

# Pulse-Train Code 여기 선형 예측 (PT-CELP) 부호화기에 관한 연구

김흥국, 하남규, 오기은, 김무영, 김상봉  
삼성종합기술원 음성연구실

## A Study on the Pulse-Train Code Excited Linear Prediction Coder : PT-CELP

H. K. Kim, N. K. Ha, K. E. Oh, M. Y. Kim, and S. R. Kim  
Speech Technology Lab. Samsung Advanced Institute of Technology

### 요약

본 논문에서는 4.16kbps의 전송률을 갖는 음성 부호화기 구조에 관하여 기술한다. 제안된 음성 부호화기는 개방 피로 피치 검출기와 이로부터 생성된 pulse train을 코드북으로 갖는 CELP 부호화기 (Pulse-Train Code Excited Linear Prediction : PT-CELP)이다. Pulse-Train codebook은 분석 프레임별로 부호화 및 복호화 양단에서 생성되며 음성의 피치 및 포먼트 정보를 내포하고 있다. 구현된 PT-CELP는 random codebook 방식의 CELP에 비해 적은 크기로 codebook을 만들 수 있으며 음성의 특징을 충분히 반영하므로 합성된 음성의 음질을 향상시킬 수 있다.

### 1 서론

일반적으로, 음성의 피치와 포먼트 값은 단구간의 음성분석 구간 내에서 변화한다. 이 정보는 음성의 주기성 뿐만 아니라 음질을 좌우하는 요소이다. Code excited linear prediction (CELP) 부호화기에서의 pitch filter 혹은 적응 코드북 (adaptive codebook)은 음성의 주기성을 모델링하며 모델링하고 남은 잔차 신호는 잡음 코드북 (random codebook) 중의 최적의 codevector로 표시된다. 그러나 음성의 주기성은 시간에 따라 변하며 주기성이 일정하다 하더라도, 즉 피치가 같은 경우에도 음성의 포먼트 구조가 변하기 때문에 적응 코드북으로 모델링되지 않은 잔차 신호는 음성의 모양을 갖게 된다. 따라서 이를 잡음 코드북으로 처리할 경우 합성음의 고품질을 기대하기 어렵다.

이러한 문제를 해결하기 위한 노력으로, pitch synchronous innovation CELP (PSI-CELP) [1]에서는 예측된 피치 주기에 맞추어 잡음 코드북을 주기신호로 만든 후 모델링에 사용한다. 이러한 경우 음성의 피치 주기성은 잘 표현되는 반면 포먼트 정보를 복구하는 데는 미흡한 점이 있다. 또한 G.723 6.4kbps 부호화기 [2]는 multi-pulse maximum likelihood quantization 방법으로 피치의 장단에 따라 regular pulse 혹은 multi-pulse로 pitch filter의 잔차 신호를 모델링한다. 이 방식은 pulse의 위치 정보를

표현하기 위해 프레임당 80bits를 할당해야한다. 따라서 4kbps 정도의 압축률을 갖는 부호화기를 구현하기 위해서는 pulse 정보의 표현에 할당된 bit를 줄여야만 한다.

본 논문에서는 피치 및 포먼트 정보를 포함하는 코드북을 갖는 CELP 구조를 제안한다. 부호화기의 개방 피로 피치 검출기 (open-loop pitch extractor)에서 얻어진 피치 정보와 linear predictive coefficient (LPC) 스펙트럼으로부터 추출된 포먼트 정보를 이용하여 펄스열을 만든다. 이 펄스열은 분석 프레임마다 재구성되며 복호화부에서도 분석부에서의 동일한 펄스열을 만들어 낼 수 있다. 이러한 구조는 random 코드북을 갖는 CELP에 비해 코드북 크기를 줄일 수 있으며 동시에 합성 음질을 향상시킬 수 있다.

### 2 PT-CELP의 구조

그림 1은 PT-CELP 부호화기의 구성도를 보여 준다. 부호화기는 음성의 선형 예측 분석, 피치 및 포먼트 추출, 펄스열 생성, 펄스열 코드북 생성, 적응 코드북 탐색, 펄스열 코드북 탐색 과정의 순서로 진행된다. 전송 파라메타로는 피치, 유·무성 정보, 적응 코드북 색인 및 이득, 펄스열 코드북 색인 및 이득이며 복호화부에서는 이 정보로부터 포먼트 추출, 펄스열 생성, 펄스열 코드북 생성, 적응 코드북과, 펄스열 코드북으로부터 LPC 필터 합성, postfiltering의 순서로 진행하여 음성을 재생해 낸다.

PT-CELP 부호화기는 8kHz 표본화율의 16bit 선형 pulse coded modulation (PCM) 데이터를 입력으로 하여 25ms의 블록 데이터에 대해 동작한다. 각각의 블록, 즉 한 프레임은 4개의 부프레임으로 나뉜다. 한 프레임의 데이터에 대해 10차 선형 예측 계수를 구한 후 line spectrum pairs (LSP) 계수로 변환하고 이 LSP 계수를 프레임당 24bits로 split vector quantizer (SVQ) [3]를 이용하여 양자화한다. 양자화된 LSP 계수로 부터 부프레임별 LSP 계수를 구한다. 부프레임별 LSP 계수를 LPC 계수로 변환한 후, 음성의 합성 필터  $\frac{1}{A(z)}$ 와 정각 가중 필터  $W(z) = \frac{A(z)}{A(z/\gamma)}$ 를 구한다. 본 연구에서는  $\gamma = 0.8$ 로 하였다.

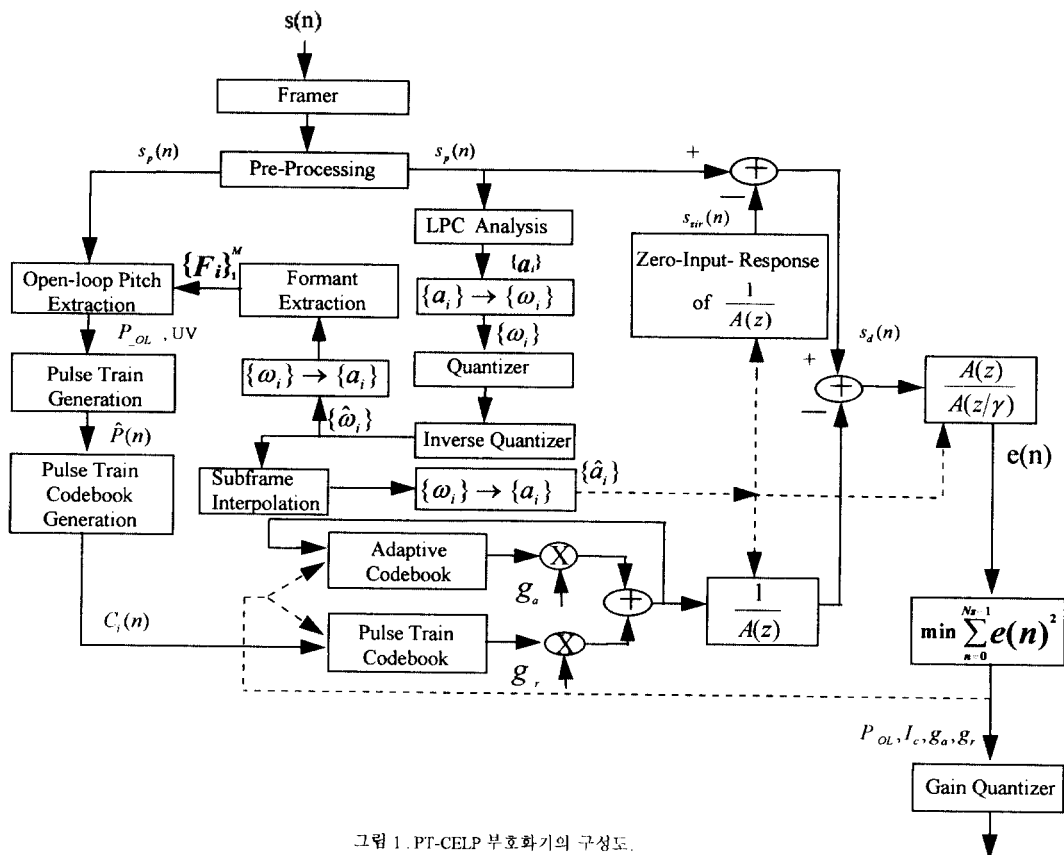


그림 1. PT-CELP 부호화기의 구성도.

적용 코드북의 피치와 펄스열 코드북을 만들기 위해 음성 신호를 cutoff 주파수가 2kHz인 2차 Butterworth 필터에 통과시킨다. 현재 프레임의 200 샘플과 다음 프레임의 50 샘플을 합쳐 250 샘플로 부터 개방 회로 피치 ( $P_{OL}$ )를 Y. Medan *et. al*에 의해 제안된 알고리즘 [4]을 기반으로 하여 구한다. 또한 LPC 스펙트럼으로부터 포먼트 정보 ( $F_i$ )를 구한다. 이렇게 추출한 개방회로 피치와 포먼트는 다음 장에서 설명할 펄스열 코드북 발생에 이용된다.

적용 코드북 탐색을 위한 탐색 구간은 개방 회로 피치 값 근처에만 진행된다. 또한 개방회로 피치의 오차를 보상하기 위해 개방 회로 피치의 정수배가 되는 피치 근처에서도 적용 코드북을 탐색하여 원음과의 차가 청각 가장 민감할던 오차의 자승 값이 최소가 되는 lag ( $P_{CL}$ )를 찾고 이를 폐쇄 회로 피치라 한다. 즉, 개방회로 피치 ( $P_{OL}$ )에 대한 적용 코드북 탐색 구간,  $I_c$  는

$$I = I_1 \cap (I_2 \cup I_3),$$

$$I_1 = \{20, 21, \dots, 146\},$$

$$I_2 = \{P_{OL} - 4, \dots, P_{OL}, \dots, P_{OL} + 3\},$$

$$I_3 = \begin{cases} \{2 \cdot (P_{OL} - 4), \dots, 2 \cdot (P_{OL} + 3)\}, & \text{if } P_{OL} < 72 \\ \{\frac{E_{OL}}{2} - 4, \dots, \frac{E_{OL}}{2} + 3\}, & \text{otherwise,} \end{cases}$$

이 되고. 전송되는 파라미터는  $P_{OL}$ 을 7 bits, 적용 코드북 탐색 결과 폐쇄 회로 피치가  $I_2$  혹은  $I_3$  범위 내에 있는 가를 나타내는 lag position bit ( $L_P$ )를 1 bit 할당한다. 개방 회로 피치와 폐쇄 회로 피치의 차를 3 bits로 표현한다. 즉,

$$L_P = \begin{cases} 0, & \text{if } P_{CL} \in I_2 \\ 1, & \text{if } P_{CL} \in I_3 \end{cases}$$

그리고

$$\Delta P = \begin{cases} P_{CL} - (P_{OL} - 4) & P_{CL} \in I_2 \\ P_{CL} - (2 \cdot (P_{OL} - 4)) & P_{CL} \in I_3 \text{ and } P_{OL} < 72 \\ P_{CL} - (\frac{E_{OL}}{2} - 4) & P_{CL} \in I_3 \text{ and } P_{OL} \geq 72 \end{cases}$$

로 표현된다. 피치 이득,  $g_a$ , 은 4 bits 선형양자화기로 양자화된다. 적용 코드북 탐색으로 찾았던 폐쇄 회로 피치와 피치 이득을 가지고 LPC 합성 필터를 통과시켜 원 음성 신호로부터의 차를 구하고 이를 펄스열 코드북으로 모델링한다.

펄스열 코드북은 부프레임마다 32개의 codevector로 구성된다. 최적의 codevector를 찾는 방법은 적응 코드북 탐색과정과 동일하며 codevector 색인  $i_c$ 와 이득  $g_r$ 를 각각 5 bits씩 할당한다.

### 3 Pulse-Train Codebook의 발생

그림 2는 펄스열 코드북을 생성하기 위한 펄스열 발생 과정을 보인다. 우선 피치와  $M$ 개의 포먼트 정보로부터 피치의 역수인 기본 주파수 ( $F_0$ )와 포먼트간의 비 ( $R_i, i = 1, \dots, M$ )를 구한다. 포먼트 정보를 나타낼 시간축 상의 펄스의 위치와 크기를  $R_i$ 를 이용하여 구한다. 구하여진 펄스열에 유·무성 정보로부터 잡음을 스펙트럼 영역에서 부가한 후, 이를 시간영역의 펄스열로 변환시킨다. 그림 3은 입력음성에 대한 펄스열의 예를 보인다. 펄스열이 입력 음성과 정확한 동기화를 이루지는 않았지만 음성의 피치 주기와 포먼트 정보를 반영함을 볼 수 있다.

200 샘플의 펄스열로부터 다음 식과 같이 펄스열 코드북을 생성한다.

$$C_i(n) = \hat{p}(n + 4i), \text{ for } i = 0, \dots, 31, n = 0, \dots, 49$$

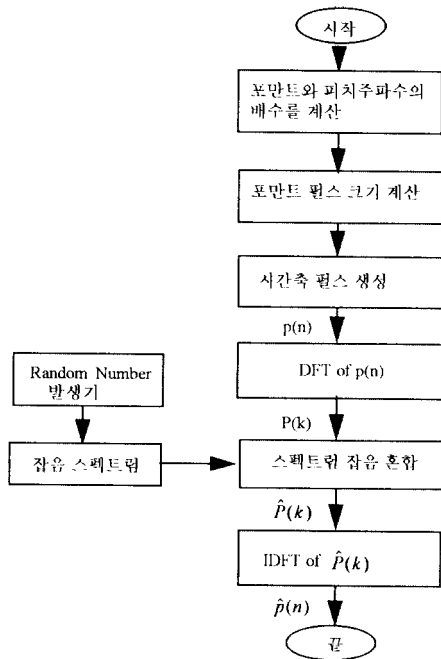


그림 2. 펄스열 생성 과정 흐름도.

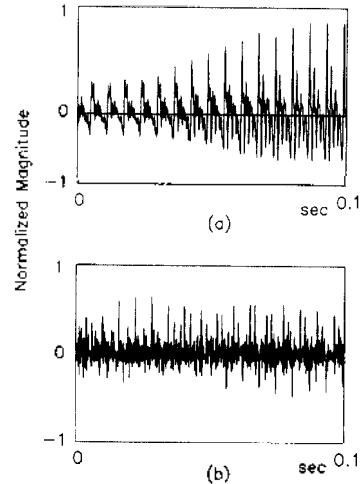


그림 3: 펄스열의 예시도. (a) 입력 음성. (b) 펄스열.

### 4 PT-CELP의 Bit 할당과 성능평가

표 1은 제안된 PT-CELP 부호화의 프레임당 bit 할당을 보여 준다. 프레임 길이가 25ms이므로 전송률은 4.16kbps가 된다. PT-CELP 부호화는 한 프레임을 분석하기 위해 25ms에 해당하는 음성 데이터가 필요하며, 개방 회로 피치 추출시 성능 향상을 위해 6.125ms의 다음 프레임 데이터가 부가적으로 필요하므로 전체 31.125ms의 algorithmic delay가 발생한다.

표 1: PT-CELP 부호화의 프레임당 할당된 BIT수.

(프레임 길이 = 25ms, 부프레임 길이 = 6.125ms)

특성	파라메타	Notation	Bits/Frame
스펙트럼 포락정보	10차 LSP	$\omega_i$	24
피치 (적용 코드북)	개방회로 피치	$P_{OL}$	7
	유·무성 정보	$UV$	1
	Lag Position	$L_P$	$1 \times 4$
	$\Delta$ 피치	$\Delta P$	$3 \times 4$
펄스열 코드북	색인	$I_c$	$5 \times 4$
	이득	$g_r$	$5 \times 4$
합계			104

PT-CELP 부호화기의 성능을 평가하기 위해서 한국인 남녀 각 3인이 서로 다른 5개의 문장을 실험실 환경하에서 발성하였다. 표 2는 4.8kbps DoD CELP [5]와 PT-CELP에 대해 각 화자별 평균 segmental signal-to-noise ratio (segSNR)를 보여 준다. 표에서 보는 바와 같이 PT-CELP는 남성화자에서, DoD CELP는 여성화자에서 각각 segSNR이 0.4dB가 높음을 알 수 있다. 이는 여성화자의 경우, 기본 주파수가 높기때문에 첫번째 포먼트의 주파수가 기본주파수에 근접하게 되어서 펄스열에 첫번째 포먼트 정보가 나타나지 않기 때문이다.

표 2: DoD CELP와 PT-CELP의 SEGMENTAL SNR (dB)의 비교.

화자 (성)	발음길이 (sec)	DoD CELP (dB)	PT-CELP (dB)
HKK (남)	16.83	7.7749	8.1175
KSK (남)	17.07	6.8289	7.3557
MYK (남)	14.04	7.6212	7.8332
HJL (여)	21.45	10.0203	9.7800
DNM (여)	15.30	8.4035	7.3866
JYP (여)	16.59	8.1041	8.0948
평균 (남)	15.98	7.4083	7.7688
평균 (여)	17.78	8.8426	8.4205

## 5 결론

본 논문에서는 펄스열을 코드북으로 하는 새로운 형태의 CELP 부호화기를 제안하였다. 펄스열은 음성의 주기성분과 포먼트 정보뿐 이용하여 프레임별로 생성되며 기존의 잡음 코드북에 비해 적은 codevector로 고품질의 음성을 합성해 낼 수 있었다. 이를 통해 구현된 4.16kbps PT-CELP 부호화기는 31.125ms의 algorithmic delay를 갖으며 4.8kbps DoD CELP에 비해 코드북 크기가 6.25%로 줄었으며 segmental SNR 관점에서 PT-CELP 부호화기의 성능은 DoD CELP의 성능과 유사함을 알 수 있었다.

## 참고 문헌

[1] T. Moriya, S. Miki, K. Mano, and H. Ohmura, "Pitch synchronous innovation CELP (PSI-CELP)," *IEICE Trans. Fundamentals*, Vol. E76-A, No. 7, pp. 1177-1180, July 1993.

[2] ITU Draft Recommendation G.723, "Dual rate speech coder for multimedia telecommunication transmitting at 5.3 & 6.4 kbit/s," Mar. 1995.

[3] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. on Speech and Audio Processing*, Vol. 1, No. 1, pp. 3-14, Jan. 1993.

[4] Y. Medan, E. Yair, and D. Chanzan, "Super resolution pitch determination of speech signal," *IEEE Trans. on Signal Processing*, Vol. 39, No. 1, pp. 40-48, Jan. 1991.

[5] J. P. Campbell, T. E. Tremain and V. C. Welch, "The DoD 4.8kbps standard ( proposed federal standard 1016 )," in *Advances in speech coding*. B. Atal et. al (Eds.), Kluwer Academic Publishers, pp. 121-133, 1991.