

# 혼합 다중대역 여기모델에 기반한 저 전송률 음성 부호화기의 설계

## Design of a Low Bit-rate Speech Coder Based on Mixed Multi-band Excitation Model

한 우 진\*, 오 영 환\*  
(Woo-Jin Han\*, Yung-Hwan Oh\*)

\*한국과학기술원 전산학과  
(접수일자: 2002년 1월 2일; 채택일자: 2002년 4월 3일)

다중대역 여기부호화 (MBE: multi-band excitation) 음성 부호화기는 고조파 대역별로 유/무성음 판단을 수행함으로써 한 프레임 내에서 유성음과 무성음이 혼합되는 경우를 잘 모델링할 수 있다. 하지만 같은 주파수 대역에서는 유성음 성분과 무성음 성분이 공존할 수 없다. 또한 유/무성음 판단 과정에서 경험에 의한 임계치와의 비교 과정이 필요하므로 원음 스펙트럼과 합성음 스펙트럼간의 오류가 큰 경우가 발생하는 단점이 있다. 본 논문에서는 모든 주파수 영역에서 유성음 성분과 무성음 성분이 혼합되는 것을 허용하는 혼합다중대역 여기 부호화 (MMBE: mixed multi-band excitation) 음성 모델을 제안하고, 모델 파라미터인 주파수 영역 혼합 합수를 임계치와의 비교없이 효과적으로 추정할 수 있는 방법을 제시한다. 제안한 음성 모델을 적용한 2.6 kbps 음성 부호화기를 구현해 본 결과, 2.9 kbps의 전송률을 갖는 MBE 음성 부호화기에 비해서 낮은 전송률에서도 더 우수한 합성음 음질을 가지는 것으로 나타났다.

**핵심용어:** 저 전송률 음성 부호화, 음성 부호화기, 다중대역 여기모델, 음원 모델링

**투고분야:** 음성처리 분야 (2,2)

MBE (multi-band excitation) coder can achieve high quality synthetic speech below 4.0 kbps. There are, however, significant differences of the fine structure between the original spectrum and the synthetic spectrum. They are mainly due to the exclusive partition of voiced and unvoiced regions in frequency domain and the decision procedure based on the experimental threshold. This paper proposes MMBE (mixed multi-band excitation) speech model to overcome drawbacks of a MBE coder. In addition, two analysis methods, which do not need any decision procedure based on a threshold, are presented. Both voiced and unvoiced components can be mixed over all the frequency axis in the MMBE speech model. To illustrate the potential of the proposed speech model, we develop a 2.6 kbps MMBE coder and compare it with a 2.9 kbps MBE coder by both objective and subjective methods. The results have shown that the proposed coder has a better performance even at a lower bit-rate compared with the MBE coder.

**Keywords:** Low bit-rate speech coding, Speech coder, Multi-band excitation model, Speech source modeling

**ASK subject classification:** Speech signal processing (2,2)

## I. 서론

최근 디지털 핸드폰이나 음성 사서함, 자동 응답 시스템 (ARS) 등 간편한 정보 전달 수단으로 음성을 사용하는 응용 분야가 급격히 늘어나고 있다. 이러한 응용 분야에서는 음성을 기억 장소에 저장하거나 전송 선로를 통해 전송하는 것이 필요하고, 이를 위해서는 디지털화된 음성 신호를 효과적으로 표현하는 과정인 음성 부호화 (speech coding)가 수행되어야 한다. 현재 다양한 종류의 음성 부호화 방법이 개발되어 사용되고 있으나, 기억 장소나 전송 선로를 효과적으로 활용하기 위해서는 적은 정보량을 갖는 저 전송률 음성 부호화 방법이 요구된다.

일반적으로 저 전송률 음성 부호화기에서는 음원 부호화 (source coding) 방법을 사용한다. 음원 부호화 방법은 인간의 음성 생성 모델 (speech production model)에 근거하여 음성을 표현하는 방법이다. 음성 생성 과정을 살펴보면 혀의 수축에 의한 날숨이 성대 (vocal cord)를 거쳐 음원에 해당하는 여기 신호 (excitation signal)가 생성되고, 다시 조음기관인 성도 (vocal tract)와 입술을 통과함으로써 최종적인 음성이 생성된다[1]. 음원 부호화 방법에 기반한 대부분의 음성 부호화기들은 낮은 전송률으로도 인간의 성도 특성을 효과적으로 표현할 수 있다고 알려진 선형 예측 (linear prediction) 방법에 의해서 성도 특성을 표현한다. 하지만 여기 신호의 경우 효과적인 모델링을 위한 많은 연구가 이루어져 왔다[2-8].

음원 부호화 방식의 음성 부호화기 중 가장 간단한 음성 모델을 사용하는 LPC-10e[2,3] 부호화기는 한 분석 구간 내에 유성음 여기 신호, 혹은 무성음 여기 신호만이 존재한다고 가정하고 여기 신호를 모델링한다. LPC-10e 부호화기는 상대적으로 낮은 전송률을 갖지만 여기 신호 표현의 단순성으로 인해 합성음의 자연성이 떨어지는 문제점을 갖고 있다. 다중대역 여기부호화 (MBE: multi-band excitation)[4-6] 부호화기에서는 전체 분석 구간에 대한 유/무성음 판단 대신 고조파 대역별로 유/무성음 판단을 수행한다. 고조파 대역별로 유/무성음 판단을 수행하는 방법은 전체 분석 구간에 대한 유/무성음 판단 방법에 비해서 더 우수한 결과를 보이며 배경 잡음이 있는 경우에도 효과적으로 동작한다는 장점이 있다.

MBE 부호화기로 생성한 합성음 스펙트럼과 원음 스펙트럼을 비교해 보면 전반적인 모양은 비슷하지만 세부적인 모양에서 큰 차이를 보인다. 이는 현재 사용되고 있는 MBE 부호화기가 한 고조파 대역 내에 유성음 성분과 무성음 성분이 혼합되어 있는 경우를 표현할 수 없기 때문

이다. 또한, MBE 부호화기에서는 경험에 의해 결정되는 임계치에 의한 이분적 유/무성음 판단을 수행하므로 유성음 대역과 무성음 대역을 잘못 판단한 경우 큰 스펙트럼 오류가 발생한다. 따라서 원음 스펙트럼과 합성음 스펙트럼간의 스펙트럼 왜곡을 감소시키기 위해서는 같은 고조파 대역에서 유성음 성분과 무성음 성분이 혼합될 수 있도록 허용하는 음성 모델이 필요하며, 또한 임계치와의 비교없이 모델 파라미터를 추정할 수 있는 방법이 필수적이다.

본 논문에서는 MBE 부호화기의 문제점들을 해결하기 위해 모든 주파수 영역에서 유성음 성분과 무성음 성분이 혼합되는 것을 허용하는 혼합다중대역 여기부호화 (MMBE: mixed multi-band excitation) 음성 모델을 제안한다. 제안한 음성 모델은 MBE 음성 모델에 기반하지만 고조파 대역별 유/무성음 판단 대신 주파수 영역에서의 유/무성음 성분 혼합 비율을 계산함으로써 어느 주파수 대역에서나 유성음 성분과 무성음 성분이 혼합될 수 있도록 한다. 또한, 제안한 음성 모델의 모델 파라미터를 임계치와의 비교없이 추정하기 위한 두 가지 방법으로 코드북 탐색법과 간략화된 추정법을 제안한다. 코드북 탐색법을 사용하면 원음 스펙트럼과 합성음 스펙트럼간의 에러를 최소화시키는 모델 파라미터를 추정할 수 있지만 연산량이 많이 필요하다. 때문에 훨씬 적은 연산량으로도 효과적으로 모델 파라미터를 추정할 수 있는 간략화된 추정법을 함께 제안한다.

본 논문의 구성은 다음과 같다. II장에서는 본 연구의 기반이 된 MBE 부호화기에 대해서 기술한다. III장에서는 본 논문에서 제안한 MMBE 음성 모델에 대해 설명하고 모델 파라미터 값을 추정하기 위한 두 가지 방법인 코드북 탐색법과 간략화된 추정법에 대해서 기술한다. IV장에서는 제안한 MMBE 음성 모델에 대한 성능 평가 결과를 기술한다. 마지막으로 V장에서는 결론 및 향후 연구 방향에 대해서 기술한다.

## II. 다중대역 여기부호화 (MBE) 음성 부호화기

본 장에서는 본 논문의 근간이 된 MBE 음성 모델 및 모델 파라미터 분석부에 대해서 간략하게 설명한 후 이에 대한 문제점에 대해서 차례로 기술한다.

### 2.1. 다중대역 여기부호화 (MBE) 음성 모델

MBE 음성 모델에서는 고조파 대역별로 유성음 성분과

무성음 성분을 혼합하여 여기 신호를 표현한다. 단, 유성음 성분과 무성음 성분은 서로 독립적인 고조파 대역에 존재한다. MBE 음성 모델을 사용한 음성 부호화기는 유성음 성분과 무성음 성분이 공존하는 음성 분석 구간을 효과적으로 표현할 수 있으며 배경 잡음에도 강인하다[4].

### 2.2. 모델 파라미터 분석부

MBE 음성 부호화기에서는 고조파 대역의 넓이와 수를 결정하는 기본 주파수와 각 고조파 대역별 유/무성음 정보, 그리고 스펙트럼 포락을 표현하기 위한 정보를 파라미터로 갖는다. MBE 음성 부호화기의 각 파라미터와 원음 스펙트럼간의 관계를 그림 1에 나타내었다. 분석부에서는 원음 스펙트럼과 합성 스펙트럼간의 거리를 나타내는 식 (1)의  $\epsilon$ 을 최소화하도록 파라미터들을 추정한다.

$$\epsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} |S_w(\omega) - \hat{S}_w(\omega)|^2 d\omega \quad (1)$$

식 (1)에서  $S_w(\omega)$ 는 원음 스펙트럼이며  $\hat{S}_w(\omega)$ 는 합성음 스펙트럼이다.  $\epsilon$ 을 최소화하는 파라미터들을 추정하기 위해서는 피치, 스펙트럼 포락, 유/무성음 정보의 모든 조합에 대한 탐색 과정이 필요하므로 많은 연산량이 필요하다. 따라서 모든 조합에 대한 탐색 과정 대신 각 파라미터들을 순차적으로 추정하는 방법을 사용한다. 먼저 변형된 자기 상관 방법으로 피치를 추정한다. 그 후 스펙트럼 포락 정보를 계산하고 마치

막으로 각 고조파 대역에 대한 유/무성음 정보를 구한다. MBE 부호화기의 경우 피치는 변형된 자기 상관 방법 및 피치 추적 (pitch tracking) 알고리즘을 적용하여 구해진다[6]. 스펙트럼 포락을 표현하는 방법은 여러 가지가 있지만 4 kbps 이하의 전송률을 갖는 저 전송률 음성 부호화기의 경우 선형 예측 방법 (linear prediction)을 사용하는 것이 일반적이다[9,10].

피치와 스펙트럼 포락 정보를 추정한 후, 전 고조파 대역을 유성음으로 가정하고 기본 주파수 간격으로 창함수 스펙트럼을 나열한 주기적 스펙트럼을 생성한다. 이렇게 생성한 주기적 스펙트럼과 원음 스펙트럼의 오차가 임계치를 초과할 경우 무성음 대역, 임계치 이하인 경우 유성음 대역으로 결정한다.  $m$ 번째 고조파 대역의 에너지를  $\xi_m$ 이라고 하면 식 (2)에 의해서 계산된다.

$$\xi_m = \frac{\int_{a_m}^{b_m} |S_w(\omega) - A_m E_w(\omega)|^2 d\omega}{\int_{a_m}^{b_m} S_w(\omega)^2 d\omega} \quad (2)$$

식 (2)에서  $a_m, b_m$ 은 각각  $m$ 번째 고조파 대역의 시작 주파수 샘플과 끝 주파수 샘플이며  $A_m$ 은  $m$ 번째 고조파 대역을 유성음 대역으로 가정하고 식 (2)를 최소화하도록 계산한 에너지 값이다. 또한  $S_w(\omega)$ 는 원음 스펙트럼이며  $E_w(\omega)$ 는 기본 주파수 간격으로 창함수 스펙트럼을 나열하여 생성한 주기적 스펙트럼이다.

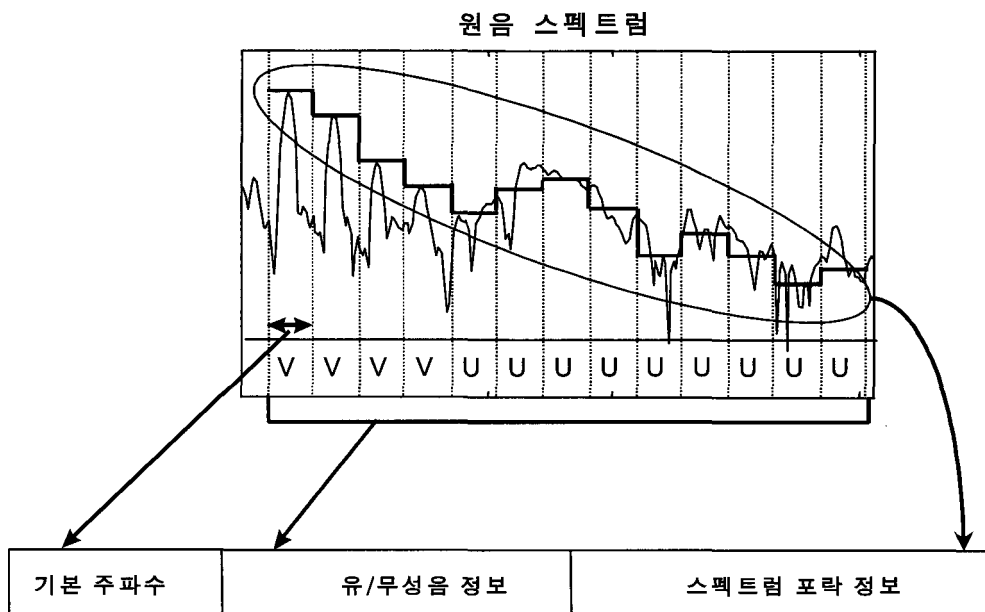


그림 1. 다중대역 여기부호화 (MBE) 모델 파라미터와 원음 스펙트럼과의 관계  
 Fig. 1. Relationship between model parameters of MBE and original spectrum.

### 2.3. 다중대역 여기부호화 (MBE) 음성 부호화기의 문제점

MBE 음성 부호화기에서는 전체 주파수 대역을 기본 주파수 너비의 균등한 대역들로 분할한 후 각 대역에 대해서 유성음 성분과 무성음 성분 중 어느 성분이 더 강한지를 판단한다. 이는 유성음 성분이 강한 대역에서는 고조파 (harmonic) 형태의 스펙트럼이 나타나며 무성음 성분이 강한 대역에서는 백색 잡음 (white noise) 형태의 스펙트럼이 나타나는 것을 반영한 것이다. 본 논문에서는 MBE 음성 부호화기가 가지는 두 가지 문제점에 대해서 기술한 후 해결 방법을 제안하고자 한다.

MBE 음성 부호화기가 갖는 첫 번째 문제점은 유성음 성분과 무성음 성분이 혼합되어 있는 주파수 대역을 표현할 수 없다는 것이다. MBE 음성 부호화기에서는 유성음 성분과 무성음 성분이 반드시 고조파 대역 경계에 의해서 분할되며 하나의 고조파 대역 내에서는 유성음 성분과 무성음 성분이 공존할 수 없다. 때문에 원음 스펙트럼과 MBE 음성 부호화기로 생성한 합성음 스펙트럼 사이에 큰 예측 오류가 발생할 수 있다.

MBE 음성 부호화기가 가지는 두 번째 문제점은 각 고조파 대역에 대한 유/무성음 판단이 경험적으로 결정되는 임계치에 의해서 이루어지므로 환경의 영향을 받기 쉽다는 것이다. MBE 음성 부호화기에서는 전체 대역을 유성음으로 가정하여 구한 주기적 스펙트럼과 원음 스펙트럼간의 에러를 구한 후 일정 임계치 이하의 에러를 갖는 대역을 유성음 대역으로 판정한다. 이 때, 임계치의 값을 어떻게 설정하느냐에 따라서 합성음의 음질이 좌우된다. 임계치의 값이 너무 작으면 상대적으로 무성음 대역으로 판단되는 경우가 많아지므로 목 선 것 같은 소리가

날 수 있다. 반대로 임계치의 값이 너무 크면 유성음 대역으로 판단되는 경우가 많아지므로 날카로운 기계음 소리가 날 수 있다. 이 문제를 해결하기 위해서 에너지의 값에 따라서 임계치의 값을 변경시켜서 사용하는 방법이 제안되었다[5]. 하지만 이 방법도 역시 경험에 의해 결정된 상수들을 사용하므로 환경의 영향을 받기 쉬우며 이전 프레임의 정보를 사용하므로 에러가 누적될 수 있는 문제가 있다.

본 논문에서는 유성음 성분과 무성음 성분이 어느 주파수 대역에서나 혼합될 수 있도록 허용하는 MMBE 음성 모델을 제안하였다. 또한 유성음 성분과 무성음 성분의 혼합비를 계산할 수 있는 분석 방법을 제안하였다. 제안한 방법을 사용하면 경험에 의한 임계치에 의존하지 않고 모델 에러를 최소화하는 혼합비를 계산함으로써 여기 신호를 표현할 수 있다.

### III. 혼합다중대역 여기부호화 (MMBE) 음성 모델과 파라미터 추정법

본 장에서는 여기 신호의 효과적인 표현을 위해서 제안한 MMBE 음성 모델을 설명하고, MMBE 음성 모델의 파라미터들을 효과적으로 추정하기 위해서 제안한 두 가지 방법인 코드북 탐색법과 간략화된 추정법에 대해서 기술한다.

#### 3.1. 혼합다중대역 여기부호화 (MMBE) 음성 모델

음성 신호를 이산 푸리에 변환 (DFT: discrete fourier transform)하여 얻은 크기 스펙트럼  $S(n)$ 은 음성 생성 모델에서 스펙트럼 포락  $H(n)$ 과 여기 신호 스펙트럼

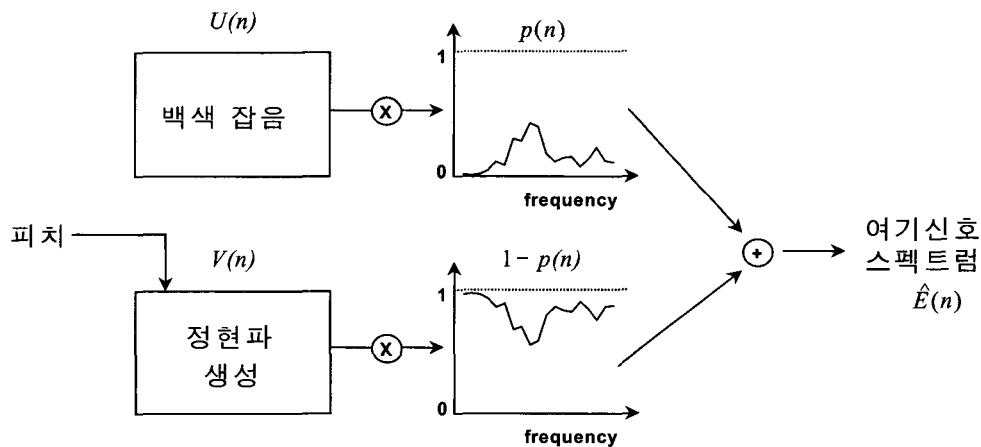


그림 2. 혼합다중대역 여기부호화 (MMBE) 음성 모델  
Fig. 2. MMBE speech model.

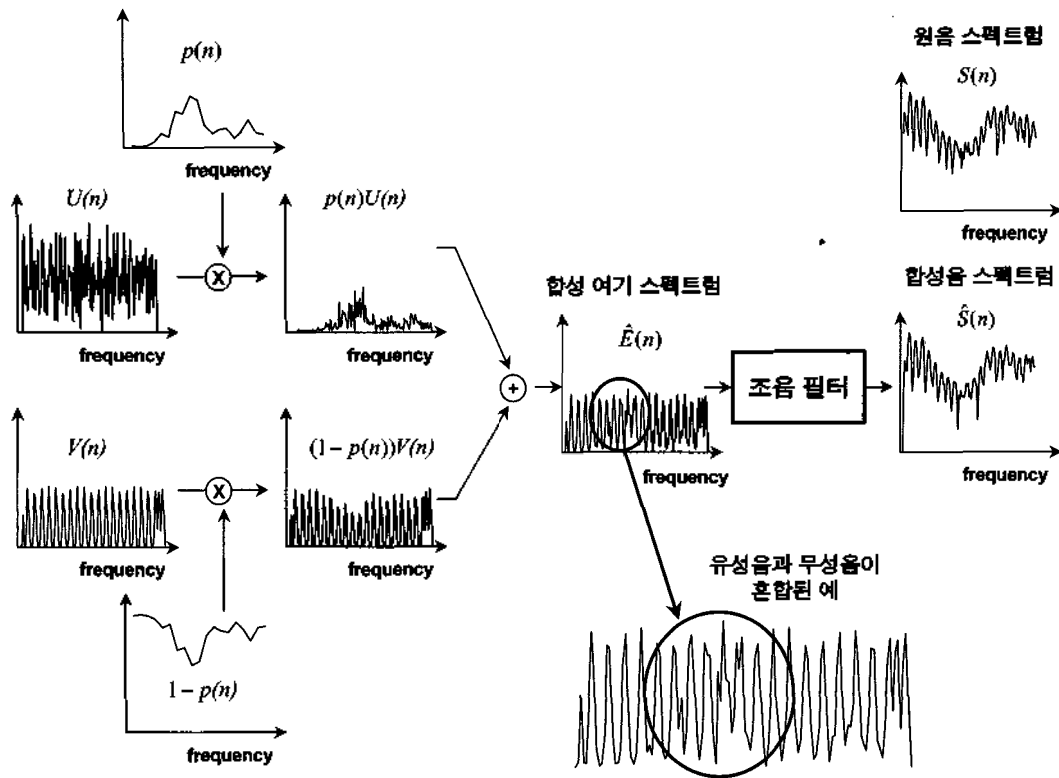


그림 3. 혼합다중대역 여기부호화 (MMBE) 음성 모델에서의 합성음 생성 과정  
 Fig. 3. Synthesis procedure of MMBE speech model.

$E(n)$ 의 곱으로 예측된다. MMBE 음성 모델에서는 식 (3), 그림 2와 같이 유성음 여기 신호 스펙트럼  $V(n)$ 과 무성음 여기 신호 스펙트럼  $U(n)$ 의 가중 합으로 여기 신호 스펙트럼  $E(n)$ 을 예측한다.

$$\hat{E}(n) = (1 - p(n))V(n) + p(n)U(n) \quad (3)$$

where  $0 \leq p(n) \leq 1, 0 \leq n < \frac{N}{2}$

식 (3)에서  $N$ 은 FFT 크기이며  $p(n)$ 은  $V(n)$ 과  $U(n)$ 의 혼합비를 나타내는 함수이다. 일반적인 음성 생성 모델에서 여기 신호 스펙트럼  $\hat{E}(n)$ 은 주파수 영역에서 고른 에너지 분포를 보이므로  $p(n)$ 의 값을 0~1로 제한한다.

예측된 여기 신호 스펙트럼을  $\hat{E}(n)$ , 예측된 스펙트럼 포락을  $\hat{H}(n)$ 이라고 하면 합성음 스펙트럼  $\hat{S}(n)$ 은  $\hat{E}(n)$ 과  $\hat{H}(n)$ 의 곱으로 표현되므로 식 (4)와 같이 쓸 수 있다. 그림 3은 MMBE 음성 모델에서 합성음 스펙트럼  $\hat{S}(n)$ 을  $\hat{H}(n)$ ,  $V(n)$ ,  $U(n)$ ,  $p(n)$ 에 의해서 표현한 것이다.

$$\hat{S}(n) = \hat{H}(n)\hat{E}(n) = \hat{H}(n)((1 - p(n))V(n) + p(n)U(n)) \quad (4)$$

MMBE 음성 모델의 파라미터는 피치, 스펙트럼 포락, 그리고 주파수 영역 혼합비를 나타내는 함수  $p(n)$ 이다.

본 논문에서는  $p(n)$ 을 효과적으로 추정하기 위한 방법으로 코드북 탐색법과 간략화된 추정법을 제안한다.

### 3.2. 코드북 탐색법

주파수 영역 혼합비  $p(n)$ 은 주파수 샘플  $n$ 에 대한 함수이나 모든  $n$ 에 대한  $p(n)$ 을 계산하는 것은 많은 연산량을 요구할 뿐 아니라 주파수 해상도 면에서 불리하다. 때문에 적당한 길이의 영역에 대해서  $p(n)$ 이 일정하다고 가정하고  $p(n)$ 을 계산한다. 본 절에서는 고조파 대역별로  $p(n)$ 을 계산하는 방법과 고정 대역별로  $p(n)$ 을 계산하는 방법을 기술한다.

#### 3.2.1. 고조파 대역별 분석법

현재 프레임의 기본 주파수를  $\omega_0$ , 1번째 고조파 대역의 시작 샘플과 끝 샘플을 각각  $a_i, b_i$ 이라고 하자. 스펙트럼 포락 함수  $\hat{H}(n)$ 이 한 고조파 대역 내에서 일정하다고 가정하면  $L$ 개의 스펙트럼 포락 값으로 표현할 수 있다. 이 때 주파수 영역 혼합비를 나타내는 함수  $p(n)$ 도 한 고조파 대역 내에서 일정하다고 가정하자. 1번째 고조파 대역의 스펙트럼 포락 값을  $a_i$ , 1번째 고조파 대역의 주파수 영역 혼합비를  $p_i$ 이라고 하면 1번째 고조파 대역에 대한

합성음  $\hat{S}(n)$ 은 식 (5)와 같이 쓸 수 있다.

$$\hat{S}(n) = a_i((1-p_i)V(n) + p_iU(n)) \quad (5)$$

where  $a_i \leq n \leq b_i$

식 (5)에서  $V(n)$ ,  $U(n)$ 은 각각 유성음 여기 신호 스펙트럼과 무성음 여기 신호 스펙트럼이다. 1번째 고조파 대역에 대한 원음 스펙트럼  $S(n)$ 과 합성 스펙트럼  $\hat{S}(n)$  간의 에러  $\epsilon$ 은 식 (6)과 같다.

$$\begin{aligned} \epsilon &= \sum_{n=a_i}^{b_i} |S(n) - \hat{S}(n)|^2 \\ &= \sum_{n=a_i}^{b_i} |S(n) - a_i((1-p_i)V(n) + p_iU(n))|^2 \end{aligned} \quad (6)$$

식 (6)을  $a_i$ 에 대해서 편미분하여 0으로 놓으면 식 (7)과 같이  $a_i$ 에 관한 식을 얻을 수 있다.

$$a_i = \frac{\sum_{n=a_i}^{b_i} S(n)(p_iU(n) + (1-p_i)V(n))}{\sum_{n=a_i}^{b_i} (p_iU(n) + (1-p_i)V(n))^2} \quad (7)$$

같은 방법으로 식 (6)을  $p_i$ 에 대해서 편미분하여 0으로 놓으면 식 (8)을 얻을 수 있다.

$$p_i = \frac{\sum_{n=a_i}^{b_i} (U(n) - V(n))(S(n) - a_iV(n))}{a_i \sum_{n=a_i}^{b_i} (U(n) - V(n))^2} \quad (8)$$

식 (7)과 식 (8)을 연립해서 풀면  $p_i$ 에 관한 식 (9)를 얻을 수 있다.

$$\begin{aligned} p_i &= \frac{A_i - B_i}{C_i - D_i} \text{ where} \\ A_i &= \sum_{n=a_i}^{b_i} S(n)V(n) \sum_{n=a_i}^{b_i} V(n)U(n) \\ B_i &= \sum_{n=a_i}^{b_i} S(n)U(n) \sum_{n=a_i}^{b_i} V(n)^2 \\ C_i &= \sum_{n=a_i}^{b_i} S(n)U(n) \sum_{n=a_i}^{b_i} V(n)(U(n) - V(n)) \\ D_i &= \sum_{n=a_i}^{b_i} S(n)V(n) \sum_{n=a_i}^{b_i} U(n)(U(n) - V(n)) \end{aligned} \quad (9)$$

식 (9)에서  $S(n)$ 은 원음 스펙트럼으로부터 얻어지는 값이므로  $V(n)$ 과  $U(n)$ 만 알 수 있으면  $p_i$ 를 계산할 수 있다. 유성음 여기 신호 스펙트럼은 기본 주파수의 배수 위치에 정규화된 창함수 스펙트럼을 나열한 주기적 스펙트럼으로 표현된다[4].  $V(n)$ 은 유성음 여기 신호 스펙트럼이므로 정규화된 창함수 스펙트럼의 중심이  $l\omega_0$  위치로 오도록 이동시켜 얻을 수 있다. 이에 비해, 무성음 여기 신호

스펙트럼은 백색 잡음 스펙트럼 형태로 나타나므로 특정 함수로 표현하기 힘들다. 때문에 본 논문에서는 무성음 여기 신호  $U(n)$ 을 예측하기 위해서 잡음 스펙트럼 코드북을 사용하는 방법을 제안한다.

잡음 스펙트럼 코드북은 각각의 코드워드가 정규화된 백색 잡음 스펙트럼인 코드북이다. 단, 코드워드의 길이가 고조파 대역의 길이로 가변하므로 중첩 코드북(overlapped codebook) 형태를 사용하였다. 중첩 코드북 자체가 정규화된 백색 잡음 스펙트럼이므로 임의의 부분을 취해도 백색 잡음 스펙트럼을 얻을 수 있다. 따라서 코드워드의 길이가 가변하는 경우에도 사용할 수 있을 뿐 아니라 코드북의 전체 크기를 줄일 수 있는 장점이 있다. 현재 분석 구간의 기본 주파수를  $\omega_0$ 라고 하면 고조파 대역의 길이  $M$ 은 식 (10)으로 계산할 수 있다.

$$M = \left\lfloor \frac{\omega_0}{2\pi} N \right\rfloor \quad (10)$$

잡음 스펙트럼 코드북의 크기를  $K$ 라고 하면 그림 4와 같이 총 코드워드의 갯수는  $K - M + 1$ 이 된다. 이때,  $\vec{C}_0 \dots \vec{C}_{K-M}$ 을 각각의 코드워드라고 하면 각 코드워드의 길이는 고조파 대역의 길이  $M$ 과 같다.  $U(n)$  자체는 직접 계산할 수 없으므로  $\vec{C}_0 \dots \vec{C}_{K-M}$  중에서 식 6을 최소화하는 코드워드  $\vec{C}_k$ 를 선택함으로써  $U(n)$ 을 근사한다. 식 9의  $U(n)$ 대신  $\vec{C}_k$ 를 대입하면 1번째 고조파 대역의 주파수 영역 혼합비  $p_1$ 을 구할 수 있다. 이 과정을  $L$ 개의 고조파 대역에 대해서 반복함으로써 주파수 영역 혼합비  $p_1 \dots p_L$ 을 구할 수 있다.

### 3.2.2. 고정 대역별 분석법

분석 구간의 길이  $M$ 이 작은 경우 창함수 스펙트럼이 점차 백색 잡음 스펙트럼과 비슷한 모양을 가지게 된다. 이때 창함수 스펙트럼과 백색 잡음 스펙트럼간의 구별이 어려우므로 분석법의 신뢰도가 떨어지는 단점이 있다. MBE 음성 부호화기에서는 이러한 문제를 해결하기 위해서 3개의 고조파 대역을 하나의 그룹으로 묶어서 분석하는 방법을 사용한다[3].

본 논문에서는 전체 주파수 대역을 고정 갯수의 대역으로 분할한 후 각 대역에 대한 분석을 수행하는 고정 대역별 분석법을 제안한다. 고정 대역별 분석법에서는 대역의 길이가 일정하게 유지되는 특징이 있다. FFT 크기를  $N$ , 고정 대역의 수를  $F$ 라고 하면  $f$ 번째 고정 대역의 시작 샘플  $a_f$ 와 끝 샘플  $b_f$ 는 식 (11)과 같이 계산할 수 있다.

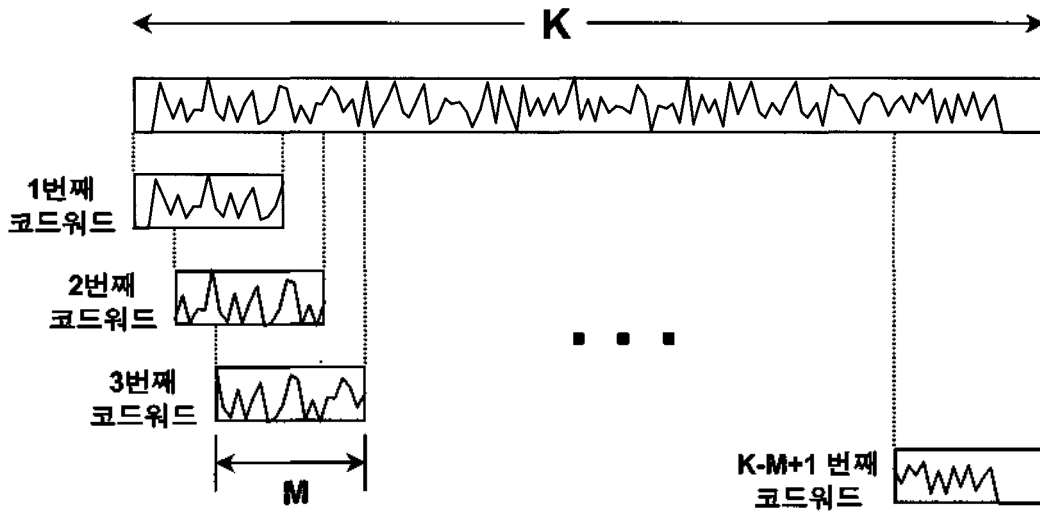


그림 4. 잡음 스펙트럼 코드북  
Fig. 4. Noise spectrum codebook.

$$\begin{aligned}
 a_f &= \left\lfloor \frac{N(f-1)}{2F} \right\rfloor \\
 b_f &= \left\lfloor \frac{Nf}{2F} \right\rfloor - 1 \\
 &\text{where } 1 \leq f \leq F
 \end{aligned}
 \tag{11}$$

식 (6)을 고정 대역에 대한 식으로 변형하면 식 (12)를 얻을 수 있다.

$$\epsilon = \sum_{n=a_f}^{b_f} |S(n) - \alpha_f((1-p_f)V(n) + p_fU(n))|^2
 \tag{12}$$

식 (12)에서  $p_f$ ,  $\alpha_f$ 는 각각 f번째 고정 대역에 대한 주파수 영역 혼합비와 스펙트럼 포락 값이다. 하지만 고정 대역의 폭이 넓은 경우 스펙트럼 포락을 정확하게 표현하지 못하는 문제가 있으므로  $\alpha_f$  대신 스펙트럼 포락 함수  $\alpha(n)$ 을 사용한다.  $\alpha(n)$ 은 모든 고조파 대역이 주기적 스펙트럼이며 한 고조파 대역 내에서는 스펙트럼 포락 값이 일정하다고 가정하여 구한 함수이다.  $\alpha(n)$ 은 식 (13)에 의해서 구할 수 있다[4].

$$\alpha(n) = \frac{\sum_{n=a_f}^{b_f} S(n)V(n)}{\sum_{n=a_f}^{b_f} V(n)^2} \quad \text{if } a_f \leq n \leq b_f
 \tag{13}$$

식 (13)의  $a_f$  과  $b_f$  은 각각 l번째 고조파 대역의 시작 샘플과 끝 샘플을 가리키는 값이다. 식 (12)의  $\alpha_f$  를  $\alpha(n)$  으로 대체하면 식 (14)를 얻을 수 있다.

$$\epsilon = \sum_{n=a_f}^{b_f} |S(n) - \alpha(n)((1-p_f)V(n) + p_fU(n))|^2
 \tag{14}$$

식 (14)를  $p_f$  에 대해서 편미분하고 0으로 놓으면 식 (15)를 얻을 수 있다.

$$p_f = \frac{\sum_{n=a_f}^{b_f} \alpha(n)(U(n) - V(n))(S(n) - \alpha(n)V(n))}{\sum_{n=a_f}^{b_f} \alpha(n)^2(U(n) - V(n))^2}
 \tag{15}$$

식 (15)를 F개의 고정 대역에 대해서 모두 적용하면 주파수 영역 혼합비  $p_1 \dots p_F$  를 구할 수 있다. 단, 식 (15)를 계산하기 위해서는  $U(n)$ 의 값을 알아야 하므로 고조파 대역별 분석법과 마찬가지로 잡음 코드북 탐색이 필요하다. 고조파 대역별 분석법에서는 주파수 영역 혼합비의 수가 고조파 대역의 수에 따라서 가변한다. 이에 비해서 고정 대역별 분석법에서는 주파수 영역 혼합비의 수가 항상 F개로 일정하므로 양자화시 유리한 장점이 있다.

### 3.3. 간략화된 추정법

코드북 탐색법은 잡음 코드북 탐색을 통해서 원음 스펙트럼과 합성음 스펙트럼 사이의 에러를 최소화하는 모델 파라미터를 추정한다. 하지만 별도의 코드북 탐색이 필요하므로 연산량이 많이 필요한 단점이 있다. 본 논문에서는 코드북 탐색법과 비슷한 성능을 가지면서도 연산량을 크게 감소시킬 수 있는 간략화된 추정법을 함께 제안한다. 제안한 방법은 간단한 가정을 들어서 잡음 코드북을 탐색하지 않고 원음 스펙트럼으로부터 직접 주파수 영역 혼합비를 추정한다. 이때, 양자화에 유리하도록 고정 대역별 분석법을 사용하였다.

f번째 고정 대역에 대한 주파수 영역 혼합비  $p_f$ 를 계산하는 경우를 생각하자.  $p_f$ 를 최적화하여 선택하였을 경우, f번째 고정 대역에 대한 원음 스펙트럼  $S(n)$ 과 합성음 스펙트럼  $\hat{S}(n)$ 의 평균이 같아진다고 가정하면 식 (16)을 얻을 수 있다.

$$\sum_{n=a_f}^{b_f} S(n) = \alpha(n) \sum_{n=a_f}^{b_f} ((1-p_f)V(n) + p_f U(n)) \tag{16}$$

식 (16)에서  $a_f, b_f$ 는 식 (11)에 의해서,  $\alpha(n)$ 은 식 (13)에 의해서 계산할 수 있다. 식 (16)을  $p_f$ 에 대해서 정리하면 식 (17)을 얻을 수 있다.

$$p_f = \frac{\sum_{n=a_f}^{b_f} (S(n) - \alpha(n)V(n))}{\sum_{n=a_f}^{b_f} \alpha(n)U(n) - \sum_{n=a_f}^{b_f} \alpha(n)V(n)} \tag{17}$$

식 (17)에서  $V(n)$ 은 창함수 스펙트럼으로부터 얻을 수 있다. 하지만  $U(n)$ 은 잡음 스펙트럼이므로 특정 함수로 가정하여 계산할 수 없다. 때문에 식 17을 계산하기 위해서는 분모 부분의  $\sum_{n=a_f}^{b_f} \alpha(n)U(n)$ 을 계산하는 방법이 필요하다. f번째 고정 대역에 k개의 고조파 대역이 포함되어 있다고 가정하자. 각 고조파 대역의 넓이는 모두 같지만 그림 5와 같이 첫 번째 고조파 대역과 k번째 고조파 대역은 고정 대역에 일부만 포함된다. 그림 5에서  $m_2 \dots m_k$ 는 각 고조파 대역의 경계를 나타낸다. 또한,  $m_1 = a_f$ 이며  $m_{k+1} = b_f + 1$ 이다. 고정 대역에 포함된 스펙트

럼 포락 값은 각각  $\alpha_1 \dots \alpha_k$ 로 나타내었다.  $\sum_{n=a_f}^{b_f} \alpha(n)U(n)$ 을 변형하여 k개의 고조파 대역에 관한 식으로 풀어서 쓰면 식 (18)을 얻을 수 있다.

$$\sum_{n=a_f}^{b_f} \alpha(n)U(n) = \alpha_1 \sum_{n=m_1}^{m_2-1} U(n) + \dots + \alpha_k \sum_{n=m_k}^{m_{k+1}-1} U(n) \tag{18}$$

$SU_i = \sum_{n=m_i}^{m_{i+1}-1} U(n)$ 이라고 놓으면 식 (18)은 식 (19)로 다시 쓸 수 있다.

$$p_f = \sum_{n=a_f}^{b_f} \alpha(n)U(n) = \alpha_1 SU_1 + \dots + \alpha_k SU_k \tag{19}$$

$U(n)$ 을 평균이 1인 잡음 스펙트럼이라고 생각하면  $U(n)$ 의 부분 집합의 평균도 1로 근사된다고 가정할 수 있다. 이때 식 (20)이 성립한다.

$$SU_i \approx m_{i+1} - m_i \tag{20}$$

식 (20)을 식 (19)에 적용하면 식 (21)을 얻을 수 있다.

$$\begin{aligned} \sum_{n=a_f}^{b_f} \alpha(n)U(n) &\approx \alpha_1(m_2 - m_1) + \dots + \alpha_k(m_{k+1} - m_k) \\ &= \sum_{n=a_f}^{b_f} \alpha(n) \end{aligned} \tag{21}$$

식 (21)을 식 (17)에 대입하면  $U(n)$ 이 제거된 식 (22)를 얻을 수 있다.

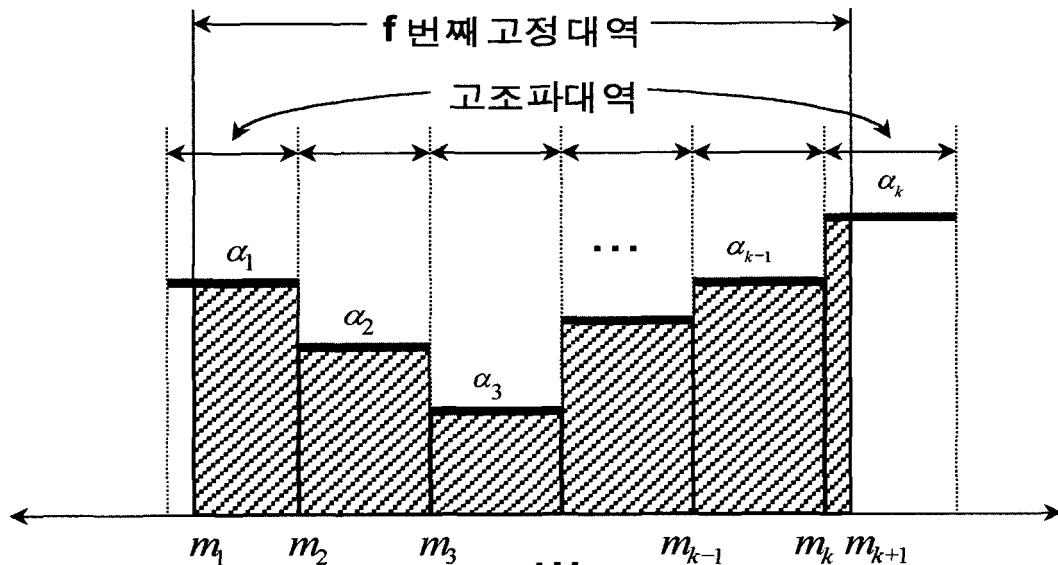


그림 5. 고조파 대역과 고정 대역의 관계  
Fig. 5. Relationship between harmonic band and fixed size band.



$$p_f \approx \frac{\sum_{n=a_f}^{b_f} (S(n) - \alpha(n)V(n))}{\sum_{n=a_f}^{b_f} \alpha(n) - \sum_{n=a_f}^{b_f} \alpha(n)V(n)} \quad (22)$$

최종적으로 식 (22)를 모든 고정 대역에 대해서 적용하면 주파수 영역 혼합비  $p_1 \dots p_F$ 를 얻을 수 있다.

### 3.4. 합성부

MMBE 음성 부호화기의 합성부에서는 분석부에서 추정된 피치 정보, 스펙트럼 포락 정보, 그리고 각 고정 대역별 주파수 영역 혼합비로부터 원래의 음성을 복구한다. 이때, 시간 영역에서 정현파들의 합으로 유성음 신호를 생성하며 주파수 영역에서 푸리에 변환을 사용하여 무성음 신호를 생성한다. 그 후 최종적으로 유성음 신호와 무성음 신호를 더하여 합성음을 얻는다. 이러한 과정은 유/무성음 혼합비율을 사용한다는 점을 제외하면 기존의 MBE 음성 부호화기와 동일하다.

## IV. 실험 및 평가

본 장에서는 III장에서 제안한 MMBE 음성 모델에 대한 실험 및 평가 결과를 기술한다. 4.1절에서는 실험을 위해 사용된 음성 자료 및 평가 방법에 관해서 기술하고, 4.2절에서는 모델 파라미터 추정법의 성능을 평가한다. 4.3절에서는 벡터 양자화 성능을 평가하고, 마지막으로 4.4절에서는 최종적으로 구현된 2.6 kbps MMBE 음성 부호화기를 기존의 2.9 kbps MBE 음성 부호화기와 비교하여 평가한다.

### 4.1. 실험 환경

제안한 MMBE 음성 모델의 성능을 평가하기 위해서 각 분석 방법을 적용한 MMBE 음성 부호화기를 구현하였다. 구현된 음성 부호화기는 8 kHz 표본 주파수로 샘플링된 데이터를 입력으로 사용하고 부호화기를 거쳐 재생된 음성 신호를 출력 파일에 저장하도록 만들어졌다.

#### 4.1.1. 실험 문장

실험을 위해서 약 34초 분량에 해당하는 10문장의 음성 데이터 (남성 5인/여성 5인)를 실험 자료로 수집하고 모든 실험에서 입력 신호로 사용하였다. 자료 음성은 SONY의 DAT (digital audio tape) deck 59ES를 사용하여 8

표 1. 실험에 사용된 시스템들

Table 1. System types used for performance evaluation (C: 코드북 탐색법 / SM: 간략화된 추정법).

시스템 명	분석법	파라미터 양자화
MBE-3	MBE+3개 대역 그룹	none
MBE-F	MBE+고정 대역별	none
MBE-Q	MBE+3개 대역 그룹	all
MMBE-AE	MMBE+고조파 대역별(CB)	none
MMBE-AF	MMBE+고정 대역별(CB)	none
MMBE-BE	MMBE+고조파 대역별(SM)	none
MMBE-BF	MMBE+고정 대역별(SM)	none
MMBE-Q	MMBE+고정 대역별(SM)	all

kHz, 16비트 PCM으로 양자화하였다. 또한, 코드북 학습을 위한 데이터로는 TIMIT 데이터 베이스에서 선별한 31.7초 분량의 음성 데이터 (남성 8인/여성 8인)를 사용하였다.

#### 4.1.2. 평가 방법

음성 부호화기의 성능을 평가할 때는 일반적으로 음질, 전송량, 연산량의 세 가지 요소가 고려된다. 본 논문에서는 MBE 음성 모델이 가지는 여기 신호 표현상의 문제를 해결하고자 MMBE 음성 모델을 제안한 것이므로 가장 중요한 평가 요소는 합성음의 음질이다. 본 장에서는 객관적 음질 평가를 위해 스펙트럼 왜곡 (SD; spectral distortion)[11]을 사용하였다. 스펙트럼 왜곡은 식 (23)과 같이 구해진다.

$$SD = \frac{2}{N} \sum_{n=1}^{N/2} [20 \log_{10} |S_x(n)| / |S_y(n)|]^2 \quad (\text{dB}^2) \quad (23)$$

식 (23)에서  $S_x(n)$ ,  $S_y(n)$ 은 각각 원음과 합성음의 스펙트럼이며  $N$ 은 FFT 크기이다). 스펙트럼 왜곡은 전체 스펙트럼 특성의 차이를 비교하고 싶을 때 사용되는 객관적 평가 방법이다.

주관적 음질 평가는 원음과 합성음을 청취자에게 직접 들려주고 평가하는 방식으로서 실제 사람에게 들리는 합성음의 음질을 평가할 수 있는 장점이 있다. 본 논문에서는 MOS (mean opinion score) 테스트를 사용하여 주관적 음질 평가를 실시하였다. MOS 테스트에서는 합성음의 음질을 excellent, good, normal, poor, bad의 5단계로 나누어 평가한다.

### 4.2. 모델 파라미터 추정 방법의 성능 평가

본 절에서는 제안한 MMBE 음성 모델의 파라미터인 주

1) 본 논문에서 구현한 시스템에서는  $N=256$ 을 사용하였다.

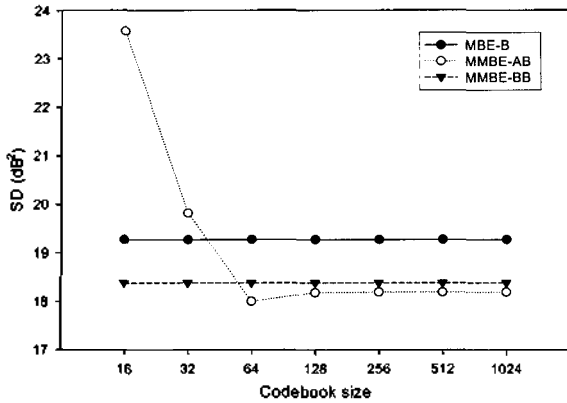


그림 6. 고조파 대역별 분석법의 성능 평가  
Fig. 6. Experimental results of harmonic band based methods.

파수 영역 혼합비를 추정하는 방법의 성능을 평가하였다. 4.2.1절에서는 고조파 대역별 분석 방법에 대한 성능을 평가하였으며 4.2.2절에서는 주파수 해상도 문제를 극복하기 위해 제안된 고정 대역별 분석 방법을 함께 적용하여 실험한 결과를 기술하였다.

#### 4.2.1. 고조파 대역별 분석법

그림 6은 MBE-B와 MMBE-AB, MMBE-BB의 성능을 비교하기 위해서 실험 문장에 대한 SD를 나타낸 그래프이다. MMBE-AB의 경우 잡음 코드북 탐색이 필요하므로 그림 6에서는 잡음 코드북의 크기가 16~1024개인 경우에 대한 결과를 보였다. MMBE-AB의 경우 잡음 코드북의 크기가 커질수록 SD가 감소하는 경향을 보였다. 또한, 잡음 코드북의 크기가 64개 이상이 되면 MBE-B에 비해서 더 작은 SD값을 갖는 것으로 나타났다. 이러한 결과에서 제안한 시스템인 MMBE-AB가 기존의 MBE 부호기에 비해서 세부 스펙트럼 구조를 더 잘 표현함을 알 수 있다. MMBE-BB의 경우도 기존의 MBE 부호기에 비해서 더 작은 SD값을 가지는 것으로 나타났다. 또한, MMBE-BB는 MMBE-AB의 잡음 코드북 크기가 64개 이상인 경우와 거의 비슷한 성능을 보였다.

#### 4.2.2. 고정 대역별 분석법

그림 7은 고정 대역의 수를 2~10으로 바꾸어 가면서 MBE-3, MBE-F, MMBE-AF, MMBE-BF의 성능을 비교한 그래프이다. 한편, MMBE-AF는 잡음 코드북의 크기를 128개로 고정시킨 후 실험하였다. MBE-3은 3개의 고조파 대역을 하나의 고조파 그룹으로 묶어서 분석하는 시스템으로 기존의 MBE 음성 부호화기에서 주파수 해상도 상의 문제를 해결하기 위해서 사용하는 방법을 적용하

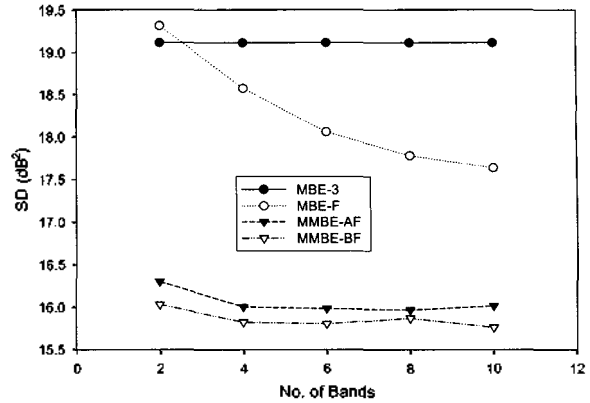


그림 7. 고정 대역별 분석법의 성능 평가  
Fig. 7. Experimental results of fixed band based methods.

였다. 하지만 그림 7에서 볼 수 있듯이 고정 대역별 분석 방법을 적용한 MBE-F가 SD면에서 더 나은 성능을 보였다. 또한 MBE-F에 비해서 MMBE-AF나 MMBE-BF가 더 나은 성능을 나타낸다. 즉, 같은 고정 대역별 분석 방법을 사용하더라도 제안한 MMBE 음성 모델이 기존의 MBE 음성 모델에 비해서 더 나은 성능을 보임을 알 수 있었다. MMBE-AF와 MMBE-BF는 큰 차이는 없었지만 전체적으로 MMBE-BF가 약간 나은 성능을 보이는 것을 알 수 있었다. MMBE-BF의 경우 별도의 잡음 코드북 탐색을 필요로 하지 않으므로 MMBE-AF에 비해서 연산량 또한 작은 장점이 있다.

#### 4.3. 구현된 음성 부호화기의 성능 평가

본 절에서는 제안한 MMBE 음성 모델을 적용한 음성 부호화기를 구현하고 기존의 MBE 음성 부호화기와 성능을 비교한다.

##### 4.3.1. 음성 부호화기 사양

본 논문에서 제안한 부분만의 성능을 비교하기 위해서 기존의 MBE 음성 모델을 사용한 2.9 kbps MBE-Q 음성 부호화기에 제안한 MMBE 음성 모델을 적용하여 2.6 kbps MMBE-Q 음성 부호화기를 구현하였다. 두 부호화기는 유/무성음 정보를 추정하는 부분을 제외하면 모두 같으므로 제안한 부분의 성능만을 비교할 수 있다. 표 2은 MBE-Q와 MMBE-Q에 대한 상세 스펙을 표로 정리한 것이다.

##### 4.3.2. 객관적 음질 평가

그림 8은 MBE-Q와 MMBE-Q의 성능을 비교한 것이다. 그림 8에서는 성능 비교를 위해서 사용한 10개의 실험 문장에 대한 SD값을 보였다. 성능 비교 결과, 모든 문장

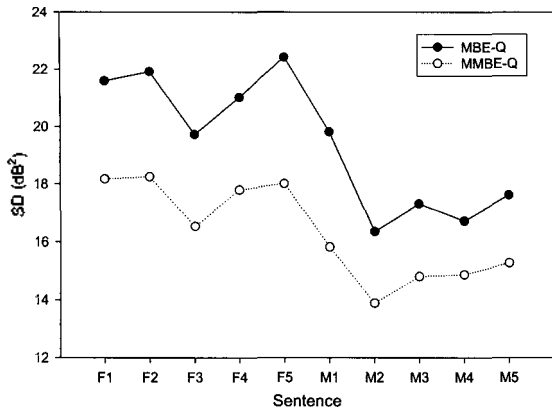


그림 8. MBE-Q와 MMBE-Q의 성능 비교  
Fig. 8. Performance comparison of MBE-Q and MMBE-Q.

표 2. MBE-Q와 MMBE-Q의 비트 할당표  
Table 2. Bit allocation of MBE-Q and MMBE-Q.

파라미터	MBE-Q (bits)	MMBE-Q (bits)
10차 LSF 계수	34	34
에너지	6	6
피치	8	8
유/무성음 정보	10	4 (벡터 양자화)
합계	58 / 2.9kbps	52 / 2.6kbps

에 대해서 제안한 시스템이 더 나은 성능을 보임을 실험적으로 검증할 수 있었다.

### 4.3.3. 주관적 음질 평가

MOS 테스트는 20인의 청취자를 대상으로 하였으며 10개의 한국어 문장을 들려준 후 합성음의 음질을 5단계로 평가하는 방식으로 진행하였다. MOS 테스트에서 MBE-Q는 평균 4.04점, MMBE-Q는 평균 4.14점을 얻었다. MOS 테스트 결과, 제안한 MMBE-Q가 더 낮은 전송률에서도 MBE-Q보다 청각적으로 더 우수한 음질의 합성음을 생성해 낼 수 있었다.

## V. 결론

본 논문에서는 MBE 음성 부호화기의 문제점을 해결하기 위해서 모든 주파수 영역에서 유성음 성분과 무성음 성분이 혼합될 수 있도록 허용하는 MMBE 음성 모델을 제안하였다. 또한 임계치와의 비교없이 주파수 영역 혼합비를 추정할 수 있는 방법으로 코드북 탐색법과 간략화된 추정법을 제시하였다. 코드북 탐색법을 사용하면 원음 스펙트럼과 합성음 스펙트럼간의 에러를 최소화하는

모델 파라미터를 추정할 수 있다. 간략화된 추정법의 경우 간단한 가정을 적용함으로써 코드북 탐색법에 비해서 훨씬 적은 연산량으로 모델 파라미터를 추정할 수 있다. 제안한 MMBE 음성 모델을 적용하여 구현한 MMBE 부호화기의 경우 주파수 영역 혼합비를 벡터 양자화함으로써 기존의 MBE 부호화기에 비해서 더 낮은 전송률을 갖도록 설계할 수 있었다. 본 논문에서는 2.6 kbps MMBE 부호화기와 2.9 kbps MBE 음성 부호화기를 각각 구현하고 두 시스템의 성능을 비교하였다.

객관적 평가 방법으로 사용한 SD의 경우 평균 16% 감소됨으로써 제안한 MMBE 부호화기가 기존의 MBE 부호화기에 비해서 더 낮은 전송률에서도 작은 왜곡을 갖는 합성음을 생성해 낼 수 있었다. 주관적 평가 방법으로 실시한 MOS 테스트에서는 제안한 MMBE 부호화기가 4.04점, 기존의 MBE 부호화기가 4.14점을 받았다. MOS 테스트의 결과로부터 제안한 MMBE 부호화기가 기존의 MBE 부호화기에 비해 낮은 전송률에서도 더 우수한 청취 음질을 갖는 합성음을 생성해 낼 수 있었다. 결과적으로, 제안한 음성 모델 및 파라미터 추정 방법을 사용하면 기존의 MBE 음성 모델을 사용한 부호화기에 비해 더 낮은 전송률에서 더 작은 스펙트럼 왜곡을 갖는 합성음을 생성해 낼 수 있다. 또한, 더 낮은 전송률에서도 청각적으로 대등한 음질의 합성음을 생성해 낼 수 있다는 것을 실험적으로 입증하였다.

현재 제안한 시스템에서는 청각적 특성이 고려되지 않은 에러 척도를 최소화하는 모델 파라미터를 추정한다. 저 전송률 음성 부호화기의 목적은 청각적 손실없이 음성 정보를 가능한 한 적은 정보량으로 표현하는 것이므로 향후 지각 가중 필터 (perceptual weighting filter) 등 청각적 특성을 고려한 에러 척도를 사용하여 모델 파라미터를 추정함으로써 주관적 평가 성능을 더 개선할 수 있을 것으로 기대된다.

## 참고 문헌

1. S. Furui, *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker, Inc., 1992.
2. J. P. Campbell, Jr., T. E. Tremain, "Voiced/Unvoiced classification of speech with applications to the U. S. government LPC-10e algorithm," *proc. of ICASSP*, 473-476, 1986.
3. T. E. Tremain, "The government standard linear predictive coding algorithm: LPC-10," *Speech Technology*, 40-49, April 1982.
4. M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): High quality speech at very low bit rates," *proc. of*

ICASSP, 937-940, 1982.

5. I. A. Gerson and M. A. Jasiuk, "Vector sum excited linear prediction (VSELP)," *IEEE Workshop on Speech Coding for Telecommunications*, 66-68, 1989.
6. D. W. Griffin and J.-S. Lim, "Multiband excitation vocoder," *IEEE Trans. on Acoustics, Speech and Signal Processing*, ASSP-36, 1223-1235, 1988.
7. D. W. Griffin, *Multi-Band Excitation Vocoder*, Technical Report No. 524, March 1987.
8. Digital Voice Systems Inc., *Inmarsat-M Voice Coding System Description: Draft Version 1.3*, 8-22, February 1993.
9. D. Rowe, W. Cowley, and A. Perkis, "A multiband excitation linear predictive speech coder," *proc. of Eurospeech*, 239-242, 1991.
10. S. Yeldener, A. M. Kondoz, and B. G. Evans, "High quality multi-band LPC coding of speech at 2.4Kb/s," *IEE Electronics Letters*, 27 (14): 1287-1289, July 1991.
11. A. M. Kondoz, *Digital Speech: Coding for Low Bit Rate Communication Systems*, John wiley and sons, 239-272, 1994.

## 저자 약력

### • 한 우 진 (Woo-Jin Han)



1995년 2월: 한국과학기술원 전산학과(학사)  
 1997년 2월: 한국과학기술원 전산학과(석사)  
 2002년 2월: 한국과학기술원 전산학과(박사)  
 2002년 2월 ~: SL2(주) 기술연구소 연구소장  
 \* 주관심분야: 음성코딩, 청각처리, 특징추출

### • 오 영 환 (Yung-Hwan Oh)



1972년: 서울대학교 공과대학(학사)  
 1974년: 서울대학교 교육대학원(석사)  
 1980년: Tokyo Institute of Technology 정보공학 전공(박사)  
 1981년 ~ 1985년: 충북대학교 컴퓨터공학과 조교수  
 1983년 ~ 1984년: University of California, Davis 연구교수  
 1995년 ~ 1996년: Carnegie-Mellon University 연구교수  
 1985년 ~ 현재: 한국과학기술원 전산학과 교수

\* 주관심분야: 음성인식, 음성합성, 음성코딩, 화자인식, 대화관리, 신경회로망, 전문가 시스템