

I. 배경잡음 제거하기

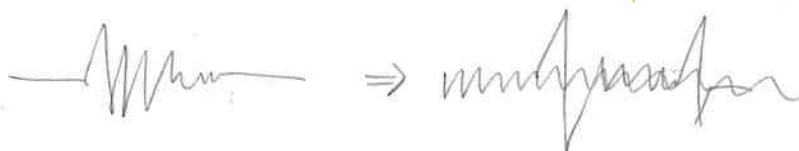
녹음환경에서는 추가되는 다양한 잡음들의 종류

① 주변환경 잡음: 음악, 자동차, 사람 등

② AC noise: 50~60 Hz의 sinusoidal 잡음

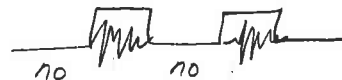
③ dc noise: base level이 올라감.

④ white noise
(백색잡음)



가정

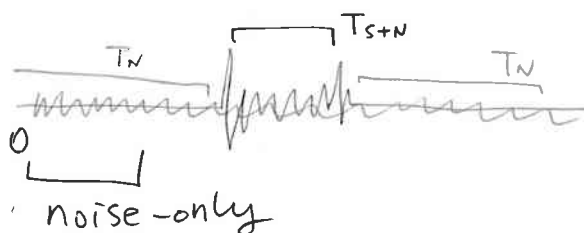
① speech signal은 시간에 따라 주파수 특성이 바뀌며 (non-stationary)
항상 존재하지 않는다.



② noise signal은 주파수 특성이 일정하다 (stationary)

③ noise-only 구간이 있다

가장 쉬운 방법으로 시작 시점으로부터 일정 시각까지는
음성이 없다고 가정한다. 예를 들어 0~0.5초는 noise-only



noise-only

그러면 noise-only 구간에서 noise spectrum을 얻을 수 있다.

관측신호 $x[t]$ 은

$$x[t] = s[t] + n[t] \Leftrightarrow X(\omega) = S(\omega) + N(\omega)$$

$$\tilde{N}(\omega) = E[F(n[t])] \approx \frac{1}{T} \sum_{t=0}^{0.5} F(x[t])$$

power spectral density를 구한다

$$\tilde{P}_{nn}(\omega) = |\tilde{N}(\omega)|^2 = \frac{1}{0.5} \sum_{t=0}^{0.5} X(\omega) \cdot X^*(\omega)$$

complex conjugate of $X(\omega)$

Spectral Subtraction

$$|\tilde{S}(\omega)|^2 = |S(\omega) + N(\omega)|^2 - |\tilde{N}(\omega)|^2 = X(\omega) \cdot X^*(\omega) - |\tilde{N}(\omega)|^2$$

Wiener filtering

$$\tilde{S}(\omega) = \underbrace{\frac{|X(\omega)|^2 - |\tilde{N}(\omega)|^2}{|X(\omega)|^2}}_{H(\omega)} X(\omega) = H(\omega) \cdot X(\omega)$$

$H(\omega)$ 를 이용하여 Linear-phase FIR filter를 구하고 적용한다.

II. EPD (end point detection)

noise-only 구간이 $0 \sim 0.5 \Rightarrow$ not always true

energy-based EPD를 하여 noise-only 구간을 구함.

이 구간에서 $|\tilde{N}(\omega)|^2$ 추정하기 가장.

① 입력 신호의 frame 별 ^{absolute} Energy를 구한다. 10ms 단위로

$$U[k] = \sqrt{\frac{1}{N_s} \sum_{t \in \text{frame } k} x^2(t)}$$

— 평균으로 나누어 주는 것과 square root를 씌워주는 것은 신호 scale에 맞춰

threshold를 쉽게 (직관적으로) 정할 수 있게 함

② initial thresholding
적당한 값을 정해

$$U[k] \geq \theta \rightarrow \text{voice}$$

$$< \theta \rightarrow \text{non-voice}$$



$$I[k] = \begin{cases} 1 & \text{if } U[k] \geq \theta \\ 0 & \text{if } U[k] < \theta \end{cases} \quad (I[k] \text{은 voice 인지 알려주는 indicator function})$$

③ $I[k]$ 에 대해서 median filtering or 적절한 smoothing
(보통 3~5 frames, 혹은)



④ $I[k]$ 에서 연속된 구간의 길이가 일정 길이 이상일 때만 voice로 (보통 100~300 ms)

⑤ 앞뒤로 extension (무성 자음과 fading-out 포함)

$$\textcircled{6} |\tilde{N}(\omega)|^2 = \frac{\sum_k (1 - I[k]) \cdot |N(k, \omega)|^2}{\sum_k (1 - I[k])}$$

