

LPC Vocoder 의 Excitation Source 개선에 관한 연구

○ 전 지 하
성균관대학교

이 근 영
전자공학과

An Enhanced Excitation Source in LPC Vocoder

Ji-Ha Jeon
Dept. of Electronics

Keun-Young Lee
Sung Kyun Kwan Univ.

Abstract

This paper describes a new technique for the generation of excitation sources in LPC system. We synthesize a speech signal using several excitation sources, according to residual signal energy and ZCR(zero Crossing Rate). One of the excitation sources mix the double differentiated glottal wave form source and noise source. As a result, we got improved speech signal than that produced by conventional LPC system.

1. 서 론

선형 예측 coding(LPC)은 음성 합성에 적합하지만 1) 유무성 판별, 2) 유성음에서 pitch 추출, 3) 합성기에서의 정확하지 못한 excitation source 등에서 발생하는 error로 인하여 자연스럽지 못한 합성음이 만들어 진다. 그림 1에 일반적인 LPC Vocoder를 나타내었다. 이 시스템은 유성음일 경우에는 pulse를 발생시키고 무성음일 경우는 noise를 발생시킨다.

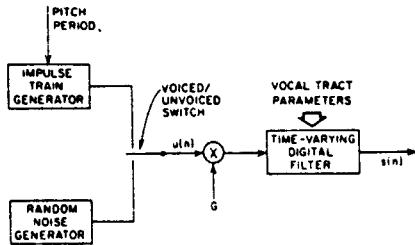


그림 1. 일반적인 LPC Vocoder 시스템

그러나, 그림 2에 나타난 것처럼 일반적인 residual 신호는 pulse를 따라 noise 성분이 혼합되어 있다. 따라서 이런 경우에는 pulse와 noise 성분을 혼합시킨 excitation source를 사용하는 것이 적절하다.

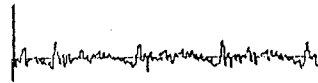


그림 2. 아 음의 residual 신호

본 논문에서는 LPC system에서 새로운 excitation source를 도입해서 음성을 합성했다. 또한 성문에서의 영향을 고려하여 impulse 대신 glottal waveform 파형을 2차 미분하여 그 파형을 pulse 성분으로 사용했다. 이렇게 함으로써 기존의 LPC system보다는 양질의 합성음을 얻을 수 있었다.

2. Pitch 주기와 Gain

Pitch 주기와 이득을 정확히 추출하기 위하여 2차 예측 계수를 가지고 residual 에러 신호에 선형 예측 방법을 적용하였다.

에러 신호를 $\delta(n)$ 이라고 하면 $\delta(n)$ 의 에너지는 다음과 같다.

$$\begin{aligned}
 E(M) &= \sum_{n=0}^{M-1} \delta^2(n) \\
 &= \sum_{n=0}^{M-1} \{e(n) - a_1 e(n-M+1) - a_2 e(n-M)\}^2 \\
 &= 1 - \frac{R^2(M-1) + R^2(M) - 2R(1)R(M-1)R(M)}{1 - R^2(1)}
 \end{aligned}$$

여기서 N은 frame의 크기이며 R(M)은 residual signal의 자기 상관 함수이다.

K(M)을

$$K(M) = R^2(M-1) + R^2(M) - 2R(1)R(M-1)R(M) \quad (1)$$

이때 Pitch 주기는 K(M)의 최대치로 부터 구해진다. excitation source의 에너지를 나타내는 B는 $B = \sqrt{1-E(M)}$ 이다

3. 음성의 분석 및 합성

(1) Parcor 방식

본 논문에서는 LPC 방법중 Parcor (Pattial Correlation) 방식을 사용해 음성을 분석 및 합성하였다.

Parcor 방식은 음성 sample들로 부터 직접 예측 계수를 구하며 system이 안정되어 있다.

Burg가 제안한 Parcor 계수값 K_i 는 다음과 같다.

$$K_i = \frac{2 \sum_{m=0}^{M-1} [e^{(i-1)}(m)b^{(i-1)}(m-1)]}{\sum_{m=0}^{M-1} [e^{(i-1)}(m)]^2 + \sum_{m=0}^{M-1} [b^{(i-1)}(m-1)]^2}$$

여기서 i 는 Parcor 계수의 차수이며 m 은 sample 수, e 는 전향 예측 신호, b 는 후향 예측 신호이다.

(2) EXCITATION SOURCE

미소 구간에서 residual 신호의 에너지 EN은

$$EN = \sum_{m=0}^{M-1} x^2(m) \quad (2)$$

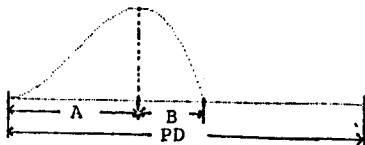
이며

영 교차율 ZCR은 $ZCR = \sum_{m=0}^{M-1} |sgn[x(m)] - sgn[x(m-1)]|$ 이다. (3)

EN은 유성음 부분에서 높고 무성음 부분에 낮으며 영 교차율 ZCR은 유성음 부분에서 낮고 무성음 부분에서 높다. 이 두 요소로 residual 신호를 판단 음성 합성시 사용되는 excitation source를 결정한다.

ZCR이 클때는 noise 성분을 preemphasis했고, EN이 클 때는 성문에서의 영향을 고려하여 glottal pulse를 2차 미분한 파형을 사용하였다.

그림 3에 glottal pulse를 나타내었다.



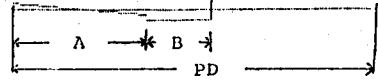
$$u(x) = a[3(\frac{x}{A})^2 - 2(\frac{x}{A})^3] \quad (\text{구간 A})$$

$$a[1 - (\frac{x-A}{B})^2] \quad (\text{구간 B})$$

그림 3. glottal pulse

그림 3에 나타난 $U(x)$ 에 나타난 식을 2번 미분한 파형을 그림 4에 나타냈다.

이때 Pitch 주기가 $PD=8ms$ 이라하면 $A=3.25ms$
 $B=1.25ms$ 로서 파형의 비율을 결정한다.



$$u''(x) = \frac{6a}{A^2} (1 - \frac{x}{A}) x \quad (\text{구간 A})$$

$$a(-\frac{2}{B}) \quad (\text{구간 B})$$

그림 4. glottal pulse 이차미분파형

4. 실험 결과 및 고찰

음성 신호를 10KHz로 sampling해서 저장하였다. 이것을 200 sample씩 커내어 분석 차수가 10차인 lattice filter에 입력시켜 분석 하였다. 이때 이 200 sample을 가지고 식 (3)을 이용해 영 교차율을 구했다. pitch 주기 PD와 residual signal energy EN 은 Lattice filter의 마지막단의 출력 residual signal로 부터 식 (1), (2)를 통해 구했다.

본 논문에서 사용한 excitation source의 회로도를 그림 5에 나타냈다.

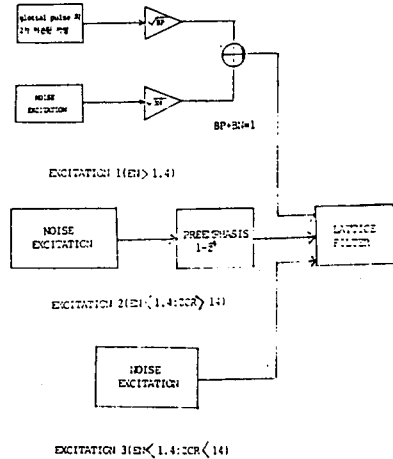


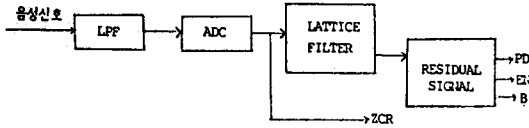
그림 5. Excitation source block diagram

이때 사용된 각 excitation source는 EN 과 ZCR에 따라서 선택된다.

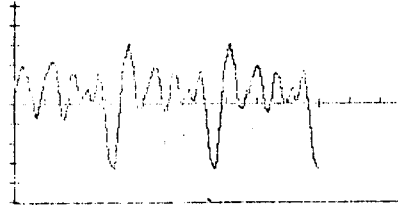
- Excitation 1: EN이 높은 경우로서 Pulse성분과 Noise성분을 혼합했다. 여기서 B 는 residual 신호로부터 구해지며 Pulse 와 Noise Source의 혼합율을 조정 한다.
- Excitation 2: EN 이 낮고 ZCR이 높은 경우 로서 이때는 noise를 preemphasis 시켜서 고역을 강조했다.
- Excitation 3: EN과 ZCR이 둘다 낮은 경우로서 이때는 noise 만을 발생 시켰다.

Excitation 1,3 에서는 각 pitch 마다 균일한 noise 발생을 시켰다. Threshold 는 $EN=1.4$ $ZCR=14$ 로 주었다. 그림 6에 음성 분석 및 합성의 전체 block를 나타냈으며

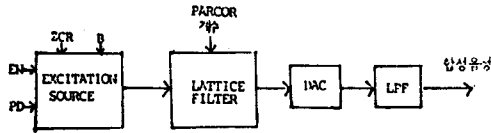
그림 7에는 residual signal 에 pulse 와 noise 가 혼합되어있는 한 frame 의 합성 파형 과 spectrum을 나타냈다.



a) 음성의 분석

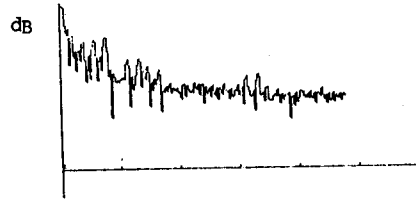


(c)



b) 음성의 합성

그림 6. 음성의 분석 및 합성 block diagram



(d)

그림 7. a) residual 신호가 pulse 와 noise 성분이 혼합되어 있는 음성파형

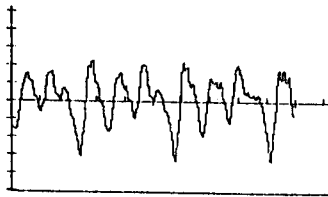
b) (a) 파형의 spectrum

c) (a) 파형의 합성파형

d) (c) 파형의 spectrum

5. 결 론

LPC Vocoder 중 에서 Parcor 방식을 이용 음성을 분석 및 합성했다. residual 신호가 pulse 와 noise 성분이 혼합된 경우 B값을 이용해 두요소의 크기를 조정해서 혼합한 excitation source 를 사용했는데, 이때 glottal pulse 를 pulse 성분으로 사용했다. 이와같이 기존의 간단한 excitation source 대신 좀 더 효율적인 excitation source 를 사용함으로써 합성음의 질을 높일수 있었다.



(a)



(b)

* 참고 문 헌 *

1. L.R.Rabiner, R.W.Shafer, "Digital Processing of Speech Signal", Prentice Hall, inc., N.J., 1978
2. J.D.Markel, A.H.Gray, JR., "Linear Prediction of Speech Signal", Springer Verlag, Berlin Heidelberg, N.Y., 1976
3. M.R.Matanssek, V.S.batalov, "A New Approach to the determination of the Glottal Waveform", IEEE, Trans., Acoustics, Speech, and Signal Processing, vol. ASSP-28, No. 6, pp616-622, dec., 1980
4. S.Y.Kwon, A.J.Goldberg, "An Enhanced LPC Vocoder with NO Voiced/Unvoiced Switch", IEEE, Trans., Acoustics, Speech, and Signal Processing, Vol. ASSP-32, No. 4, pp 851-858, august., 1984
5. S.Saito, K.Nakata, "Fundamentals of Speech Signal Processing", Academic Press, 1985