

# 이중 전송률(2.4/4.0 kbps)을 갖는 개선된 하모닉-CELP 음성부호화기

정희원 김 경 민,\*\*\*, 윤 성 완\*, 최용수\*\*, 박영철\*, 윤대회\*, 강태익\*\*

## Improved Harmonic-CELP Speech Coder with Dual Bit-Rates(2.4/4.0 kbps)

Kyong-Min Kim\*\*\*\*, Sung-Wan Yoon\*, Yong-Soo Choi\*\*\*, Young-Cheol  
Park\*\*, Dae-Hee Youn\*, Tae-Ik Kang\*\*\* Regular Members

요 약

본 논문에서는 기존에 제안한 바 있는 EHC(Efficient Harmonic-CELP) 부호화기에 기반한 이중 전송률(2.4/4.0 kbps)의 개선된 하모닉-CELP(Improved Harmonic-CELP: IHC) 음성 부호화기를 제안한다. 제안된 IHC는 유/무성음 판별에 따라 유성음에서는 하모닉 추정, 하모닉 양자화, 하모닉 합성 및 잡음 혼합을 통한 자연성 제어 등의 과정을 통해, 무성음에서는 고속 CELP 방식을 통해 음성신호를 부호화/복호화한다. 또한 제안된 IHC는 EHC의 유/무성음 판별과 하모닉 추정 성능을 개선한다. 피치 이득과 에너지 등을 이용한 유/무성음 판별은 프레임 내의 에너지 변화 등에 의해 재판별된다. 하모닉 부호화에서는 점점 추출과 델타 조정 방법을 이용하여 정확한 하모닉 추정을 하고, 혼합하는 잡음의 대역과 양을 적절하게 조절하여 자연성을 제어한다. 제안된 IHC 부호화기에 성능평가 결과, 비교 부호화기인 HVXC 연산량의 약 40%이내의 연산량으로 우수한 음질을 보임을 확인하였다.

ABSTRACT

This paper presents a dual-rate (2.4/4.0 kbps) Improved Harmonic-CELP(IHC) speech coder based on the EHC(Efficient Harmonic-CELP) which was presented by the authors. The proposed IHC employs the harmonic coding for voiced and the CELP for unvoiced segments. In the IHC, an initial voiced/unvoiced estimate is obtained by the pitch gain and energy. Then, the final V/UV mode is decided by using the frame energy contour. A new harmonic estimation combining peak picking and delta adjustment provides a more reliable harmonic estimation than that in the EHC. In addition, a noise mixing scheme in conjunction with an improved band voicing measurement provides the naturalness of the synthesized speech. To demonstrate the performance of the proposed IHC coder, the coder has been implemented and compared with the 2.0/4.0 kbps HVXC(Harmonic eXcitation Vector Coding) standardized by MPEG-4. Results of subjective evaluation showed that the proposed IHC coder and produce better speech quality than the HVXC, with only 40% complexity of the HVXC.

I. 서 론

최근 들어 음성 부호화기의 응용 분야가 다  
양해지면서 저전송률 부호화기에 대한 요구가

\* 연세대학교 전기전자공학과 미디어통신신호처리 연구실(yocello@mcp.yonsei.ac.kr), \*\* 연세대학교 정보기술학부

\*\*\* LG전자 기업통신연구소(cando@lge.com), \*\*\*\* 삼성전자 TN총괄 무선사업부(fearless@korea.com)

논문번호 : 020248-0522, 접수일자 : 2002년 5월 22일

※본 연구는 LG전자의 지원으로 수행되었습니다.

증대되었다. 이에 따라 4.0 kbps 이상의 전송률에선 CELP(Code Excited Linear Prediction)[1] 방식의 부호화기가, 4.0 kbps 미만의 전송률에서는 하모닉 부호화를 이용한 부호화기 [2-4]가 제시되어 사용되고 있다. 하지만 4.0 kbps 미만의 전송률에서 CELP 부호화 알고리즘은 양자화 비트가 부족하여 음질이 저하되는데 반해 하모닉 부호화 알고리즘은 적은 비트로도 음질에 큰 영향을 주는 유성음의 주기성을 잘 모델링한다. 이에 따라 4.0 kbps이하의 저전송률에서 유성음을 잘 표현할 수 있는 하모닉 부호화 방법과 무성음을 잘 표현할 수 있는 CELP 부호화 방법을 혼합한 하모닉-CELP 부호화 알고리즘이 제시되었다[5-6]. 이러한 혼합형 부호화 알고리즘의 성능개선을 위해 유/무성음이 혼재해 있는 전이구간을 위한 부호화 연구도 활발히 진행되고 있다 [7].

하모닉-CELP 부호화기의 대표적인 예로는 MPEG-4의 음성부호화 표준안인 HVXC(Harmonic Vector eXcitation Coder) [5]가 있다. HVXC는 가변차원 하모닉 벡터의 양자화와 고속 하모닉 합성, 실수 피치를 이용한 하모닉 크기 추정, 잡음 혼합을 통한 자연성 제어 등을 특징으로 한다. 그러나 실수 피치로 하모닉 추정시 많은 연산량이 요구되고, 전대역에 고정된 비율로 잡음을 혼합하므로 효율적인 자연성 제어를 하지 못하는 단점이 있다 [6], [8].

본 논문에서는 HVXC에 비해 적은 연산량과 우수한 음질을 갖는 하모닉-CELP 부호화 알고리즘을 제안한다. 본 논문에서 제안하는 개선된 하모닉-CELP (Improved Harmonic-CELP: IHC) 부호화기는 기존에 제안한 EHC [6], [8] 구조를 기반으로 무성음구간은 RP-VSELP(Regular Pulse-Vector Sum Excitation Linera Prediction) [9]를 사용하고, 유성음구간은 하모닉 부호화를 사용한다. 성능개선을 위해 유/무성음 판별과 하모닉 추정, 자연성 제어 부분에 대해 개선된 알고리즘을 적용하였고, 프레임 구조를 개선하여 이중 전송률에서 동작할 수 있도록 양자화 성능을 최적화하였다.

제안된 IHC 부호화기의 성능평가를 위해 2.4/4.0 kbps로 비트 할당하여 컴퓨터 모의 실험을 실시한 결과, 제안된 부호화기는 HVXC의 40%에 해당하는 연산량으로 더 나은 음질을 합성해냈다.

본 논문의 II 장에서는 하모닉-CELP 부호화기의 기본 구조를 살펴보고, III 장에서는 제안된 부호화기의 구성과 유/무성음 판별, 개선된 하모닉 추정과 자연성 제어 알고리즘에 대해 자세히 기술한다. IV 장에서는 제안된 부호화기의 성능평가 결과를 보여주며, 마지막으로 V 장에서 결론을 맺는다.

## II. 하모닉-CELP 음성 부호화기

대부분의 하모닉-CELP 부호화기는 LPC(Linear Predictive Coding) 모델을 사용하고, 여기 신호에 대해 하모닉 방법 또는 CELP 부호화한다. 하모닉 부호화 과정은 하모닉 추정, 하모닉 양자화, 하모닉 합성, 자연성 제어로 이루어진다. 즉 LPC 잔차 신호에 대해 스펙트럼을 구하여 피치 추정 결과를 기반으로 하모닉 추정을 한다. 추정된 하모닉 벡터들은 가변 차원을 가지므로 이를 고정 차원으로 변환하여 양자화하거나 코드북을 가변 차원으로 만들어 벡터 양자화한다. 하모닉 합성 때에는 하모닉 벡터들을 역 양자화하고 위상은 예측하여 정현파의 합 또는 역 고속 푸리에 변환을 이용한 여기 신호 합성을 한다. 또한 합성신호의 자연성을 제어하기 위해 잡음혼합 방법 등을 사용하기도 한다. 무성음 구간에는 일반적인 CELP 알고리즘이나 고속 CELP 알고리즘을 사용한다.

하모닉-CELP와 같이 유/무성음에 따라 서로 다른 부호화 방법을 사용하는 음성부호화기는 유/무성음 판별이 잘못되는 경우, 부적합한 부호화 방법이 적용되기 때문에 급격한 음질저하가 일어난다. 또한 MBE(Multi Band Excitation)[2]나 HVXC [5]와 같이 실수 피치를 기반으로 하는 하모닉 추정을 하는 경우 연산량이 증가하고, 정수 피치를 사용하면 연산량이 적지만 하모닉 추정 오차가 많이 발생한다 [11]. 또한 하모닉 부호화의 특성상 합성 신호의 음질이 명료도가 높은 반면 자연성이 결여되는 단점에 대한 처리도 중요하다.

HVXC의 경우, 실수 피치를 이용한 하모닉 부호화와 전통적인 CELP 알고리즘의 사용으로 연산량이 많이 소요되고, 모드에 따라 고정된 비율로 잡음을 혼합함에 따라 다양한 음성 구간의 특성을 반영하지 못하는 단점이 있다.

### III. 제안된 하모닉-CELP 음성 부호화

본 논문에서는 기존에 제안했던 하모닉-CELP 부호화 모델인 EHC를 개선하고, 이중 전송률(2.4/4.0 kbps)을 지원하는 IHC(Improved Harmonic-CELP)를 제안한다. 제안된 IHC는 유/무성음 판별에 따라 무성음에서는 고속 CELP 알고리즘인 RP-VSELP [9]를 사용하고, 유성음에서는 정수 피치를 사용하여 고속 하모닉 부호화를 수행한다.

제안된 IHC의 부호화 과정은 그림 1과 같다. 입력 신호는 프레임 단위로 LPC 분석을 거치고, LPC는 LSP로 변환되어 양자화된다. LPC 분석 후, 잔차신호 영역에서 EVRC(Enhanced Variable Rate Codec) [12]의 피치추정 알고리즘을 하모닉-CELP에 맞게 보완해서 피치를 추정한다 [8]. 그리고, 피치 추정으로부터 얻은 피치 이득과 입력 신호를 이용하여 유/무성음 판별을 한다. 유성음 구간의 하모닉 부호화는 하모닉 벡터 추정 과정과 잡음 혼합을 위한 대역별 유성음 정도를 구하는 과정으로 이루어진다. 그러나 낮은 전송률(2.4 kbps)일 때에는 양자화의 효율을 고려하여 대역별 유성음 정도를 구하는 과정이 생략된다. 각각 구해진 파라미터들은 양자화되어 복호화기로 전송된다.

제안된 IHC 부호화기의 복호화 과정은 그림 2와 같다. 역 양자화 과정을 통해 복원된 각각의 파라미터를 사용하여 여기신호를 구한다. 높은 전송률(4.0 kbps)의 경우 합성된 여기 신호에 전송된 대역별 유성음 정도를 이용하여 잡음을 발생시켜 혼합하고, 낮은 전송률일 때는 정해진 방법에 의해 잡음을 발생시켜 혼합한다. 잡음이 혼합된 최종 여기 신호는 LPC 합성 필터와 포스트 필터를 거쳐 최종적인 음성 신호가 된다.

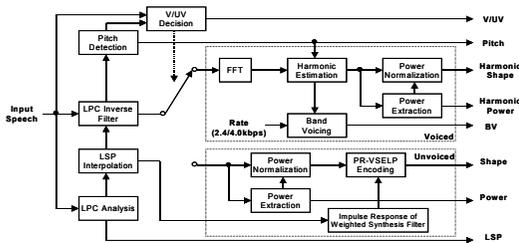


그림 1. 제안된 IHC 부호화기.

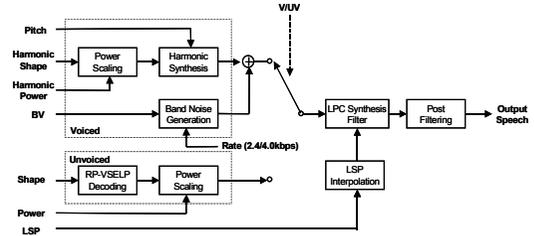


그림 2. 제안된 IHC 복호화기.

#### 3.1 유/무성음 판별

하모닉-CELP 부호화 방법에선 유/무성음 판별을 잘못 하였을 경우 합성신호에 불연속적인 파형으로 인한 잡음이 발생하여 음질이 크게 저하된다. 특히 무성음 구간을 유성음으로 판별하면 유성음 모델링에 의해 프레임 내의 에너지 분포가 고르게 되지 않아 합성신호에서 불연속적인 임펄스가 발생되고, 듣기에 거북한 불연속 잡음이 생긴다..

본 논문에서는 기존 EHC처럼 음성 신호의 에너지, 피치 예측 이득, 영교차율(Zero-Crossing Rate: ZCR), PKN(PeaKiNess) 등을 이용하여 1차 판별을 하고 [8], 잘못 판별했을 가능성이 있는 경우에 대해 2차 판별을 하여 정확한 유/무성음 판별이 이루어지도록 한다. 이를 위해 무성음이라 할 수 있을 만큼 에너지가 작은 구간에서 유성음이라 판별된 프레임에 대해 에너지 변화를 살피고, 최종적으로 PKN와 피치 예측 이득  $\beta$ 를 고려하여 다음과 같이 수정 과정을 수행한다.

$$\text{if} ( E < \gamma_0 \text{ and } \text{voicing} = 1 ) \{ \\ \text{if} ( E1/E2 < \rho \text{ or } E2/E1 > \rho ) \\ \text{if} ( PKN < \zeta_1 \text{ or } \beta < \zeta_2 ) \text{voicing} = 0 \\ \}$$

여기서  $E, E1, E2$ 와  $PKN$ 과  $ZCR$ 은 아래 식과 같이 주어지고,  $\gamma_0, \rho, \zeta_1, \zeta_2$ , 는 실험적으로 결정되며 본 논문에서는 각각 200, 2.0, 1.5, 0.6으로 한다.

$$E = \sqrt{\sum_{n=0}^{N-1} s^2(n) / N}, \quad (1)$$

$$E1 = \sqrt{\sum_{n=0}^{N/2-1} s^2(n) / (N/2)}, \quad (2)$$

$$E2 = \sqrt{\sum_{n=N/2}^{N-1} s^2(n) / (N/2)}, \quad (3)$$

$$PKN = \frac{\sqrt{\frac{1}{N} \sum_{n=0}^{N-1} r^2(n)}}{\frac{1}{N} \sum_{n=0}^{N-1} |r(n)|}, \quad (4)$$

$$ZCR = \frac{1}{2N} \sum_{n=1}^{N-1} |sgn[s(n)] - sgn[s(n-1)]|, \quad (5)$$

여기서  $s(n)$ ,  $r(n)$ ,  $N$ 은 각각 음성 신호, 잔차 신호, 프레임 길이이다.

각 단계의 유/무성음 판별 성능을 그림 3에 예시하였다. 그림 3에서 (a)는 원래 음성 파형을 나타내고 (b)는 1차 유/무성음 판별 [8]에 의한 합성 신호의 파형, (c)는 2차 유/무성음 판별에 의한 합성 신호의 파형을 나타낸다. (b)의 3400 샘플 부근과 5000 샘플 부근에서 무성음에 대해 유성음으로 판별한 부분이 발생하여 임펄스 형태의 불연속적인 파형이 나타났으나, (c)에서는 개선된 알고리즘을 통해 유/무성음 판별을 정확히 함으로써 불연속적인 파형이 나타나지 않도록 제어할 수 있음을 확인할 수 있다. 이로 인해 불연속 잡음이 제어되어 합성 신호의 음질도 향상되었다.

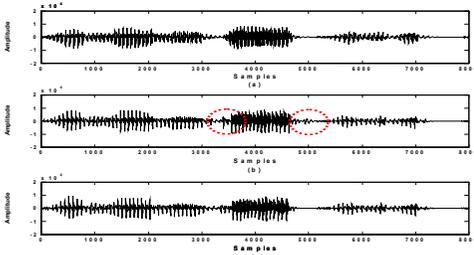


그림 3. 유/무성음 판별 성능 비교 :  
 (a) 원신호의 파형,  
 (b) 1차 유/무성음 판별에 의한 합성신호의 파형,  
 (c) 2차 유/무성음 판별에 의한 합성신호의 파형.

### 3.2 하모닉 추정

하모닉 추정 방법으로는 스펙트럼의 최고점을 추출하는 침점 추출(peak picking) 방법 [11]과 원 스펙트럼  $X_w(m)$ 과 합성 스펙트럼  $\widehat{X}_w(m)$ 의 오차 에너지  $E_l$ 을 최소화하도록 하

모닉 크기  $A_l$ 을 구하는 MMSE (Minimum Mean Squared Error) 방법 [2]등이 있다. 특히 MMSE 방법은 스펙트럼 상에서 합성을 통해 분석을 하므로 주파수 축에서의 분석-합성(analysis-by-synthesis) 부호화로도 볼 수 있는데 침점 추출 방법보다 연산량은 많지만 성능이 우수하다고 알려져 있다. 다음은 그 과정이다. 합성 신호의 스펙트럼은 하모닉 크기  $A_l$ 과 분석 윈도우 스펙트럼의 곱으로 주어지므로 원 신호의 스펙트럼과의 오차에너지를 최소화하는  $A_l$ 을 구한다. 식 (6)은 스펙트럼 오차에너지를 나타내고, 식 (7)은 합성 신호의 스펙트럼을 나타낸다.

$$E_l = \sum_{m=a_l}^{b_l} \{ |X_w(m)| - |\widehat{X}_w(m)| \}^2, \quad (6)$$

$$|\widehat{X}_w(m)| = A_l \left| W_R \left( \left[ \frac{N_2}{N_1} m - \frac{N_2}{2\pi} \omega_0 l + 0.5 \right] \right) \right|, \quad 1 \leq l \leq L, \quad L = \left\lceil \frac{\pi}{\omega_0} \right\rceil = \left\lceil \frac{P_0}{2} \right\rceil. \quad (7)$$

식 (6)에서  $X_w(m)$ 은 분석 윈도우가 씌워진 LPC 잔차 신호의  $N_1$  포인트 FFT 스펙트럼이다. 식 (7)에서  $W_R(m)$ 은 2 개 이상의 피치 주기 ( $P_0$ )를 포함할 수 있는, 길이가  $N_R$ 인 분석 윈도우  $w_R(n)$ 의  $N_2 (\gg N_1)$  포인트 FFT 스펙트럼이며,  $[x]$ 은 실수  $x$ 의 정수 부분을 취하는 연산을 나타낸다.  $a_l$ 과  $b_l$ 은  $l$ 번째 하모닉 대역의 시작과 끝을 의미하고,  $L$ 은 하모닉 개수이다. 식 (6)을  $A_l$ 에 대해 편미분  $\frac{\partial E_l}{\partial A_l} = 0$ 을 취하면 스펙트럼 오차 에너지를 최소화하는  $A_l$ 이 식 (8)과 같이 구해진다.

$$A_l = \frac{\sum_{m=a_l}^{b_l} |X_w(m)| \left| W_R \left( \left[ \frac{N_2}{N_1} m - \frac{N_2}{2\pi} \omega_0 l + 0.5 \right] \right) \right|}{\sum_{m=a_l}^{b_l} \left| W_R \left( \left[ \frac{N_2}{N_1} m - \frac{N_2}{2\pi} \omega_0 l + 0.5 \right] \right) \right|^2} \quad (8)$$

이때 비트 제한이나 연산량 제한 등으로 인해 입력 음성 신호에 대한 피치 추정의 정확도가 한정되어 있다면  $\omega_0 = \frac{2\pi}{pitch}$ 의 정수배  $w_0 \times l$ 로 표현되는 합성 신호의 하모닉 중심 주

과수와 원 신호의 하모닉 중심 주과수의 차이가 고주과수 대역으로 갈수록 커지게 되어 하모닉 추정이 부정확해진다. 따라서 식 (9)와 같이  $\Delta_l$ 을 이용하여 고주과수 대역의 하모닉 간격을 조정함으로써 스펙트럼 오차 에너지가 최소가 되도록 하모닉 크기를 구하는 방법인 델타 조정(DA) 방법 [11]을 통해 고주과수 대역에서 정확하게 하모닉을 추정할 수 있다.

$$A_l = \frac{\sum_{m=a_l}^{b_l} |X_w(m + \Delta_l)| \left| W_R \left( \left[ \frac{N_2}{N_1} m - \frac{N_2}{2\pi} \omega_0 l + 0.5 \right] \right) \right|}{\sum_{m=a_l}^{b_l} \left| W_R \left( \left[ \frac{N_2}{N_1} m - \frac{N_2}{2\pi} \omega_0 l + 0.5 \right] \right) \right|^2},$$

$$-d_l \leq \Delta_l \leq d_l, \quad d_l = \left\lfloor \frac{\alpha \omega_0}{L-1} (l-1) \right\rfloor, \quad 1 \leq l \leq L, \quad (9)$$

식 (9)에서 대역별 조정폭의 변화량을 나타내는 상수  $\alpha$ 는 0.5 이하이며 실험적으로 결정된다.

그러나 피치 변이가 일어나는 구간 등에서는 저주과수 대역에서도 하모닉 간격이 어긋나서 음질에 큰 영향을 주는 저주과수 대역의 하모닉 추정 오차가 발생하기도 한다. 본 논문에서는 저주과수 대역에서 발생하는 스펙트럼 오차를 줄이기 위해 합성 신호와 원 신호 스펙트럼의 하모닉 피크를 일치시키는 침점 추출(Peak Picking: PP) 방법을 결합하여 사용한다. 즉 각 하모닉 대역의 피크를 일치시키는 PP 방법과 DA 방법의 하모닉 간격 조정값,  $\tau_l$ 과  $\Delta_l$ 을 순차적으로 이용하는 PP-DA 방법을 통해 전주과수 대역에 걸쳐 정확한 하모닉 추정을 한다.

$$A_l = \frac{\sum_{m=a_l}^{b_l} |X_w(m + \tau_l + \Delta_l)| \left| W_R \left( \left[ \frac{N_2}{N_1} m - \frac{N_2}{2\pi} \omega_0 l + 0.5 \right] \right) \right|}{\sum_{m=a_l}^{b_l} \left| W_R \left( \left[ \frac{N_2}{N_1} m - \frac{N_2}{2\pi} \omega_0 l + 0.5 \right] \right) \right|^2}, \quad (10)$$

식 (10)에서  $\tau_l$ 은 합성 신호의 스펙트럼에서 각 하모닉 피크 위치에 해당하는  $w_0 \times l$ 의  $\pm(1 \sim 2)$  범위 내에서 원 신호 스펙트럼  $X_w(m)$ 의 최대값에 위치하도록 결정된다.

그림 4는 PP 방법의 사용 여부에 따른 하모닉 크기 추정 오차 정도를 나타낸다. DA 방법만을 사용해 하모닉 추정을 한 (a)는 저주과수 대역의 스펙트럼 오차가 존재함을 볼 수 있다. 이러한 저주과수 대역의 하모닉 크기 오차는 음질을 크게 왜곡시키므로 PP 방법을 사용하여

(b)에서처럼 저주과수 대역의 오차를 줄여 정확하게 하모닉을 추정할 수 있다.

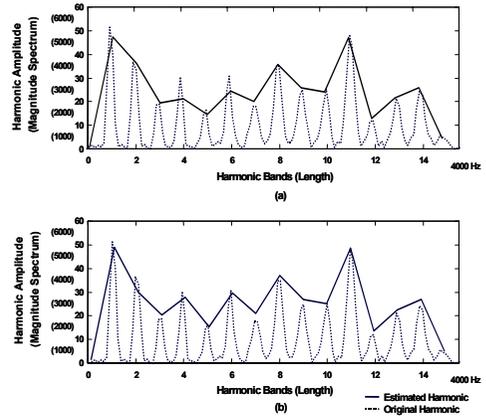


그림 4. 피크피킹(PP) 방법 사용 여부에 따른 하모닉 크기 추정 비교 :  
(a) DA 방법을 사용한 하모닉 추정.  
(b) PP-DA 방법을 사용한 하모닉 추정.

### 3.3 자연성 제어

실제 음성 신호를 관찰해 보면 유성음 구간이어도 고주과수 대역으로 갈수록 보통 하모닉(유성음) 성분이 줄고 잡음(무성음) 성분이 늘어난다. 그러나 하모닉 합성을 할 때는 하모닉 피크 사이의 간격을 일정하게 유지시켜주기 때문에 실제로 무성음 성분이 혼합되어 있는 음성 신호의 주과수 특성과 달라지게 된다. 이렇게 유/무성음이 혼합되어 있는 음성 신호를 유성음 성분만으로 표현하게 되면 명료도는 높아지나 자연성이 떨어진다. 자연스러운 음질을 위해 해신 하모닉 대역별 유성음 정도 (Band Voicing: BV)를 고려한 백색 잡음을 혼합하는 방법을 사용한다 [6], [8]. 그러나 BV을 잘못 추정하는 경우 오히려 과도한 잡음이 추가되어 잡음에 의해 음질이 저하될 수 있으므로 잡음이 혼합되는 주과수 대역과 양에 주의하여야 한다.

본 논문에서는 하모닉 대역별 유성음 정도  $BV_l$ 을 원 신호 스펙트럼과 합성 신호 스펙트럼의 정규화된 오차 에너지  $E_l$ 를 사용하여 식 (11)과 같이 정의한다. 하모닉 추정과 마찬가지로 BV 추정에서도 PP-DA 방법을 사용한다. BV를 적용한 각 대역별 하모닉 크기값은 식 (12)와 같다.

$$\begin{aligned}
 BV_l &= 1 - E_l \\
 &= 1 - \frac{\sum_{m=a_l}^{b_l} | |X_w(m+\Delta_l+\tau_l)| - |\widehat{X}_w(m)| |^2}{\sum_{m=a_l}^{b_l} |X_w(m+\Delta_l+\tau_l)|^2},
 \end{aligned} \tag{11}$$

여기서  $\tau_l$ 과  $\Delta_l$ 는 앞 절에서 설명한 하모닉 간격의 조정 값이다.

$$\begin{aligned}
 A_{v,l} &= BV_l \cdot A_l, & 1 \leq l \leq L \\
 A_{w,l} &= (1 - BV_l) \cdot A_l \cdot N_l, & 1 \leq l \leq L \\
 A_l &= A_{v,l} + A_{w,l}, & 1 \leq l \leq L
 \end{aligned} \tag{12}$$

여기서  $A_l$ 은 각 대역별 하모닉 합성을 위한 스펙트럼 크기이고,  $N_l$ 은 대역별로 첨가되는 잡음,  $A_{v,l}$ 과  $A_{w,l}$ 은 각각 유성음성분과 잡음혼합을 위한 하모닉 크기를 나타낸다.

그림 5는 하모닉 대역별 유성음 정도(BV)를 구할 때 DA 방법만을 사용한 경우와 PP-DA 방법을 사용한 예를 비교하여 보여준다. 그림에서 (a)는 LPC 잔차 신호의 스펙트럼을 나타내고, (b)는 (a)의 스펙트럼으로부터 BV를 추정하기 위해 DA 방법과 PP-DA 방법을 사용한 결과를 나타낸다. 대체적으로 하모닉 성분이 강한 저주파수 대역에서 BV는 1에 가까운 값을 갖고 하모닉 성분이 어긋나거나 잡음 성분이 존재하는 대역에서는 BV값이 0에 가까워지는 현상을 발견할 수 있다. 또한 (b)에서 DA 방법만을 이용하여 BV 추정된 경우 2 kHz 이하의 저주파수 대역에 속하는 6, 7, 8 번째의 하모닉에서 하모닉 성분이 강함에도 불구하고 BV 값이 낮게 추정되는 결과를 나타냈다. 이러한 경우 LPC 합성필터를 통해 저주파수 대역의 잡음이 증폭되어 전체적으로 과도한 잡음을 혼합하는 결과를 얻게 되므로 음질을 크게 저하시킨다. 그러나 PP-DA 방법으로 추정하면 2 kHz 이하의 낮은 주파수 대역에서도 DA 방법으로 추정된 BV보다 비교적 유성음 정도를 잘 표현할 수 있음을 볼 수 있다. 즉 하모닉 추정에서와 같이 PP 방법을 통해 과도한 잡음 혼합 원인인 저주파수 대역의 BV 추정 오차를 줄일 수 있고, 결과적으로 적절한 잡음 혼합을 통해 합성 신호의 자연성을 제어할 수 있다.

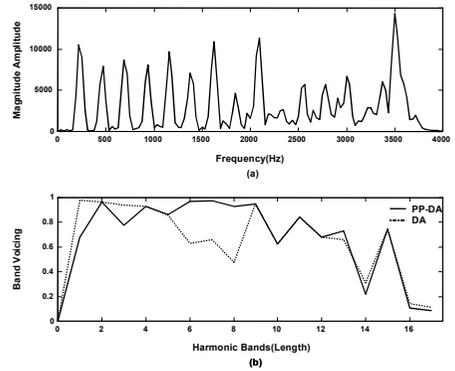


그림 5. 하모닉 대역별 유성음 정도(BV) 비교 :  
 (a) LPC 잔차 신호의 스펙트럼,  
 (b) PP-DA 방법과 DA 방법에 의한 BV추정.

이러한 BV정보는 양자화되어 전송되는데, 낮은 전송률일 때에는 충분한 양자화 비트를 할당할 수 없으므로 BV을 전송하지 않고 복호화기에서 BV을 만들어 잡음을 혼합함으로써 양자화 효율을 증대시킨다.

무성음에서 유성음으로 변하는 구간에 대한 BV은 식 (13)과 같다.

$$BV_l = \begin{cases} 0.98 & , 1 \leq l < \lfloor \frac{L}{2} \rfloor \\ 0.9 - \beta_1(l - \lfloor \frac{L}{2} \rfloor) & , \lfloor \frac{L}{2} \rfloor \leq l \leq L \end{cases} \tag{13}$$

여기서  $\beta_1$ 는 0.03으로 정하고,  $\lfloor x \rfloor$ 는 실수  $x$ 보다 크지 않은 최대 정수를 뜻한다.

유성음이 지속되는 구간에서는 2 kHz 이하 대역에서 잡음 성분이 없다고 식 (14)와 같이 BV를 정하여 가정해 잡음을 혼합한다.

$$BV_l = \begin{cases} 1.0 & , 1 \leq l < \lfloor \frac{L}{2} \rfloor \\ 0.85 & , \lfloor \frac{L}{2} \rfloor \leq l < \lfloor \frac{3L}{4} \rfloor \\ 0.85 - \beta_2(l - \lfloor \frac{3L}{4} \rfloor) & , \lfloor \frac{3L}{4} \rfloor \leq l \leq L \end{cases} \tag{14}$$

여기서  $\beta_2$ 는 0.05로 정한다.

#### IV. 비트 할당 및 성능 평가

제안된 IHC 부호화기를 이중 전송률(2.4/4.0 kbps)을 갖도록 비트 할당하여 성능평가를 하였다. 성능평가를 위해서 8 kHz로 샘플링되고 16 비트로 선형 양자화된 음성 신호를 사용했다. 프레임은 20 ms이고, 부프레임은 5 ms씩 4개로 분할되는데 LPC 분석과 피치 추정, 하모

닉 추정을 위해 10 ms 길이의 예견 구간을 갖는다. 유성음은 프레임 단위로, 무성음은 부프레임 단위로 부호화 및 복호화된다.

제안된 IHC의 2.4/4.0 kbps 전송률에 대한 비트할당은 표 1과 같다. 4.0 kbps에서 하모닉 크기를 양자화할 때는 음질에 더 큰 영향을 주는 저주파수 대역인 1.5 kHz 이하 대역의 각 0.5 kHz 대역마다 9, 8, 8 비트를 더 할당하여 2.4 kbps로 양자화한 후 발생하는 양자화 오차를 줄였다. 대역별 유성음 정도는 4.0 kbps일 때만 4 비트로 백터 양자화하고 2.4 kbps일 때는 양자화 실험 결과 대역별 유성음 정도에 대해 4비트 이하로 비트 할당을 하는 것보다 하모닉 정보에 대해 비트 할당하는 것이 더 좋은 성능을 나타냈으므로 양자화하지 않았다.

제안된 IHC의 연산량을 평가하기 위해 HVXC의 연산량과 비교하였다. 표 2에서 각 부호화기의 두 전송률에 대한 연산량을 부동 소수점 시뮬레이션을 통한 프로파일 기능을 사용하여 ms 단위로 나타냈다. 전체적으로 전송률과 관계없이 IHC 부호화기의

표 1. 제안된 2.4/4.0 kbps IHC 부호화기 비트 할당.

	2.4 kbps		4.0 kbps	
	UV	V	UV	V
LPC	15	24	27	27
V/UV	1	1	1	1
Pitch	-	7	-	7
BV	-	-	-	4
Power	4×4	6	5×4	6
Shape	4×4	10	8×4	10+25
Total/20ms	48	48	80	80
kbps	2.4	2.4	4.0	4.0

표 2. HVXC와 IHC 부호화기의 연산량 비교.

Rate (kbps)	HVXC (ms)		IHC (ms)	
	Enc.	Dec.	Enc.	Dec.
2.0/2.4	2632.5	990.6	977.8	336.1
	(10%)	(10%)		
4.0/4.0	3367.2	1023.2	1155.7	338.9
	(10%)	(10%)		

표 3. MOS 평가 결과.

	female	male
Original	4.58	4.54
IHC 2.4kbps	2.84	3.28
IHC 2.4kbps	3.11	3.65
HVXC 2.0kbps	2.76	2.83
HVXC 4.0kbps	3.20	3.50

연산량은 HVXC의 40%이하이다. 두 부호화기의 경우 최적화되지 않은 C 코드 상태로 연산량을 비교하였지만 제안된 IHC 부호화기가 HVXC보다 연산량 면에서 훨씬 효율적이라는 사실을 알 수 있다.

제안된 2.4/4.0 kbps IHC 부호화기를 비교 부호화기인 HVXC와 상대적으로 음질 평가하기 위해 MOS(Mean Opinion Score) 평가를 실시하였다. 음성 샘플로는 조용한 환경에서 녹음되고 8 kHz로 샘플링된 8 초 길이의 서로 다른 남녀 화자 각각 8 문장씩 총 16 개 문장을 사용하였고, 청취 대상은 성인 남녀 20명으로 하였다. 원음도 청취 대상에 포함하였으므로 절대적인 MOS 보다는 상대적 차이가 의미가 있다. MOS 평가 결과는 표 3에 나타났다.

MOS 평가 결과 제안된 IHC는 여성화자의 경우 HVXC와 비슷한 음질을 나타냈고, 남성화자의 경우 HVXC보다 우수한 음질을 나타냈다.

### V. 결론

본 논문에서는 이중 전송률(2.4/4.0 kbps)을 지원하면서 음질과 연산량 면에서 향상된 하모닉-CELP 부호화기인 IHC 부호화기를 제안하였다. 제안된 부호화기는 음성 신호의 에너지 변화에 따른 유/무성음 재판별 과정을 수행하고, 점점 추출 방법으로 델타조정 방법을 보완하여 더욱 정확한 하모닉 추정과 하모닉 대역별 유성음 정도 추정을 한다. 향상된 유/무성음 판별과 하모닉 부호화를 통해 제안된 2.4/4.0 kbps IHC 부호화기는 비교부호화기인 HVXC에 비해 40% 이내의 연산량을 가지면서 HVXC에 비해 여성화자의 경우 비슷한 음질을, 남성화자

의 경우 우수한 음질을 갖는 실험 결과를 나타냈다. 현재는 개선된 잡음 혼합 모델과 전이 구간 모델에 대해 연구 중이고, 고정 소수점 DSP (Digital Signal Processor)을 이용한 실시간 구현을 진행하고 있다.

### 참고문헌

[1] M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction(CELP): High Quality Speech at Very Low Bit Rates," *IEEE Proc. ICASSP 85*, pp.151.1-25.1.4, 1985.

[2] D. W. Griffin and J. S. Lim, "Multi-Band Excitation Vocoder," *IEEE Trans. on Acoust. Speech and Signal Proc.*, vol.36, No.8, pp.1223-1235, 1988.

[3] W. B. Kleijin and K. K. Paliwal, *Speech Coding and Synthesis*, Elsevier, 1995.

[4] A. V. McCree, T. P. Barnwell III, "A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding," *IEEE Trans on Acoust. Speech and Signal Proc.*, vol.3, No. 4, pp.242-250, 1995.

[5] ISO/IEC FCD 0.1 Subpart 2, "Information Technology Very Low Bit Rate Audio-Visual Coding," 1998.

[6] Y. S. Choi, S. K. Ryu, Y. C. Park, and D. H. Youn, "Efficient Harmonic-CELP Based Hybrid Coding of Speech at Low-Bit Rate," *Proc. of ICSLP 2000*, vol. II, pp.210-213, 2000.

[7] T. Wang, K. Koishida, V. Cuperman and A. Gersho, "A 4kb/s Hybrid MELP/CELP Coder with Alignment Phase Encoding and Zero Phase Equalization," *Proc. of ICASSP 2000*, pp.1379-1382, 2000.

[8] 최용수, 김경민, 윤대희, "효율적인 하모닉-CELP 구조를 갖는 저 전송률 음성부호화기," *한국음향학회지*, 제20권, 제5호, pp.3

5-47, 2001년 7월.

[9] Y. S. Choi, S. W. Park, and D. H. Youn, "Fast Vector-Sum Codebook Search Method for Low Bit Rate Speech Coding," *IEE Electronic Letters*, vol.33, No.6, 1997.

[10] Douglas B. Paul, "The Spectral Envelope Estimation Vocoder," *IEE E Trans. on Acoust. Speech and Signal Proc.*, vol.29, No.4, pp.785-794, 1981.

[11] Y. S. Choi and D. H. Youn, "Fast Harmonic Estimation Using a Low Resolution Pitch for Low Bit Rate Harmonic Coding," *Proc. of Eurospeech 2001*, vol.2, pp.2447-2450, Sep. 2001.

[12] Proposed TIA/EIA/PN-3292 Standard - Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems, Official Ballot Version, Qualcomm Inc., Apr. 1996.

김 경 민(Kyung-Min Kim)

정회원



1999년 8월 : 연세대학교 전자공학과(공학사)  
 2001년 8월 : 연세대학교 전기전자공학과(석사)  
 2001년 9월~현재 : 삼성전자 TN총괄 무선사업부 근무

<주관심분야> 음성부호화, DSP실시간 구현

윤 성 완(Sung-Wan Yoon)

정회원

한국통신학회 논문지 제26권 4B호 참고

최 용 수(Yong-Soo Choi)

정회원

한국통신학회 논문지 제26권 1B호 참고

박 영 철(Young-Cheol Park)                   정회원  
한국통신학회 논문지 제24권 9B호 참고

윤 대 희(Dae-Hee Youn)                   정회원  
한국통신학회 논문지 제26권 1B호 참고

강 태 익(Tae-Ik Kang)                   정회원



1979년: 연세대학교 전자 공  
학과(공학사)  
1981년: 연세대학교 전자 공  
학과(석사)  
1990년: Polytechnic Uni  
versity 전기공학과(박사)  
1981~1984년 : 육군사관

학교 조교수  
1990년~현재:LG전자   기업통신연구소   책임연구  
원

<주관심분야> 디지털영상/음성신호처리, VoIP, A  
TM