Extraction and Representation of Features, Spring 2011

Lecture 6: Voice/Sound Activity Detection and De-Noising

Zheng-Hua Tan

Multimedia Information and Signal Processing
Department of Electronic Systems
Aalborg University, Denmark
zt@es.aau.dk

Outline

- Voice activity detection
  - Features
  - Classifiers
- De-noising
  - Spectral subtraction
  - Wiener filter
  - Non-local means de-noising

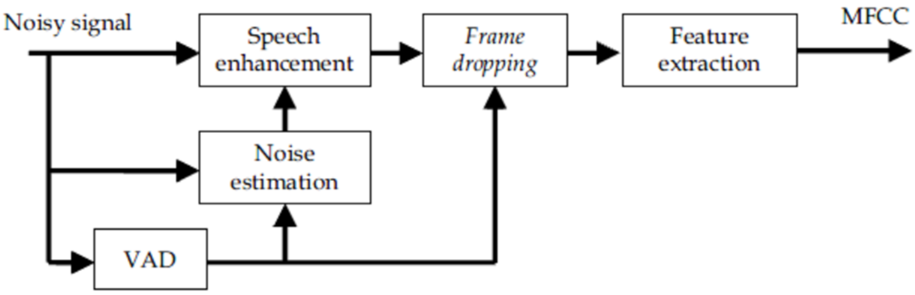
Voice activity detection

- To detect the presence or absence of speech in a segment of an acoustic signal.
- The detected non-speech segments can subsequently be abandoned to improve the overall performance of these systems.

Applications

- Wireless communications
- Real-time speech communication over the Internet
- Hearing aids devices
- Speech and speaker recognition
- Noise reduction

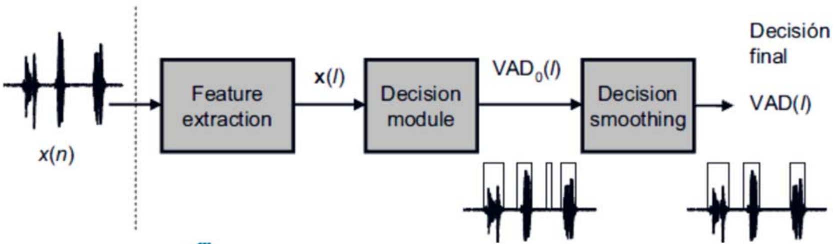
VAD (and de-noising) application



Techniques

- Noise robust features
  - Energy, Pitch detection, Spectrum analysis, Zero-crossing rate, Periodicity measure , MFCC, Entropy
- Decision rules/classifiers
  - Thresholding, SVM, GMM, decision tree, …

$$\frac{P(\mathbf{x}|H_1)}{P(\mathbf{x}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)}{P(H_1)}$$



Energy

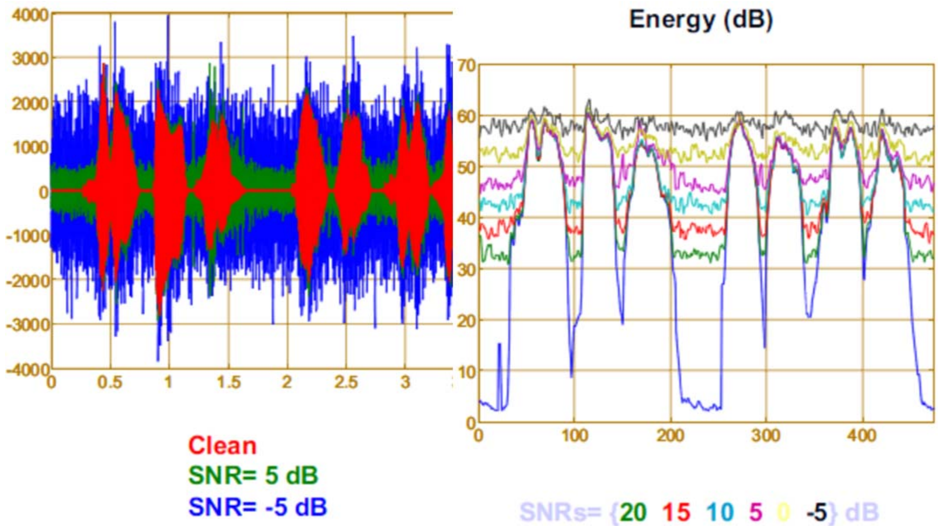
- One of the most common features
- *Et* at the *t*-th frame is computed as the logarithm of the signal energy; for N-length Hamming-windowed speech samples

$$E_t = \log \sum_{n=1}^{N} s_n^2$$

- The ratio of log-energy of the input frame to that of noise (*En):*

$$\frac{E_t}{E_n}$$

Energy



(Ramírez, 2007)

Zero crossing rate (ZCR)

- The number of times the signal level crosses zero
- ZCR ratio of the input frame to noise

$$\frac{Z_t}{Z_n}$$

where *Zt* denotes the ZCR of the input frame, and *Zn* denotes that of noise

GMM modeling and likelihood ratio

- A log-likelihood ratio of speech GMM to noise GMM for input frames is used for the GMM feature.
- The feature calculated as follows, where θ_s and θ_n denote the model parameter set of GMM for the speech and noise, respectively

$$\log(p(x_t \mid \theta_s)) - \log(p(x_t \mid \theta_n))$$

Performance evaluation

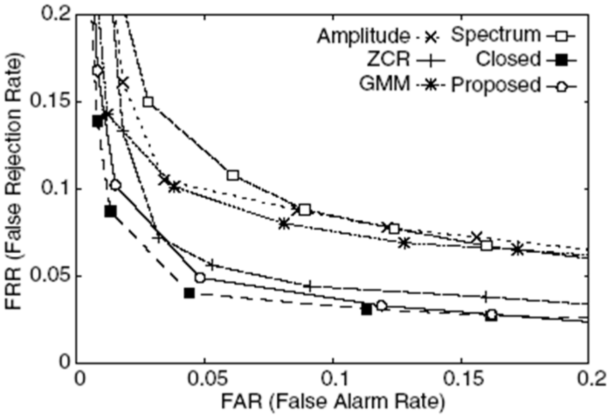
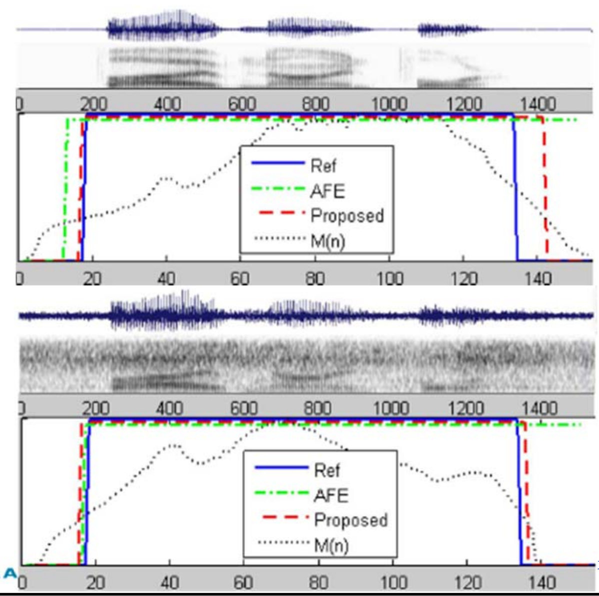


Figure 3: Sensor room:10db

(Kida, IS2005)

Performance evaluation



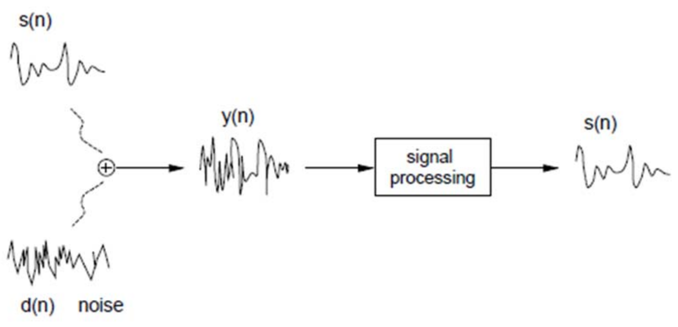
(Tan, 2010)

Outline

- Voice activity detection
  - Features
  - Classifiers
- De-noising
  - Spectral subtraction
  - Wiener filter
  - Non-local means de-noising

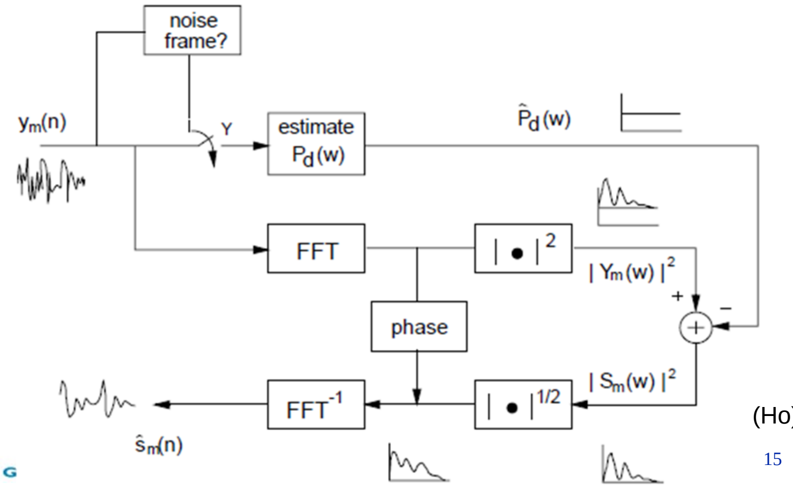
De-noising/enhancement

- Recover $s(n)$ from $y(n) = s(n) + d(n)$



Spectral subtraction

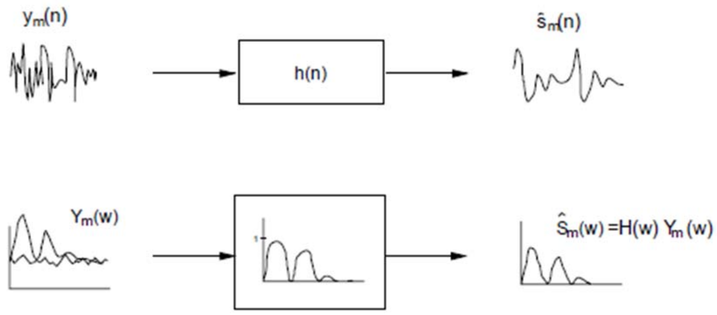
- Subtracting noise power spectrum from noisy signal power spectrum



(Ho)

Wiener filtering

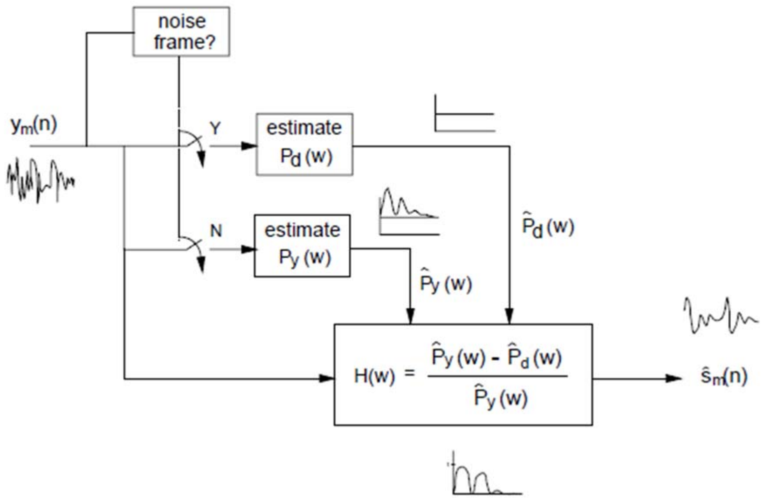
- Concept:



$$H(w) = \frac{P_s(w)}{P_y(w)}$$

- $H(\omega)$ weights spectrum according to SNR at different frequencies (Ho)

Wiener filtering



(Ho)

Iterative wiener filtering

- Estimating $\hat{P}_s(\omega)$ by $\hat{P}_y(\omega) - \hat{P}_d(\omega)$ may not be good
- Can do better by computing $\hat{P}_s(\omega)$ from the Wiener filter output
- Algorithm:

$\hat{P}_s(\omega)_0 = \hat{P}_y(\omega) - \hat{P}_d(\omega)$

$i = 0$

repeat

$$H(\omega)_i = \frac{\hat{P}_s(\omega)_i}{\hat{P}_s(\omega)_i + \hat{P}_d(\omega)}$$

$$S_m(\omega)_{i+1} = H(\omega)_i\, Y_m(\omega)$$

$$\hat{P}_s(\omega)_{i+1} = |S_m(\omega)_{i+1}|^2$$

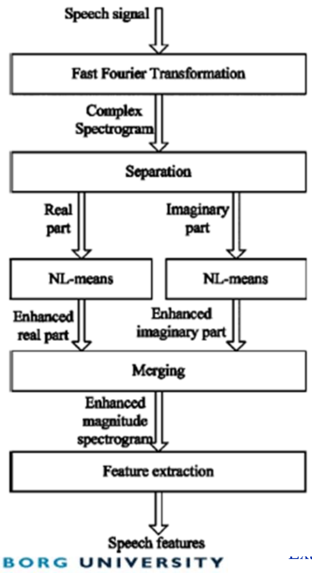
i=i+1

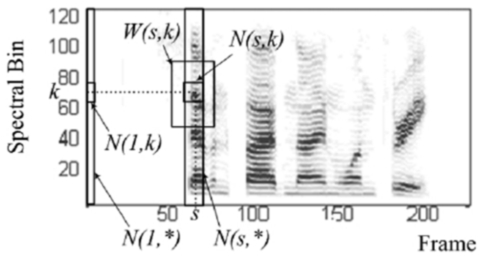
until convergence

Wiener filtering

- Wiener filtering is an optimum filter in the mean-square error sense
- Wiener filtering, assuming known signal and noise spectra, gives an upper bound in performance

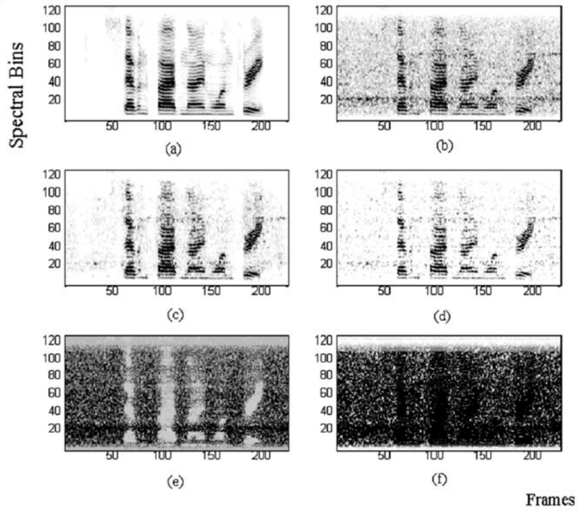
Non-local means de-noising





(Xu, 2008)

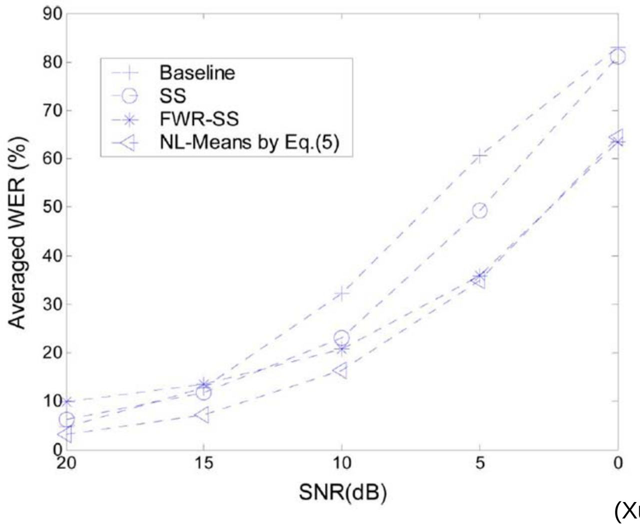
Non-local means de-noising



(Xu, 2008)

Fig. 3. Comparisons of magnitude spectrograms between the NL-means and SS methods. (a) Clean speech. (b) Clean speech corrupted by the "car" noise with 5-dB SNR. (c) Enhanced by NL-means. (d) Enhanced by SS. (e) Noise removed by NL-means. (f) Noise removed by the SS.

Non-local means de-noising



(Xu, 2008)

Fig. 4. WER (%) (averaged over different noise types) comparisons for a range of SNR values.

Summary

- Voice activity detection
  - Features
  - Classifiers
- De-noising
  - Spectral subtraction
  - Wiener filter
  - Non-local means de-noising

AALBORG UNIVERSITY