I. Exitation Signal Modeling

$e[t] = a[t] * x[t]$

↳ glottal pulse :

$F_0 = \frac{1}{\text{pitch}}$

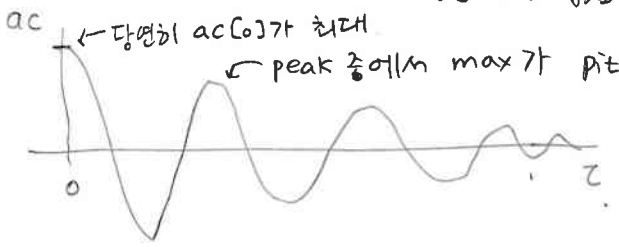
LPC analysis의 residual signal을 glottal pulse로 모델링하자

① pitch 찾기.

pitch 간격으로 가장 correlation이 높다

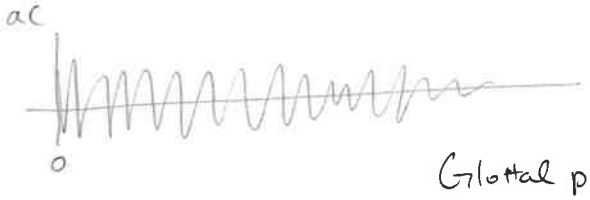
autocorrelation $ac(\tau) = \sum_{t=\tau+1}^{N_F} x[t] \cdot x[t-\tau]$

* 평균계산 없음. * τ가 큰수록 끝에지는 term 개수작음



(voiced)

* -3. autocorrelation은 residual이 아닌 (x)
preemphasis (x)
x[t] (o)



(unvoiced, 거의 Gaussian random)

Glottal pulse가 아니므로 pitch가 없다

⊛ V/UV/silence detection이 선행되어야 함

a) $\frac{ac[1]}{ac[0]} \gtrless \theta$ → UV / → V decision rule

θ = 0.5 정도

정확도는 떨어짐

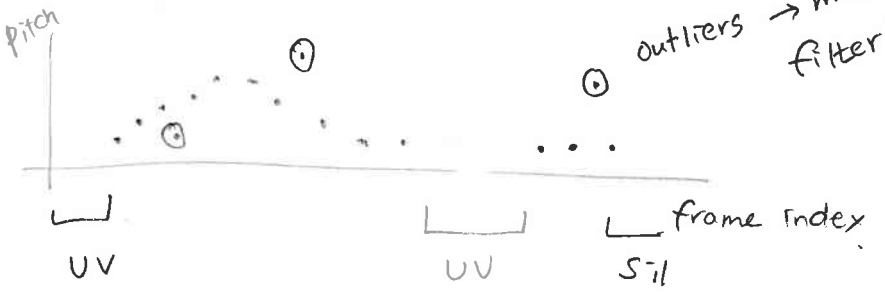
energy threshold와 같이사용

Voiced는 Vocal cord 통과하면서 high frequency energy 더 많이 흡수됨

UV는 ac[0]나 ac[1]이나 차이가 크지 않음

(주의) no preemphasis

b) project 2에서 학습한 V/UV/S classifier 사용



② glottal pulse 생성

type 1 :

type 2 : (추천)

π 2π 3π 4π

type 3 :

↳ 가장 glottal pulse에 가깝지만 간격 정하기 힘듦

③ Concatenation to previse exitation : 시작점 정하기

e[k-1, :]

e[k, :]

붙이기

이전 frame의 last pulse

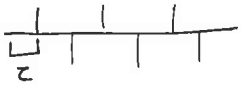
현재 frame의 첫번째 pulse

p[k-1] or p[k] ? → 해보기

③-2 Concatenation은 오류가 누적됨

따라서 residual과 가장 잘 align되는 시작점을 찾는 방법도 있음

$r[t] = a[t] * \alpha[t]$ (residuals)

$\hat{e}(t, \tau) =$  $\tau = 0 \sim \frac{p}{2}$

$\tau^* = \arg\max \sum \hat{e}[t, \tau] \cdot r[t]$ cross-correlation with residual

or

$\arg\max \sum \text{sign}(\hat{e}[t, \tau]) \cdot \text{sign}(r[t])$ 부호들의 cross-correlation

④ normalization

생성된 pulseel energy는 residual signal과 같도록 Scaling

$$g[k] = \sqrt{\frac{1}{N_f}\sum r^2(t)}$$ (residual의 standard deviation)

$$\hat{e}[k,t] \leftarrow g[k]\frac{\hat{e}[k,t,\tau^*]}{\sqrt{\frac{1}{N_f}\sum \hat{e}^2[k,t,\tau^*]}}$$ (또는 frame gain $g[k]$를 다른 parameter로 저장

⑤ Unvoiced ?

$$\hat{e}[k,t] \leftarrow g[k] \cdot \text{rand.gauss}(0,1)$$

⑥ Encoding 6bytes/frame = 600 bytes/sec

Flagvuv[k] byte : voiced, unvoiced, silence

G[k] : byte or short — excitation gain or frame gain

p[k] : short — pitch

offset[k] : byte — τ^*

[decoding]