

논문 2005-42SP-2-14

# 전이구간 부호화를 이용한 2.4 kbit/s 다중모드 음성 부호화 방법

## (Method of a Multi-mode Low Rate Speech Coder Using a Transient Coding at the Rate of 2.4 kbit/s)

안 영 옥\*, 김 중 학\*\*, 이 인 성\*\*\*, 권 오 주\*\*\*\*, 배 문 관\*\*\*\*

( Yeong-uk Ahn, Jong-hak Kim, Insung Lee, Oh-ju Kwon, and Mun-Kwan Bae )

### 요 약

현재 개발된 4 kbit/s 이하의 저 전송율 음성부호화 시스템은 STC(Sinusoidal Transform Coding)나 MBE (Multi-band Excitation Coding)에 바탕을 두고 있다. 이러한 저 전송율 부호화기들은 대표적인 전이구간 신호인 유성음의 시작점과 끝점에서의 혼합신호(onset signal, offset signal), 비주기적인 신호(non-period signal) 등은 정확히 표현하지 못하기 때문에 자연스런 음질을 만들어 내지 못한다. 본 논문에서는 유성음에는 하모닉 모델, 무성음에서는 스토캐스틱 모델, 전이구간에는 하모닉 기반의 비주기적인 펄스의 위치를 추적하는 방식을 사용하여 효과적으로 전이구간을 모델링 하는 방법과 2.4 kbit/s 다중모드 부호화방법을 제안한다. 제안한 방법은 원본신호에서 선형예측 부호화 방법으로 추출된 잔여신호를 신호의 성격에 따라 모델을 달리하는 방법이며, 각각의 신호의 성격에 따라 좋은 성능을 나타내는 모델을 사용하였다. 또한 효율적인 전이구간 모델링 방법의 도입으로 저 전송율에서 CELP(Code Excitation Linear Predictive) 부호화 방식에 의해 시간축에서 합성되는 여기신호와 선형위상을 이용한 하모닉 부호화 방식에 의해 주파수축에서 합성되는 여기신호를 효율적으로 결합이 가능하다는 것이 제안된 2.4 kbit/s 다중모드 부호화기의 장점이다. 제안된 방법의 2.4kbit/s 다중모드 부호화기는 미국 연방 표준부호화기인 2.4 kbit/s MELP(Mixed Excitation Linear Prediction) 부호화기보다 더 좋은 성능을 나타낸다.

### Abstract

The low rate speech coders under 4 kbit/s are based on sinusoidal transform coding (STC) or multiband excitation (MBE). Since the harmonic coders are not efficient to reconstruct the transient segments of speech signals such as onsets, offsets, non-periodic signals, etc, the coders do not provide a natural speech quality. This paper proposes method of a efficient transient model and a multi-mode low rate coder at 2.4 kbit/s that uses harmonic model for the voiced speech, stochastic model for the unvoiced speech and a model using aperiodic pulse location tracking (APPT) for the transient segments, respectively. The APPT utilizes the harmonic model. The proposed method uses different models depending on the characteristics of LPC residual signals. In addition, it can combine synthesized excitation in CELP coding at time domain with that in harmonic coding at frequency domain efficiently. The proposed coder shows a better speech quality than 2.4 kbit/s version of the mixed excitation linear prediction (MELP) coder that is a U.S. Federal Standard for speech coder.

**Keywords :** Speech coding, Vocoder, MELP, Speech compression, Multi-mode coder

### I. 서 론

현재 4 kbit/s 이상의 음성 부호화기는 CELP (Code

Excitation Linear Predictive) 코더가 주류를 이루고 있지만, 4 kbit/s 이하에서의 CELP 부호화 방법은 고음질의 음성을 재생하기 어렵기 때문에 다른 부호화 방법이 요구된다. 그래서 4 kbit/s 이하의 저 전송율에서는 STC<sup>[1]</sup> (Sinusoidal Transform Coding)나 MBE<sup>[2]</sup> (Multi-Band Excitation coding)와 같은 스펙트럼 영역에서의 음성 부호화 방식이 좀 더 나은 음질을 얻기 위한 한 방법으로 사용된다. 하모닉 코딩에 기반을 둔 이러한 스펙트럼 영역의 음성 부호화기들은 저 전송율에서 유성음의 주기적인 신호를 표현하기 위한 효과적인 방법이다. 그러나 하모닉 코딩은 음성의 시작점과 끝점에서의 유성

학생회원, \*\* 정회원, \*\*\* 평생회원, 충북대학교 전자공학과

(Department of Radiowave Engineering Chungbuk National University)

\*\*\*\* 정회원, 국방과학연구소

(Agency for Defense Development)

※ 본 연구는 국방과학 연구소 연구과제

(과제번호:03-06-03)지원으로 수행되었습니다.

접수일자: 2004년9월22일, 수정완료일: 2004년11월29일

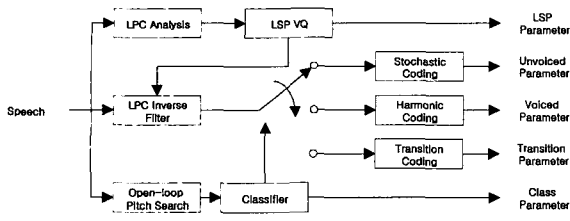


그림 1. 제안된 다중모드 저전송율 코더의 부호화기  
Fig. 1. Encoder of a proposed multi-mode low rate coder.

음과 무성음의 혼합신호(onset signal, offset signal), 이전 혹은 미래 프레임과의 연속적인 주기성을 보장하지 못하는 비 주기적 신호 등의 전이구간의 신호를 표현하기에는 적당하지 못한 모델이기 때문에 자연스러운 음질을 만들어 내기가 어렵다. 그래서 저 전송율에서 음질을 향상시키기 위한 하모닉 및 스토캐스틱 부호화의 혼합 기술을 사용한 MELP<sup>[3]</sup> (Mixed Excitation Linear Prediction)와 같은 새로운 음성 부호화기들이 제안되었고,<sup>[4]</sup> 그 혼합 방법은 유성음 세기 정도에 따라 정현파 및 잡음 발생정도를 달리하는 방법에 바탕을 둔다. 그러나 이러한 방법은 고정 대역분석에 의한 스펙트럼 왜곡을 발생시키고 전 대역에 걸친 정확한 잡음 스펙트럼 정보를 표현하기에는 불충분하다. 또한, 전이 음성에 대한 정확한 표현을 위한 고려도 불충분하다. 그러므로 음성 스펙트로그램으로부터 추출된 잡음 성분과 비주기적인 신호를 나타내기 위한 새로운 혼합 방법<sup>[5]</sup>이 시도되었다. 하지만 4 kbit/s이하의 저 전송율에서의 고려가 없었을 뿐만 아니라, 주파수축 합성 방법인 하모닉 모델만으로는 역시 잡음 스펙트럼 정보를 표현하기에는 불충분하다.

본 논문에서는 새로운 혼합 방법의 다중모드 부호화기와 하모닉-CELP 혼합 부호화 방법에 적합한 효율적인 전이구간 부호화 방법을 제안한다. 제안된 다중모드 부호화기는 유성음 프레임에 대해서는 정현파 모델을 이용하여 여기신호를 모델링하고 그 특성 파라미터들을 전송, 복원하는 방법을 사용하였으며, 무성음 프레임에 대해서는 256개의 배열 요소를 가지는 순환 코드북을 이용하여 여기신호를 모델링하였다. 마지막으로 제안된 하모닉-CELP 혼합 부호화 방법에 적합한 효율적인 전이구간 부호화 방법을 이용하여 전이구간 프레임에 대하여 여기신호를 부호화하였다. 제안된 전이구간 부호화 방법은 효율적인 파라미터 검출과 양자화 방법을 포함한다. 또한 하모닉-CELP 부호화 방법의 단점인 서로 다른 축(시간축과 주파수축)에서의 합성방법의 결함으

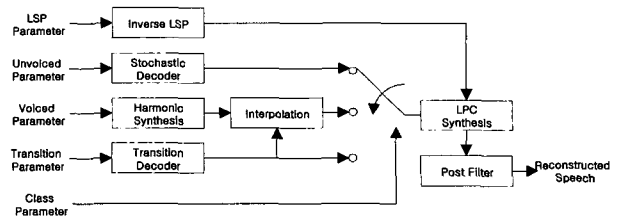


그림 2. 제안된 다중 모드 저전송율 코더의 복호화기  
Fig. 2. Decoder of a proposed multi-mode low rate coder.

로 인하여 유성음과 무성음의 경계점에서 발생하는 원음에 대한 합성음의 왜곡 현상을 효율적으로 보완하여 자연스런 결합이 가능하도록 한다. 제안된 방법은 2.4 kbit/s의 저 전송율의 부호화기에 적용하여 MELP 부호화기와 성능을 비교하였다.

본 논문은 총 VI장으로 구성되어 있다. II장에서는 다중모드 부호화기의 전체 구조와 세 가지 모드에 대하여 각각 적용한 부호화 방법에 대하여 설명하고, III장에서는 다중모드 부호화기의 비트 할당에 대하여 설명한다. IV장에서는 실험 결과에 의해 얻어진 파형을 비교 분석하며, V장에서는 제안된 방법을 이용한 부호화기와 동일한 전송율을 갖는 MELP 부호화기의 성능에 대하여 PESQ과 A/B 테스트를 통해 비교 평가를 한다. 마지막으로 VI장에서 결론을 맺는다.

## II. 다중모드 부호화기의 구조

다중모드 부호화기는 분류기에 의해 20ms의 프레임 길이마다 유/무성음 그리고 전이구간 음성으로 분류하여 각각의 음성의 특성에 따라 모델을 달리한다. 각 분류 코더의 입력 신호는 LPC 잔여신호가 된다. 이러한 다중모드 부호화기와 복호화기의 구조는 그림 1과 2에 도시하였다. 우선 입력된 음성은 선형예측(Linear Prediction) 분석을 하며, 여기서 얻은 LPC는 LSP(Line Spectrum Pairs)로 변환된 다음, 전형적 2단계 벡터양자화와 상호예측 벡터양자화 중 선택적으로 양자화 된다. 최종적으로 양자화 된 LSP계수를 보간하여 입력 음성신호를 역 LPC 필터링 시킴으로써 잔여 신호를 추출하게 된다.

추출된 잔여신호는 유/무성음과 전이구간을 판별하는 분류기에 의해 각 대응 코더에 의해 부호화된다. 분류 값과 이에 따른 각 코더의 파라미터 값 그리고 LSP 인덱스가 복호기에 전송되며, 복호기는 이 분류 값에 따라 유성음 구간일 경우 파형 합성기와 파형 보간 과정

을 거치며, 무성음 구간에서는 시간영역 합성과정을 거친다. 또한, 일반적으로 무성음 프레임과 유성음 프레임 사이에 존재하는 전이 구간에서는 시간 축에서 합성된 무성음과 주파수 축에서 합성하여 보간을 하는 유성음 사이의 자연스런 결합을 위한 데이터의 저장과 보간 과정을 거친다. 최종적으로는 LPC 합성필터와 후단여파기를 거쳐 복원 신호를 만들어 낸다.

1. 기존의 하모닉 부호화

다중모드 부호화기의 유성음 부분은 정현파 모델을 사용한다. 정현파 모델에서의 여기신호 표현은 다음과 같이 표현될 수 있다.<sup>[6]</sup>

$$s(n) = \sum_{k=1}^K A_k \cos(w_k n + \phi_k) \quad (1)$$

여기서,  $A_k$ 와  $\phi_k$ 은 주파수가  $w_k$ 인 정현파 성분에 대한 크기와 위상을 나타내며,  $K$ 는 정현파형의 개수를 나타낸다. 유성음 구간의 여기신호는 대부분의 음성 신호 정보가 하모닉 성분으로 이루어져있으므로, 적절한 스펙트럴 기본 모델을 이용하여 근사화 시킬 수 있다. 저전송을 부호화기의 특성상 비트 할당의 여유가 많지 않으므로 선형위상 합성 방법을 이용한 근사 모델을 사용하기 위하여 식 (2)의 합성 여기신호를 정의한다.

$$\hat{s}^l(n) = \sum_{k=1}^{K_l} A_k^l \cos(\theta_k^l(n)) \quad (2)$$

여기서,  $l$ 과  $K_l$ 은 프레임 번호와 각 프레임에서의 하모닉 개수를 나타낸다. 이 정현파 모델에서 복호화기에서 합성을 위해 전송 되어야 할 정보는  $l$ 번째 프레임의 하모닉 크기를 나타내는  $A_k^l$ 과 피치 각주파수를 나타내는  $w_0$ 이다. 하모닉 크기를  $A_k$ 를 구하기 위해 LPC 잔여신호는 헤밍윈도우에 의해 윈도우되고, 적절한 계산량, 복잡도, 주파수 해상도를 갖는 512-point FFT를 사용하여 원신호의 스펙트럼을 구하여 기준 모델을 구한다. 이 기준 모델을 이용하여 식 (3)의 값이 최소가 되는 하모닉 크기  $A_k$ 를 페 루프 검색 방법으로 결정한다.

$$e_k = \sum_{i=a_k}^{b_k} (X(i) - A_k B(i))^2 \quad (3)$$

여기서,  $X(i)$ 는 원신호의 스펙트럼 크기,  $B(i)$ 는 헤밍윈도우의 스펙트럴 크기를 뜻한다.  $a_k$ 와  $b_k$ 는  $k-1$

번째 하모닉 주파수와  $k+1$ 번째 하모닉 주파수에 해당하는 FFT의 경계 값이다. 여기서 구한 하모닉 크기들의 개수는 약 10에서 60개로 가변적이기 때문에 벡터 양자화하기 위해서는 오버 샘플링을 구현하는 차원 변환기를 필요로 하게 된다. 이때 사용된 가변 차원 양자화는 가변차원으로 추출된 하모닉 크기 값들에 DCT (Discrete Cosine Transform) 변환을 취하고, 양자화시 비정방형 변환 벡터 양자화 방법을 사용하여 양자화한다.

이와는 달리 위상은 전송되지 않는 정보이며, 식 (4)와 같이 복호화기에서 선형위상을 합성할 수 있다.<sup>[5]</sup>

$$\theta_k^l(n) = kw_0^{l-1}n + \frac{k(w_0^l - w_0^{l-1})}{2N}n^2 + \phi_k \quad (4)$$

여기서,  $N$ 은 프레임 길이를 나타내며,  $\phi_k$ 는  $k$ 번째 하모닉의 초기위상이며, 제안된 부호화기에서는 정규분포를 갖는 랜덤 변수를 사용한다. 이와 같은 선형위상의 합성이 가능한 이유는 인간의 청각 시스템은 위상의 연속성이 보존되는 동안 선형위상에 둔감하여, 부정확한 또는 완전히 판이한 이산위상을 허용할 수 있다. 이러한 지각적 특성은 저 전송을 코딩에 있어 하모닉 모델의 연속성에 대한 중요한 조건이 된다. 따라서, 합성 위상은 측정된 위상을 대체할 수 있게 된다.<sup>[7]</sup>

복호화기에서의 합성은 부호화기로부터 전송된 스펙트럴 파라미터와 피치 파라미터 값을 사용하여 수행하게 된다. 우선 기준 파형을 합성하기 위해, 스펙트럴 파라미터를 역 영자화 과정을 통해 하모닉 크기들을 추출한다. 그런 다음 식 (2)와 (4)에서 제시한 선형 위상 방법을 사용하여, 각 하모닉 크기들에 해당하는 위상 정보를 만들어낸 후, 기준 파형을 만들어 낸다. 일반적으로 8 kHz로 샘플링된 음성의 경우에 유성음의 경우 피치 주기는 120-point 미만이기 때문에 피치 주기를 표현하기 위한 기준파형은 128-point면 한 피치 주기에 대한 표현이 가능하다. 그러므로 적절한 계산량을 위해 128-point IFFT를 통해 기준 파형을 만들어 낸다. 이렇게 만들어진 기준 파형은 피치 정보를 포함하지 않은 상태이기 때문에 순환 형태로 재구성한 다음, 피치 주기로부터 얻은 식 (5)의 오버 샘플링 비율로 피치 변화를 고려하여 식 (6)과 같이 보간하고, 샘플링하여 최종 합성 여기신호를 얻어낸다.<sup>[8]</sup>

$$ov = \frac{256}{2T_p} = \frac{256/4}{T_p/2} = \frac{64}{K} \quad (5)$$

$$p_{ov}[n] = \sum_{i=0}^n \left( \frac{N-i}{N} ov^{i-1} + \frac{i}{N} ov \right) \quad (6)$$

여기서,  $ov$ 는 오버 샘플을,  $p_{ov}[n]$ 은 샘플링 위치를 나타내고,  $T_p$ 는 피치주기를 나타낸다.

순환 파형을 만들어 내는 과정 중 프레임간의 부드러운 연속성을 보장하기 위해 현 순환 파형의 시작 지점을 바꾸는 방법이 필요하게 된다. 이를 위해 식 (7)의 *offset* 값을 정의하며, 식 (7)을 이용하여 식 (8)과 (9)와 같이 각 프레임 순환 파형의 끝 지점을 한 개의 기준 파형의 구간 길이인 128로 맞추므로써 다음 프레임의 첫 지점과 자연스럽게 이어질 수 있도록 하게 된다.<sup>[8]</sup>

$$offset = 128 - mod(L - 128) \quad (7)$$

$$w^{l-1}(n) = w^{l-1}(mod(n, 128)) \quad (8)$$

$$w^l(n) = w^l(mod(offset + n, 128)) \quad (9)$$

여기서  $L$ 은  $N$ 개의 샘플을 복원시키기 위해 오버 샘플에 사용되는 데이터 개수이며,  $mod(x, y)$ 는  $x$ 를  $y$ 로 나눈 나머지 값을 나타낸다. 또한  $w^l(n)$ 은  $l$ 번째 순환 파형,  $w^{l-1}(n)$ 은  $l-1$ 번째 기준 파형을 나타내며, 식 (8)과 (9)에서의  $n = 0, \dots, L - 1$ 이다.

## 2. 일반적인 스토케스틱 부호화

다중모드 부호화기의 무성음 부분은 여기신호가 잡음과 같은 특성을 나타내기 때문에 256개의 배열 요소를 가지는 순환 코드북을 사용하고<sup>[9]</sup>, 분석/합성 방법을 이용하여 여기신호를 부호화한다. 이러한 순환 코드북의 사용은 메모리 크기를 줄일 수 있는 장점이 있으며, 왜곡 측정치가 최소가 되는 이득 모양 벡터를 찾아내는 방법을 사용한다.<sup>[10]</sup>

## 3. 하모닉-CELP 부호화에 적합한 전이구간 부호화

전이구간 모델은 하모닉 부호화 방법과 함께 주파수 축에서 합성하는 방식으로 제안되어왔다.<sup>[5]</sup> 하지만 하모닉 부호화 방식은 잡음과 같은 신호의 부호화에 방식에 적합하지 않다는 것은 널리 알려져 있다. 특히, 저 전송율에서 고정된 대역폭에 대한 하모닉 부호화 방법은 합성된 신호의 원음에 대한 왜곡을 발생시킨다. 하지만 잡음과 같은 신호에 성능이 좋은 CELP 부호화 방식의 사용은 하모닉 부호화 방법보다는 효율적이지만 서로 다른 축(시간축과 주파수축)에서의 합성방법의 결합으

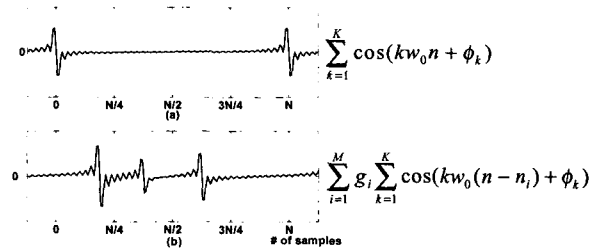


그림 3. 비주기적인 펄스의 순차적인 모델링

Fig. 3. Sequential modeling of an aperiodic pulse signal.

로 인하여 유성음과 무성음의 경계점에서 발생하는 원음에 대한 합성음의 왜곡 현상을 피할 수 없다. 본 절에서는 이를 보완하기 위한 효율적으로 전이구간 모델링 방법에 대한 상세한 설명을 한다. 또한 2.4 kbit/s의 전송율의 다중 모드 부호화기에 적용하기 위한 효율적인 파라미터 검출 방법과 양자화 방법에 대해서도 논한다. 마지막으로, 제안된 방식의 전이구간 부호화 방법의 분류를 위한 전이구간 검출 방법을 제안한다.

다중모드 부호화기의 전이구간 부호화 방법을 소개하기 전에, 음성의 시작점과 끝점에서의 유성음과 무성음의 혼합신호와 비주기적인 신호를 대표적인 전이구간 신호라 정의한다. 이러한 전이신호들은 부분적으로 펄스 신호가 존재하지만, 이 펄스 신호들은 대부분 주기적이지 않다. 저 전송율 부호화 방법에서 주기적인 신호에 대한 부호화는, 하모닉 모델이 좋은 성능을 내며, 잡음과 같은 신호에 대한 부호화는 스토케스틱 모델이 좋은 성능을 내지만 전이신호에 대한 부호화에는 적합하지 못하다. 그러므로 저 전송율 하모닉-CELP 부호화 방법에 적합하며 비주기적인 펄스신호의 모델링을 위한 효과적인 전이구간 모델링 방법을 제안한다.

비주기적인 신호를 하모닉 기반의 펄스신호로 모델링하기 위하여 전이구간 신호의 주기를 한 프레임의 길이  $N$ 이라고 가정하면, 식 (4)에서  $w_k^{l-1} = w_k^l = 2\pi k/N$ 이라고 할 수 있다. 그러면 식 (4)는 식 (10)과 같이 정리 될 수 있다.

$$\theta_k^l(n) = \frac{2\pi kn}{N} + \phi_k \quad (10)$$

주어진 식 (2)와 (10)에 의해 하나의 프레임에 하나의 펄스가 존재하는 그림 3. (a)와 같은 신호를 만들어 낼 수 있다. 이 펄스는  $n = 0$ 인 위치에 존재하며 주기는  $N$ 이다. 또한  $\phi_k$ 는 펄스의 모양에 영향을 끼치는 인자이다. 예를 들어  $\phi_k$ 가 0이라면 펄스의 모양은 우함수의 형태를 가지며,  $\phi_k$ 가  $-\pi/2$ 이라면 펄스의 모양은 기함

수의 형태를 가진다. 그림 3. (a)에서는  $\phi_k = \pi/3$  일때의 예를 들고 있다. 한 프레임 내에 하나의 펄스만 존재하는 그림 3. (a)와 같은 신호는 프레임내에 어느 곳이든 쉬프트 파라미터  $n_i$ 를 사용하여 프레임내 어떠한 곳으로도 이동이 가능하다. 결론적으로,  $M$ 개의 펄스 이득과  $M$  개의 쉬프트 파라미터를 갖는 신호들의 합으로 표현된 그림 3. (b)와 같은 신호로 모델링이 가능하다.

그림 3의 예에서는 전이구간과 같이 비주기적인 펄스 신호를 하모닉 기반의 모델을 이용하여 모델링이 가능하다는 것을 보여주고 있다.

전이구간 모델은 디코더에서 식 (11)와 (12)에 의해 정의된 여기신호를 이용하여 합성한다.

$$\hat{s}^l(n) = \sum_{i=1}^M g_i \sum_{k=1}^{K_i} A_k^l \cos(\theta_k^l(n, n_i)) \quad (11)$$

$$\theta_k^l(n, n_i) = \frac{2\pi k(n - n_i)}{N} + \phi_k \quad (12)$$

여기서  $M$ 은 펄스의 개수,  $n_i$ 는 쉬프트 파라미터,  $g_i$ 는 이득 파라미터이다. 또한 위상 파라미터  $\phi_k$ 는 펄스의 모양에 영향을 끼치는 점에서는 일반적인 정현파 모델의 위상파라미터와는 구분이 되어야 할 것이다.

본 논문에서 제안한 다중 모드 부호화기는 협대역 부호화기로서 약 300~3400 Hz를 대상으로 부호화를 실시하고, 프레임 길이와 같은 기본 피치 주기 ( $N=160$ )는 고정 되어 있으므로  $K_i$ 은 68을 사용하였고, 펄수의 개수  $M$ 은 4개를 찾았다.

가. 파라미터 검출

전이구간 모델은 일반적인 정현파 모델에 추가적으로 이득 파라미터와 쉬프트 파라미터가 존재한다. 같은 전송율을 가지는 부호화기에서 추가적인 정보가 존재한다는 것은 기존의 파라미터의 손실을 감수해야한다는 것을 의미한다. 그래서 전이 구간 모델은 하모닉 모델에서 전송하는 피치주파수에 대한 정보를 전송하지 않고 고정된 피치주파수인  $w_k = 2\pi k N$ 를 사용한다.

또한, 여기신호의 모든 이득 파라미터  $g_i$ 에 동일한 스펙트럼 포락선(spectral envelope)  $A_k$ 를 사용하기 위하여, 프레임 내에서 스펙트럼 크기가 천천히 변한다고 가정한다. 그리고 스펙트럼 포락선  $A_k$ 는 인간 청각 신경의 인지적 특성을 고려하여, 임계대역(critical band)을 가칭주파수 안에서 식 (13)와 같은 바크(Bark) 대역으로 표현이

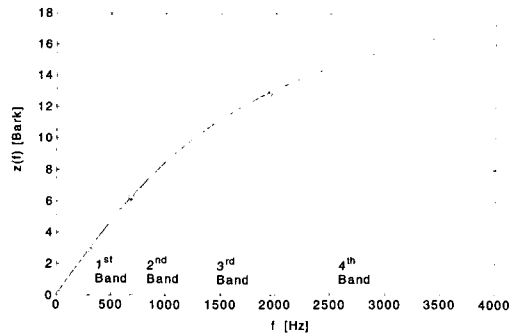


그림 4. 4개의 대역으로 나눈 바크대역  
Fig. 4. The divided Bark spectrum into four band.

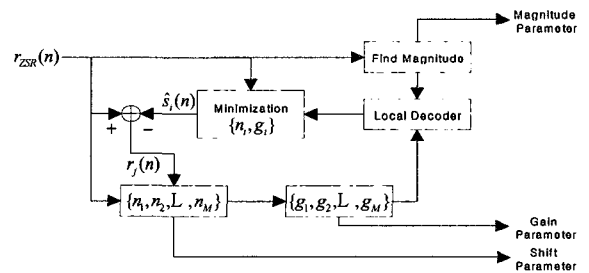


그림 5. 전이구간 모델링 블록도  
Fig. 5. The block diagram of the transition frame modeling.

가능하다.<sup>[11]</sup> 이러한 바크 대역은 24~26개의 대역으로 나뉜다. 그러나 제안한 부호화 방법은 협대역에서 적용하기 위한 방법이므로 본 논문에서는 약 3~16 Bark까지 만을 고려하여, 그림 4와 같이 협대역에서 4개의 대역을 나누어 각 대역별로 스펙트럼 포락선의 평균값을 추출하여 벡터 양자화를 하였다. 물론 좀 더 많은 대역으로 나누어 양자화를 한다면 약간의 성능의 향상은 있겠지만 2.4 kbit/s의 저 전송율을 감안하여 펄스의 위치와 이득에 좀더 많은 비트를 할당하는 것이 전체적으로 더 좋은 성능을 내기 때문에, 4비트 벡터 양자화를 하는 하모닉 크기는 4개의 대역으로 나누었다.

$$z(f) = 13 \arctan(0.00076f) + 3.5 \arctan\left(\left(\frac{f}{7500}\right)^2\right) \quad [Bark] \quad (13)$$

위상 파라미터  $\phi_k$ 는 펄스의 모양을 좌우한다. 이 파라미터는 다른 파라미터에 비해 비교적 중요도가 떨어지므로 전송을 하지는 않는다. 그러므로 위상 파라미터는 정규분포를 갖는 랜덤 변수를 사용하였다.

전이구간 모델에서는 기존의 하모닉 모델의 파라미터는 중요도를 낮추는 반면 쉬프트 파라미터와 이득 파라미터의 중요도는 높였다. 그림 5는 전이구간에서의 파라미터 예측 방법을 도시하였다. 전이 프레임에서의 목적 신호  $r_{ZSR}(n)$ 은 LPC 잔여신호의 ZSR(Zero

표 1. 쉬프트 파라미터  $n_1$ 의 위치

Table 1. location of a shift parameter  $n_1$ .

	location of a shift parameter	# of bits
pulse 1	0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155	5bit

State Response)이며,  $\hat{s}_i(n)$ 과  $r_j(n)$ 은 식 (14)<sup>[5]</sup>와 (15)을 따른다.

$$\hat{s}_i(n) = g_i \sum_{k=1}^K A_k \cos(\theta_k(n, n_i)) \quad (14)$$

$$r_j(n) = r_{ZSR}(n) - \sum_{i=1}^j \hat{s}_i(n) \quad (15)$$

여기서  $M \geq j \geq i$ 이며,  $r_j(n)$ 은  $j$ 개의 누적된 펄스 신호와 LPC 잔여신호의 ZSR의 차에 의해 생기는 잔차 신호이다. 그리고  $r_j(n)$ 은 반복적인 과정을 통해 펄스 이득과 위치를 찾기 위한 목적 신호가 된다.  $M$ 개의 쉬프트 파라미터와 이득 파라미터를 추출하기 위해서 펄스의 위치  $n_i$ 의 초기치는  $r_{i-1}(n)$ 에서 찾는다. 또한, 첫 번째 펄스의 위치를 찾기 위한  $r_0(n) = r_{ZSR}(n)$ 이 된다.  $r_{i-1}(n)$ 에서 검색된 쉬프트 파라미터를 이용하여, 식 (16)을 이용하여  $g_i$ 의 초기치를 찾는다.

$$g_i = r_{i-1}[n] - \hat{s}_i[n] \quad (16)$$

찾아진 초기  $n_i$ ,  $g_i$ 는 지역 디코더에서 합성을 하며, 이것은 다시 식 (17)의 MSE(Mean Square Error)에 의해 최적화 된다.

$$E_i = \sum_{n=0}^{N-1} [r_{ZSR}(n) - g_i \sum_{k=1}^K A_k \cos(\theta_k(n, n_i))]^2 \quad (17)$$

여기서  $E_i$ 는  $i$ 번째 반복에 의해 생기는 MSE이다.  $M$ 번의 반복에 의해 모든  $n_i$ ,  $g_i$ 가 찾아지면 양자화를 위해 분석/합성 방법에 의해 다시 한번 최적화를 거쳐 최적의 이득과 위치를 검색한다.

전이 모델의 첫 번째 펄스  $n_1$ 는 표 1을 참고하여 검색하게 된다. 표 1에 의해 찾아진  $n_1$ 는 다시 1bit를 추가로 할당하여  $\pm 1$ 의 위치를 다시 검색하여 최적화를 한다. 그 다음의 펄스의 위치들은 첫 번째 펄스를 기준으로 오른쪽 혹은 왼쪽, 순방향 혹은 역방향, 홀수 혹은

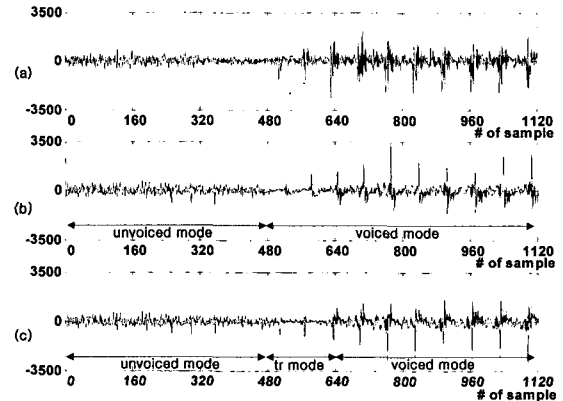


그림 6. 전이구간 모델에 의한 프레임간 결합  
Fig. 6. The combination of the frames by the transitional modeling.

짝수의 인자를 가지고 최적의  $n_i$ 를 찾게 된다. 또한, 최적화과정을 거친 이득 파라미터  $g_i$ 는  $\{g_1, g_2, \dots, g_M\}$ 의 평균을 제거한 값을 로그 도메인에서 벡터 양자화를 한다.

나. 효율적인 모드간 결합을 위한 전이구간 부호화

하모닉-CELP 혼합 부호화 방법은 잡음과 같은 특성을 갖는 무성음 프레임에서는 CELP 부호화 방법을 이용하여 시간축에서 합성을 하며, 주기적인 신호의 특성을 갖는 유성음 프레임에서는 하모닉 부호화 방법을 이용하여 주파수축에서 합성을 하는 방법이다. 그러나 이렇게 서로 특성이 다른 시간축 합성과 주파수축 합성에 의한 방법은 합성된 여기 신호의 결합 시에 서로 자연스럽게 이어지게 하지 못한다. 특히 2.4 kbit/s 부호화기의 특성상 유성음에 대한 주파수축 합성은 위상 정보 없이 선형 위상에 의한 합성을 하기 때문에 스토케스틱 모델을 사용한 이전 프레임의 지연 정보들을 이용할 수 없어 독립적인 합성을 하게 된다. 하지만 제안된 전이구간 부호화 방식은 하모닉 기반의 모델이지만 시간축에서 여기신호를 합성을 하기 때문에 CELP 여기신호와 시간축에서 자연스럽게 결합이 되며, 하모닉 여기신호와 상호 상관 값과 위상에 대한 정보를 이용하여 자연스럽게 결합이 가능하다.

그림 6. (a)는 목적 신호이며 (b)는 하모닉-CELP 부호화 방식에 의해 합성된 여기신호, (c)는 제안된 전이구간 부호화 방법을 이용하여 다중모드 부호화 방식에 의한 합성 여기신호이다. (b)와 같이 무성음과 유성음만의 모드에 의한 부호화기는 첫 번째 유성음 프레임은 하모닉 크기의 보간을 할 수 없으므로 삼각 윈도우에

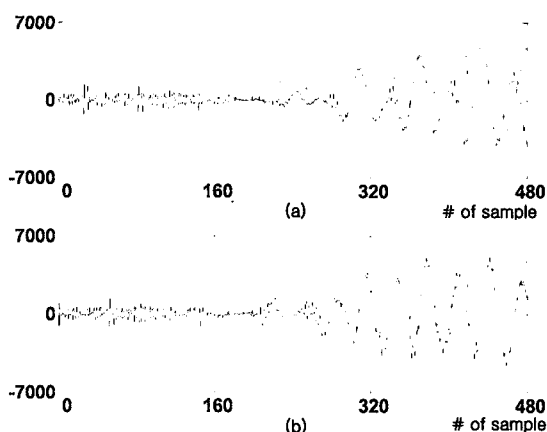


그림 7. 전이구간 모델의 도입에 따르는 합성신호 비교  
Fig. 7. Compare synthesized signal according to transient coding.

의한 유성음 합성과 이전 스토케스틱 여기신호를 이용한 추가적인 보상을 하게 된다. 그래서 첫 번째 유성음 프레임은 급격하게 변하는 펄스의 크기를 잘 반영하지 못할 뿐만 아니라 펄스의 위치 또한 제대로 반영하지 못하는 단점이 있다. 하지만 그림 6. (c)에서는 제안된 전이구간 모델의 추가로 전이구간 여기신호는 원본 신호의 펄스의 위치와 크기를 잘 반영하고 있다. 또한 이 전이구간 모델은 하모닉 기반의 여기신호 합성을 하기 때문에 다음 프레임의 유성음 합성을 위한 정확한 위상 정보의 전달이 가능하다. 이 지연에 대한 추가적인 위상정보는 이전 전이구간 프레임과 현재 유성음 프레임의 여기신호 간의 상호상관 값이 최대가 되는 값과 식 (18)를 이용하여 구할 수 있다.

$$\phi_k = w_k C_{\max} \quad (18)$$

여기서  $C_{\max}$ 는 이전 전이프레임의 여기신호와 현재 유성음의 여기신호의 최대 상호상관 값을 갖는 거리로써 정수 값만을 가진다.

결론적으로 2.4 kbit/s의 부호화기에서 전이구간 모델의 도입으로 추가적인 위상 정보의 전송 없이도 서로 다른 모델을 사용하는 무성음 프레임과 유성음 프레임 간의 자연스러운 결합이 가능하다. 이에 대한 결과를 그림 7에 보여주고 있다.

그림 7의 (a)와 (b)에서 합성된 음성 신호는 실선, 원본 음성은 점선으로 표시하여 3개의 프레임에 대하여 하모닉-CELP 부호화에 의한 합성 신호와 다중모드 부호화에 의한 합성신호를 비교하고 있다. 처음 프레임은 모두 CELP 부호화를 하였으며, 두 번째 프레임은 (a)에서는 하모닉 부호화 방법, (b)에서는 전이구간 부호화

방법을 사용하였다. 마지막 프레임에 대해서는 모두 하모닉 부호화 방법을 사용하였다. (a)의 두 번째 프레임에서 하모닉 부호화 방법의 합성은 정규분포를 갖는 랜덤 위상을 사용하기 때문에 원본음성에 대해서 모든 프레임에 일정한 지연을 가지도록 합성을 하는 것이 불가능하다. 하지만 (b)에서는 두 번째 프레임에 전이구간 부호화 방법을 사용하였기 때문에 비교적 정확한 지연 정보를 가지고 합성이 되며 이전 프레임인 CELP 부호화 방법과도 잘 결합이 된다. 또한 다음 프레임의 하모닉 부호화를 위한 지연 정보는 제안된 식 (18)에 의해 인코더로부터 추가 정보 없이 디코더에서 독립적으로 생성이 가능하며, 이 정보를 이어지는 프레임으로 전달이 가능하다. 결론적으로 제안한 방법에 의하여 전이구간 프레임 이후의 유성음 프레임에서 하모닉 부호화 방법에 의해 합성된 신호들도 이전 프레임과 연계하여 원본신호에 대하여 일정한 지연을 가지며 합성된다.

#### 다. 전이구간 검출

모드 선택 파라미터는 매 프레임 마다 결정된다. 첫 번째로 합성된 스펙트럼과 원본 스펙트럼의 유사성 및 신호의 파워 값, LPC 잔여신호의 파워로 정규화된 최대 자기 상관 값, 영 교차율(Zero Crossing Rate) 값을 사용하여 유성음과 무성음을 결정하고,<sup>[12]</sup> 두 번째로 전이구간에 대한 모드 선택이 이루어진다. 전이구간 검색에는 프레임내의 에너지 변화를 나타내는 에너지 비율과 프레임간의 주기적인 성격의 변화를 나타내는 이전 프레임과의 피치주기의 변화를 사용하여 결정한다. 식 (21)에서 전이구간 검색에 사용한 에너지 비율을 정의하고 있다. 프레임의 중간을 기준으로 LPC 잔여신호의 ZSR에 대한 좌우 에너지 비율을 나타내며,  $ER$ 이 1에 가까울수록 좌우 에너지의 비율이 비슷하며, 0에 가까울수록 에너지의 변화가 심하다는 것을 나타낸다. 이것은 모델링 하려는 전이신호가 좌우 에너지 비율의 편차가 심한 신호의 검출을 목적으로 하기 때문이다.

$$E_1 = \sum_{n=0}^{(N/2)-1} r_{ZSR}^2(n) \quad (19)$$

$$E_2 = \sum_{n=N/2}^{N-1} r_{ZSR}^2(n) \quad (20)$$

$$ER = 1 - \frac{E_1 - E_2}{E_1 + E_2} \quad (21)$$

프레임간의 피치 주기의 변화는 유성음과 무성음이 만나는 지점과 비주기적인 신호가 존재하는 곳에서도

심하게 나타난다. 본 논문에서는 유성음에서는 현재 프레임 음성의 피치와 이전 프레임 음성의 피치와 유사성을 유지하여야 한다는 Inmarsat 표준<sup>[13]</sup>을 응용하여, 폐구간 피치 검색에 의해 구해진 현재 프레임의 피치 변화량이 다음의 식 (22)과 (23)를 만족하는 프레임을 선택한다.

$$P_c - 0.8P_p \leq P_d \tag{22}$$

$$P_c - 1.2P_p \geq P_d \tag{23}$$

여기서  $P_p$ 는 이전 프레임의 피치,  $P_c$ 는 현재 프레임의 피치,  $P_d$ 는 피치의 변화량이다.

### III. 비트 할당

양자화 과정은 크게 LSP 및 여기신호 양자화 과정으로 구성된다. 우선 LSP 양자화 과정은 3 단계 분할 벡터 양자화(3-Stage Split Vector Quantization)를 적용하였다. 하지만 3단계 양자화는 모든 모드에 적용이 되는 것이 아니라 무성음 모드에만 적용이 되며, 나머지 모드에 대하여는 2단계 분할 벡터 양자화를 적용하여 16 비트를 할당하였고, 모드 선택에는 2 비트를 할당하였다. 따라서 총 48 비트중 30비트가 여기신호 표현을 위해 할당된다. 3가지 모드에 따라 여기신호의 비트 할당이 달라진다. 무성음 모드에서는 각 프레임을 2 개의 부 프레임으로 나누어 각 부 프레임 마다 코드북의 형태 8 비트와 이득 4 비트씩을 할당하였다. 유성음 모드에서는 하모닉 크기는 2 단계 DCT 양자화를 사용하였으며, 이에 대한 이득은 5 비트를 할당하였고, 잔차 하모닉을 다시 8 비트 벡터 양자화를 하였다. 또한, 피치는 7 비트를 할당하여 양자화를 하였다. 마지막으로, 전이구간에서는 쉬프트 파라미터의 첫 번째 펄스에는 6 비트 그리고 나머지 3개의 펄스에는 각각 4 비트를 18 비트를 할당하였다. 이득은 8 비트 벡터 양자화를 하였고, 변화가 심하지 않고 바크대역에 의해 4 개의 대역으로 분할된 하모닉 크기에는 4 비트를 할당하였다.

### IV. 실험

실험의 결과를 보이기 위하여, 제안된 부호화기에 사용된 테스트 음성은 “시골의 밤하늘에 반딧불이 반짝반짝 빛납니다.”라는 3.5 초의 길이를 갖는 8000 Hz로 샘플링 된 협대역 남자 음성을 테스트 음성으로 사용하여

표 2. 모드별 비트 할당

Table 2. Bit allocation of modes.

전송율		2.4 kbit/s		
10차 LSP 양자화 값		1차 상태 : 6비트		
		2차 상태 : 10비트		
V/UV /S/TR		2 비트		
여기 신호		무성음		유성음
	LSP	6비트		
	10ms	형태 : 8비트	피치 : 7비트	하모닉 크기 : 4비트
		이득 : 4비트	DCT1 : 5비트	
	10ms	형태 : 8비트	DCT2 : 5비트	이득 : 8비트
이득 : 4비트		이득 : 5비트		
		잔차 : 8비트		
총 비트수		48비트	48비트	48비트

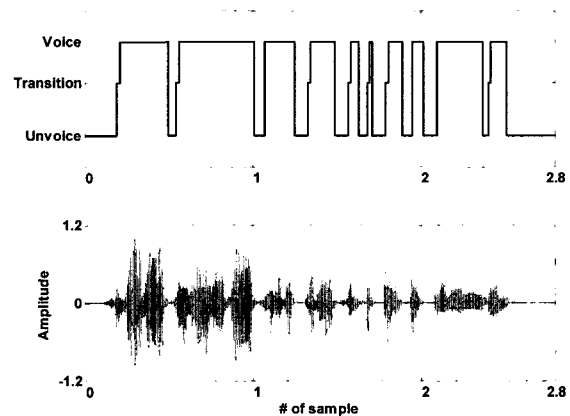


그림 8. 입력 음성에 대한 모드 선택

Fig. 8. Mode selection for input speech.

분석 후 합성을 시행하였다. 제안된 부호화기의 실험 결과는 크게 두 가지로 요약을 하였다. 전체 테스트 음성에 대하여 MELP 부호화기와 제안된 부호화기의 파형을 비교 분석하였고, 부분적인 전이구간에 대하여 전이모형을 사용하였을 때와 사용하지 않았을 때의 파형을 비교 분석하였다.

그림 8에서는 입력 음성에 대한 모드 선택 결과를 보여주고 있다. 전이구간의 모드 선택에 사용한 알고리즘의 문턱 값들은 실험적인 결과로 얻었다. 실험적인 문턱 값이기 때문에, 모든 음성에 대하여 항상 정확한 대표적인 전이 음성만을 검출하지는 못한다. 하지만 테스트 음성에서는 II장에서 설명한 대표적인 전이구간 음성에 대한 검출이 비교적 정확하게 이루어졌으며, 모드 결합을 위한 무성음과 유성음 사이의 프레임을 주로 전이구간으로 선택하여 전체적인 음질의 향상을 이루었다. 테스트 음성의 총 176프레임 중에서 유성음 102 프레임, 무성음 67 프레임, 전이구간 7프레임을 차지한다.



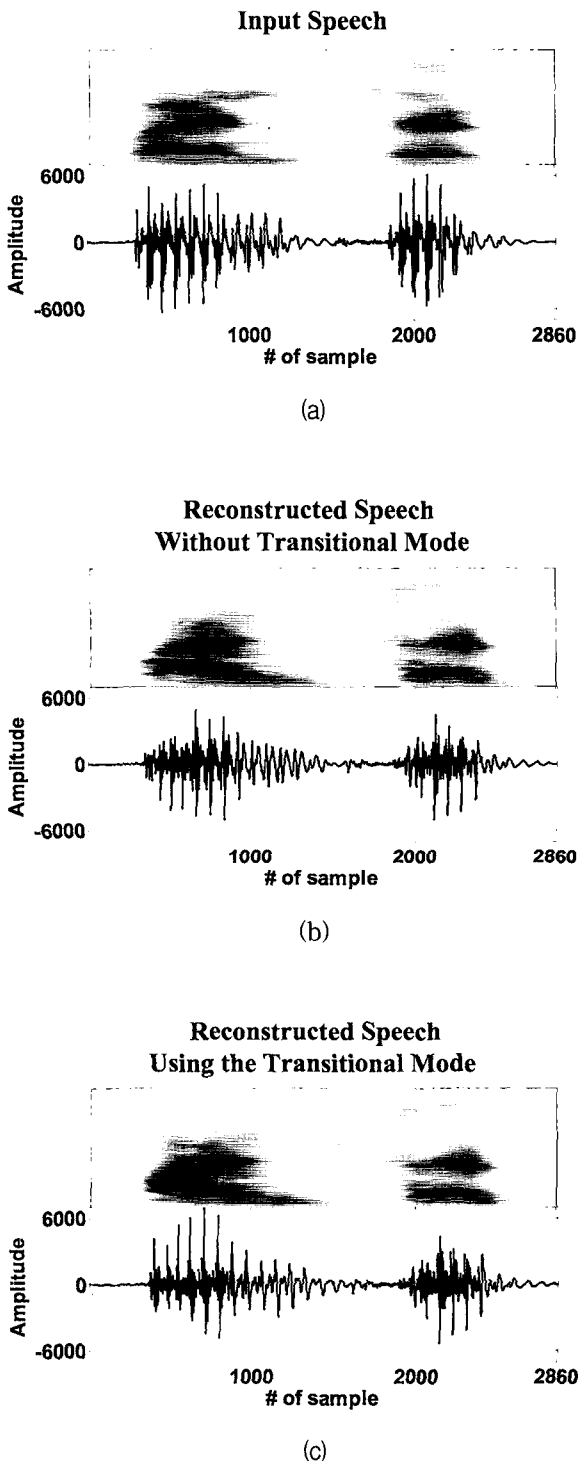


그림 9. 전이 구간을 포함한 복원된 음성 비교  
 Fig. 9. Compare two reconstructed speech including transitional frame.

파형의 좀더 자세한 관찰을 위하여 전이프레임(2400~24160 샘플)을 포함하고 있는 23680~26540 샘플을 확대한 그림 9을 도시하였다.

그림 9에서는 테스트 음성의 가장 마지막 전이구간이 있는 일부분을 확대하여 입력 음성 (a), 제안된 코더에서 전이구간이 없이 유/무성음만을 사용하여 합성하였을 때의 음성 (b), 마지막으로 다중모드를 사용한 제안된 부호화 방법에 의한 합성 음성 (c)를 비교하고 있다. 그림 9에서 320~480 샘플(그림 8에서 24000~24160인 부분)의 음성이 시작하는 부분이 전이구간으로 선택된 프레임이다. 입력 음성의 에너지가 급변하는 이러한 신호는 (b)에서 유성음으로 선택이 되어 모델링을 하게 된다. 이전 프레임과의 보간을 하는 하모닉 모델의 특성상 급변하는 펄스를 잘 따라가지 못하는 것을 시간 축에서 관찰 할 수 있으며, 주파수 축에서 선 반향(Pre-echo) 왜곡이 특히 낮은 주파수 축에서 일어나고 있음이 확인된다. 하지만 그림 (c)에서는 전이구간의 낮은 주파수에서의 선 반향 현상이 감소하였음을 관찰할 수 있으며 높은 주파수에서는 원본 신호에 충실하게 노이즈 성분도 또한 포함하고 있다. 시간 축에서의 급변하는 펄스 또한 잘 표현하고 있다. 단지 전이 프레임 다음에 연속된 유성음 프레임에서 하모닉 모델의 특성상 이전 전이구간의 랜덤 위상을 저장하여 선형위상 합성을 하므로 시간 축에서의 파형의 모양이 반전이 되었지만 이것은 펄스의 위치에 대한 변화가 아니므로 위상의 변화에 둔감한 인간의 청각 특성을 고려할 때 음질 저하에 미치는 영향은 적다.

비록 전이구간 음성은 전체 176개의 프레임 중에 7개의 프레임으로 약 4%를 차지하지만, 전체 음질에는 항상에는 그 이상의 영향을 미친다.

그림 10에서는 제안된 다중모드 코더와 MELP 코더를 테스트 음성을 통한 단편적인 비교가 가능하다. 시간 축에서의 신호의 비교로는 음질의 영향을 알아보기 힘들다는 면에서 주파수 축에서의 분석을 중심으로 한다. (b)와 (c)를 비교해보면 주파수 축에서의 전체적인 스펙트로그램은 유사하지만, 부분적으로 약 4000, 9200, 14200 샘플 근처에서 MELP 코더에 의해 복원된 음성은 급격히 에너지가 낮아진다. 이는 합성음성의 자연스러움을 떨어트리고 기계적인 음질을 나타내는 것에 기여를 하게된다. 또한, 프레임간의 혹은 프레임내의 하모닉 포락선이 원음의 변화를 잘 따라가지 못하고, 상수값을 유지하는 것도 관찰된다.

### V. 성능 평가

제안된 부호화기의 성능 평가를 위하여 PESQ

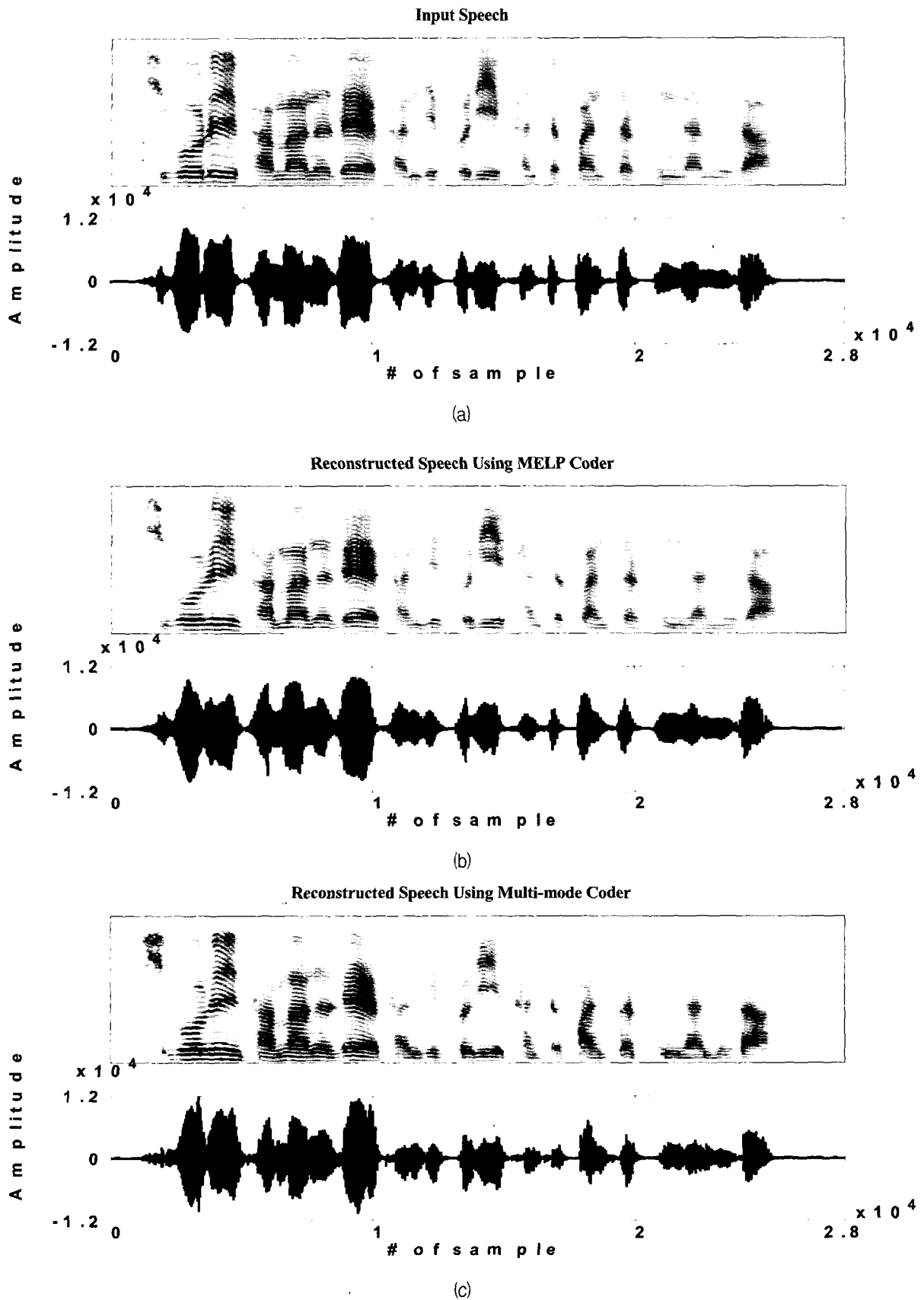


그림 10. 제안된 코더와 MELP 코더의 복원된 신호의 비교  
 Fig. 10. Compare the reconstructed signal of the proposed coder with MELP coder.

표 3. PESQ MOS 시험 결과  
Table 3. PESQ MOS test results.

분류	여자	남자	전체
2.4 kbit/s MELP	3.100	3.382	3.241
2.4 kbit/s without transitional mode	3.230	3.337	3.284
2.4 kbit/s proposed coder	3.319	3.443	3.381

표 4. A/B 테스트 시험 결과  
Table 4. A/B test results.

분류	여자	남자	전체
2.4 kbit/s MELP	32%	47%	39.5%
2.4 kbit/s proposed coder	68%	53%	60.5%

(Perceptual Evaluation of Speech Quality)<sup>[14]</sup>를 이용한 객관적 음질 평가를 하였다. 표 3은 2.4 kbit/s의 동일한 전송율을 갖는 MELP 부호화기, 제안된 부호화기에서 유/무성음의 모드만을 사용하였을 때의 부호화기, 마지막으로 제안된 다중모드 부호화기에 대한 음질 시험 결과를 보여준다. 시험 결과에서 보여주듯 제안된 부호화기는 같은 전송율에서 MELP 부호화기보다 평균MOS 0.14의 성능 향상을 나타내고 있다. 표 4는 MELP부호화기와 제안된 다중모드 부호화기와의 A/B 테스트 결과를 보여주고 있다. A/B 테스트는 발화자가 다른 한국어 문장 남자와 여자 각각 10문장을 사용하여, 15명의 훈련되지 않은 청취자를 대상으로 시행하였다. 이 결과에서도 PESQ와 유사한 상관성을 보이며 여성 음성에서 특별히 좋은 성능을 나타내었다. 특히, MELP 부호화기에 비해서 남자 음성보다 여자음성에서 더 좋은 성능을 나타내는 것은 여자 음성의 피치주기가 짧기 때문으로 생각된다. 왜냐하면, 피치 주기가 짧으면 하모닉 개수가 더 적어지고 하모닉 크기의 잔차 양자화에 의해 더 정확하게 표현되는 대역은 더 넓어지기 때문이다. 또한, 전이구간 모델에 의한 다중모드 부호화기의 성능은 유/무성음 모드만 사용하였을 때에 비해 고른 성능 향상을 보인다. 일반적으로 실시간 구현이 가능한 하모닉 부호화, CELP 부호화와 CPU time을 이용하여 상대적인 계산량의 비교를 통하여 전이구간 부호화 방법도 실시간 구현이 가능한 것으로 관찰된다.

## VI. 결 론

본 논문에서는 군사용 통신 등에서 이용되고 있는 2.4 kbit/s의 다중모드 부호화기와 하모닉-CELP 혼합 부호화 방법에 적합한 효율적인 전이구간 부호화 방법을 제안하였다. 다중모드 부호화 방법은 일반적으로 사

용하는 하모닉 부호화 방법과, CELP 부호화 방법을 사용하였으며, 이에 적합한 제안된 전이구간 부호화 방법을 적용하였다. 이러한 다중 모드 부호화 방법의 구조적인 장점은 시간 축 모델인 스토캐스틱 모델과 주파수 축 모델인 하모닉 모델의 자연스런 혼합을 가능하게 해준다. 모델링 측면에서는 기존의 모델로 표현하기 어려운 전이구간에 대하여 제안된 부호화 방법을 사용함으로써 음질이 향상되었다. 제안된 부호화기의 객관적 음질 평가의 일환으로 PESQ MOS 시험과 A/B 테스트를 하였다. 시험 결과로 MOS 0.14의 음질 향상과 21% point 더 높은 선호도를 보이는 것으로 확인하였다. 제안된 부호화 방법은 전이구간의 효율적인 적용으로 하모닉 코더의 단점을 보완함으로써 원본 음성의 좀 더 정확한 표현이 가능하고, 유성음 프레임과의 위상 연속성을 유지하여 자연스러운 음질 보장과 더불어 더욱 명료한 음질의 향상을 이루었다.

## 참 고 문 헌

- [1] T. Quatieri, R. McAulay, "Speech transformations based on a sinusoidal representation.", IEEE Transactions on Signal Processing, Vol. 34, pp. 1449-1464, Dec. 1986.
- [2] J. C. Hardwick, J. S. Lim, "A 4.8kbps multi-band excitation speech coder", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 1, pp. 374-377, 1988.
- [3] A. McCree, Truong Kwan, E. B. George, T. P. Barnwell, V. Viswanathan, "A 2.4 kbit/s MELP coder candidate for the new U. S. Federal Standard", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 1 pp. 200-203, May 1996.
- [4] C. Laflamme, R. Matmti, J. P. Adoul, "Harmonic-stochastic excitation (HSX) speech coding below 4 kbit/s", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 1, pp. 204-207, May 1996
- [5] Chunyan Li, V. Cuperman, "Enhanced harmonic coding of speech with frequency domain transition modeling ", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 2, pp. 581-584, May 1998.
- [6] Lajos Hanzo, C. Somerville, J. Woodard, Voice Compression and Communications, John Wiley & Sons, Inc., Publishers., pp. 531-564, 2001.
- [7] E. Shlomot, V. Cuperman, A. Gersho, "Combined harmonic and waveform coding of speech at low

bit rates", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 2, pp. 585-588, May 1998.

[8] M. Nishiguchi, J. Matsumoto, "Harmonic and noise coding of LPC residuals with classified vector quantization", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 1, pp. 484-487, May 1995.

[9] L. Yang, T. R. Fischer, S. Kang, I. Lee, "Codebook optimization in variable-rate CELP coders with sparse codebooks", IEEE Workshop on Speech Coding Proceedings, pp. 85-86, Sep. 1995.

[10] A. M. Kondoz, Digital Speech, John Wiley & Sons, Inc., Publishers., pp. 174-209, 1994.

[11] W. B. Kleijn, K. K. Paliwal, Speech Coding and Synthesis, Elsevier Science Publishers, pp. 399-431, 1995.

[12] M. Nishiguchi, A. Inoue, Y. Maeda, J. Matsumoto, "Parametric speech coding-HVXC at 2.0-4.0 kbps", IEEE Workshop on Speech Coding Proceedings, pp. 84-86, June 1999.

[13] DVS1. Inmarsat-M Voice Codec. Issue 3.0 ed., Augst 1991.

[14] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ) : An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codec", Feb. 2001.

저 자 소 개



안 영 옥(학생회원)  
 2003년 충북대학교 반도체공학과 졸업.  
 2005년 충북대학교 전자공학과 석사 졸업.  
 <주관심분야 : 음성/오디오 부호화, DSP, 적응필터>



김 종 학(정회원)  
 1998년 충북대학교 전자공학과 졸업.  
 2000년 충북대학교 전자공학과 석사 졸업.  
 2002년 충북대학교 전자공학과 박사 수료.

2002년~현재 (주) C&S Technology 선임연구원  
 <주관심분야 : 음성/오디오 부호화, 음성 압축, DSP>



이 인 성(평생회원)  
 1983년 연세대학교 전자공학과 졸업.  
 1985년 연세대학교 전자공학과 석사 졸업.  
 1992년 Texas A&M University 전기공학과 박사 졸업.

1986년~1987년 한국통신 연구개발단 전임연구원.  
 1993년~1995년 한국전자통신연구원 이동통신기술연구단 선임연구원.  
 1995년~현재 충북대학교 전기전자공학부 교수.  
 <주관심분야 : 음성/영상 신호 압축, 이동통신, 적응필터>



권 오 주(정회원)  
 1989년 2월 경북대학교 전자공학과 학사  
 1993년 2월 경북대학교 전자공학과 석사  
 2002년 2월 경북대학교 전자공학과 박사

1997년 9월~1999년 1월 Matra Marconi Space UK 연구원  
 1993년 3월~현재 국방과학연구소 선임연구원  
 <주관심분야 : 위성통신, 디지털통신신호처리, 다중반송파신호처리, SDR, OFDM>



배 문 관(정회원)  
 1990년 2월 경북대학교 전자공학과 학사  
 1992년 2월 경북대학교 전자공학과 석사  
 1993년 3월~현재 국방과학연구소 선임연구원

<주관심분야 : 멀티미디어 통신시스템, 위성통신, 음성 및 영상부호화 등>