

## 다중 펄스와 다중 대역 이차 장구간 예측을 이용한 CELP 음성 부호화기

### A CELP Speech Coder Using Secondary Long Term Prediction with Multi-Band Pass Filtered Multi-Pulses

서 정 태\*, 최 용 수\*\*, 강 홍 구\*\*\*, 윤 대 회\*\*

(Jeong-Tae Seo\*, Yong-Soo Choi\*\*, Hong-Goo Kang\*\*\*, Dae-Hee Youn\*\*)

※이 논문은 1996년도 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

#### 요 약

본 논문에서는 낮은 비트율 CELP 음성 부호화기의 장구간 예측기의 성능 향상 방법을 제안한다. 비트율을 낮추기 위해서는 분석 구간의 길이가 길어져야하며 이에 따라 장구간 예측기의 성능이 저하되어 장구간 예측 후에도 준 주기성 성분이 상당량 존재하므로 백색 잡음으로 구성된 통계 코드북만으로는 이를 모델링하기 어려워진다. 제안 방법에서는 다중 대역 필터와 다중 펄스열을 이용하여 한 번 더 필터링(이차 장구간 예측)함으로써 장구간 예측 후의 신호가 통계 코드북에 적합한 백색 잡음 형태로 되도록 모델링한다. 제안된 방법의 성능을 평가하기 위해 4.8 kbps 비트율로 양자화한 후, 기존에 제안된 같은 전송률의 MBCELP와 DoD-CELP와 비교하였다. 실험 결과 제안된 방법이 기존 부호화기들에 비해 주/객관적인 음질에서 우수한 성능을 보여준다.

#### ABSTRACT

This paper describes a good method to enhance the long-term predictor which is one of critical problem in low-rate CELP coder. There exists quasi-periodic component after long-term prediction as longer the length of analysis frame, so it is hard to model this signal with stochastic codebook which consists of white gaussian noise. In this paper, the error signal after long-term prediction is band-pass filtered and it is modelled by multi-pulse codebook. To demonstrate the performance of the proposed method, the proposed speech coder is quantized with 4.8 kbps and it is compared with MBCELP(multi-band CELP) and DoD-CELP. The proposed method shows better quality than reference coders in terms of both subjective and objective quality.

#### I. 서 론

최근 들어 시/공간적 제한을 초월하여 서로의 의사를 전달할 수 있는 이동 통신이 통신 산업의 궁극적인 목표로 인식되고 있다. 이동 통신을 통해 주고받는 인간의 의사 전달 수단 중 가장 중요한 정보는 음성이다. 그러나 음성을 디지털로 표현하기 위해서는 많은 비트가 필요하므로 한정된 채널용량을 효율적으로 사용하기 위해 낮은 전송율에서도 성능이 우수한 음성 부호화 알고리즘의 연구는 필수적이다.

CELP(Code Exited Linear Prediction)[1]는 4.8 kbps 내

외의 전송율에서 양호한 음질을 얻을 수 있으므로 현재 이동 통신 시스템 음성 부호화 방식으로 가장 주목받고 있는 방법이다[2]. CELP의 중요한 특징은 여기 신호를 모델링하기 위한 장구간 예측기와 통계 코드북 개념이라 할 수 있다. 장구간 예측기는 유성음의 여기 신호에 존재하는 피치 성분을 모델링하기 위한 것으로서 일반적으로 과거의 여기 신호로부터 주기에 따라 변화되는 적응 코드북이 이용되어 왔다. 그러나 전송율을 낮추기 위해서는 분석 구간의 길이가 길어져야하므로[6] 코드북의 벡터 차원이 증가하게 되고 이에 따라 장구간 예측기의 성능 또한 급격히 저하된다. 이는 장구간 예측 후의 잔차 신호에 준 주기성 성분들이 많이 남아있게 되어 백색 잡음만으로 이루어진 통계 코드북 만으로는 이를 효과적으로 모델링하기 어렵기 때문에 발생한다. 이와 같은 문제점을 극복하기 위해, 유성음으로 판정되는 분석 구간에서는 펄

\*충주산업대학교 제어계측공학과

\*\*연세대학교 전자공학과

\*\*\*AT&T Labs-Research 연구원

접수일자: 1997년 7월 30일

스형 코드북을, 무성음으로 판정되는 구간에서는 통계 코드북을 사용하는 방법[4], 그리고 피치 지연 시간의 해상도를 정수 단위에서 실수 단위로 증가시키는 방법[5] 등이 제안되었다. 이와 같은 방법들의 공통점은 주기적 피치 성분을 효율적으로 모델링하여 음질을 개선시키는 것이다.

원 음성의 LPC 여기 신호는 잡음성 신호 성분과 주기적인 성분이 주파수 영역에 따라 다르게 분포한다[6]. 이와 같은 관찰을 근거로 하여, 조화 부호화(harmonic coding)[7]와 CELP 방식을 혼합하여 여기 신호를 잘 표현할 수 있도록 하는 MBCELP(Multi-Band CELP)[8]가 제안되었다. 이 방법은 장구간 예측 과정에서 구한 지연 간격을 주기로하는 임펄스 열(impulse train)을 분할된 몇 개의 대역 필터에 통과시켜 잔존하는 주기성을 제거하는 방법이다. 그러나, MBCELP 방식으로 낮은 전송율을 갖도록 구현하기 위해서는 제한된 정보량으로 인해 분할 대역의 수를 제한해야 한다. 따라서, MBCELP에서 처리 동일한 크기를 갖도록 펄스를 구성하는 방법은 평탄하지 않은 LPC 여기 신호의 스펙트럼을 잘 표현하지 못한다[8]. 이것을 시간 영역에서 해석하면, 장구간 예측 후의 잔차 신호에는 주기적인 피크 성분을 주위로 여러 개의 펄스들이 군집되어 있으므로 균일한 임펄스 열로 구성된 여기 신호로는 주기적 성분을 효과적으로 제거하기 어렵다는 것을 뜻한다.

본 논문에서는 군집되어 있는 비균일 크기의 임펄스 열을 효과적으로 모델링하기 위한 방법을 제안한다. 먼저, 잔차 신호의 대역을 분할하여 각 대역 특성에 잘 부합하는 다중 펄스를 찾는다. 다중 펄스는 구성 형태에 따라 단조 감소 크기를 갖는 방법과 몇 개의 랜덤 위치와 크기를 갖는 방법으로 나누어 실험한다. 부호화기 성능을 평가하기 위한 비교 기준 부호화기로 DoD-CELP[9]와 MBCELP[6]를 사용한다. 제안한 방법들은 유성음 구간에 존재하는 펄스들을 잘 모델링함에 따라 객관적인 평가기준에서 뿐만 아니라 주관적인 청취 평가에서도 우수한 성능을 얻을 수 있었다.

본 논문의 전체적인 구성은 다음과 같다. II 장에서는 다중 펄스 다중 대역 이차 장구간 예측을 이용한 CELP 부호화를 제안하고, III 장에서는 모의 실험 및 주/객관적 음질에 대해서 언급한다. 마지막으로 제 IV 장에서 결론을 맺는다.

II. 다중 펄스 다대역 여기신호 모델링 (MPMBCELP)

CELP와 같은 분석-합성(analysis-by-synthesis) 부호화기의 음질은 여기 신호 모델링 방법에 따라 크게 좌우된다. 특히 피치에 해당되는 주기적 성분을 표현하기 위한 장구간 예측기는 합성 음질에 큰 영향을 미치는 것으로 알려져 있다[6]. 또한, 전송율을 낮추기 위해 분석 부프레임의 길이를 늘리는 경우에는 한 프레임내에 두 개 이상

의 피치가 존재하게 되는 반도수가 잦아짐에 따라 장구간 예측 성능은 더욱 떨어진다. 이를 해결하기 위해 장구간 예측 성능을 향상시키기 위한 다양한 방법들이 시도되어 왔으며[5][6], 최근에 제안된 MBCELP에서는 장구간 예측 후의 신호를 여러 주파수 대역으로 분할한 후 다시 한 번 장구간 예측(이차 장구간 예측)을 함으로써 성능 향상을 꾀하였다.

그러나, 낮은 전송율에서 동작하는 부호화기를 구현하기 위해서는 대역수를 제한하여야만 한다. 이러한 경우에는 피치 주기를 갖는 단순한 임펄스열만으로는 피치 주위에 남아있는 펄스 성분을 효율적으로 제거하지 못하게 되므로 성능 향상에 큰 도움을 주지 못한다. 본 논문에서는 MBCELP의 이차 장구간 예측에 사용한 단순한 임펄스 열을 다중 펄스 구조를 갖는 다대역 여기 신호로 확장한 MPMBCELP(multi-pulse multi-band CELP)를 제안한다. 제안한 음성 부호화기는 그림 1과 같다.

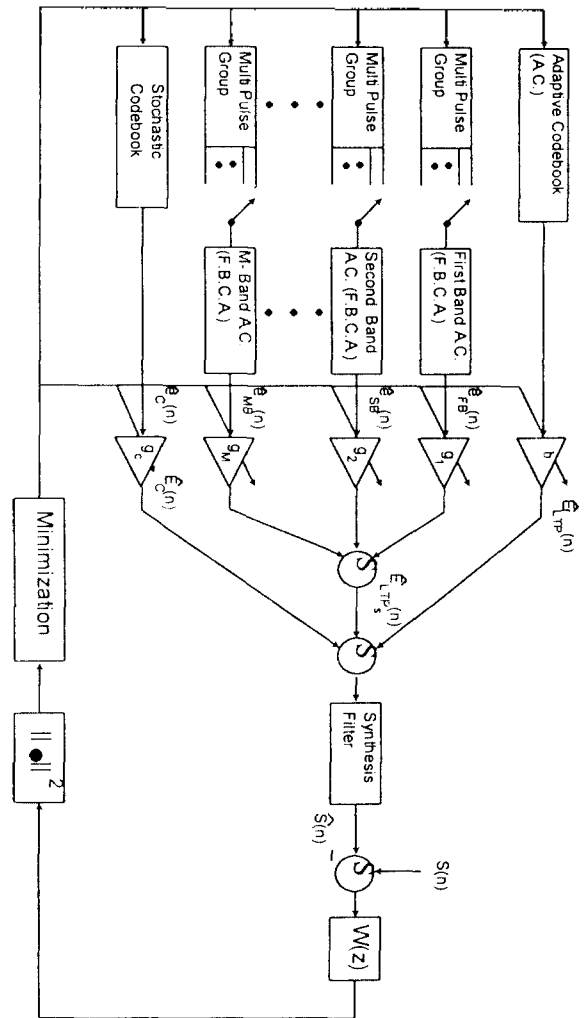


그림 1. 다중 펄스 다대역 CELP (MPMBCELP) 부호화기의 블록도  
Fig. 1 MPMBCELP speech encoder

그림 2에 보인 장구간 예측 과정 후 잔차 신호의 주파수 영역 특성을 보면 고주파 영역에서 보다 저주파 영역에서 주기적인 성분이 강하게 나타남을 알 수 있다. 따라서 각 대역에 따라 다른 특성을 갖는 여기 신호를 사용하면 합성음의 주관적인 청각 특성을 높일 수 있다[6]. 다중 펄스 다대역 부호화 방법에 사용된 이차 장구간 예측 과정은 일차 장구간 예측에서 얻어진 피치 정보를 근간으로 하는 펄스열을 구성함으로써 이루어진다.

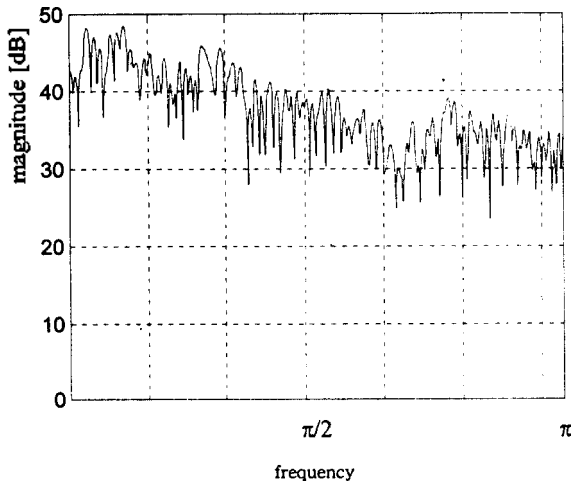


그림 2. 일차 장구간 예측 후 잔차 신호의 스펙트럼  
Fig. 2 spectrum of residual signal after 1<sup>st</sup> LTP

대역 분할을 위한 필터는 주파수 대역  $B_i$ 에 대응되는 *sinc* 함수로 표현된다.

$$\text{rect}_i(n) = i \sin c(in/M) - (i-1) \sin c((i-1)n/M), \quad (1)$$

$$1 \leq i \leq M, 0 \leq n < N.$$

여기서  $M$ 은 대역 수이며,  $N$ 은 필터의 길이이다. *sinc* 함수의 길이는 성능과 계산량을 고려하여 80 정도로 제한할 수 있다. 분할 대역의 개수는 전송율에 따라 결정되는데 4.8 kbps 전송을 구현을 위해서는 두 개로 고정하는 것이 타당하다.

다중 펄스를 구성하는 최적 방법은 MPLPC[10]에서 제안된 방법과 같이 분석 부프레임마다 가중 오차를 최소화하는 펄스의 위치와 크기를 찾는 것이다. 그러나, 이 방법은 많은 부가 비트와 계산량을 필요로 하므로 비트율을 4.8 kbps로 맞추기 어려워진다. 따라서, 본 논문에서는 비트와 계산량 면에서 효율적인 두 가지 다중 펄스 구성 방법을 제안한다.

### 2.1 다중 펄스 구성 방법

#### A. 랜덤 크기와 위치를 갖는 다중 펄스

일차 장구간 예측 후 잔차 신호의 형태는 피크 주위에

상대적으로 큰 값이 집중되어 있는 형태를 갖는다. 또한, 피치 사이의 성분들은 통계적 코드북을 사용하면 모델링이 가능하므로 이차 장구간 예측의 목표는 피크 주위에 집중되어 있는 값만을 제거하는 것이다. 이차 장구간 예측에 사용할 펄스 개수는 목표 비트율에 따라 선택적으로 조절할 수 있으며 이때 성능 또한 펄스 개수에 따라 변하게 된다. 그림 3은 펄스 개수에 따른 성능 변화를 구간 신호 대 잡음비(segSNR; segmental signal to noise ratio)를 사용하여 구한 결과이다.

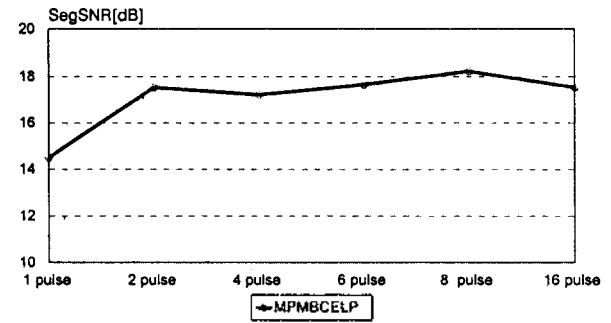


그림 3. 펄스 개수에 따른 음성신호의 성능 변화  
Fig. 3 performance comparison according to pulse number

2 개의 펄스에 대해서도 만족할 만한 성능을 보임을 알 수 있다. 또한, 첫번째 펄스의 크기는 1로 고정하고, 두번째 펄스의 크기는 -1과 -0.5로 제한함으로써 가중 합성 필터링의 컨벌루션 과정에서 발생하는 곱셈 연산을 간단한 덧셈과 뺄셈만으로 표현하도록 하여 고속연산 수행이 가능하도록 한다. 이와 같은 제한 조건하에, 4 비트를 할당하여 다중 펄스군을 구성하였으며, 각각의 펄스에 대해 장구간 예측이 끝난 잔차 신호와 합성 필터를 통과한 다대역 신호 사이의 가중 오차를 최소화 되도록 하는 다중 펄스를 선택한다.

#### B. 단조 감소 크기를 갖는 다중 펄스

앞 절에서 언급한 다중 펄스 모델링이 비교적 우수한 성능을 보이지만 펄스의 수를 확장하므로써 얻게되는 음질 향상에 비해서 비트 및 계산량의 증가가 큰 단점이 있다. 이러한 문제점을 극복하기 위해 본 절에서는 간단한 선형 함수를 이용하여 잔차 신호의 포락선 형태를 근사화할 수 있는 다중 펄스 구성 방법을 제안한다. 그림 4에서 관찰할 수 있듯이, 일반적으로 장구간 예측 후의 잔차 신호는 주기적 성분에서 다음 주기적 성분까지 단조 감소하는 포락선 형태를 갖는다.

이러한 형태에 가장 유사한 특성을 갖는 함수는 지수 함수의 형태로 주어지나 실시간 구현이 어려울 뿐만 아니라 계산량이 증가하게 된다. 따라서, 포락선의 형태를 비교적 잘 모델링할 수 있고 구현이 쉬운 일차 함수를 이

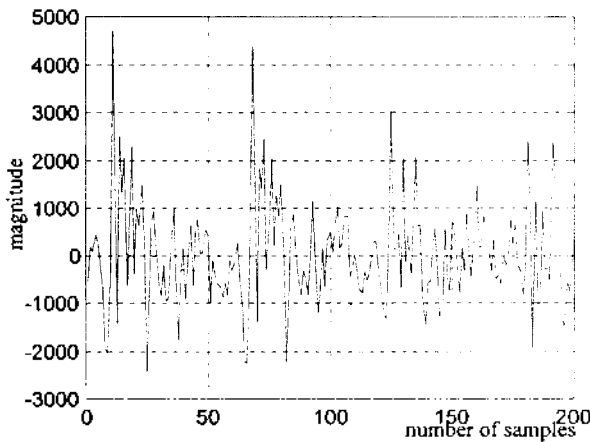


그림 4. 장구간 예측후의 잔차신호  
Fig. 4 residual signal of LTP

용하여 다중 펄스를 구성한다. 일차 장구간 예측에서 구한 피치 주기의 위치에 놓이게 되는 주 펄스의 크기를  $A$ 로 놓았을 때, 피치 구간의 끝이 0의 값을 갖도록 하는 함수는 다음 식과 같이 나타낼 수 있다.

$$A_p(n) = \max \{A(1 - n/\min(P, L)), 0\} \quad (2)$$

여기서,  $P$ 는 피치 값,  $A_p(n)$ 는 펄스 크기,  $A$  다중 펄스 크기의 시작 값,  $L$ 은 다중 펄스 범위이다.  $A_p(n)$ 의 기울기를 결정하는 주요 변수인  $A$ 가 증가할수록 피치 펄스를 제외한 다중 펄스의 크기는 상대적으로 작아지게 되므로 기존의 MBCELP와 유사한 결과를 얻을 수 있는 반면  $A$ 가 감소할수록 다중 펄스의 크기는 피치 펄스와 거의 유사하게 되므로 합성 음질이 떨어지게 된다. 실험을 통해  $A$ 가 1일 경우가 최적의 성능을 얻을 수 있었다.

MPMBCELP 부호화기의 이차 장구간 예측과정은 일차 예측과는 달리 다음과 같이 두 단계로 나뉘어서 수행된다.

- 단계-1: 원신호로부터 개방 루프(open loop) 방법으로 피치를 추정한다.  
단계-2: 추정된 피치를 중심으로 주위 몇 샘플만 비교하여 최적 위치를 검색한다.

MPMBCELP 모델에서 장구간 예측 후 수행되는 다대역 여기 신호를 이용한 준주기성 제거 과정에서 만들어지는 여기 신호는 장구간 예측 과정에서 구한 최적 위치로 결정되는 피치 주기로 구성된다. 그러나, 기존의 CELP 방식과 같이 저장된 전 프레임 여기 신호와의 상관 관계를 이용하여 구한 최적 위치를 피치로 사용하게 되면, 원 음성 신호의 피치에 비해 주기가 정수배로 확장되거나 (pitch doubling) 줄어든 피치(pitch halving)를 얻을 수 있다[11]. 이와 같이 변환된 피치로 다중 펄스 다대역 여기

신호를 구성하게 되면, 분석 부프레임의 길이가 피치에 비해 큰 경우에는 합성 음질에 나쁜 영향을 끼친다[12]. 그러므로, MPMBCELP 부호화기에서는 먼저 원 음성 신호로부터 피치를 추정하여 피치 더블링이나 피치 반감이 발생되지 않도록 한 다음, 추정된 피치 주위의 몇 샘플만을 검색하여 최적 위치를 찾도록 하였다. 추정 피치 주위의 몇 샘플을 비교함으로써 장구간 예측 과정에서 소요되는 계산량을 줄일 수 있음은 물론 부호화 시간상의 잇점도 얻을 수 있다. 초기 피치 추정이 원 음성 신호의 실제 피치와 다소 차이가 있다하더라도 탐색 과정에서 보정될 수 있기 때문에 초기 피치 추정 과정의 정확성은 크게 중요하지는 않다.

## 2.2 다중 펄스 다대역 여기신호 모델변수의 추정

장구간 예측 과정을 통하여 주기적 특성이 제거된 다중 펄스 다대역 여기 신호들은 장구간 예측 과정과 유사하게 계산된다. 앞 절에서 설명한 것과 같이 구성된 다중 펄스 신호를 청각 특성이 가중된 합성 필터에 통과 시킨 후 잔차 신호의 피크 성분과 가장 잘 대응되는 최적 위치를 찾아 두번째 여기 신호로 이용한다. 다중 대역 통과 필터의 임펄스 응답 함수는 미리 계산한 후 저장하면, 부호화 과정에 소요되는 계산량을 줄일 수 있다.

분할된 각 대역의 다중 펄스 다대역 여기 신호 모델 변수들은 다음과 같이 단계적으로 변수를 추정한다. 가장 저주파수 대역에 대해 다중 펄스 여기 신호를 구성한 다음, 이 신호를 이용하여 잔차 신호로부터 기여 성분을 제거한다.

두번째 대역 여기 신호에 사용되는 목적 신호는 첫번째 대역에 의한 기여 성분이 제거된 신호를 사용한다. 이와 같은 반복과정을 분할된 대역 전체에 대해 수행함으로써 각 대역특성에 적합한 다중 펄스 다대역 여기 신호를 구성한다. 그 추정 방법은 다음과 같다[6].

적용 코드북에서 다대역 여기 신호 벡터  $V_j$ 의 최적 위치는 위치  $j$ 에 대해서 다음 식을 최대로 함으로써 계산할 수 있다.

$$E_j = \frac{(XY_j^T)^2}{Y_j Y_j^T}, \quad j = P_{\max} \rightarrow P_{\min} \quad (3)$$

여기서  $X$ 는 필터  $1/A(z/\alpha)$ 으로 기준 여기 신호를 필터링함으로써 얻어지는  $1 \times L$ 의 크기를 갖는 기준 신호 벡터이며,  $Y_j = V_j H$ 로 나타낸다. 여기서,  $H$ 는 필터  $1/A(z/\alpha) = W(z)/A(z)$ 의 임펄스 응답 행렬에 의해 주어진다. 그리고  $V_j$ 는 다음과 같다.

$$V_j = [v_j(0) \ v_j(1) \ \dots \ v_j(L-1)] \\ v_{j-1}(n) = v_j(n-1), \quad 1 \leq n < L \quad (4)$$

신호 벡터  $V_j$ 와 식 (3)의 분자는 계산량 감소를 위해 다음과 같은 반복 과정을 사용해 구할 수 있다.

$$\begin{aligned}
 Y_{j-1} &= [y_{j-1}(0) \ y_{j-1}(1) \ \dots \ y_{j-1}(L-1)] \\
 &= Y_{j-1} H \\
 &= v_{j-1}(0) [h(0) \ h(1) \ \dots \ h(L-1)] \\
 &\quad + [0 \ y_j(0) \ y_j(1) \ \dots \ y_j(L-2)]
 \end{aligned}
 \tag{5}$$

여기서  $h(n)$ 은  $n \geq L_h$  ( $L_h \ll L$  혹은  $L_h = 20 \sim 30$ )에 대해서 무시할 수 있으므로 (5)식은 다음과 같이 표현될 수 있다.

$$\begin{aligned}
 Y_{j-1} &= v_{j-1}(0) [h(0) \ h(1) \ \dots \ h(L_h) \ 0 \ \dots \ 0] \\
 &\quad + [0 \ y_j(0) \ y_j(1) \ \dots \ y_j(L-2)]
 \end{aligned}
 \tag{6}$$

또한, (3)식의 분모는 다음과 같이 표현할 수 있다.

$$\begin{aligned}
 Y_{j-1} Y_{j-1}^T &= Y_j Y_j^T - y_j(L-1) \\
 &\quad + v_{j-1}(0) \{ v_{j-1}(0) + 2 \sum_{i=1}^{L_h} h(i) y_j(i-1) \}
 \end{aligned}
 \tag{7}$$

여기에서  $E_h$ 는 임펄스 응답의 에너지로서 아래와 같이 주어진다.

$$E_h = \sum_{i=1}^{L_h} h^2(i)
 \tag{8}$$

최적 위치  $j$ 에 대해서 이득  $g_j$ 는 다음과 같은 식으로 얻어질 수 있다.

$$g_j = (XY_j^T)/(Y_j Y_j^T)
 \tag{9}$$

### 2.3 통계 코드북

이차 장구간 예측 과정이 끝난 신호는 인접 샘플간의 상관 관계 및 주기적인 피크 성분이 효과적으로 제거 되어 백색 잡음에 가까운 특성을 지닌다. 본 논문에서는 이와같은 신호의 모델링을 위해 CELP계열 부호화기에서 이용하는 N차원의 가우시안 랜덤 벡터를 사용하는 대신, +1 혹은 -1로 만들어지는 이진 코드로 별도 코드북 구성이 필요하지 않는 이진 코드북을 사용한다. 이진 펄스로 모델링하므로써 계산량을 현저히 줄일 수 있을 뿐만 아니라, 이진 펄스로 구성된 코드북 탐색 후 실제로 부호화하는 과정에서 코드북을 저장할 필요가 없으므로 메모리를 줄일 수 있고, 이진 코드북은 전송과정의 비트 에러에 대해서 가우시안 랜덤 코드북보다 원 코드북과의 오차가 적으므로 채널 에러에 대해서도 강인하다[13].

원래 BCELP(Binary CELP)[13][14]에서 이진 코드북 구성 방법은 이진 코드 벡터를 +1 혹은 -1의 크기를 갖는 펄스들을 RPE(Regular Pulse Excitation)[10]에서 사용한 것과 같이 일정 간격으로 배치한 다음, 가능한 모든 이진 펄스의 조합 전체를 탐색하므로써 최적 여기 벡터를 구한다. 본 논문에서는 이와같은 BCELP의 성능을 향상하기 위해 위상 조정으로 이진 코드북의 구조를 단순화시킨 방법(SBCELP; simplified BCELP)을 이용하였다[14].

BCELP에서와는 달리 SBCELP에서는 성능을 향상시키기 위해서 RPE에서처럼 위상 조정(phase adaptation)을 도입하였으며 원 알고리즘에서 사용한 4 샘플 위상 간격의 반인 2 샘플만 천이 가능하도록 위상을 조정하였다. 4가지 위상을 갖는 경우와 2가지 위상을 갖는 경우를 비교해 보면 0.5 dB 정도의 segSNR차이를 보이지만, 2가지 위상만 갖는 경우에는 계산량 측면에서 효과적이다.

## III. 실험 및 성능평가

### 3.1 양자화

먼저, 선형 예측 계수(LPC)는 비교적 안정성 판단이 쉽고, 양자화 과정이 간단한 LSP(Line Spectrum Pair)[15]를 사용하였다. 본 논문에서는 스칼라 양자화기가 갖는 연산량 감소 효과와 동일 성능에서 필요한 비트가 상대적으로 적은 벡터 양자화기의 장점을 조합하여 34 비트 내에서 우수한 성질을 유지하는 스칼라-벡터 분리 양자화기로 LSP 계수를 양자화 하였다.

일차 및 이차 장구간 예측의 탐색구간을 20~146샘플로 한정하였으므로, 별도의 양자화 과정 없이 7비트로 장구간 예측 피치 값을 표현하여 전송할 수 있으며, 이득에 대해서는 5비트로 표현 가능한 32레벨 양자화기를 사용하였다. 백색 잡음 통계 코드북에 사용할 이득의 양자화 방법도 합성 음질에 중요한 영향을 미친다. 본 논문에서는 적은 비트로도 일정 성능을 유지하기 위해 음성 신호의 동적 변화에 가변하여 양자화 테이블을 적응적으로 변화시키는 양자화 방법을 사용하였다.

표 1. 4.8 kbps MPMBCELP 비트 할당

Table 1. 4.8 kbps MPMBCELP bit allocation

Parameters	Duration(ms)	Bits	Bits/sec
LPC(LSP)	25	34	1360
Pitch1 + Pitch2	2×12.5	7 + 3	400
$\beta$ (LTP gain)	12.5	4	320
Multipulse codebook	12.5	4	320
Pos.1(MB excit.1)	12.5	7	560
Pos.2(MB excit.2)	12.5	0	0
$g_1$ (MB excit. gain1)	12.5	4	320
$g_2$ (MB excit. gain2)	12.5	4	320
$l_i$ (codebook index)	12.5	10	800
$g_i$ (codebook gain)	12.5	5	400
Total	25	120	4800

### 3.2 실험 및 평가

모의 실험을 위해 준비된 음성 데이터는 주변 잡음이 없는 환경에서 남녀 화자 각 2명이 발음한 문장을 8KHz 샘플링 하여 사용하였다. 전체 음성 데이터의 길이는 약

30 초 정도 구간을 갖는다.

본 논문에서 제안한 이차 장구간 예측과정을 사용한 음성 부호화기의 성능 비교를 위해, 현재 4.8 kbps 부호화기의 표준 방식으로 평가받고 있는 DoD-CELP[9]와 기존의 MBCELP[6] 시스템을 사용하였다. 본 논문의 주요 목적은 장구간 예측 과정을 정확히 수행함으로써 잔차 신호의 통계적 특성을 백색 잡음과 유사한 형태로 만들기 위한 것이다. 따라서 제안된 방법의 이차 장구간 예측 과정 성능을 비교하는 것이 평가 기준의 한 방법이 될 것이다. 그림 5는 기존의 태역 분할 방법[6]과 본 논문에서 제안한 방법의 이차 장구간 예측의 성능을 비교하기 위한 한 구간의 정규화된 상관함수의 결과이다. 일차 및 이차 장구간 예측 후 남은 잔차 신호는 음성 신호의 주기적 특성이 제거된 특성을 갖는다. 그러나, 장구간 예측기의 비효율성으로 상당한 양의 주기성분이 존재하게 된다. 그림 5의 (a)에 표시된 MBCELP 방법[6]의 결과에서 알 수 있는 것과 같이 장구간 예측후에 한번 더 이차 장구간 예측을 수행하였음에도 불구하고 여전히 피치 주기에 상당히 큰 피크값을 갖는 것을 알 수 있다. 이와 같은 피크 성

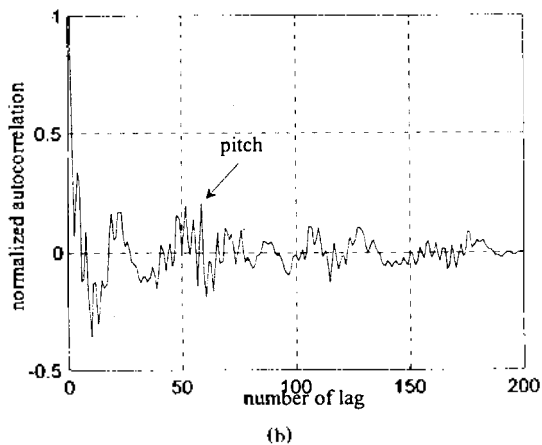
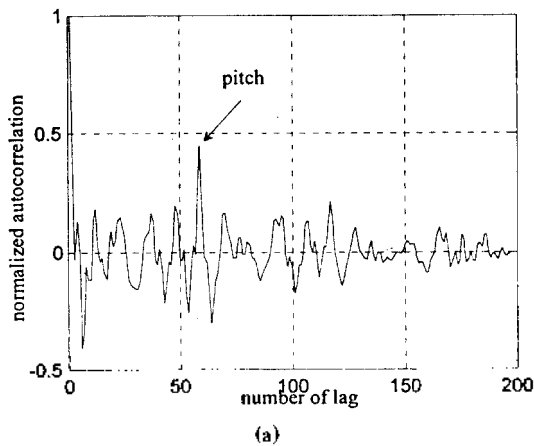
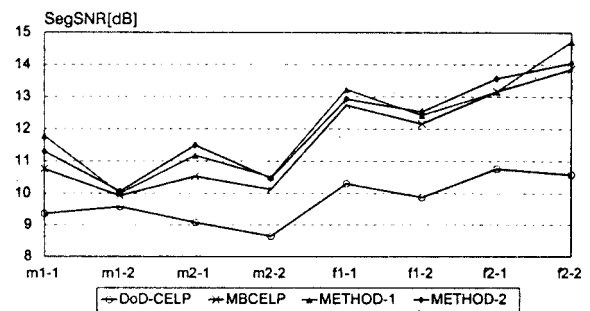


그림 5. 이차 장구간 예측 후 잔차 신호의 상관 함수의 한 예  
(a)MBCELP와 (b)MPMBCELP  
Fig. 5 autocorrelation function of residual signal after 2'nd LTP  
(a)MBCELP (b)MPMBCELP

분은 통계적 코드북에 의한 여기신호 모델링 과정에 나쁜 영향을 미치게 되고 전체적인 부호화기의 성능을 크게 감소시킨다. 그러나, 그림 5의 (b)에 나타난 다중 펄스 다대역 여기신호는 이러한 피크 성분이 제거되어 백색 잡음에 가까운 상관관계 함수를 얻을 수 있었다.

제안한 MPMBCELP의 다중 펄스 모델링 방법과 MBCELP[6]에 대해, 원 음성 신호와 그 음성을 합성하기 위한 여기 신호, 각 부호화 방법에 의해 합성된 복원 신호, 원 음성 신호와 복원된 신호 사이의 오차 신호를 관찰하여 보면, 본 논문에서 제안한 방법들이 오차 신호의 에너지가 크게 줄어드는 것을 알 수 있다. 전체 음성 구간 중, 본 논문에서 제안한 이차 장구간 예측의 성능 개선 효과로 인해 상대적으로 무성음 구간보다 유성음 구간에서 오차가 적어지는 것을 알 수 있다.



METHOD-1: 경험적인 다중 펄스를 이용한 MPMBCELP  
METHOD-2: 일차 선형함수로 구성된 MPMBCELP  
m1-1: 첫 번째 남성화자의 첫 번째 문장  
f2-2: 두 번째 여성화자의 두 번째 문장

그림 6. 각 음성 신호에 대한 SegSNR의 변화  
Fig 6. SegSNR comparison of test speech signal

제안한 부호화 방법은 기존 부호화 방법인 DoD-CELP에 비해 약 2-3dB 정도, MB-CELP에 비해 약 1dB 정도 향상된 결과를 얻을 수 있었으며, 이와 같은 결과는 본 논문에서 제안한 부호화 방법이 이차 장구간 예측의 도입을 통해 유성음 구간을 잘 표현해 줄 수 있기 때문이다. 음성 부호화기의 성능 평가를 위해서는 객관적 측도에 의한 결과보다는 합성음의 청취를 통해 주관적으로 평가하는 MOS(Mean Opinion Score) 결과가 더 의미 있는 평가 방법으로 알려져 있다[3]. MOS 평가는 20명을 대상으로 수행되었으며, 청취 평가가 지나치게 나쁘거나 좋은 사람의 데이터를 제외한 나머지 점수를 평균해서 표 3에 나타내었다.

본 논문에서 제안한 부호화 방식은 객관적 및 주관적 측도에서 우수한 결과를 얻을 수 있었다. 그러나, 밴드 분할과 이차 장구간 예측을 거쳐야 하는 부호화 알고리즘의 복잡성으로 인한 계산량은 문제점으로서 앞으로 많은 성능 개선이 요구된다.

표 2. MOS 평가 결과  
Table 2. MOS TEST

부호화 방법	평 균
DoD-CELP	3.45
MB-CELP	3.40
MPMBCELP	3.60

IV. 결 론

본 논문에서는 4.8 kbps 정도의 전송율에서 우수한 음질을 유지하는 CELP 계열 음성 부호화 알고리즘에 관한 연구를 수행하였다. CELP 계열 부호화기에서는 유성음 구간의 여기 신호가 음질에 가장 큰 영향을 주므로 합성 음의 음질은 장구간 예측기의 성능에 따라 크게 좌우된다.

본 논문에서는 장구간 예측 성능을 향상시키기 위해, 대역 분할 방법을 통한 이차 장구간 예측 과정을 도입한 부호화 방법을 제안하였다. 먼저, 일차 장구간 예측 후의 잔차 신호를 대역 분할한 후 몇 개의 펄스 군으로 구성된 펄스열로 모델링 하여 성능 개선을 시도하였다. 피치와 피치사이에는 펄스가 선형 함수의 형태로 감소된다는 특징을 이용하여 펄스열을 간단한 일차 함수로 모델링함으로써 성능을 향상시킬 수 있었다. 제안된 부호화 방법의 예로서 4.8 kbps MPMBCELP를 구현하였으며, 4.8 kbps DoD-CELP와 MBCELP와 음질을 비교 평가하였다. 실험 결과, 장구간 예측 후의 잔차 신호에 남아 있던 주기적 성분의 상관 관계가 기존의 MBCELP에 비해 크게 줄어들었음을 확인할 수 있었으며, 이러한 효과로 인하여, 합성 음질의 경우에도 SegSNR 측면에서 평균 1~2dB 정도, MOS 측면에서 0.15 정도 향상된 결과를 얻을 수 있었다.

참 고 문 헌

1. M. R. Shoroder, B.S. Atal, "Code-Excited Linear Prediction(CELP) High-Quality Speech at Very Low Bit Rates," *Proc. ICASSP*, pp. 937-940, 1985.
2. B. S. Atal and B.E. Caspers, "Beyond multipulse and CELP towards high quality speech at 4kb/s," in: B.S. Atal et al., eds., *Advances in Speech Coding*, Kluwer, Amsterdam, pp. 191-201, 1991.
3. L. R. Rabiner, R. W. Shafer, *Digital Processing of Speech Signals*, Prentice-Hall Inc., 1978.
4. W. Granzow, B.S. Atal, K.K. Paliwal, J. Schroeder, "Speech Coding at 4kb/s and lower using single-pulse and stochastic models of LPC excitation," *IEEE ICASSP*, pp. 217-220, 1991.
5. J. S. Marques and Tribolet, "Pitch prediction with fractional delays in CELP coding," *Proc. EUROSPEECH*, pp. 509-512, 1989.
6. G. Yang, H. Leich and R. Boite, "Multiband Code-Excited

Linear Prediction (MBCELP) for Speech Coding", *Signal Processing*, Vol. 31, pp. 215-227, 1993.

7. J. C. Hardwick and J.S. Lim, "The Application of the IMBE speech coder to mobile communications," *IEEE ICASSP*, pp. 249-252, 1991.
8. 최용수, 강홍구, 안명섭, 안동순, 윤대희, "대역 분할 장구간 예측을 이용한 CELP 부호화기" 한국 음향학회지 제14권 4호, pp. 38-45, 1995.
9. J. P. Campbell, T.E. Tremain, V.C. Welch, "DoD 4.8kbps Standard (Proposed Federal Standard 1016)," *Advances in Speech Coding*, Kluwer, 1991.
10. P. Kroon, E. F. Deprettere, R. J. Sluyter, "Regular Pulse Excitation: A Novel Approach to effective And Efficient Multipulse Coding of Speech," *IEEE Trans. ASSP*, Vol. ASSP-34, pp. 1054-1063, Oct. 1986.
11. Federal Standard 1016, Feb. 1991.
12. J. P. Campbell, T.E. Tremain, V.C. Welch, "An Expandable Error-protected 4800 bps CELP Coders(U.S. Federal Standard 4800 bps Voice Coder)", *IEEE ICASSP*, pp. 735-738., 1989.
13. R. A. Salami, "Binary Code Excited Linear Prediction (BCELP): New Approach to CELP coding of Speech without codebooks", *Electronics letters*, Vol. 25, No. 6, 16th March, 1989.
14. R. Boite, H. Reich and Gao Yang, "Simplification and Improvement of The Binary Coded Excited Linear Prediction (BCELP) for Speech Coding", *Proc. EUSIPCO*, pp. 1211-1214, 1990. Barcelona, Spain.
15. F. K. Soong, B.H. Juang, "Line Spectrum Pair(LSP) And Speech Data Compression," *Proc. ICASSP*, pp. 1.10.1-1.10.4, 1984.

▲서 정 태(Jeong-Tae Seo)



1981년~1985년: 연세대학교 전자공학  
학과(공학사)  
1985년~1987년: 연세대학교 대학원  
전자공학과(공학석사)  
1990년~1995년: 연세대학교 대학원  
전자공학과(공학박사)  
1995년~현재: 충주산업대학교 제어  
계측 공학과 조교수

※주관심분야: 음성 및 영상 데이터 압축, Fractal

▲강 홍 구(Hong-Goo Kang)

1985년~1989년: 연세대학교 전자공학과(공학사)  
1989년~1991년: 연세대학교 대학원 전자공학과(공학석사)  
1991년~1995년: 연세대학교 대학원 전자공학과(공학박사)  
1996년~현재: AT&T Labs-Research 연구원

※주관심분야: 음성 데이터 압축

▲최 용 수(Yong-Soo Choi)

1989년~1993년: 연세대학교 전자공학과(공학사)

1993년~1995년: 연세대학교 대학원 전자공학과(공학석사)

1995년~현재: 연세대학교 대학원 전자공학과 박사과정  
재학중

※주관심분야: 음성 데이터 압축

▲윤 대 회(Dae-Hee Youn)

현재: 연세대학교 전자공학과 교수

한국음향학회지 14권 4호 참조