# 적응 예측에 의한 음성 합성에 관한 연구

°김 인 철, 방 성 일, 진 연 강
단국대학교 전자공학과

# A Study on Speech Synthesis by the Adaptive Prediction

In Cheol KIM, Sung Ill BANG, Yeon Kang CHIN
Dept. of Electronic Eng., DanKook Univ.

## ABSTRACT

APC (Adaptive Predictive Coding), one of the Digital Speech Coding Methods, requires lower bite-rate than other time-domain waveform coding methods, for it has an another predictor (Long-term predictor).

In this study, We obtained each parameters in Korean vowels and sentence by using APC method. Then, We synthesized the speech with these parameters.

## 1. 서 론

아나로그 신호를 디지틀화 하는것은 신호를 효율적으로 재생 할 수 있고, 암호화가 쉬우며, 신호를 쉽게 저장 할 수 있는 장점이 있다. 특히, 음성 신호의 디지틀화는 음성의 저장과 디지틀 전송, 음성 합성 장치, 화자(Speaker) 인식 및 확인, 음성 인식 등의 분야에 적용 가능 하다[1,2].

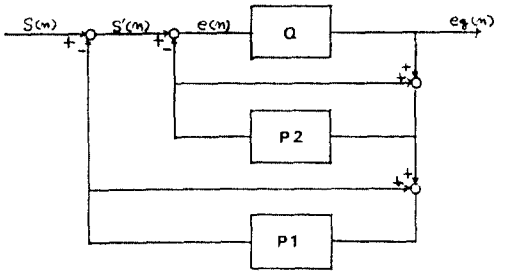
음성 신호의 부호화 방법에는 크게 파형 부호화 방법과 인간의 발성 기관을 모델링 하는 Source Coding 방법으로 구분되고, 파형 부호화 방법은 시간 영역 부호화 방법과 주파수 영역 부호화 방식으로 나누어 진다[3].

본 연구에서는 음성 신호의 pitch 성분을 제거하기위한 또 하나의 예측기를 갖는 적응 예측 부호화(Adaptive Predictive Coding) 방법에 의해 한국어 단모음과 3초 정도의 길이를 갖는 문장으로부터 각 파라미터들을 구하고, 이 파라미터들을 이용해서 음성을 합성했다.

## 2. 적응 예측 부호화 방식의 구조

적응 예측기의 구조는 <그림1>과 같이 long-term 예측기 P1(z), short-term 예측기 P2(z), 그리고 quantizer Q의 3가지로 구성되며, 이 두 예측기의 역할은 음성 신호의 여유분을(redundancy) 제거하는 것이다[4,5].

long-term 예측기 P1(z)은 음성 신호의 준 주기적 성질(quasi-periodic)에 기인한 long term 여유분을 제거하기 위한 것이고, short-term 예측기 P2(z)는 연속된 음성 샘플들 사이의 상관 관계에 의해 short-term 여유분을 제거 하는 것이다.




<그림1> 적응 예측기의 구조

P1(z)와 P2(z)는 각각 다음과 같다.

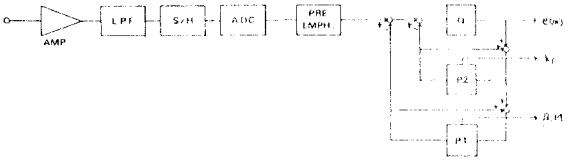
$$P1(z) = \beta z^{-M} \quad (1)$$

$$P2(z) = \sum_{i=1}^{p} a_i z^{-i} \quad (2)$$

여기서, $\beta$ 와 $\alpha_i$는 예측기의 계수 이고, p는 short-term 예측기의 차수, M은 pitch 주기 또는 pitch 주기의 정수배의 값을 가지며 일반적으로 M>p 이다[3,5].

## 3. 음성 신호의 부호화

적응 예측 방식에 의한 음성 신호의 부호화는 <그림2>와 같은 과정으로 수행 했다.




<그림2> 음성 부호화 과정

(1) Pre-emphasis

입력된 음성 신호는 (3)식으로 주어지는 필터를 이용해 Pre-emphasis 한다[6].

$$P(z) = 1 - \mu z^{-1} \qquad (3)$$

$\mu$가 ±1에 가까워 지면 ± 6 db/oct. 의 이득을 갖는다.
본 연구 에서는 계산의 편리를 위해 1로 했다.

(2) 피치 주기 M의 산출

피치 주기를 검출하는 방법은 여러가지가 있으나 덧셈과 뺄셈 연산 만을 사용해 구할 수 있는 AMDF (Average Magnitude Difference Function)를 사용한다[7].
AMDF 는 (4)식으로 주어지며 AMDF 의 값을 최소로 하는 i 값이 M 값이 된다.

$$AMDF(i) = \sum_{i=1}^{N} \left| s(n) - s(n-i) \right| \qquad (4)$$

(3) $\beta$ 값의 산출

최적의 $\beta$ 값은 long-term 예측기의 예측 오차 신호 S'n의 평균 제곱 오차(Mean Square Error)를 최소로 하는 값으로 결정된다.
최적의 $\beta$ 값은 (5)식으로 주어진다[4].

$$\beta = \frac{< Sn \cdot Sn-M >}{< Sn^2-M >} \qquad (5)$$

여기서, < > 는 시간 평균(time average)을 나타낸다.

(4) short-term 예측기

short-term 예측기의 계수를 구하는 방법은
1. Autocorrelation Method
2. Covariance Method
3. Lattice Method 등이 있다[1,9,10]

여기서는, correlation 함수를 구하는 중간 과정을 거치지 않고 입력 음성 신호로 부터 직접 반사 계수(Reflaction Coeff.)를 구해 예측 오차 신호를 얻을 수 있는 lattice 방법을 사용했고, lattice 필터의 구조는 2-multiplier 구조를 사용했다.
반사계수 ki는 조화 평균 방법(Harmonic Mean)을 사용해 계산 했다[1,9].

## 4. 음성 합성

음성 합성의 음성 부호화의 역 과정으로 20 m sec. 간격으로 계산되는 예측기 계수 $\beta$, ki, M 과 예측 오차 신호 e(n) 에 의해 합성 되며, 음성 합성 과정은 <그림3>과 같다.


<그림3> 음성 합성 과정

De-emphasis는 Pre-emphasis의 역 과정으로 (6)식과 같은 De-emphasis 필터로 행한다.

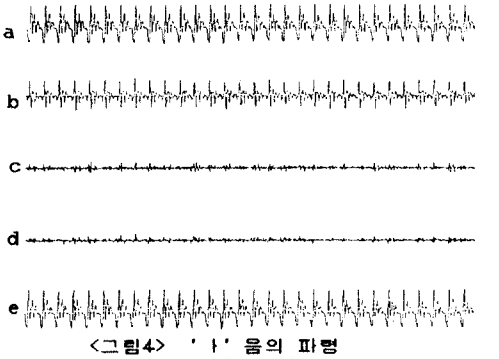
$$P(z) = \frac{1}{1 - \mu z^{-1}} \qquad (6)$$

## 5. Simulation 및 결과

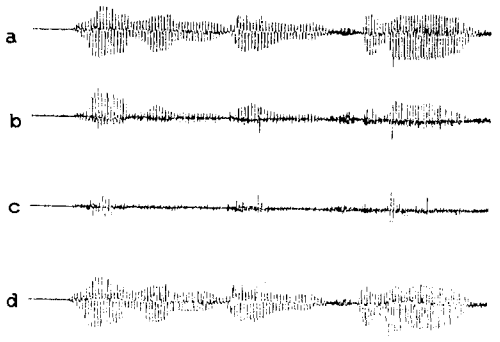
short-term 예측기의 차수는 12차로 simulation 했다.
표본화 주파수는 8KHz 로 했고 6차 Butterworth 필터를 이용해 3.4KHz로 음성 신호를 대역 제한 한 후 12 bits로 A/D 변환 했다.
한 frame을 20m sec.(160 샘플)로 했고, 각 계수들은 20m sec. 간격으로 구해진다.

simulation은 IBM-PC/XT에서 C와 assembly 언어로 행했으며 시간은 약간 소요 되었다.

<그림4>는 한국어 단모음 'ㅏ' 에 대한 각 부분의 파형이며 <그림5>은 '나는 당신을 사랑합니다'라는 문장을 중 일부의 파형이다.




<그림4> 'ㅏ' 음의 파형
a. 원래의 음성 신호
b. pre-emphasis된 음성 신호
c. long-term 예측 오차
d. 전체 예측 오차
e. 합성된 음성 신호




<그림5> 문장음의 파형
a. 원래의 음성 신호
b. pre-emphasis된 음성 신호
c. 전체 예측 오차
d. 합성된 음성 신호

## 6. 결 론

위의 적응 예측 부호화 방식을 이용해 한국어 단모음과 문장을 합성 했으며, 합성 음질은 좋은 편이었다.

long-term 예측기 계수 β는 유성음 구간 에서는 0.9 - 1.1 사이의 값을 가졌고, 무성음 구간에서는 0.1 - 0.3 사이의 값을 가졌다.

이 적응 예측 부호화 방식을 사용하면 낮은 bit-rate에서 양질의 음성합성기를 구성 할 수 있을 것이다.

## References

[1] L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, 1978.

[2] R.W. Schafer and J.D. Markhoul, Speech Analysis, IEEE Press, 1978.

[3] J.L. Flanagan, et al., "Speech Coding," IEEE Trans. Comm., vol.COM-27, no.4, pp.710-737, Apr. 1979.

[4] B.S. Atal and M.R. Schroeder, "Adaptive Predictive Coding of Speech Signals," Bell Syst. Tech. J., pp.1973-1986, Oct. 1970.

[5] J.D. Gibson, "Adaptive Prediction in Speech Differential Encoding Systems," Proc. IEEE, vol.68, no.4, pp.488-525, April 1980.

[6] J.D. Markel and A.H. Gray, Jr., "A Linear Prediction Vocoder Simulation Based upon the Autocorrelation Method," IEEE Trans. ASSP, vol.ASSP-22, no.2, pp.124-134, April 1974

[7] M.J. Rose, et al., "Average Magnitude Difference Function Pitch Extractor," IEEE Trans. ASSP, vol.ASSP-22, no.5, pp.353-362, Oct. 1974.

[8] J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, vol.63, no.4, pp.561-580, April 1975.

[9] J. Makhoul, "Stable and Efficient Lattice Methods for Linear Prediction," IEEE Trans. ASSP, vol.ASSP-25, no.5, pp.423-428, Oct. 1977

[10] J.I. Makhoul and L.K. Cosell, "Adaptive Lattice Analysis of Speech," IEEE Trans. ASSP, vol.ASSP-29, no.3, pp.654-659, June 1981.