

광대역 음성부호화기를 위한 매칭퍼슈잇 알고리즘과 CELP 방법을 이용한 고대역 부호화 방법

Highband Coding Method Using Matching Pursuit Estimation and CELP Coding for Wideband Speech Coder

정 규 혁*, 안 영 옥**, 김 중 학*, 신 재 현*, 서 상 원*, 황 인 관*, 이 인 성*
(Gyu-Hyeok Jeong*, Yeong-Uk Ahn**, Jong-Hark Kim*, Jae-Hyun Shin*, Sang-Won Seo*
In-Kwan Hwang* In-Sung Lee*)

*충북대학교 전자공학과, **코아로직

(접수일자: 2005년 9월 6일; 채택일자: 2005년 10월 13일)

본 논문에서는 대역분할 광대역 음성부호화기와 이를 위한 고대역 부호화 방법과 구조를 제안한다. 제안하는 광대역 음성부호화기는 광대역 입력 음성신호를 저대역 신호 (0Hz-4kHz)와 고대역 신호 (4kHz-8kHz)로 나눈다. 저대역 신호는 G.729 Annex E로 부호화하고, 고대역 신호는 4kbps의 전송률로 제안하는 방법으로 압축된다. 고대역 신호는 LPC 분석 후 신호특성에 따라 모드를 분류된다. stationary 모드에서는 매칭퍼슈잇 알고리즘과 CELP 방법으로 부호화하는 다단계 구조의 혼합 여기신호모델이 적용되며, nonstationary 모드에서는 CELP 방법으로 부호화된다. 제안한 광대역 음성부호화기의 성능을 주관적 방법으로 G.722 48kbps SB-ADPCM, G.722.2 12.85kbps ACELP와 비교를 하였다. 제안한 부호화기는 G.722보다 나은 성능을 보이고, G.722.2보다 나쁘지 않은 성능을 가지는 것을 확인하였다.

핵심용어: 광대역 음성부호화기, 매칭퍼슈잇 알고리즘, CELP, 스펙트럴 크기, 위상, VQ

루고분야: 음성처리 분야 (2.2)

In this paper, a split bandwidth wideband speech coder and its highband coding method are proposed. The coder uses a split-band approach, where the wideband input speech signal is split into two equal frequency bands from 0-4kHz and 4-8kHz. The lowband and the highband are coded respectively by the 11.8kb/s G.729 Annex E and the proposed coding method. After the LPC analysis, the highband is divided by two modes according to the properties of signals. In stationary mode, the highband signals are compressed by the mixture excitation model; CELP algorithm and MP (Matching Pursuit) algorithm. The others are coded by the only CELP algorithm. We compare the performance of the new wideband speech coder with that of G.722 48kbps SB-ADPCM and G.722.2 12.85kbps in a subjective method. The simulation results show that the performance of the proposed wideband speech coder has better than that of 48kbps G.722 and no better than that of 12.85kbps G.722.2.

Keywords: Wideband Speech Coder, Matching Pursuit(MP), CELP, Spectral Magnitude, Phase, VQ

ASK subject classification: Speech Signal Processing (2.2)

1. 서론

최근 네트워크 기술의 발달과 더불어 좀 더 넓은 음성 대역폭의 사용이 가능해지면서 현대역 음성부호화기에

비해 음질이 우수한 광대역 음성부호화기에 많은 관심이 모아지고 있다. 광대역 음성부호화기는 이동통신이나 기존 전화망과의 인터페이스가 필요 없는 분야를 중심으로 다양한 분야에서 사용될 것으로 보인다. 이와 같이 수요가 증가함에 따라 국제 표준화 기구에서는 1988년부터 광대역 음성관련 표준화가 활발하게 진행되어왔다[1]. 최근 ITU-T의 Q.9/WP3/SG16에서는 가변 비트를 광대

역 음성부호화기의 표준화를 진행중에 있다[2]. 가변 비트를 광대역 음성부호화기는 패킷 스위칭 (Packet Switching)방식의 통신망에서 7kHz 대역폭 통신을 기본으로 하며, 음성부호화기의 패킷을 여러 계층으로 나누어 주어진 채널 환경에서 Congestion 발생이나 패킷손실의 발생할 경우 최적의 통신이 가능하도록 저대역에 해당하는 패킷만을 이용함으로써 협대역 통신을 가능하게 한다.

대역분할 구조의 광대역 음성부호화기[3-5]는 계층적 구조를 위해 고대역과 저대역으로 나누고 독립적으로 부호화한다. 저대역은 기존의 네트워크 장치들과의 호환성을 위해 표준화된 협대역 음성부호화기를 사용하며, 고대역 신호의 부호화는 일반적으로 파라미터 기반의 방식이나 Transform 방식을 사용한다. 하지만 고대역 부호화시 비트전송율이 낮아짐에 따라 생기는 주파수 왜곡으로 음질 저하가 발생한다[6]. 이같은 문제 해결을 위한 접근방법 중에 대표적인 DCT 계수의 효율적인 양자화 방법이 많이 연구되었다[7-10]. 본 논문에서는 고대역에 존재하는 하모닉 성분을 효율적으로 부호화한다면 음질 향상을 이룰 수 있다는 가정 하에 Transform 방식과 달리 저전송률 협대역 음성부호화기에서 성능이 입증된 HVXC (Harmonic Vector eXcitation Coder)와 비슷한 하모닉-CELP 방법을 고대역 신호의 특성에 맞게 수정 적용하여 새로이 접근한다. 고대역 신호의 특성을 연구하여 정현파 모델과 CELP에 기반을 둔 효율적인 부호화 방법과 구조를 제안하였고, 신호의 특성에 맞는 부호화 구조를 결정하여 최종 광대역 음성부호화기를 개발하였다. 이때 저대역 부호화기는 G.729 Annex E[11]를 사용한다.

본 논문에서는 고대역 신호의 특성에 적합한 정현파 모델과 CELP 방법의 효율적인 부호화 구조를 중심으로 다음과 순서로 전개한다. 2장에서는 제안한 광대역 음성부호화기의 전체적인 구조에 대해 설명하고, 3장에서는 고대역 음성부호화기의 구조와 제안하는 모델링 방법에 대해 설명한다. 4장에서는 파라미터들의 양자화 방법과 비트할당에 대해 설명한다. 5장에서는 성능 평가를 6장에서 결론을 맺는다.

II. 제안한 광대역 음성부호화기의 구조

제안하는 광대역 음성부호화기의 전체적인 구조는 그림 1과 같다. 매 20ms 프레임마다 8kHz의 대역폭을 가지는 광대역 음성신호에 대하여 QMF를 통과시켜 저대

역 신호 (0-4kHz)와 고대역 신호 (4-8kHz)로 분리하고 각 신호는 다운샘플링된다. 여기서 사용된 QMF는 통과대역은 0.48π 이며, 전이대역은 0.04π 을 가지는 96탭의 FIR 필터이고, 합성시 생기는 알리아싱과 스펙트럼 크기의 왜곡을 최소화하도록 설계되었다. 저대역 신호는 G.729 Annex E[11]로 부호화 한다. 고대역 신호는 LPC 분석과 정현파 모델링이 가능하도록 π 만큼 쉬프트시키고, 제안하는 부호화 방법에 의해 압축된다. 이때, 비트 할당의 효율을 높이고 정확한 피치 주기를 예측을 위해 저대역 전송 파라미터인 피치 정보를 이용한다.

복호화기에서는 역 과정을 통해 광대역 음성신호가 된다. 이 신호를 7kHz의 차단 주파수를 갖는 저역 필터를 통과시켜 최종 출력을 얻는다. 15.8kbps의 전송률을 가지며, 고대역 전송률은 4kbps가 된다. 제안된 광대역 음성부호화기는 20ms 프레임 길이를 사용하며 5ms의 look-ahead 와 6ms의 QMF 지연을 가지고 있어 총 31ms의 지연값을 갖는다.

III. 제안한 고대역 음성부호화기

3.1. 고대역 부호화기 구조

그림 2는 고대역 부호화기의 블록도이다. 광대역 음성신호의 4-8kHz의 대역폭에 해당하는 정보가 포함된 고대역 신호는 매 20ms 마다 6차 LPC 분석을 거친다. LP 잔여신호는 모드 판별기에서 신호의 특성에 따라 stationary 모드와 nonstationary 모드로 구별된다. 동일한 부호화 구조를 위해 stationary 모드에서는 매칭퍼슈잇으로 분석/합성 과정을 거치고 이때의 잔차 신호는 순환 코드북으로 모델링하는 다단계 구조를 가지고, nonstationary 모드에서는 펄스 코드북으로 모델링한 후 펄스 코드북의 기여도를 뺀 신호를 순환 코드북으로 모델링하는 다단계 구조로 모델링된다.

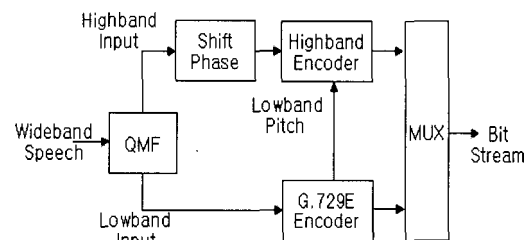


그림 1. 제안하는 광대역 음성부호화기 구조
Fig. 1. Structure of proposed wideband speech encoder.

3.2. 모드 판별기

모드 판별기는 10ms의 부프레임에서 동작하며, 저대역 신호의 특징을 고려하여 정지성(stationary) 신호에서는 stationary 모드로 분류되고, 그렇지 않은 구간에서는 nonstationary 모드로 분류한다. [11]에서 소개된 고대역 신호의 모드 판별과 같이 '저대역 신호가 주기적일 때에만 고대역 신호에도 주기적이다'라는 가정을 전제한다. 주기적인 신호를 모델링하는 정현파 모델의 사용여부를 결정하는 것은 음질에 큰 영향을 미친다. 이를 위해 고대역의 주기적인 특성의 강약과 하모닉의 유무 정도를 검출해내는 다음의 알고리즘을 새로이 보완 적용한다: 1차 및 2차 예측 계수의 평균(Average 1st and 2nd Prediction Coefficient: APC)과 정규화 된 주기적 상관값(Normalized Correlation of a periodic lag: NC). 이는 주기적인 신호를 모델링하는 정현파 모델의 사용여부를 결정하는 중요한 평가 요소가 된다.

음성 샘플들간의 상관성이 저대역 신호보다 떨어지는 특성을 가지는 고대역 신호 특성을 감안하여 1·2차 예측 계수들의 평균값(APC)을 사용하였다. 이는 음성 샘플간의 상관성이 클 경우 큰 값을 가지게 된다. 또한 보다 정확한 저대역의 피치정보를 이용하여 LP 잔여신호의 주기적 특성의 강약의 정도를 나타내는 정규화된 주기적 상관값(NC)을 사용하므로써 stationary 구간의 신호를 검출한다.

$$APC = \frac{\sum_{n=0}^{HFRM-2} \zeta(n)\zeta(n+1) + \sum_{n=0}^{HFRM-3} \zeta(n)\zeta(n+2)}{2 \sum_{n=0}^{HFRM-1} \zeta^2(n)} \quad (1)$$

$$NC = \frac{\sum_{n=0}^{T_p-1} r(n)r(n-T_p)}{\sum_{n=0}^{T_p-1} r^2(n-T_p)} \quad (2)$$

여기서 $HFRM$ 은 한 프레임인 160샘플을 의미하고, T_p 는 저대역의 피치로부터 계산된 고대역 신호의 피치 주기를 나타낸다. 또한 $\zeta(n)$ 은 전처리된 음성신호, $r(n)$ 은 LP 잔여신호를 나타낸다.

모드선택에 적용되는 알고리즘은 다음과 같이 정리된다.

• 저대역에서

- 잔여신호의 또렷함 비율 (Sharpness Rate: SR)[12]
- 영교차율 (Zero-crossing Rate: ZR)[12]
- 전처리된 음성 신호의 1차 예측 계수 (First-order Prediction Coefficient: FPC)[10]

• 고대역에서

- 영교차율 (Zero-crossing Rate: ZR)[12]
- 잔여신호의 에너지 비율 (Energy Rate: ER)[10]
- 1차 및 2차 예측 계수의 평균 (Average 1st and 2nd Prediction Coefficient: APC)
- 정규화된 주기적 상관값 (Normalized Correlation of a periodic lag: NC)

3.3. stationary 모드의 부호화

[13]에서 제안된 정현파 모델에서는 FFT 피크 검출방법으로 파라미터를 검출하고, 스펙트럴 크기는 선형 보간을 하고, 위상은 큐빅(cubic) 보간을 한다. 이는 분석 프레임의 반에 해당하는 지연과 과거 파라미터의 사용이 필수적이다. 고대역 패킷은 채널 상태에 따라 전달 유무가 결정되므로 과거 파라미터와의 보간은 없어야하며, 최종 광대역 음성 출력 시 저대역 음성과의 합성을 위해 불필요한 지연을 피해야한다. 따라서 이와 같은 조건에 부합되면서 최적의 성능을 가지기 위해 FFT보다 뛰어난 해상도를 가지고 있는 매칭퍼슈잇을 이용하여 고대역에

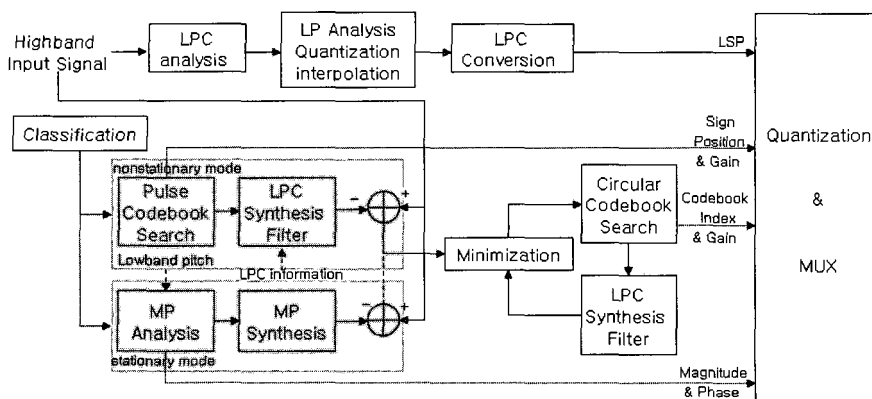


그림 2. 제안된 고대역 음성부호화기
Fig. 2. Block diagram of highband speech coder.

맞는 새로운 정현파 모델을 사용한다. nonstationary 모드 프레임과의 부드러운 연결과 둘째단의 목적신호 생성을 위한 시간맞춤을 맞추기 위해 위상을 전송파라미터로 사용하고, 이에 필요한 비트수는 둘째단에서의 CELP의 모양이득과의 결합 벡터 양자화와 저대역의 피치정보를 사용함으로써 얻는다. 이는 효율적인 정현파 파라미터 추출 뿐 아니라 피치정보에 할당되는 비트도 절약하는 장점이 있다.

고대역의 기본 주파수는 식 (3)으로 계산되어진다.

$$w_k^L = w_0^L \left(k + \left[-\frac{T_L^L}{2} \right] \right) - \pi \quad (3)$$

여기서 w_0^L 은 저대역 신호의 기본 각주파수, T_L^L 은 저대역 신호의 피치 주기, $[\]$ 은 가우스 정수를 나타낸다.

그림 3은 다단계 구조를 가지는 부호화 블록이다. 첫째단에서는 정현파 파라미터 추출을 위한 스펙트럼 분석 도구로 매칭퍼슈잇을 사용한다. 매칭퍼슈잇은 목적신호로부터 에너지가 가장 큰 성분들을 반복적으로 추출하여 보다 적응적으로 목적신호의 순음을 검출할 수 있다[14]. 정현파 사전을 가지는 매칭퍼슈잇은 식 (4)를 반복과정을 통해 최소가 되는 해를 구한다.

stationary 모드에 적용된 매칭퍼슈잇은 [15]와 [16]에서 제안된 모델과 다르다. 큰 지연 값을 가지는 시간 변이 항이 포함된 [15]에서의 모델은 최종 광대역 음성 출력 시 저대역 신호와 합해야하기에 실용적이지 못하다. 또한 LP 잔여신호를 목적신호로 분석하는 [16]에서의 모델은 MP 분석의 기여도를 뺀 CELP의 목적신호 생성에서 오차가 생겨 음질에 악영향을 끼친다. 본 논문에서는 효율적 부호화를 위해 시간 변이 이득 항을 포함하지 않으면서 파라미터 추출 시 선형예측 스펙트럼 응답을 이용하여 고대역 신호를 목적신호로 반복과정을 거치는 매칭퍼슈잇을 적용한다.

식 (4)는 매칭퍼슈잇의 반복과정을 위한 왜곡 측정 함수이다. 여기서 $S_0^k(n)$, $h(n)$ 은 영 상태 고대역 신호와 선형예측 스펙트럼 응답이고, $S_{i-1}^k(n)$ 은 k 번째 프레임에서 $i-1$ 번째 반복단계의 목적신호이다. 가중치 합

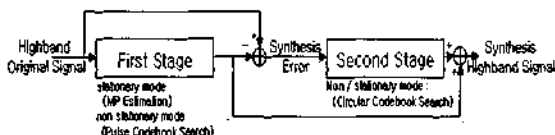


그림 3. 고대역 부호화기에 적용된 다단계 구조
Fig. 3. Structure of multi-stage applying to highband coder.

수 $w(n)$ 은 헤밍윈도우이다.

$$E_i^k = \sum_{n=0}^{N-1} \{w(n)(S_{i-1}^k(n) - h(n)A_i^k \cos(w_i^k n + \phi_i^k))\}^2$$

$$= \sum_{n=0}^{N-1} \{w(n)(S_{i-1}^k(n) - h(n)\{A_i^k \cos(w_i^k n) \cos(\phi_i^k) + A_i^k \sin(w_i^k n) \sin(\phi_i^k)\})\}^2 \quad (4)$$

$$S_i^k(n) = S_{i-1}^k(n) - h(n)A_i^k \cos(w_i^k n + \phi_i^k)$$

식 (4)에서 w_i^k 를 기본 주파수의 배수로 고정시키고, 식을 간단히 하기 위해서 다음과 같은 식을 정의한다.

$$a_i^k = A_i^k \cos(\phi_i^k), \quad b_i^k = -A_i^k \sin(\phi_i^k) \quad (5)$$

$$hc(n) = h(n) \cos(w_i^k n), \quad hs(n) = h(n) \sin(w_i^k n) \quad (6)$$

$$fcc = \sum_{n=0}^{N-1} \{w(n)hc(n)hc(n)\}, \quad fcs = \sum_{n=0}^{N-1} \{w(n)hc(n)hs(n)\}$$

$$fss = \sum_{n=0}^{N-1} \{w(n)hs(n)hs(n)\}, \quad frc = \sum_{n=0}^{N-1} \{w(n)S_{i-1}^k(n)hc(n)\}$$

$$frs = \sum_{n=0}^{N-1} \{w(n)S_{i-1}^k(n)hs(n)\} \quad (7)$$

식 (5)~(7)을 이용하여 왜곡 측정 함수(E^k)를 최소화 하는 해를 구하면 스펙트럴 크기와 초기위상을 아래 식과 같이 구할 수 있다[16].

$$A_i^k = \sqrt{(a_i^k)^2 + (b_i^k)^2}, \quad \phi_i^k = -\tan^{-1}\left(\frac{b_i^k}{a_i^k}\right) \quad (8)$$

합성 여기신호는 2.5절의 양자화기로 양자화된 파라미터로 식 (8)과 같이 표현되고[13], 영 상태 고대역 신호와의 차를 통해 둘째단의 목적신호가 생성된다. 둘째단에서는 정현파 모델과 같은 분석 프레임단위로 nonstationary 모드와 같은 순환 코드북을 사용하여 부호화된다.

$$\hat{e}(n) = \sum_{l=1}^L \hat{A}_l \cos(l\hat{w}_0 n + \hat{\phi}_l) \quad (9)$$

여기서 $\hat{A}_l, \hat{\phi}_l$ 은 주파수가 $l\hat{w}_0$ 인 l 번째 양자화된 스펙트럴 크기와 위상을 나타낸다.

3.4. nonstationary 모드의 부호화

nonstationary 모드에서는 펄스 코드북과 순환 코드북을 사용한 CELP 방법이 그림 3과 같은 구조로 적용된

표 1. ACELP의 고정 코드북
Table 1. Fixed codebook of ACELP.

Sign	Possible Position	Bit Allocation	
		Position Bits	Sign Bits
+,-1	0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75	4	1
+,-1	2 7 12 17 22 27 32 37 42 47 52 57 62 67 72 77	4	1
+,-1	4 9 14 19 24 29 34 39 44 49 54 59 64 69 74 79	4	1

다. 모드 판별 과정에서 stationary 모드를 엄격히 제한함으로써 생기는 nonstationary 모드에서 주기적 신호 모델의 경우를 위해 세 개의 펄스를 첫째단에서 +, - 값을 가지는 펄스를 찾는 CELP 방식의 펄스 코드북 탐색으로 검출한다. 3개의 트랙에서 각 트랙당 하나의 펄스를 찾아낸다. 각 트랙에서 펄스가 존재하는 위치는 표 1과 같다.

펄스 코드북 탐색은 식 (10)과 같이 왜곡측정치가 최소가 되는 파라미터를 찾는다.

$$E_{\xi} = |\overline{S}_w - GHV_{\xi}^{pul}|^2 \quad (10)$$

\overline{S}_w^k 는 가중된 영 상태 고대역 신호로 가중된 음성에서 가중된 선형예측 합성필터의 영 입력 응답을 제거한 벡터이고, G 는 펄스 코드북 이득값, V_{ξ}^{pul} 는 ξ 번째 대수 코드 벡터이다. H 는 선형예측 스펙트럼 응답인 $h(n)$ 으로 구성된다. 행렬 H 는 digonal은 $h(0)$ 로 구성되고, lower digonal은 $h(0), h(1), \dots, h(79)$ 를 사용하여 계산된 대칭 행렬로 구성된다.

식 (10)를 최소화하는 최적의 코드벡터는 식 (11)를 최대화하는 하는 것이다[11].

$$\tau_{\xi} = \frac{C_{\xi}^2}{\epsilon_{\xi}} = \frac{(d^T V_{\xi}^{pul})^2}{(V_{\xi}^{pul})^T \Phi V_{\xi}^{pul}} \quad (11)$$

$$d = H \cdot X, \quad \Phi = H^T \cdot H$$

벡터 d 는 목적신호 \overline{S}_w^k 와 선형예측 스펙트럼 응답 $h(n)$ 의 상관관계이고, 행렬 Φ 는 $h(n)$ 의 상관관계 행렬이다. 여가신호 벡터 V_{ξ}^{pul} 은 영이 아닌 펄스의 개수로 이루어진 대수 구조의 코드북이고, 검색에 필요한 벡터 d 와 행렬 Φ 를 미리 계산되어 매우 빠른 검색이 가능하다[17].

펄스 코드북의 기여도가 제거된 신호는 순환 코드북으로 재 모델링된다. 코드북은 256개의 배열 요소를 가지며 사인값을 포함하여 총 512개의 모양벡터가 사용된다. 다음과 같은 왜곡 측정치가 최소가 되는 이득과 모양 벡터를 분석/합성과정을 반복하여 찾는다[18].

$$E_{\xi} = |e_{syn} - GHV_{\xi}^{cir}|^2$$

$$e_{syn} = S_w - S_{syn}^{pul} \quad (12)$$

여기서 e_{syn} 는 가중된 영 상태 응답 음성 \overline{S}_w 와 펄스 코드북의 기여분 S_{syn}^{pul} 와의 차이이다. G 는 순환 코드북의 이득을 나타내고, V_{ξ}^{cir} 은 ξ 번째 순환 코드북 벡터이다.

최적 코드북 인덱스를 검색하는 과정에서 모든 코드북들을 검색하는 것이 아니라 목적신호와 순환 코드북간의 상관도가 높은 모양벡터들을 미리 검색한 후, 미리 검출된 32개의 후보 모양벡터에서 최적의 모양벡터를 선택함으로써 연산량을 줄인다. 파라미터들은 벡터양자화를 통해 전송된다.

3.5. 양자화 방법

선형예측계수는 LSP 파라미터로 변환되고, 변환된 LSP는 2단 벡터 양자화를 통해 첫째 단에서 8bit와 둘째 단에서 4bit의 인덱스 정보를 전송한다.

그림 4에서는 NTT 한국어 DB에서 임의의 20문장을

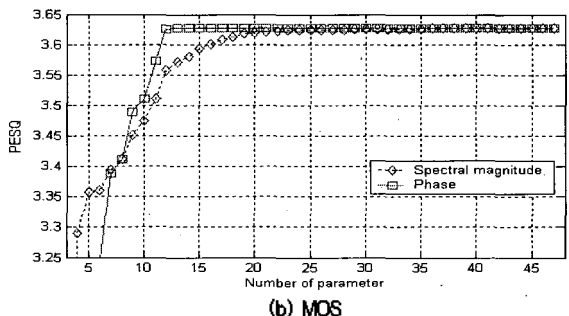
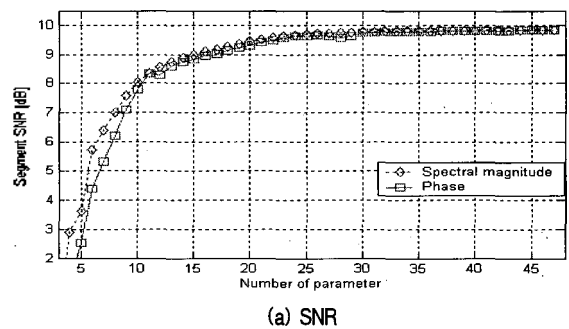


그림 4. 정현파 파라미터 개수에 따른 결과
Fig. 4. Results according to numbers of sinusoidal parameter.

표 2. 4kbps 고대역 음성부호화기의 비트 할당
Table 2. Bit allocation of the 4kbps highband coder.

Case		1		2		3	
Bit Rates		4 kbps					
LSPs		12 bits					
Mode		2 bits					
Frame		1st Subframe (stationary mode)	2nd Subframe (stationary mode)	1st Subframe (stationary mode)	2nd Subframe (nonstationary mode)	1st Subframe (nonstationary mode)	2nd Subframe (nonstationary mode)
MP Index	Magnitudes	10 bits	10 bits	10 bits			
	Phases	7 bits	7 bits	7 bits			
	Joint gains	7 bits	7 bits	7 bits	9 bits	9 bits	9 bits
Pulse Codebook	Signs				1×3 bits	1×3 bits	1×3 bits
	index				4×3 bits	4×3 bits	4×3 bits
Circler Codebook	Signs	1 bits	1 bits	1 bits	1 bit	1 bit	1 bit
	index	8 bits	8 bits	8 bits	8 bits	8 bits	8 bits
Total		80					

매칭퍼슈잇으로 분석 합성하여 합성시 사용된 정현파 파라미터 수에 따라 SNR과 MOS값을 도기한 것이다. 실험 결과값에 따라 최적의 양자화 개수를 스펙트럴 크기는 15개, 위상은 8개로 고정하여 양자화한다.

위상 양자화기는 위상을 크기가 1인 단위원 위의 극 좌표계에서 각 차원의 목적위상과 양자화 위상의 차를 고대역 입력신호의 선형예측 스펙트럼 응답에 따른 가중치를 주어 벡터양자화하고[16], 양자화되지 않는 위상에 대해서는 랜덤위상을 사용해 합성한다.

스펙트럴 크기 양자화기는 고정차원의 DCT를 이용한다. 각각의 스펙트럴 크기값들은 평균 에너지에 의해서 정규화되고, DCT변환을 이용하여 2단 벡터 양자화한다. 즉, 정규화된 스펙트럴 크기는 첫째단에서 DCT로 변환되어 4개의 스펙트럴 크기 후보 인덱스를 찾고, 4개의 후보인덱스는 둘째단에서의 IDCT로 변환하고 정규화 이득과 곱해 스펙트럴 크기와 비교한다. 왜곡이 최소가 되는 인덱스를 검출하고 이를 복호화기로 전송한다.[16]

이득 양자화기는 각 모드에서의 두개의 이득값을 결합하여 벡터 양자화한다. stationary 모드에서는 스펙트럴

크기의 정규화 이득값과 순환 코드북 이득값을, nonstationary 모드에서는 펄스 코드북 이득값과 순환 코드북 이득값을 분석 및 합성구조로 벡터 양자화한다.

그림 5에서 결합 이득 벡터 양자화 과정을 도시하고 있다. 첫째단에서는 식 (13)이 최소가 되는 이득 인덱스를 32개 검출하고, 둘째단에서는 식 (14)이 최소가 되는 코드북 이득 인덱스를 검색한다.

$$E = |S_w - H(G_{1st}^{joint} V_{1st})|^2 \quad (13)$$

$$E = |S_w - H(G_{1st}^{joint} V_{1st} + G_{2nd}^{joint} V_{2nd})|^2 \quad (14)$$

여기서 S_w 은 목적신호인 영 상태 고대역 신호이고, H 는 선형예측 스펙트럼 응답 행렬이다. 그리고 G^{joint} 는 2차원을 가지는 결합 이득 코드북을 나타내고, V_{1st} , V_{2nd} 는 각 단에서의 여기 벡터들이다.

3.6. 비트할당

각 부프레임별로 모드가 선택되어지므로 가능한 경우의 수는 네가지이다. 각 부프레임의 모드가 stationary 만 있을 경우는 case 1, nonstationary만 있을 경우는 case 2, 그리고 순서에 관계없이 nonstationary 모드와 stationary 모드가 같이 있으면 case 3으로 정의한다. 제안하는 고대역 부호화기의 비트 할당은 표 2에 나타난다. 여러 분석과 청취 실험을 통해 최적화하였다.

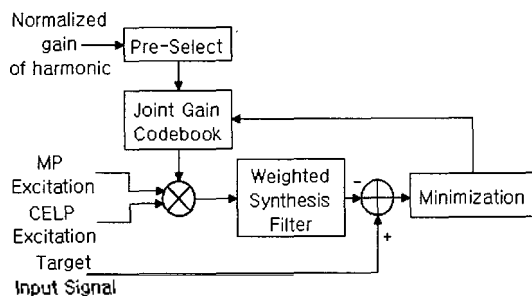


그림 5. 결합 이득 양자화기 구조
Fig. 5. Structure of joint gain quantizer.

IV. 성능 평가

그림 6과 7에서 NTT DB 광대역 음성의 고대역 입력신호와 합성신호의 전체 혹은 일부분에 대하여 비교를 하고 있다. 고대역 신호의 합성 결과는 그림 6을 통해 알 수 있다. 그림 7은 각 모드의 분석 및 합성한 결과를 320샘플 블록으로 도시한 것이다. stationary-nonstationary 모드나 nonstationary-stationary 모드로 전환되는 경계에서 부드러운 연결이나 각 모드의 분석 및 합성 결과를 시간 파형과 스펙트럼의 비교를 통해 알 수 있다.

고대역 신호의 음질에 영향을 미치는 중요한 정보는 고대역에서 비교적 낮은 주파수 대역이며 부호화 시 이 부분의 정보를 제대로 전달하는 것이 중요하다. 그림 6과 7을 통해 각 모드의 합성된 고대역 신호 일부분의 스펙트럼과 시간파형을 보면 주파수 축에서 낮은 주파수 대역을 잘 표현되고 있음을 알 수 있다. 이와 반대로 높은 주파수 대역은 하모닉 존재 유무와 관계없이 입력신호에 비해 왜곡이 일어나고 있음을 알 수 있다. 이러한

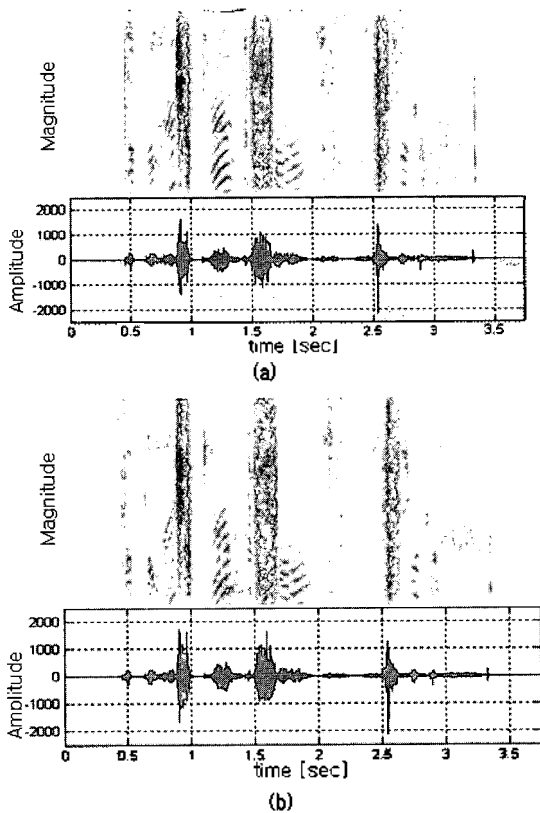


그림 6. 고대역 입력신호와 합성신호의 파형과 스펙트럼
(a) 고대역 입력신호
(b) 고대역 합성신호

Fig. 6. Waveform and spectrum of highband input signal and synthesized signal.
(a) highband input signal.
(b) highband synthesized signal.

부분의 영향으로 기계적인 음성이 발생할 수 있으나, 최종 출력 전에 7 kHz의 차단 주파수 (cutoff frequency)를 갖는 저역 통과 필터를 통과시킴으로써 제거된다.

제안한 15.8 kbps 광대역 음성부호화기의 성능을 위해 주관적 A/B 선호도 테스트 방법으로 측정한다. 비교 평가 부호화기는 표준화가 진행중인 가변 비트를 광대역 음성부호화기의 스펙[2]과 같다. 전체적인 성능 평가 규격은 다음과 같다.

- 대상: 훈련되지 않은 16명
- 음성 데이터: NTT 한국어 DB 20문장 (남자 10문장 여자 10문장)
- 비교평가 대상 부호화기: G.722 48 kbps (Test 1) [19] G.722,2 12.85 kbps (Test 2)[20]
- 순서 - 비교평가 대상 부호화기와 제안 부호화기의 순서를 임의적으로 정함
 - 남자와 여자 음성을 반복적으로 배열

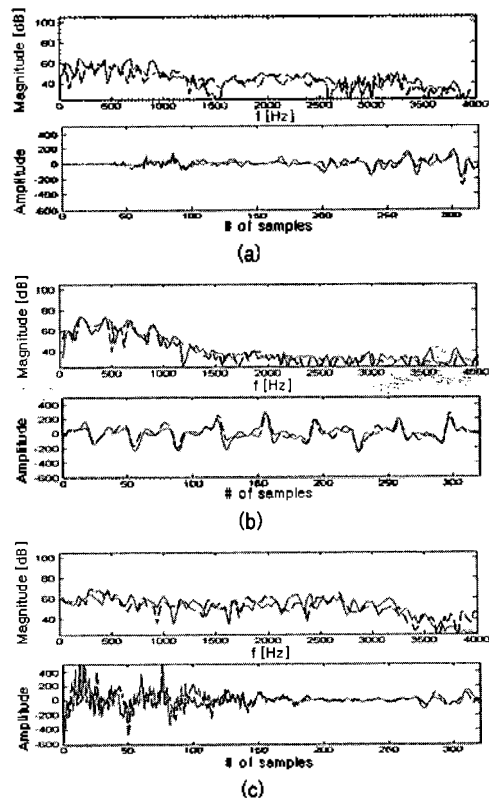


그림 7. 고대역 신호의 구간별 파형과 스펙트럼: 입력신호(실선), 합성신호(점선)

- (a) nonstationary 모드와 stationary 모드
- (b) stationary 모드와 stationary 모드
- (c) nonstationary 모드와 nonstationary 모드

Fig. 7. Waveform and spectrum according to highband segment : input signal(solid), synthesized signal(dash).
(a) nonstationary-stationary mode.
(b) stationary-stationary mode.
(c) nonstationary-nonstationary mode.

표 3. 광대역 음성부호화기의 주관적 성능평가 결과
Table 3. Performance evaluation result of the wideband speech coder.

Coder	Test 1		Test 2	
	G.722 (48kbps)	Proposed coder (15.8kbps)	G.722.2 (12.85kbps)	Proposed coder (15.8kbps)
Female	35.0%	65.0%	44.4%	55.6%
Male	42.5%	57.5%	54.4%	45.6%
Total	37.7%	62.3%	49.4%	50.6%

성능 평가 결과는 표 3과 같다. Test 1에서 제안하는 광대역 음성부호화기는 G.722 (48kbps)보다 여자 음성에 대하여 30.0% 높은 선호도를 보였으며, 남자 음성에서 대하여 15.0%의 높은 선호도를 보인다. 또한 전체적으로 성별에 상관없이 62.3%라는 높은 선호도를 보였다. Test 2에서 제안하는 광대역 음성부호화기는 G.722.2 (12.85kbps)보다 여자 음성에 대하여 11.2% 높은 선호도를 보였지만, 남자 음성에 대하여는 8.8% 낮은 선호도를 보였고, 전체적으로 1.2% 높은 선호도를 보였다. 제안된 부호화기가 전체적으로 남자 음성보다 여자 음성에서 더 좋은 음질을 내는 것은 대체적으로 고대역의 낮은 주파수의 하모닉 개수가 적기 때문이다. 이것은 하모닉의 개수가 적으면서 고대역의 하위대역 부분에 에너지가 집중되어 있을 경우, 스펙트럴 크기 양자화 오차가 작아져 매칭퍼슈잇에 의해 합성된 여기신호가 더 정확하게 표현된다. 반면 고대역의 전대역에 하모닉 성분의 에너지가 고르게 퍼져있는 경우, 매칭퍼슈잇 분석에 의해 모델링하지 못하는 신호가 많아 부호화 효율이 떨어지게 되지만, 매칭퍼슈잇의 차신호를 CELP 방법으로 모델링하는 다단계 구조를 사용함으로써 제한된 대역폭의 단점을 보완한다.

V. 결론

본 논문에서는 대역분할 구조의 광대역 음성부호화기를 위한 고대역 부호화방법을 제안하였다. 광대역 입력 신호는 대역분할되어, 저대역 음성신호는 기존의 네트워크 장치들과 호환성을 유지하도록 G.729 Annex E로 부호화한다. 고대역 신호는 정현파 모델과 CELP방법을 바탕으로 고대역 신호 특성에 적합한 새로운 정현파 모델을 제안하였고, 다단계 구조를 가지는 정현파 모델과 CELP 방법의 효율적인 조합으로 데이터의 압축율을 높였다. 제안한 광대역 음성부호화기는 A/B 테스트를 통하여 G.722 (48kbps)보다 우수하고, G.722.2 (12.85kbps)보다

나쁘지 않은 음질을 나타내는 것을 확인하였다.

제안하는 고대역 부호화 방법은 stationary 모드에서는 하모닉 개수나 위상의 개수를 조절하거나 들쭉단에서의 코드북 크기를 증가시킴으로써 추가적인 비트할당이 가능하고 제한된 대역폭의 단점을 극복할 수 있다. nonstationary 모드에서는 검색하는 펄스의 개수나 트랙의 증가, 그리고 펄스 위치의 증가를 통해 추가적인 비트할당이 가능하게 된다. 향후 제안한 고대역 부호화기를 바탕으로 계층적 구조의 멀티 코덱 설계가 가능하다.

감사의 글

본 연구는 2003년도 한국 과학재단의 특정 기초연구 사업(과제번호 R01-2003-000-11620-0)의 지원으로 수행 되었습니다.

참고 문헌

1. 이미숙, "광대역 코덱의 기술 및 표준화 동향," TTA Journal, 65-71, Mar. 2004.
2. ITU-T SG16 Q.9, "Report of Q.9/16 meeting," Nov. 2004.
3. T. Nomura, M. Iwadare, M. Serizawa and K. Ozawa, "A bitrate and bandwidth scalable CELP coder," IEEE International Conference on Acoustics, Speech and Signal Processing, 1, 341-344, May 1998.
4. K. Koishida, V. Cuperman and A. Gersho, "A 16-kbit/s bandwidth scalable audio coder based on the G.729 standard," IEEE International Conference on Acoustics, Speech and Signal Processing, 2, 1149-1152, Jun. 2000.
5. Sung-Kyo Jung, Kyung-Tae Kini and Hong-Goo Kang, "A bit-rate/bandwidth scalable speech coder based on ITU-T G.723.1 standard," IEEE International Conference on Acoustics, Speech and Signal Processing, 1, 285-288, May 2004.
6. Lajos Hanzo, F. Clare, A. Somerville and Jason P. Woodard, *Voice Compression and Communications* (John Wiley & Sons Ltd., New York, 2001), 531-564.
7. 송재중, 박호중, 김무영, 김도석, 김정수, "광대역 신호 압축기를 위한 주파수 대역 특성에 선택적인 양자화 방법," 음향학회지 20 (7), 76-82, 2001.
8. 이우석, 박호중, 손창용, 이영범, "대역폭 계층 구조의 광대역 음성 부호화기 개발," 음향학회지 23 (6), 481-487, 2004.
9. B. Kovesi, D. Massaloux and A. Sollaud, "A scalable speech and audio coding scheme with continuous bitrate flexibility," IEEE International Conference on Acoustics, Speech and Signal Processing, 1, 273-

276, May 2004.

10. 오연선, 신재현, 이인성, "MLT 여기신호를 이용한 광대역 음성 부호화기 설계," 음향학회지 24 (5), 230~237, 2005.
11. ITU-T Recommendation. G.729 Annex E, "11.8kbit/s CS-ACELP speech coding algorithm", Sep. 1998.
12. 3GPP C.S0030-0, "Selectable mode vocoder service option for wideband spread spectrum communication system", Dec. 2001.
13. R. McAulay and T. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," IEEE Transactions on Signal Processing, 34, 744-754, Aug. 1986.
14. S. G. Mallet and Zhifeng Zhang, "Matching pursuit with time-frequency dictionaries", IEEE Transactions on Signal Processing, 41, 3397-3415, 1993.
15. E. B. George and M. J. T. Smith, "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," IEEE Transactions on Signal Processing, 5, 389-406, Sep. 1997.
16. 안영욱, 정규혁, 김종화, 양용호, 이인성 "정현파 모델 부호화기를 위한 MP(Matching Pursuit)알고리즘과 파라미터 양자화기" 음향학회지 24 (7) 402-409, 2005.
17. Kyung jin Byun, Hee Bum Jung, Minsoo Hahn and Kyung Soo Kim, "A Fast ACELP Codebook Search Method," IEEE International Conference on Acoustics, Speech and Signal Processing, 1, 422-425, Aug. 2002.
18. A. M. Kondo, *Digital Speech*(John Wiley & Sons Ltd., New York, 1994), 174-212.
19. ITU-T Recommendation. G.722, "7 kHz audio-coding within 64 kbit/s," Nov. 1988.
20. ITU-T Recommendation. G.722.2, "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband," Jan. 2001.

저자 약력

• 정규혁 (Gyu-Hyeok Jeong)



2004년 2월: 충북대학교 전기전자공학과(공학사)
 2004년 3월~현재: 충북대학교 전파공학과 (공학석사)
 *주관심분야: 음성/오디오 부호화, 통신신호처리

• 안영욱 (Yeong-Uk Ahn)



2003년 2월: 충북대학교 반도체공학과 (공학사)
 2005년 2월: 충북대학교 전파공학과 (공학석사)
 2005년 2월~현재 (주) 코아르직
 *주관심분야: 음성/오디오 부호화, 통신신호처리, 적응필터

• 김종학 (Jong-Hark Kim)



1998년 2월: 충북대학교 전자공학과 학사
 2000년 2월: 충북대학교 전파공학과 석사
 2000년 3월~현재: 충북대학교 전파공학과 박사과정
 *주관심분야: 음성/오디오 부호화, 영상압축, 적응필터

• 신재현 (Jae-Hyun Shin)



1997년 2월: 충북대학교 전자공학과 학사
 2003년 8월: 충북대학교 전파공학과 석사
 2003년 9월~현재: 충북대학교 전파공학과 박사과정
 *주관심분야: 음성 신호처리, LSP 양자화, 적응필터

• 서상원 (Sang-Won Seo)



2004년 2월: 충북대학교 전기전자공학과 학사
 2004년 3월~현재: 충북대학교 전파공학과 석사과정
 *주관심분야: 음성/오디오 부호화, 통신신호처리

• 황인관 (In-Kwan Hwang)



1979년 2월: 이주대학교 전자 공학과 졸업
 1990년 6월: Polytechnic University, E.E. Dep 박사
 1997년 3월: 한국전자통신연구원 위성망 연구실장 책임 연구원
 1997년 3월~현재: 충북대학교 전기전자컴퓨터공학부 부교수
 *주관심분야: 이동통신, WBro, UWB, RFID, Cognitive Radio

• 이인성 (In-Sung Lee)



1983년 2월: 연세대학교 전기공학과 (공학사)
 1985년 2월: 연세대학교 전기공학과 (공학석사)
 1992년 2월: Texas A&M University 전기공학과 (공학박사)
 1993년 2월~1995년 9월: 한국전자통신연구원 이동통신 기술연구소장
 1995년 10월~현재: 충북대학교 전기전자공학부 정교수
 *주관심분야: 음성/영상 신호 압축, 이동통신, 적응필터