noise reduction based on gain processing in the frequency domain:**
  - **Real time processing :**
    - **Low delay: < 30 ms (including acquisition frame), *e.g.* 24 ms max for SFT 294 database**
    - **"Reasonable" computation coast: < 20 WMOPS (typical at *Fs* = 16 kHz), *e.g.* 12 WMOPS max for SFT 294**

  ➔ **Realist for implementation in terminals or distributed in the network**

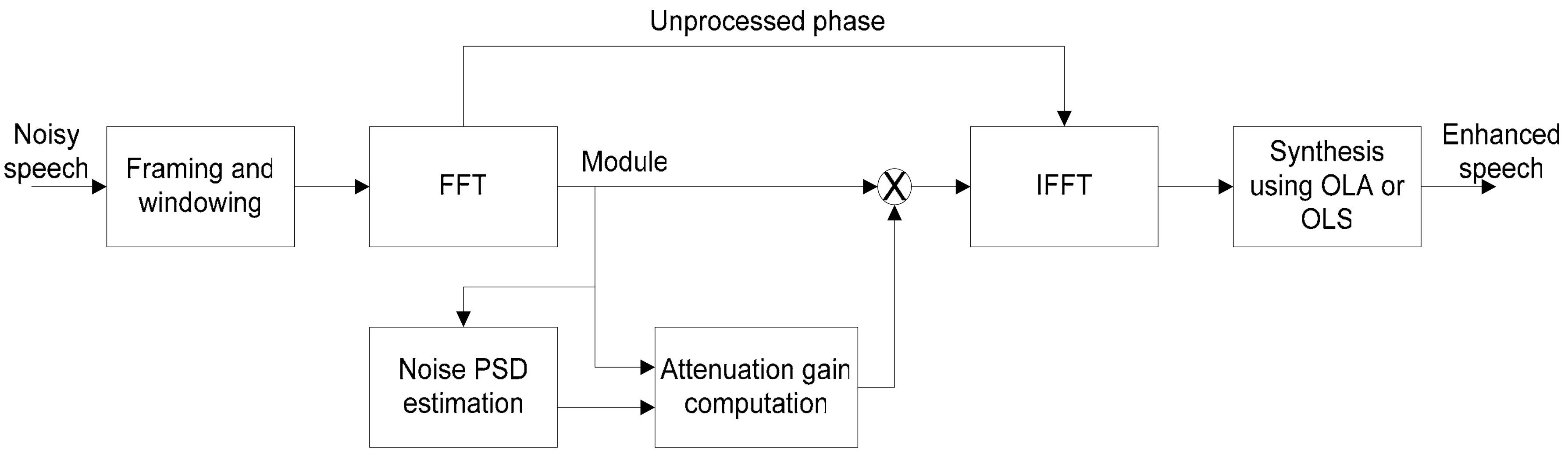
- **More complicated methods out of the scope:**
  - **Techniques based on model with training (HMM, etc.)**
  - **Multi-sensor approaches:**
    - **Using spatial properties: *e.g.* fixed & adaptive microphones arrays**
    - ***Blind Source Separation* (BSS): *e.g.* Time-Frequency separation or sparsity of signals**
    - **With "noise only reference": based on the knowledge of the corrupting signal**

# Outline

- Scope of the presentation
- Classical speech enhancement techniques
- Tuning for real-world communications
- Processing used for ETSI STF 294

# Speech enhancement principle

- **Characteristics**
  - **Block processing**
  - **Frequency domain implementation**
    - **Module processed by spectral attenuation**
    - **Noisy phase unprocessed**



# Basics

- Hypothesis
  - Additive model
  - Stationarity of speech and noise over frame duration
  - Speech and noise are independents

- Signal representations
  - Time domain $x(t) = s(t) + n(t)$

    $x(t) \rightarrow$ Noisy speech

    $s(t) \rightarrow$ Desired signal

    $n(t) \rightarrow$ Background noise
  - Frequency domain

$$|X(p,k)|e^{i\Phi_X(p,k)} = |S(p,k)|e^{i\Phi_S(p,k)} + |N(p,k)|e^{i\Phi_N(p,k)}$$

- Clean speech estimation: $\hat{S}(p,k) = G(p,k)X(p,k)$

**Wiener filter:** $G_w(p,k) = 1 - \frac{1}{SNR_{post}(p,k)} = \frac{SNR_{prio}(p,k)}{1 + SNR_{prio}(p,k)}$

# Signal-to-Noise Ratio estimation

- **Theoretical SNR estimators**
  - ***a posteriori* SNR**

$$SNR_{post}(p,k) = \frac{|X(p,k)|^2}{E\{|N(p,k)|^2\}}$$

  - ***a priori* SNR**

$$SNR_{prio}(p,k) = \frac{E\{|S(p,k)|^2\}}{E\{|N(p,k)|^2\}}$$

➔ **But in practice we know only** $X(p,k)$

  - **We must estimate:**

$E\{|N(p,k)|^2\}$ **and** $E\{|S(p,k)|^2\}$

# Signal-to-Noise Ratio estimation

- **Practical SNR estimators**
  - **Noise PSD**
    - **During speech pauses only (needs VAD)**

$$\hat{\gamma}_n(p,k) = \lambda \hat{\gamma}_n(p-1,k) + (1-\lambda)\,|X(p,k)|^2$$

Fogetting factor: $0 < \lambda < 1$

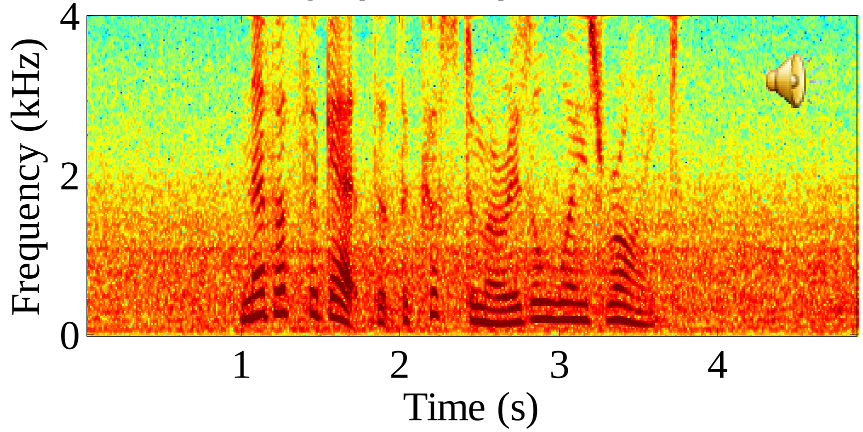
    - **Continuous noise estimation (Minimum Statistics like) [Martin 94]**
  - ***a posteriori* SNR:** $S\hat{N}R_{post}(p,k) = \dfrac{|X(p,k)|^2}{\hat{\gamma}_n(p,k)}$
  - ***a priori* SNR (Decision-Directed approach) [Ephraïm & Malah 84]**

$$S\hat{N}R_{prio}(p,k) = \beta \frac{|\hat{S}(p-1,k)|^2}{\hat{\gamma}_n(p,k)} + (1-\beta) Max\left(S\hat{N}R_{post}(p,k) - 1, 0\right)$$

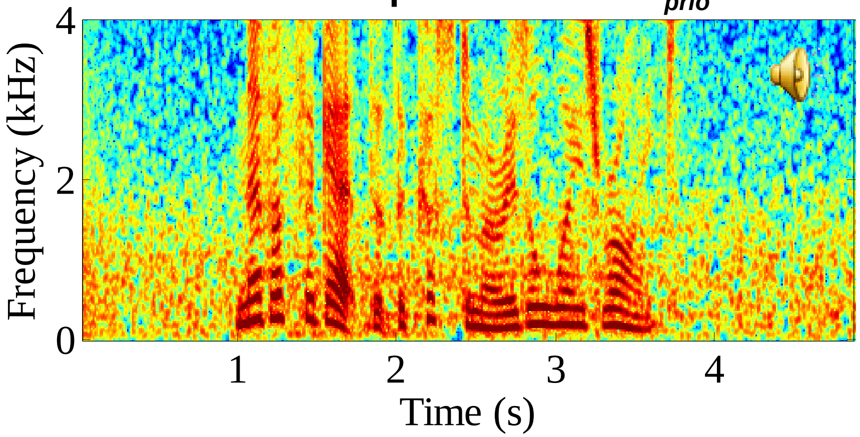
Typically, $\beta = 0.98$

# Importance of Decision-Directed approach: example

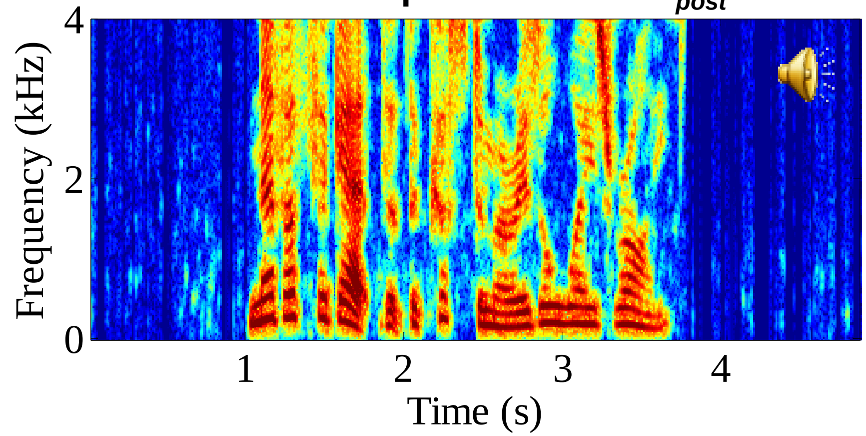
**Noisy speech spectrum**




**Gain computed with *SNR*$_{prio}$**




**Gain computed with *SNR*$_{post}$**



# Outline

- ❑ Scope of the presentation
- ❑ Classical speech enhancement techniques
- ❑ Tuning for real-world communications
- ❑ Processing used for ETSI STF 294

# Tuning for real-world communications

- ❑ **Ambient noise is a part of the communication**
  - ➢ Example: can't you talk without shouting?!!!
    - Hands-free in car:
    - Perfect noise reduction (clean speech):
    - More realist tuning (12 dB NR):
  - ➢ In some cases, background sounds can enrich the communication
  - ➔ Improve the listening comfort by reducing the noise without totally suppress it
- ❑ **The problem of noise reduction is not still solved**
  - ➢ ➘ noise ⇨ ➚ speech distortion
  - ➔ Compromise noise reduction level / desired signal distortion
  - ➢ This compromise involves various tunings parameters

# Outline

- Scope of the presentation
- Classical speech enhancement techniques
- Tuning for real-world communications
- Processing used for ETSI STF 294

# Processing used for ETSI STF 294

- **Algorithms**
  - **All algorithms based on short term spectral attenuation (Wiener filtering) with Decision-Directed SNR estimators**
  - **Difference between processings consist only in the choice of tuning parameters and of noise estimation procedure: ➔ taking into account typical behaviors of noise reduction algorithms**

- **Parameter 1**
  - **Aim: consider 2 families of noise PSD estimation**
    - **With noise estimation using VAD: efficient at moderate to high SNR**
    - **Continuous noise estimation: alternative for low SNR and tracking long term variation of noise during speech**

# Processing used for ETSI STF 294

- Parameter 2
  - Impact of the filter resolution
    - "Smooth" noise reduction filter: gain function limited to 65 coefficients (constraint applied in the time domain)
    - "Sharp" filter (257 coefficients)
  - ➔ Compromise between noise reduction sharpness (efficient in spectral valleys) and distortion of speech
- Parameter 3
  - Maximum noise reduction level
    - Moderate: threshold of -9 dB
    - More aggressive: threshold of -18 dB
  - ➔ Associated with parameter 2, set the dynamic of the noise reduction filter

# Typical example as conclusion

- Case of opposing tunings

Condition : car noise, handset

Noisy speech:

Processed, smooth filter, NR level of 9 dB:

Processed, sharp filter, NR level of 18 dB:

➔ Intermediate behaviours available in the database

# Thank you for the attention