

단순화된 다중 모드 방법을 이용한 음성 부호화기

正會員 姜泓求*, 徐廷泰*, 李寅誠**, 車日煥*, 尹大熙*

A Speech Coder using Simplified Multi-mode Method

Hong Goo Kang*, Jeong Tae Seo*, In Seong Lee**, Il Whan Cha*, Dae Hee Youn* Regular Members

要 約

본 논문에서는 음성의 구간 특성에 따라 다른 부호화 방법을 사용하여 4kbps 이하에서 우수한 성능을 갖는 음성 부호화기를 제안한다. 단구간 평균 에너지와 장구간 예측후의 잔차 신호 에너지를 이용하여 평형 상태 유성음 구간과 그 외 구간을 모드-A와 모드-B로 구분한 후, 모드 변화에 따라 각각에 적합한 여기 신호를 적용한다. 주기적인 특성이 강하게 나타나는 모드-A 구간에는 제한된 크기를 갖는 구조화된 다중 펄스로부터 완전 탐색 방법으로 여기 신호를 생성하며, 특성이 변하는 유성음 및 무성음 구간으로 구분되는 모드-B에는 모드 변화에 따라 백색 가우시안 혹은 펄스형 잡음으로 구성된 코드북을 사용한다. 제안된 방법은 유성음 구간에서 명료성을 유지하며, 천이 구간 및 무성음 구간에서도 일정한 음질을 유지하여 전체적으로 우수한 성능을 나타내었다. 또한, 프레임의 길이를 줄이고 모드 구분을 단순화시킴으로써 부호화 지연 시간과 계산량을 줄일 수 있었다.

ABSTRACT

This paper proposes a good speech coder which applies different coding method according to the characteristic of segment at bit-rates below 4 kbps. Speech signal is divided with 2 modes such as stationary voice(mode-A) and etc.(mode-B) using the average energy of the short-time speech and the residual after long term prediction. Then, we made a proper excitation signal according to the mode changes. The full search method with several constrained pulses is used at mode-A because it represents the characteristic of voiced efficiently. White Gaussian or pulse-like codebook are used at mode-B according to the mode changes.

The proposed method has good performance because it has transparency sound at voiced region and reasonable quality at transition and unvoiced region. In addition, it can be reduced computational load as reducing the frame size and simplifying the method of mode decision.

*연세대학교 전자공학과

Dept. of Electronics Eng., Yonsei Univ.

**한국전자통신연구소 신호처리 연구실

Electronics and Telecommunications Research Institute

論文番號 : 95113-0320

接受日字 : 1995年 3月 20日

본 연구 중 일부 내용은 한국전자통신연구소(ETRI)의 연구비 지원에 의한 결과임.

I. 서 론

음성 관련 분야 중 신호를 압축한 후 왜곡없이 복원하기 위한 부호화 연구는 디지털 음성 통신 분야에서 매우 중요한 역할을 담당하고 있다. 이러한 연구는 디지털 셀룰라, 음성 저장 및 전송 시스템 등을 중심으로 활발히 연구되고 있으며, CELP(Code Excited Linear Prediction)⁽¹⁾, MPE(Multi-Pulse Excitation)⁽²⁾, RPE(Regular Pulse Excitation)⁽³⁾등과 같이 선형 예측(linear prediction)을 이용하는 분석-합성(analysis-by-synthesis)방법이 주로 사용되고 있다. 최근에는 4kbps 내외의 전송률을 갖는 half-rate에서도 기존 시스템과 비슷한 성능을 유지하기 위한 연구를 통해 음성 부호화 기술은 급격한 발전을 이루고 있으며, 이를 바탕으로 각 국에서는 표준안을 결정하기 위한 작업이 진행중이다⁽⁴⁾.

선형 예측 방법을 이용하여 전송률을 절반 이하로 낮추기 위해 프레임의 길이를 늘리는 경우에는⁽⁴⁾ 장구간 예측 성능이 저하되어 구간에 따라 잔차 신호의 백색 잡음 특성이 사라지게 되므로 기존의 코드북 만으로는 성능을 유지하기 어려워진다. 이를 해결하기 위해 전처리 단에서 음성 구간의 특성에 따라 몇가지 모드로 구분한 후, 각각에 대해 독립적인 부호화 방법을 사용하는 다중 모드(multi-mode)를 방법이 연구되고 있다⁽⁵⁾⁽⁶⁾⁽⁷⁾. 미국 디지털 셀룰라의 half-rate 표준안 후보로 제출된 M-LCELP(Multi-mode Learned CELP)⁽⁸⁾에서는 음성 신호를 4가지 모드로 구분한 후, 유성음의 특징을 나타내는 구간에서는 일정 간격으로 구성된 펄스 코드북을 그리고, 무성음 구간에서는 백색 가우시안으로 구성된 코드북을 다단으로 구성하는 방식을 취하고 있다. 그러나, 이 방법은 음성 구간을 분류하기 위한 알고리즘이 복잡하며, 부호화 방법 또한 복잡한 다단 방식을 사용함으로써 계산량이 많다는 단점이 있다.

본 논문에서는 프레임 에너지와 잔차 신호 에너지를 이용하는 간단한 구분 알고리즘을 통해 음성을 정상 상태의 유성음(모드-A)과 그 외 구간(모드-B)으로 구별하고 모드 변화에 따라 효율적인 여기 신호를 적용하여 이를 극복하고자 한다. 논문의 구성은 다음과 같다. 2장에서는 대표적인 다중 모드 방법인 M-LCELP⁽⁸⁾에 대해 살펴보고, 3장에서는 제안된 단순화된 다중 모드 방법을 서술한다. 4장에서는 4kbps 이하의 평균 전송률

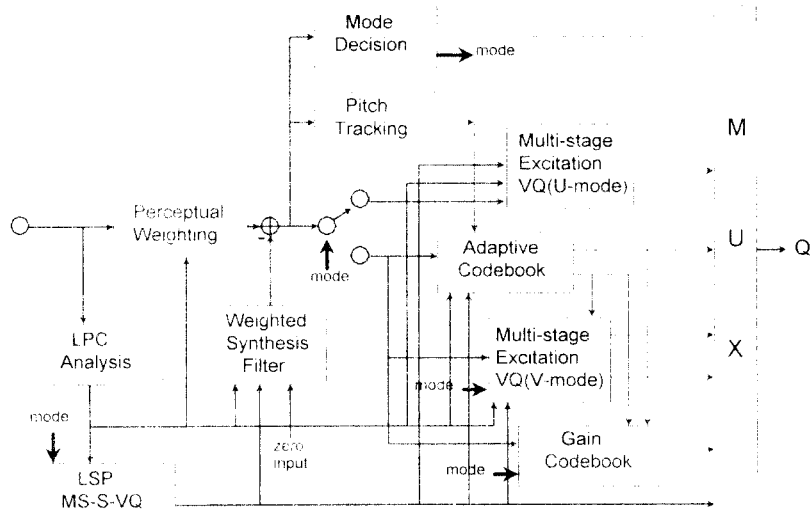
을 유지하기 위한 비트 할당 방법과 이를 연구하여 평균 3.4kbps로 구현된 부호화기의 주/객관적 성능 평가에 대해 알아본다. 성능 평가는 표준안으로 제정된 방법과의 비교가 객관적으로 타당하므로 4.8kbps VSELP⁽¹³⁾와 DoD-CELP⁽¹⁴⁾를 사용한다. 마지막으로 5장에서 결론을 맺는다.

II. M-LCELP

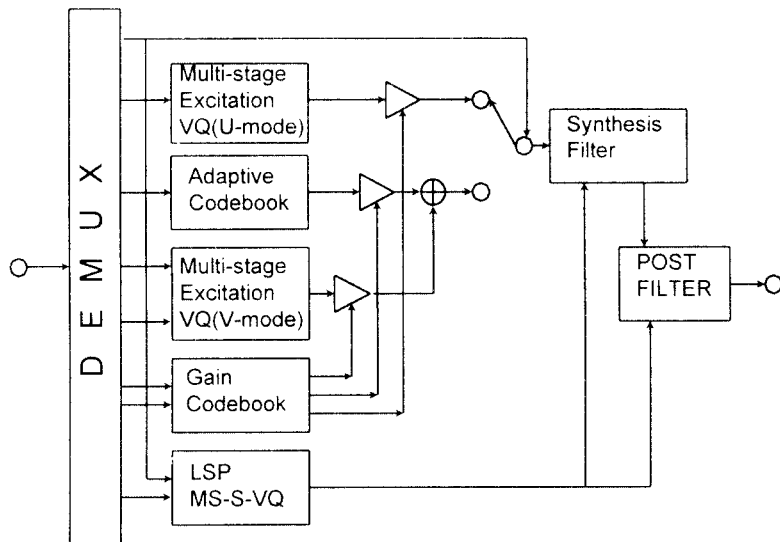
M-LCELP(Multi-mode Learned CELP)⁽⁸⁾는 복미 디지털 셀룰라의 표준 방법을 위해 제안된 방법으로서 그림 1과 같은 분석 및 합성 구조를 지닌다.

입력된 음성 신호는 선형 예측(linear spectrum)을 통해 계수를 구하고, 이로부터 청각 특성을 반영하기 위한 심리 가중 필터에 통과한다. 구해진 선형예측 계수는 양자화 이득을 얻기 위해 LSP(Line Spectral Pair)로 변환되어 전송된다⁽⁹⁾. 모드 결정(mode decision)블럭에서는 3개의 임계값으로 구성된 피치 예측 이득을 이용하여 0, 1, 2, 3의 네가지 모드로 구분한다. 결정된 모드 정보에 따라 각 구간의 여기 신호 생성 방법이 달라진다. 모드-1, 모드-2, 모드-3에서는 간속비를 2로 갖는 일정 간격의 펄스(regular pulse) 코드북을 2단으로 구성하여 여기 신호로 사용하며, 무성음으로 이루어지는 모드 0에서는 장구간 예측을 하지 않고, 3단으로 구성된 비정기적 펄스 코드북을 사용하여 여기 신호를 생성한다. 코드북을 다단으로 구성함에 따른 계산량과 성능과의 trade-off 관계에 의해 첫번째 단에서는 몇 개의 후보를 미리 선택하고, 두번째 단에서는 선택된 코드와의 조합을 통해 최적의 코드북을 찾는 방법을 사용한다. 최적 코드북을 찾기 위한 방법은 코드북으로부터 생성된 여기 신호로부터 합성 신호를 얻은 후, 이 신호와 가중필터에 통과된 입력 음성 신호와의 오차가 가장 적은 여기 신호를 선택하는 과정으로 이루어진다. 결국 전송될 변수는 각 구간의 모드 정보, 여기 신호 구성을 위한 코드북 인덱스와 이득, 그리고 각 모드에 따라 다르게 할당된 LSP 양자화 정보이다. 복호화 단계에서는 전송된 정보로부터 해당 변수를 얻고, 이를 합성 필터에 통과시켜 합성음을 얻는다.

M-LCELP에서 위와 같이 여기 신호를 복잡하게 생성하는 이유는 전송률을 4kbps로 낮추기 위해 프레임 길이를 40 msec로 확장하였기 때문이다. 프레임의 길



(a)



(b)

그림 1 M-LCELP 부호화기 및 복호화기
Fig. 1. Encoder and decoder of the M-LCELP

이가 길어지면 음성의 구간 특성 변화가 심해지므로 스펙트럼 포락선 성분을 정확히 분리하기 어려워진다. 따라서, 구분한 모드 각각에 대해서도 여기 신호를 좀 더 세밀하고, 정교하게 처리해야만 좋은 음질을 유지할 수

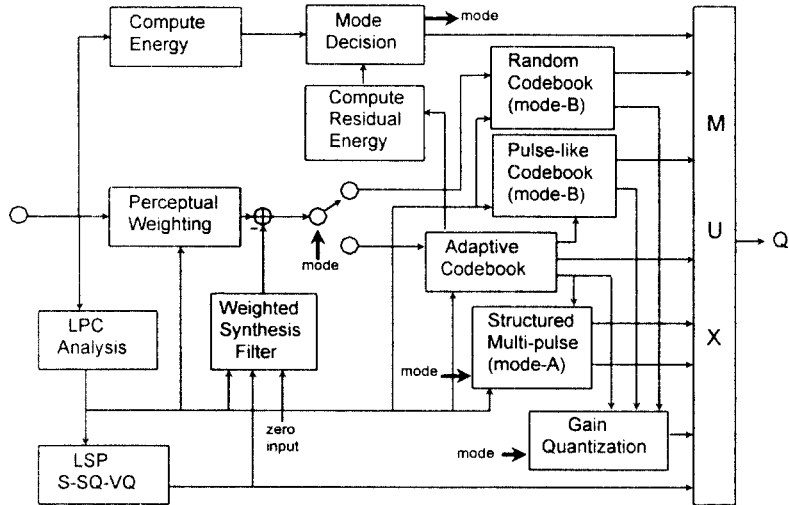
있다. 또한, 프레임의 길이가 길어짐에 따라 부호화 지연 시간(coding delay)이 길어진다는 단점도 발생한다. 본 논문에서는 이러한 문제를 극복하기 위해 프레임의 길이를 축소하고, 모드 변화에 따라 여기 신호를 단

순화하여 성능 개선을 시도한다.

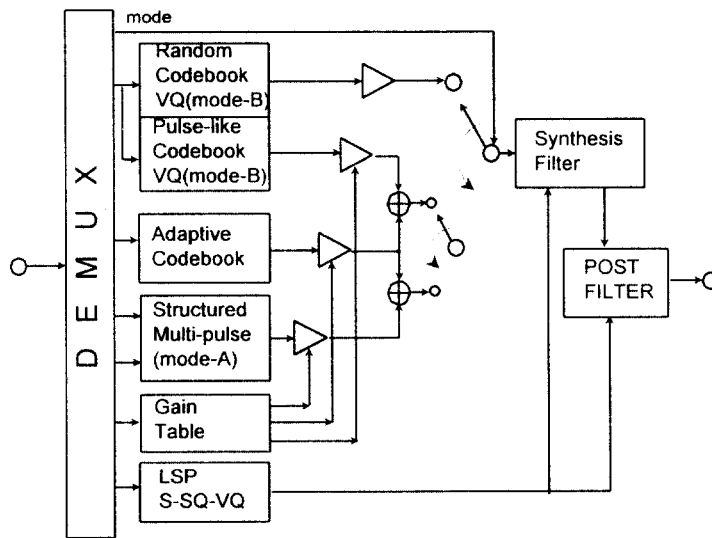
Ⅲ. 단순화된 다중 모드 방법

위에서 서술한 M-LCELP의 문제점을 해결하기 위해

본 논문에서 제안된 단순화된 다중 모드 방법의 블록도는 그림 2와 같다. 모드 결정을 위해 음성 신호의 단구간 에너지와 장구간 예측 후의 잔차 신호(residual signal) 에너지를 사용하며, 이로부터 2가지로 모드를 구분한다. 평형 상태의 유성음 구간에서는 구조화된 다



(a)



(b)

그림 2. 제안 방법의 부호화기 및 복호화기
Fig. 2. Encoder and decoder of the proposed method

중 펄스를 사용하여 여기 신호를 생성하며, 그 외 구간은 모드 변화에 따라 백색 가우시안 잡음 혹은 펄스형으로 구성된 코드북을 사용한다. 최적 여기 신호를 선택하는 방법은 기존 방법과 동일하다⁽¹⁾.

본절에서는 모드 구분 방법과 여기 신호 모델링 방법으로 구분하여 제안된 알고리즘을 설명한다.

3-1. 모드 구분법

다중 모드 방법에서는 구간 신호의 특성을 구분하기 위한 전처리 단계가 반드시 필요하다. 기존 방법에서는 분석 프레임의 길이가 길어짐에 따라 복잡한 변수를 통해 모드를 구분했지만, CELP 알고리즘 특성 상 모드 구분을 정교하게 하여도 큰 이득을 얻기 어렵다. 그러나, 구조화된 다중 펄스 모델과 같이 유성음 구간에서는 좋은 특성을 유지하지만, 그 외 구간에 대해서는 문제가 발생하는 방법에 대해서는 간단한 모드 구분을 통해 성능 개선을 유도할 필요가 있다. 따라서, 본 논문에서는 프레임의 길이를 줄인 후, 단구간 입력 신호 에너지와 잔차 신호 에너지와 같이 단순하면서도 구간의 특징을 비교적 효율적으로 반영할 수 있는 변수를 이용하여 모드를 구분하였다.

가. 단구간 에너지

입력된 일정 구간의 음성 신호는 (1)식에 주어진 n 번째 프레임에서의 단구간 에너지를 이용하여 유성음 구간과 그 외 구간으로 구분한다.

$$E_s(n) = \sum_{m=-\infty}^{\infty} s^2(m)w(n-m) \quad (1)$$

여기서, $s(m)$ 은 입력 음성 신호이며, $w(n)$ 은 창(window) 함수이다.

실험을 통해 임계값(THD_1)을 구한 후, 단구간 에너지가 임계값보다 작을 때에는 모드-B로 판별하며, 클 경우에는 다음 절에 주어진 단구간 잔차 신호 에너지를 이용하여 다시 세분한다.

$$E_r(n) = \begin{cases} > |THD_1|, & \text{유성음 구간} \\ < |THD_1|, & \text{무성음/목음 구간 : 모드-B(2)} \end{cases}$$

나. 잔차 신호 에너지

(2)식에서 유성음으로 판별된 구간에서는 다시 잔차 신호의 에너지를 계산하여 장구간 예측 성능을 조사한다. 특성이 변하는 유성음 부분의 잔차 신호에는 펄스형

성분이 복잡하게 존재하므로 평형 상태의 유성음에 비해 에너지가 크다. 이런 구간은 비슷한 크기의 펄스 성분이 여러개가 발생하므로 여기 신호 발생시 무성음 구간과 같이 취급하는 것이 효율적이다. 이를 구분하기 위해 (3)식과 같이 정규화된 잔차 신호의 에너지를 사용한다.

$$E_r(n) = \frac{1}{N} \sum_{m=-\infty}^{\infty} r^2(m)w(n-m) \quad (3)$$

여기서 $r(n)$ 은 장구간 예측 후의 신호이고, $w(n)$ 은 창(window) 함수이며, N 은 부프레임 길이이다. 단구간 에너지 판별법에서와 마찬가지로 임계값(THD_2)은 실험을 통해 얻는다.

$$E_r(n) = \begin{cases} > |THD_2|, & \text{평형 상태의 유성음 구간: 모드-A} \\ < |THD_2|, & \text{특성이 변하는 유성음 구간: 모드-B(4)} \end{cases}$$

구간 에너지와 잔차 신호의 에너지로 부터 무성음 구간과 특성이 변하는 유성음 구간은 동일한 모드로 구분되므로 결국 모드는 두 종류로 구성된다. 그림 3은 음성 신호를 사용하여 실제로 모드를 구분한 예를 나타낸다.

그림 3에서 (a)는 원음, (b)는 (a)로부터 프레임마다 구한 에너지 그리고, (c)는 잔차신호이며, (d)는 잔차 신호의 에너지를 부프레임 단위로 구한 값이다. (b)에서 점선으로 표시된 THD_1 보다 작은 값을 갖는 영역은 모드-B로 선택되며, 큰 값을 갖는 영역에 대해서는 부프레임에 대해 다시 한번 에너지를 구한다. (d)에 주어진 임계값 THD_2 는 특성이 변하는 유성음 구간을 판단하기 위해 주어진 값이다. ①로 표시된 곳은 특성이 변하는 유성음 구간을 나타내는 부분으로서 본 논문에서 모드-B로 판단하는 부분이다.

3-2. 모드 결정에 따른 부호화 방법

합성음의 성능을 높이기 위해서는 모드 구분에 따라 각각에 적합한 여기 신호를 발생하는 것이 중요하다. 유성음 구간에서는 잔차 신호에 존재하는 피치 성분을 모델링하는 것이 효과적이며, 무성음 구간에서는 백색 잡음 특성을 반영하는 것이 효과적이다. 본 논문에서는 낮은 전송률에서도 이러한 조건을 만족할 수 있도록 하는 여기 신호 생성 방법과 이전 구간의 모드 정보를 이용하여 모드 변화에 따라 추가 정보 없이도 여기 신호를 효과적으로 응용할 수 있는 방법을 제안한다.

가. 평형 상태 유성음 : 모드-A

유성음 구간에서의 명료성을 높이기 위해서는 피치 성

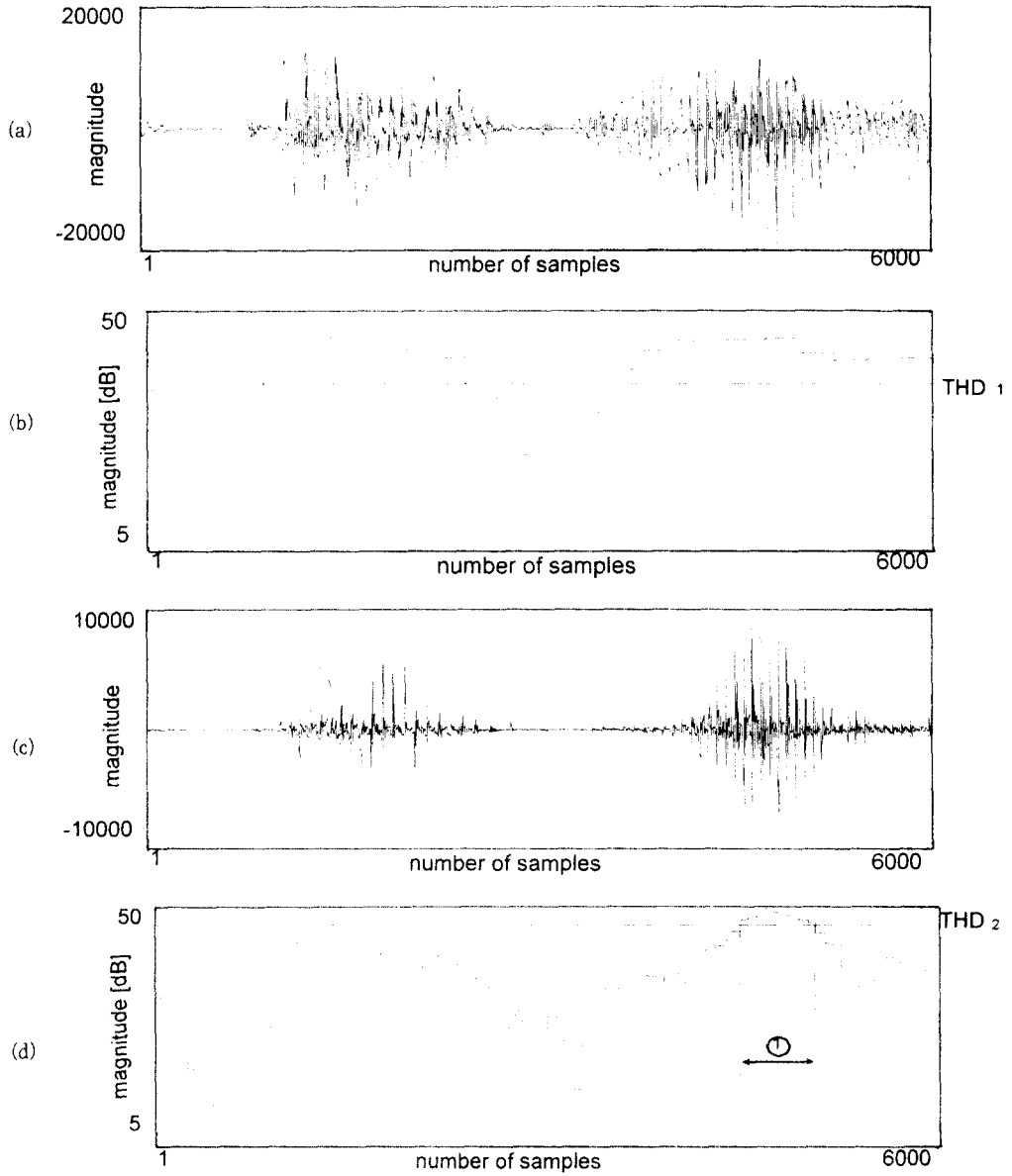


그림 3. 모드 구분 예

(a) 원음 (b) 단구간 에너지($E_s(n)$) (c) 잔차 신호 (d) 잔차 신호 에너지($E_r(n)$)

Fig. 3. An example of mode decision

(a) original signal (b) short-time energy($E_s(n)$) (c) residual signal (d) residual signal energy($E_r(n)$)

분을 정확히 모델링하는 것이 중요하다. 이를 위해 그림 4와 같이 구조화된 펄스 조합으로 부터 최적의 펄스 위치와 간격을 결정하는 구조화된 다중 펄스 방법을 사용한다^[4]. 구조화된 다중 펄스 방법은 MPE(Multi Pulse Excitation)^[2]에서와 같이 다중 펄스열을 도입하여 장구간 예측 후의 신호에 존재하는 펄스 성분을 모델링하는 방법으로서 펄스 개수와 이득 정보를 미리 고정시킨 후, 오차가 최소가 되는 펄스의 위치를 찾아 전송하는 것이다. 펄스수는 부프레임의 길이에 따라 조정하며, 이득 정보는 실시간 처리에 편한 수로 미리 양자화된 값을 사용한다. 따라서, 전송률 및 연산량은 펄스 위치 탐색 간격에 따라 결정된다. 다음은 펄스 개수와 탐색 영역, 그리고 펄스의 양자화 레벨에 다음과 같은 제한을 두는 방법이다.

제한조건-1: 펄스 개수는 M개로 고정한다.

제한조건-2: 첫번째 펄스 크기는 "1"이며, 부프레임의 어느 위치에나 존재할 수 있다.

제한조건-3: 두번째 이후의 펄스 크기는 $\pm 2^n$ 으로 양자화하고, 최적 위치는 이전 펄스부터의 일정 간격에 대해 완전 탐색(full search)을 통해 결정한다.

전송률과 여기신호 생성 방법과의 상관 관계에 따라 프레임의 길이는 20 msec에서 30 msec 내외로 결정한다. 펄스 위치 정보를 표시하기 위해서는 2의 거듭 제곱 값으로 하는 것이 효율적이다. 전송률을 낮추고, 고속 알고리즘을 적용하기 위해서는 부프레임에 사용할 최대 펄스의 개수를 제한해야 한다. 일반 사람의 피치 간격과 부프레임 길이를 고려할 때 피치 성분은 세 개를 넘지

않으며, 두 개일 경우에도 만족할 만한 성능을 얻을 수 있음을 실험을 통해 관찰하였다^[8]. 두 개의 펄스로 여기 신호를 구성할 경우에는 두번째 펄스의 탐색 영역(D)와 크기(α) 정보만 필요하며, 고속 반복 알고리즘^[6]을 사용하면 연산량을 매우 감소시킬 수 있다. 펄스 개수 뿐만 아니라 두번째 펄스 탐색 영역 및 양자화 레벨도 성능 및 전송률 면에서 중요한 변수이다. 같은 전송률에서는 양자화 레벨보다는 탐색 영역이 더 중요한 변수이며, 제한된 크기에서는 양자화 레벨이 4단계 이상 증가하더라도 별 차이가 없다^[8]. 두번째 펄스의 탐색 영역은 이전 펄스로부터 32 내외의 간격으로 제한하더라도 성능이 크게 저하되지 않았으므로^[8], 탐색 영역 D에는 5 비트를 할당한다. 또한, 두번째 펄스의 크기는 2비트로 양자화 하여 4가지 값, (+1, -1, 0.5, -0.5)을 갖도록 제한한다.

나. 기타 구간

펄스의 모양이 일정하게 유지되는 유성음 구간에서는 위와 같은 구조화된 펄스를 이용하는 방법이 효율적이다. 그러나, 유성음에서 다른 형태의 유성음으로 변하는 부분이나 무성음 구간에서와 같이 랜덤 성분이 강하게 나타나므로 2개의 펄스로는 모델링이 어려워진다^[8]. 여러 개의 펄스 조합으로 코드를 구성하면 이를 해결할 수 있지만, 전송률 및 계산량이 증가하게 되므로 벡터 양자화를 통한 코드북을 사용하는 것이 효율적이다. 또한 특성이 변하는 유성음 구간에서는 백색 가우시안 보다는 DoD-CELP^[14]에서 사용한 코드와 같이 0이 많이 포함된 펄스 형태의 코드북을 이용하여 여기 신호를 구성하는 것이 평효성을 유지하는데 효과적이며, 연산량도 줄일 수 있다^[8].

다중 모드의 또다른 장점은 장구간 예측기의 사용 여부를 결정할 수 있다는 것이다. 일반적으로 무성음 구간에서는 장구간 예측을 하더라도 성능이 향상되지 않으므로 장구간 예측 지연 시간과 이득에 비트를 할당할 필요가 없어진다. 따라서, 평균 전송률을 더욱 낮출 수 있다. 그러나, 본 논문의 모드-B에서와 같이 유/무성음 특징을 나타내는 구간을 모두 하나의 모드에 포함하는 경우에는 이 기능을 일관적으로 적용하기 어렵다. 특성이 변하는 유성음 구간에서 장구간 예측을 수행하지 않으면, 음질 저하가 심하기 때문이다. 모드를 따로 구분한다면 해결 가능하겠지만, 모드 구분을 위한 정보가 1

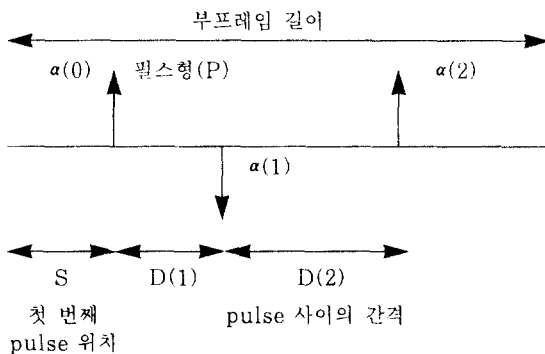


그림 4. 세 개의 펄스로 구성된 구조화된 다중 펄스
Fig. 4. Structured multi-pulse with three pulses

비트 추가되므로 전송률이 늘어난다. 추가 정보 없이 장 구간 예측 여부를 결정할 수 있는 한 방법은 매 구간마다 모드가 변화되는 상태를 살피는 것이다.

그림 5는 모드 변화에 따른 장구간 예측 사용 여부

및 여기 신호 생성 방법을 나타낸 상태도이다. 그림에서 볼 수 있듯이 모드-B로 판별된 구간에서는 이전 구간의 모드를 검토하여 이전 구간이 모드-A일 경우에는 장구간 예측을 수행하고 모드-B일 경우에는 장구간 예측을

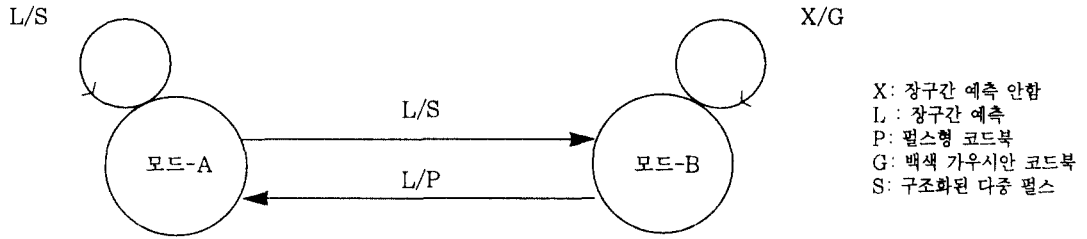


그림 5. 모드 변화에 따른 상태도
Fig. 5. State diagram with varying the mode information

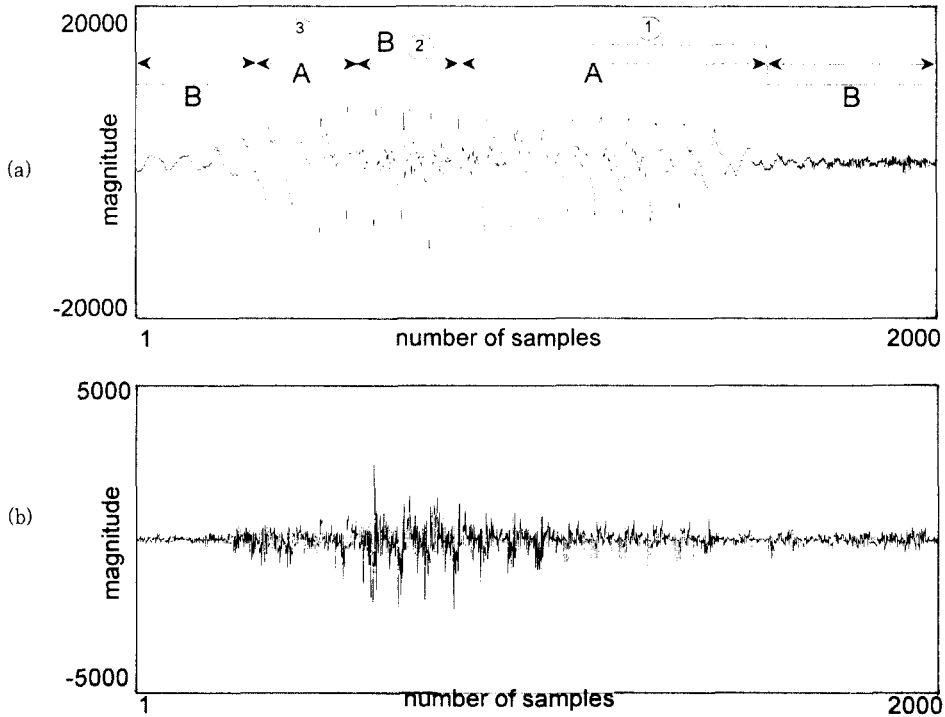


그림 6. 모드 변화에 따라 생성된 여기 신호
(a) 원음 신호와 모드 구분 정보 (b) 생성된 여기 신호
Fig. 6. Generated excitation signal according to the mode changing
(a) original signal and mode decision (b) generated excitation signals

수행하지 않음을 알 수 있다. 그림 6은 모드 변화에 따라 생성된 여기 신호를 나타낸 그림이다. ①로 표시된 구간은 유성음 구간이 연속되는 모드-A로서 구조화된 다중 펄스 방법에 의해 피치 성분이 정확히 나타내는 것을 알 수 있다. ② 구간은 유성음의 특징이 변하는 모드-B 구간으로서 펄스형 코드북에 의해 다수의 펄스 성분으로 비교적 정확히 구성된다. ③ 구간은 무성음에서 유성음으로 변하는 천이 구간으로서 다소 여기 신호 구성이 정확하지 못하지만 주관적 성능을 좌우하는 요인은 아니다.

IV. 실험 및 결과 고찰

제안한 부호화기의 성능 평가를 위해 모의 실험을 수행하였다. 실험에 사용된 데이터는 조용한 환경에서 남녀 화자 각 2명이 발음한 문장을 8KHz 샘플링하여 사용하였으며, 길이는 약 30초 정도이다. 프레임 길이는 32ms로 구성하였으며, 부프레임은 프레임 길이의 1/4로 하였다. 프레임 길이를 2의 거듭 제곱 형태로 한 이유는 평형 상태 유성음에 적용하는 펄스 위치 정보를

효율적으로 표시하기 위해서이다. 이와 같은 환경에서 평균 4kbps 이하의 전송률을 갖기 위해 표 1과 같이 비트를 할당하였다. 실험 결과 제안한 부호화기의 평균 전송률은 약 3.4kbps로 나타났으며, 문장에 따라 차이는 있지만, 최대 전송률은 4kbps를 넘지 않았다.

성능 비교를 위해 4.8kbps 전송률을 갖는 VSELP^[13]와 DoD-CELP^[14]를 사용하였으며, 평가 방법은 객관적 평가 방법으로 구간 신호대 잡음비(SegSNR)와 신호대 잡음비(SNR)를 사용하였으며, 주관적 평가 방법으로는 A-B Test 및 MOS(Mean Opinion Score)를 사용하였다.

주관적 평가는 연구원 10명을 대상으로 실시하였으며, 평가 결과는 다음과 같다.

표 2 객관적 성능 평가

	SegSNR	SNR
VSELP	13.60	10.35
DoD-CELP	9.40	8.30
제안 방법	11.65	9.82

표 1. 제안 방법의 비트 할당

변 수	bit/8ms 부프레임			bit/32ms 부프레임		
	평형 상태 유성음 (모드-A)	기타 구간 (모드-B)		평형 상태 유성음 (모드-A)	기타 구간 (모드-B)	
		→모드-B	→모드-B		→모드-B	→모드-B
LPC 계수(LSP)				30	30	30
LTP 지연시간	7	7		28	28	
LTP 이득	3	3		12	12	
구간 특성 정보	1	1	1	4	4	4
	S=6 D=5 P=2			52		
통계 코드북		9	9		36	36
이득	3	3	3	12	12	12
Total (kbps)				138 (4.3)	122 (3.9)	82 (2.6)

표 3. 주관적 성능 평가(A-B Test)

VSELP	제안 방법	DoD-CELP	제안 방법
44.5%	55.5%	47.0%	53.0%

표 4. 주관적 성능 평가(MOS Test)

VSELP	DoD-CELP	제안 방법
3.15	3.07	3.28

제안된 방법은 VSELP 방법에 비해 객관적 수치는 낮지만, 명료성이 뛰어나므로 좋은 평가를 받은 것으로 생각된다. 그러나, 무성음에서 유성음으로 변하는 부분과 유성음이 지속되는 부분에서는 다소 음질 저하 현상이 발생하였다. 이는 펄스 모델이 단순화되고 코드북의 크기가 제한됨에 따라 발생하는 것으로서 좀 더 낮은 전송률에서는 고려해야할 문제라고 판단된다.

V. 결 론

본 논문에서는 구간 음성의 특성에 따라 다른 부호화 방법을 사용하여 4kbps 이하에서 우수한 성능을 갖는 음성 부호화기를 제안하였다. 음성 신호의 단구간 에너지와 장구간 예측 후의 에너지를 이용하여 평형 상태 유성음(모드-A)과 기타 구간(모드-B)으로 나누고 모드 변화에 따라 각각 다른 부호화 방법을 적용하였다. 모드-A 구간에는 구조화된 다중 펄스를 이용하였고, 모드-B 구간에는 이전 프레임의 모드에 따라 백색 가우시안 잡음 혹은 펄스형 잡음을 코드북으로 구성하여 사용하였다. 또한, 모드-B 중 무성음 혹은 목음이 지속되는 구간에서는 장구간 예측을 수행하지 않았으며, 이로부터 전송률 및 계산량 이득을 얻었다.

제안한 방법에 사용한 구조화된 펄스 모델링 방법은 고속 알고리즘을 통해 실시간처리가 용이하고, 피치 성분을 정확히 모델링함에 따라 명료성이 뛰어나다. 무성음 구간과 특성이 변하는 유성음 구간에서는 백색 가우시안과 펄스형 잡음으로 구성된 코드북을 각각 사용함으로써 과거 구조화된 펄스만을 사용한 방법의 문제점이었던 음질 왜곡을 해결할 수 있었다. 전체 성능 평가 결과 4.8 kbps VSELP에 비해 구간 신호대 잡음비는 약

1dB 정도 낮았지만 주관적 성능 평가 결과는 기존 두 방법에 비해 비슷하거나 우수하였다. 그러나, 유성음이 지속되는 구간과 무성음에서 유성음으로 변하는 천이구간에서의 음질 저하 현상은 다소 문제점으로 남아 있다. 또한, 구간 특성에 따라 전송률이 변하므로 일정한 전송률을 요하는 경우에는 적용이 어렵다는 단점이 있다.

참고문헌

1. M. R. Schroeder, B. S. Atal, "Code Excited Linear Prediction(CELP) High-Quality Speech at Very Low Bit Rates," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 25.1.1~25.1.4, Florida, Mar., 1985.
2. B. S. Atal, J. R. Remde, "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp.614-617, Paris, May, 1982.
3. P. Kroon, E. F. Depretive, R. J. Sluyter, "Regular Pulse Excitation : A Novel Approach to Effective And Efficient Multipulse Coding of Speech," *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-34, pp.1054-1063, Oct., 1986.
4. *Speech and Audio Coding for wireless and Network Applications*, Kluwer Academic Publisher, 1993.
5. K. Ozawa, M. Serizawa, T. Miyano and T. Nomura, "M-LCELP Speech Coding at 4kbps," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp.269-272, Adelaide, May, 1994.
6. J. Martin, B Wachter, "A CODEC Candidate for the GSM Half-Rate Speech Channel" *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Adelaide, May, 1994.
7. K. Swaminathan, et al, "Speech and Channel CODEC Candidate for the Half-Rate Digital Cellular" *Proc. Int. Conf. Acoust., Speech, Signal processing*, pp.1265-268, Adelaide, May, 1994.

8. 강홍구, 서정태, 차일환, 윤대회, "구조화된 다중 펄스 열을 이용한 낮은 전송률 음성 부호화기" 한국 통신학회 논문지, vol. 20, No. 6, Jul., 1995..
9. P. Kabal and R. P. Ramachandran, "The Computation of Line Spectral Frequencies Using Chebyshev Polynomials", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-34, 1419-1426, Dec., 1986.
10. S. Wang, A. Gersho, "Phonetic Segmentation for Low Rate Speech Coding," *Advances in Speech Coding*, Kluwer Academic Publisher, 1991.
11. F. K. Soong and B. H. Juang, "Line Spectrum Pair and Speech Data Compression", *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp.1.10.1-1.10.4., San Diego, Mar., 1984.
12. K. K. Paliwal, and B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24bits/frame," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp.661-664, Toronto, May, 1991.
13. I. A. Gerson, M. A. Jasiuk, "Vector-Sum Excited Linear Prediction(VSELP) Speech Coding at 8Kbps," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp.641-644, Albuquerque, Apr., 1990.
14. J. P. Campbell, T. A. Tremain, V. C. Welch, "DoD 4.8Kbps Standard(Proposed Federal Standard 1016)," *Advances in Speech Coding*, Kluwer Academic Publisher, 1991.
15. J. Markel, A Grey, *Linear Prediction of Speech*, Springer-Verlang, 1976.



姜泓求(Hong Goo Kang) 정회원

1989년 2월 : 연세대 전자공학과 졸업
 1991년 2월 : 동 대학원 전자공학과 졸업(공학석사)
 1991년 3월~현재 : 동 대학원 박사 과정

*주관심 분야 : 음성 부호화, 오디오 부호화, 실시간 신호처리 시스템



徐廷泰(Jeong Tae Seo) 정회원

1985년 2월 : 연세대 전자공학과 졸업
 1987년 2월 : 동 대학원 전자공학과 졸업(공학석사)
 1990년 3월~현재 : 동 대학원 박사 과정

*주관심 분야 : 음성 부호화, 영상 압축, 적응 디지털 신호처리

李寅誠(In Seong Lee) 정회원

한국통신학회 논문지 Vol. 20, No. 2, Feb., 1995. 참조.

尹大熙(Dae Hee Youn) 정회원

한국통신학회 논문지 Vol. 20, No. 2, Feb., 1995. 참조.



車日煥(Il Whan Cha) 정회원

1959년 2월 : 연세대 전기공학과 졸업(공학석사)
 1983년 8월 : 동 대학원 전자공학과 졸업(공학박사)

1969년~1970년 : 영국 University of Southampton Institute of Sound and Vibration 연구원

1986년 9월~1987년 7월 : University of Iowa 교환교수
 1988년 8월~현재 : 연세대학교 전자공학과 교수

*주관심 분야 : 건축 음향, 소음진동, 초음파, 음성 부호화, 오디오 부호화.