한국음향학회지 제20권 제5호 pp. 35~47 (2001)

효율적인 하모닉-CELP 구조를 갖는 저 전송률 음성 부호화기

Efficient Harmonic-CELP Based Low Bit Rate Speech Coder

최 용 수*, 김 경 민**, 윤 대 희**
(Yong-Soo Choi*, Kyung-Min Kim**, Dae-Hee Youn**)

*LG전자 네트워크연구소, **연세대학교 신호처리연구센터
(접수일자: 2000년 10월 2일; 수정일자: 2001년 1월 11일; 채택일자: 2001년 5월 4일)

본 논문에서는 하모닉 부호화기와 CELP (Code Excited Linear Prediction) 부호화기의 장점을 고려한 효율적인 저 전송률 하모닉-CELP 음성 부호화기를 제안한다. 제안된 하모닉-CELP 부호화기에서는 프레임 단위 유/무성음 판별에 따라 무성음 구간에서는 고속 CELP 방식으로 부호화하고 유성음 구간에서는 개선된 하모닉 부호화를 수행한다. 제안된 부호화기는 무성음 부호화를 위한 RP-VSELP (Regular Pulse Vector Sum Excited Linear Prediction), 유성음 부호화를 위한 간단한 정수 피치 검색, 정수 단위 피치에서의 고속 하모닉 추정, 가변 차원 하모닉 벡터 양자화, 주파수 해상도를 반영한 인지 가중치, 고속 하모닉 합성, 대역별 유성음 정도에 따른 자연성 제어, 다중 모드 등을 주요한 특징으로 하며, 이러한 특징들로 인해 기존의 HVXC (Harmonic Vector eXeited Coder) 부호화기에 비해서 매우 낮은 복잡도를 갖는다. 주관적인 음질 평가 결과, 제안된 2.4 kbps 하모닉-CELP 부호화기는 낮은 지연과 적은 계산량으로 양호한 음질을 얻을 수 있음을 확인하였다.
핵심용어: 하모닉-CELP 부호화, RP-VSELP, 피치 검색, 고속 하모닉 추정, 인지 가중, 자연성 제어, 다중 모드
투고분야: 음성처리 분야 (2.2)

This paper describes an efficient harmonic-CELP speech coder by taking advantages of harmonic and CELP coders into account. According to frame voicing decision, the proposed harmonic-CELP coder adopts the RP-VSELP coder as a fast CELP in case of an unvoiced frame, or an improved harmonic coder in case of a voiced frame. The proposed coder has main features as follows: simple pitch detection, fast harmonic estimation, variable dimension harmonic vector quantization, perceptual weighting reflecting frequency resolution, fast harmonic synthesis, naturalness control using band voicing, and multi-mode. These features make the proposed coder require very low complexity, compared with HVXC coder. To demonstrate the performance of the proposed coder, a 2.4 kbps coder has been implemented and compared with reference coders. From results of informal listening tests, the proposed coder showed good quality while requiring low delay and complexity.
Keywords: *Harmonic-CELP coding, RP-VSELP, Pitch deteotio, Fast harmonic estimation, Perceptual weighting, Naturalness control, Multi mode*
ASK subject classification: *Speech signal processing (2.2)*

책임저자: 최용수 (cando@lge.com)
153-023 서울시 금천구 가산동 60-39 LG전자(주)
LG전자 디지털 네트워크 연구소 미디어기기실
(전화: 02-818-4875; 팩스: 02-818-4802)

I. 서 론

디지털 컴퓨터와 통신 시스템이 생활 저변으로 확대됨에 따라 디지털화된 음성 신호의 사용이 점점 더 일반화되고 있다. 음성 파형 자체를 정확하게 재생하기 위해서는 많은 비트가 필요하고 이것은 시스템의 경제성과 효율성을 저하시킨다. 따라서 가능하면 적은 비트를 사용하여 주관적인 음질에 충실한 효과적인 음성 부호화 알고리듬이 필요하다.

코드북을 사용하여 여기 신호를 표현하는 CELP (Code Excited Linear Prediction)[1]는 8 kbps 이하의 저 전송률에서 지금까지 제안된 가장 효율적인 음성 부호화 알고리듬으로 평가되고 있다. 그러나 3 kbps 이하의 전송률에서는 심각한 음질 저하를 보여준다. 반면에 이러한 전송률에서도 주파수 영역에서 기본 주파수의 하모닉 성분으로 여기 신호를 표현하는 하모닉 부호화기는 우수한 음질을 갖는다. 일반적으로 부호화기의 특성상, 낮은 전송률에서도 CELP는 잡음성 신호를 잘 재생하고, 하모닉 부호화기는 주기적인 신호를 잘 표현한다.

이러한 사실에 근거하여 4 kbps 이하의 낮은 전송률에서 하모닉 부호화기와 CELP의 장점을 결합한 혼성 부호화 방법들이 제안되었다[2-4]. 이러한 하모닉-CELP 부호화기들은 선형 예측 부호화에 기반을 두고 있으며, LPC 여기 신호를 유/무성음으로 구분하여 무성음 구간에서는 CELP를 사용하고, 유성음 구간에서는 하모닉 부호화기를 사용하는 시간과 주파수 영역 부호화기가 혼재된 PWI (Prototype Waveform Interpolation)[2], MDRC (Mixed Domain Residual Coding)[3], HVXC (Harmonic Vector eXcited Coder)[4] 등이 대표적인 예이다. 이러한 혼성 부호화는 유/무성음 구간의 신호 특성에 적합한 부호화 방법을 사용하여 적은 부호화 비트로도 우수한 음질을 얻을 수 있기 때문에 최근 많은 연구가 진행되고 있다. 그러나, 기존의 하모닉-CELP 혼성 부호화기들은 무성음 부호화를 위한 CELP는 전통적인 방법[1]을 그대로 사용하기 때문에 계산량이 많다. 또한, 유성음 부호화를 위한 하모닉 부호화를 보면, PWI와 MDRC는 피치 길이의 DFT (Discrete Fourier Transform)와 정현파 보간법으로 인한 많은 계산량을 필요로 한다. HVXC는 가변 차원 하모닉 벡터 양자화, 고속 하모닉 합성 등 우수한 기능을 갖고 있지만, 상당한 계산량과 버퍼 지연을 갖는 MBE (Multi Band Excitation)[5,6] 방법을 기본으로 하는 실수 피치 추정 과정을 포함하고 있으며, 모드에 따라 고정된 비율의 잡음을 혼합하므로 음성의 구간별 특성을 제대로 반영하지 못하고 있다.

본 논문의 목적은 위에서 언급한 최근의 저 전송률 음성 부호화의 연구 경향을 바탕으로 하여 계산량과 음질에서 효율적인 저 전송률 하모닉-CELP 부호화기를 제안하는 것이다. 이러한 목적을 위하여 본 논문에서 제안된 하모닉-CELP 부호화기는 기존의 혼성 부호화기가 갖는 장점을 최대한 살리면서 무성음 부호화에 적합한 고속 CELP 알고리듬과 개선된 하모닉 부호화 알고리듬을 사용한다. 제안된 하모닉-CELP 부호화기는 프레임 단위 유/무성음 판별에 따라 무성음 구간에서는 고속 CELP 방식인 RP-VSELP (Regular Pulse Vector Sum Excited Linear Prediction)[7-9]로 부호화하고 유성음 구간에서는 개선된 하모닉 부호화를 수행한다. 따라서 제안된 혼성 부호화기는 두 방식을 결합한 형태이므로 적절한 계산량을 필요로 하며 양호한 음질을 얻을 수 있다.

무성음 부호화를 위한 RP-VSELP는 벡터-합 코드북을 이루는 기저 벡터를 상호 정규직교 특성을 갖는 레귤러 펄스 열로 구성하여 기존의 VSELP가 갖는 채널 에러에 대한 강인함, 우수한 음질 등의 장점들을 유지하면서 코드북 검색 과정의 계산량을 크게 감소시킨다. 또한, RP-VSELP는 반복적인 폐회로 훈련 과정을 통해서 코드북 최적화를 수행할 수 있다.

유성음 부호화를 위한 개선된 하모닉 부호화는 PMSVQ (Predictive Multi Stage Vector Quantization) 방식의 LSP (Line Spectrum Pair) 양자화, 시간 영역 프레임 유/무성음 판별, 간단한 정수 피치 검색, 정수 단위 피치에서의 신뢰할 만한 고속 하모닉 분석, 가변 차원 하모닉 벡터 양자화, 청각 특성을 반영한 인지 가중치, 고속 하모닉 합성, 대역별 유성음 정도에 따른 자연성 제어, 다중 모드 등을 주요한 특징으로 한다.

제안된 하모닉-CELP 부호화기의 성능을 평가하기 위하여 2.4 kbps 비트 할당을 하여 모의 실험을 하였다. 비공식적인 청취 실험 결과, 제안된 혼성 부호화기는 낮은 지연과 적은 계산량으로 양호한 음질을 얻을 수 있음을 확인하였다.

본 논문의 구성은 다음과 같다. II 장에서는 본 논문의 이론적 배경에 대해 설명하고 III 장에서는 제안된 하모닉-CELP 부호화기의 전체 구성, LSP 양자화, 피치 검색, 유/무성음 판별, RP-VSELP, 하모닉 추정 및 양자화 그리고 자연성 제어 알고리듬에 대해서 자세히 기술한다. IV 장에서는 제안된 부호화기의 성능을 평가하기 위한 2.4 kbps 비트 할당을 보여 주며, 계산량과 음질

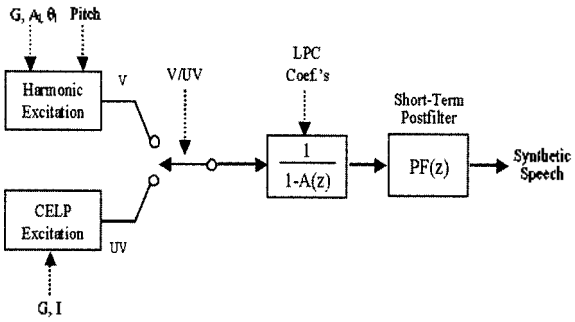


그림 1. 하모닉-CELP 복호화기
Fig. 1. Harmonic-CELP decoder.

평가를 통한 성능을 고찰한다. 마지막으로 V 장에서 결론을 맺는다.

II. 이론적 배경

저 전송률에서 CELP는 신호의 크기와 위상을 동시에 부호화하기 때문에 합성음의 자연성이 좋다. 반면에 하모닉 부호화기는 합성음의 인지도는 높으나 자연성이 결여된 기계적인 음질을 갖는다. 이러한 이유는 CELP 부호화기는 음질의 자연성에 관련된 위상 정보를 충실히 복원하고 부호화 오차 신호도 잡음성 신호인 반면에, 하모닉 부호화기는 부호화 오차가 순음 성분이고, 낮은 전송률에서는 위상 예측을 하므로 자연성에 관련된 분산 위상 정보를 손실하기 때문이라고 볼 수 있다. 부호화기 특성상 저 전송률에서 CELP는 파찰음, 파열음 등을 포함한 잡음성 신호를 잘 재생하고, 하모닉 부호화기는 주기적인 펄스 신호를 잘 표현한다[10].

이러한 사실에 근거하여 4 kbps 이하의 저 전송률에서 하모닉 부호화기와 CELP의 장점을 결합한 하모닉-CELP 부호화 방법들[2-4]이 제안되었으며, 이 방법들은 유/무성음 구간의 신호 특성에 적합한 부호화 방법을 사용하여 적은 부호화 비트로도 우수한 음질을 얻을 수 있기 때문에 최근 많은 연구가 진행되고 있다. 이러한 혼성 부호화기들은 그림 1에서와 같이 선형 예측 부호화 모델에 기반을 두고 있으며, 단구간 상관관계가 제거된 LPC 여기 신호를 유/무성음으로 구분하여 무성음 구간에서는 CELP를 사용하고 유성음 구간에서는 하모닉 부호화기를 사용하는 형태이다. 그림 1에서 G, A_l, θ_l, I은 각각 이득, 하모닉 크기, 위상, 코드북 인덱스이다. 또한, 그림 1에서 알 수 있듯이, 이러한 하모닉-CELP 부호화기는 LPC 보코더에서 유성음일 경우는 하모닉 여기 신호를, 무성음일 경우는 CELP 여기 신호를 사용하는 형태의 개선된 LPC 보코더로 볼 수 있다.

지금까지 제안된 하모닉-CELP 부호화기들 중에서 PWI[2], MDRC[3], HVXC[4]가 대표적인 예이며, 대개 CELP 부호화기는 동일하고, 하모닉 부호화 방법에 차이가 있다. 이 중에서 현재 HVXC는 MPEG-4의 낮은 전송률 음성 부호화 표준안으로 채택되었다. 그러나 이들 하모닉-CELP 부호화기들은 무성음 부호화를 위한 CELP는 전통적인 방법을 사용하고 있기 때문에 계산량이 많다. 또한, 유성음 구간에서의 하모닉 부호화를 보면, PWI와 MDRC에는 시간 영역에서 피치 주기의 여기 신호를 벡터 양자화하여 하모닉 크기와 위상을 동시에 부호화하고 복호화 단에서는 이를 피치 길이만큼 DFT한 후 주파수 영역에서 정현파 보간법[5,6]으로 여기 신호를 재생하기 때문에 부호화 단과 복호화 단 양쪽에서 많은 연산을 필요로 한다. HVXC는 여기 신호에 윈도우를 취하여 FFT (Fast Fourier Transform)한 후, 주파수 영역에서 MBE [5,6] 방법으로 하모닉 크기를 추정하고, 위상을 예측하여 정현파 합성을 한다. HVXC는 가변 차원 하모닉 벡터 양자화, 그리고 IFFT (Inverse FFT)와 WOLA (Weighted OverLap Add)를 이용한 고속 하모닉 합성 등 우수한 기능을 갖고 있다. 그러나 HVXC는 MBE의 3 프레임 전방향/역방향 추적을 기반으로 하는 실수 피치 추정 과정을 포함하고 있다. 이것은 많은 계산량뿐만 아니라 상당한 버퍼 지연을 의미한다. HVXC는 자연성 제어를 위해 프레임 모드에 따라 미리 고정된 비율의 잡음을 혼합하므로 다양하게 변하는 음성의 주파수 대역별 잡음특성을 제대로 반영하지 못하고 있다.

III. 제안된 하모닉-CELP 혼성 부호화

본 장에서는 2 장에서 설명한 하모닉-CELP 부호화 모델에 기본을 두고 계산량과 음질에서 효율적인 하모닉-CELP 부호화기, 일명 EHC (Efficient Harmonic-CELP) 부호화기를 제안한다. 그림 2와 3은 제안된 EHC 부호화기와 복호화기이다. 제안된 EHC 부호화기는 프레임 단위로 시간 영역에서 유/무성음을 판별하고 이에 따라 무성음 여기 신호 구간에서는 고속 CELP 방식인 RP-VSELP [7-9]로 부호화하고 유성음 여기 신호 구간에서는 개선된 하모닉 부호화를 수행한다.

제안된 EHC 부호화기 및 복호화기의 기본 구조는

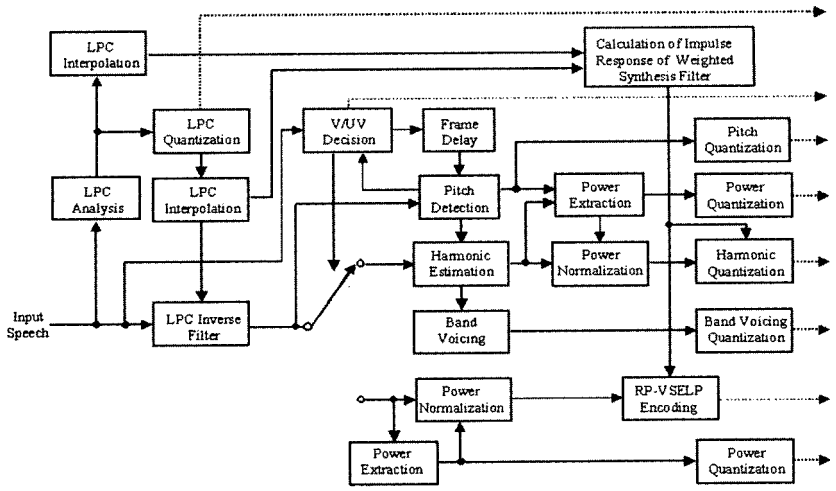


그림 2. 제안된 EHC 부호화기
Fig. 2. Proposed EHC encoder.

HVXC와 유사하다. 그러나 HVXC와는 달리 무성음을 위한 CELP는 전통적인 방법을 그대로 사용하지 않고 코드북 훈련을 통해 음질을 향상시킬 수 있는 고속 CELP 방식인 RP-VSELP를 사용하고 있으며, 유성음 구간에서는 낮은 지연을 갖는 간단한 피치 검색 알고리듬과 낮은 피치 해상도에서도 신뢰성 있는 하모닉 분석 방법을 사용하고 있다. 또한 청각의 주파수 해상도를 고려한 하모닉 크기 벡터 양자화와 입력 신호의 특성에 따른 대역별 유성음 정도에 따른 잡음 혼합 등이 HVXC와는 다른 특징이다.

제안된 EHC 부호화기는 그림 2와 같이 크게 LPC 분석, 피치 추정, 유/무성음 판별, 하모닉/RP-VSELP 부호화로 구성되어 있고, 복호화기는 그림 3과 같이 정현파 보간/RP-VSELP 복호화, LPC 합성 그리고 적응 포스트필터로 구성되어 있다.

부호화 과정을 살펴보면 우선, 입력된 음성 신호를 고역 통과 필터를 사용하여 DC 성분을 제거하고 프레임 단위로 LPC 분석을 한다. LPC는 LSP로 변환된 후 벡터 양자화되고 부프레임 단위로 보간된다. 보간된 LPC를 사용하여 LPC 분석 필터를 구성하여 잔차 신호를 구한다. 잔차 신호의 자기상관 값으로부터 개회로 피치를 구하고, 음성 신호와 개회로 피치 이득을 이용하여 유/무성음을 판별한다. 유/무성음 정보에 따라, 무성음인 경우에는 부프레임 단위로 개회로 이득을 구하고 고속 RP-VSELP를 사용하여 파형을 부호화한다. 유성음인 경우에는 스펙트럼 하모닉을 추정하고, 이를 개회로 이득으로 정규화한 후 가변 차원 벡터 양자화한다. 또한 하모닉 대역에서 유성음 정도를 구하여 벡터 양자화한다.

복호화 과정은 역 양자화 과정을 거친 복원된 파라미터

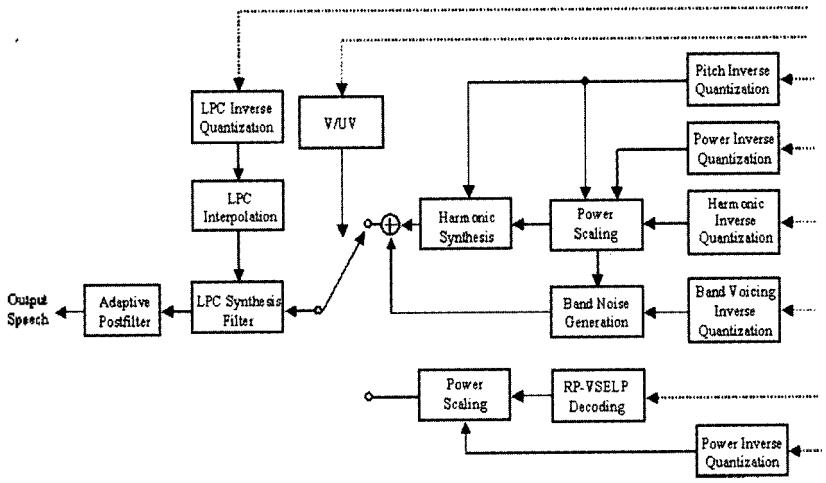


그림 3. 제안된 EHC 복호화기
Fig. 3. Proposed EHC decoder.

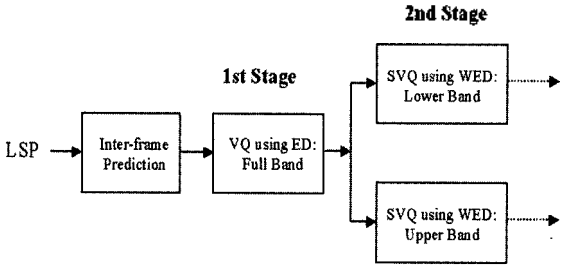


그림 4. LSP 양자화 과정
Fig. 4. Simplified LSP quantization process.

를 사용하여 무성음일 경우에는 RP-VSELP 복호화를 통해 여기 신호를 재생한 후 이득을 곱한다. 유성음일 경우에는 이득으로 정규화한 후 정현파 보간을 통해 유성음 여기 신호를 재생하고 하모닉 대역 유성음 정보에 따라 잡음을 혼합하여 최종적인 LPC 여기 신호를 재생한다. 여기 신호는 LPC 합성 필터를 통해 음성 신호로 합성되고, 마지막으로 적응 포스트필터를 거쳐 최종적인 합성음을 출력한다.

3.1. LSP 양자화

LSP는 PMSVQ (Predictive Multi stage Split Vector Quantization) 방식[11]에 기본을 둔 알고리듬을 사용한다. 이 방법에서는 먼저, 과거 프레임의 LSP를 이용하여 프레임간 예측을 하여 오차를 구한 다음, 첫 번째 단계에서 그 오차를 전체 차원 벡터 양자화한다. 그리고 두 번째 단계에서는 첫 번째 단계의 잔차 성분을 분리 벡터 양자화를 수행한다. 그림 4는 본 논문에서 사용한 PMSVQ 방식의 LSP 양자화 과정을 간략하게 보여 주고 있으며, 그림 5는 보다 상세한 과정을 나타내고 있다.

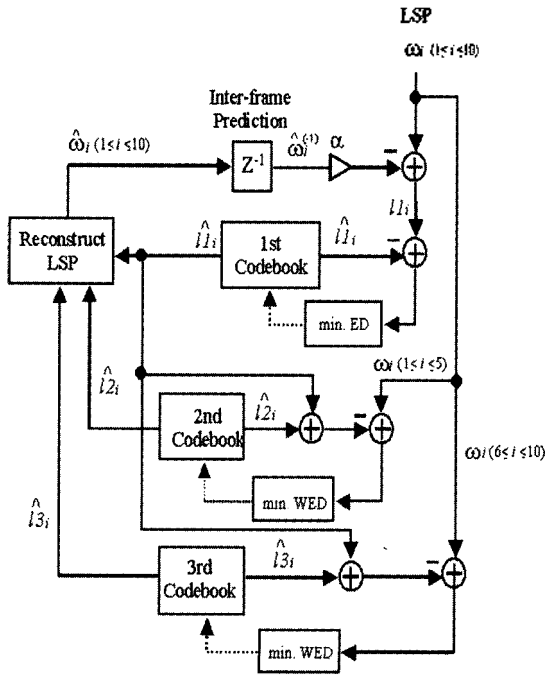


그림 5. LSP 양자화 상세 과정
Fig. 5. Detailed LSP quantization process.

LSP 양자화 성능은 일반적으로 LPC 스펙트럼 왜곡(Spectral Distortion: SD)을 사용하여 평가되며, 일반적으로 명료한 음질을 갖기 위한 전체에 대한 SD 기준은 1 dB 정도로 알려져 있다[13]. 그러나 이것은 엄격히 말해서 유성음 구간에 적합한 기준이며 무성음 구간에서는 이 보다 약간 높은 값이 허용된다[14]. 따라서 음질에 중요한 영향을 미치는 유성음 구간에서의 양자화 성능이 중요하다.

표 1은 긴 실험 데이터에 대해서 위에서 설명한 PMSVQ의 LSP 양자화 성능을 나타내는 SD를 보여 주고 있다. 여기서 사용된 PMSVQ는 프레임간 상관관계가 제거된 LSP 오차 신호를 1 단계에서는 10 차원 VQ를 하고 2 단계에서는 5 차원씩 분할하여 VQ를 수행한다. 비트 할당은 (1 단계: 전체 차원, 2 단계: 저차원 + 고차원)으로 표시하였고, 전체 및 유성음과 무성음 구간에 대해서 따로 SD를 구하였다. 표 1에서 할당된 비트 수가 증가할수록 그에 비례하게 성능이 향상되고, 2 단계보다는 1 단계에 더 많은 비트를 할당하고, 2 단계에서도 고차원보다는 저차원에 비트를 더 많이 할당할 때 양자화 성능이 향상됨을 관찰할 수 있다. 표 1에서 *표는 주어진 비트 수에서 최적 할당을 나타내며, 그 성능을 그림 6에 나타내었다.

LSP 양자화 코드북은 긴 훈련 데이터로부터 추출한 LSP 주파수 열로부터 LBG 알고리듬[12]을 사용하여 유/무성음 구간에 따라 다르게 설계하였다. 실험 결과, 유/무성음 구별 없이 설계된 코드북을 사용하였을 경우와

표 1. 비트 할당에 따른 PMSVQ 방식의 LSP 양자화 성능
Table 1. Performance of the PMSVQ LSP quantization vs bit allocation.

Bits	Bit Allocation	SD (dB)		
		Total	Voiced	Unvoiced
26	*9, 9+8	0.85	0.89	0.76
	8, 9+9	1.18	0.92	1.69
25	*9, 8+8	0.90	0.95	0.81
	8, 9+8	0.90	0.95	0.81
24	*8, 8+8	0.98	1.02	0.89
23	*8, 8+7	1.26	1.12	1.53
	7, 8+8	1.18	1.21	1.12
22	*8, 7+7	1.36	1.24	1.59
	7, 8+7	1.52	1.37	1.81
21	*7, 7+7	1.61	1.48	1.86

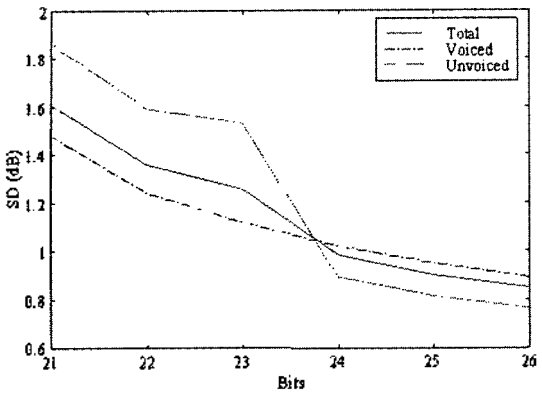


그림 6. 최적 비트 할당에 따른 PMSVQ 방식의 LSP 양자화 성능
Fig. 6. Performance of the PMSVQ LSP quantizer with an optimal bit allocation.

비교해서 약 0.2 dB 정도 SD 차이를 보였다. 그림 6에서 비트 수가 증가할수록 전체 및 유성음 구간에 대한 스펙트럼 왜곡은 점차 감소하고, 무성음 구간에 대해서는 24 비트 이상에서 급격히 감소함을 알 수 있다.

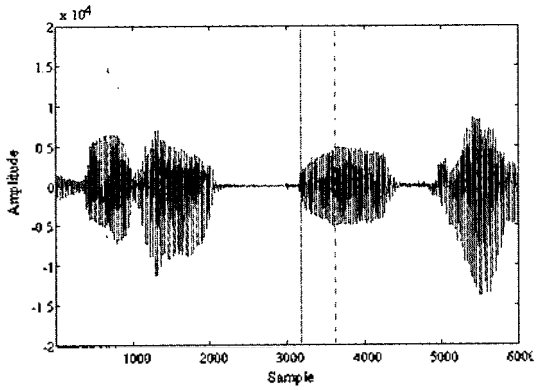
3.2. 피치 검색

하모닉 부호화기에서 피치는 음질과 지연에 중요한 영향을 미치므로 여러 파라미터 중에서 가장 중요하다. 일반적으로 CELP 부호화기에서 피치 두 배 (doubling) 및 절반 (halving) 현상은 적응 코드북을 사용하기 때문에 음질에 큰 영향을 주지 않으므로 피치 검색 방법은 비교적 간단하다. 그러나 하모닉 계열 부호화기에서는 피치 그 자체로도 음질에 심각한 영향을 주며 다른 부호화 파라미터의 분석에도 영향을 준다. 따라서 하모닉 부호화기에서는 피치를 정확하게 검색하기 위해서 상당한 버퍼 지연과 계산량을 필요로 하는 각각 3 프레임 (60 ms)의 전방향/역방향 피치 추적을 한다. 본 논문에서는 낮은 지연을 갖는 전방향 추적을 기본으로 하는 EVRC (Enhanced Variable Rate Codec)[15]의 피치 검색 알고리듬을 EHC 부호화기에 사용한다. 이 때 제안된 혼성 부호화기에 적합하도록 지역 최대치 비교 방법[11]과 프레임 단위 유/무성음 판별에 따른 기준 윈도우의 선택 방법을 도입한다. 기준 윈도우 선택법은 유성음 프레임이 연속되면 첫 번째 윈도우에서 얻어진 지연을 기준 값으로 하고, 그렇지 않으면, 두 번째 윈도우의 지연을 기준 값으로 한다. 즉, 연속되는 유성음 구간에서는 과거 프레임 피치의 영향을, 그렇지 않으면 현재 프레임 피치의 영향을 반영하는 것이다.

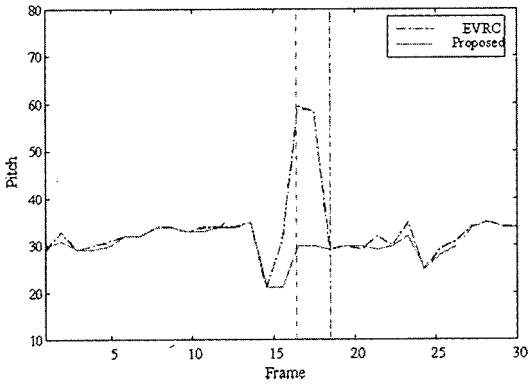
본 논문에서는 EVRC와 제안된 혼성 부호화기의 피치 검색 알고리듬의 성능을 비교하였다. 실험 결과, 남성 화자에 대해서는 성능의 차이가 거의 없었으나 여성 화자의 경우에는 유성음이 시작되는 구간에서 성능에 차이가 있었다.

그림 7은 여성 화자의 음성에 대한 EVRC와 제안된 피치 검색 방법의 결과이다. 그림에서 알 수 있듯이 제안된 혼성 부호화기의 피치 검색 알고리듬은 하모닉 부호화기의 음질에 심각한 영향을 미치는 유성음 구간에서 피치 두 배 현상을 효과적으로 제거한다.

일반적으로, 개회로 방법으로 추정된 피치는 응용되는 부호화기의 특성에 맞게 폐회로 방법으로 구성된 보다 정확한 추정 단계를 거친다. 폐회로 방법은 대개 합성을 통한 분석 (Analysis-by-Synthesis: AbS) 형식이며 CELP에서는 시간 영역에서 장구간 코드북 검색으로 구현되며, 하모닉 부호화기에서는 주파수 영역에서 스펙트럼 유사



(a) 음성 신호
(a) Speech signal



(b) 피치 컨투어
(b) Pitch contour

그림 7. EVRC와 제안된 EHC 부호화기의 피치 검색 비교
Fig. 7. Example of pitch detection by EVRC and the proposed EHC coder.

도 비교로 구현된다. 이때 피치 파라미터의 해상도는 실수 단위까지 확장되며 나머지 파라미터 분석에 상당한 영향을 미친다. 그러나 CELP 복호화기와는 달리 보간을 통해 합성음을 재생하는 하모닉 복호화기의 경우에는 피치의 해상도보다는 피치 컨투어가 음질에 중요한 역할을 하며[16], 합성음의 주관적 음질에 영향을 미치는 피치 해상도는 남성 화자보다 여성 화자 음성이 높다[17].

본 논문에서는 양자화되지 않은 정수 단위 개회로 피치를 사용하여 하모닉 추정을 하고, 피치를 46 이하는 한 샘플 단위로 47 이상은 두 샘플 단위로 하여 6 비트 스칼라 양자화한다: P_0 = {20, 21, ..., 46, 48,..., 118, 120}. 정수 단위의 피치에서 하모닉 추정의 정확도를 높이기 위한 알고리듬은 3.5.1 절에서 설명한다.

3.3. 유/무성음 판별 및 다중 모드 결정

유/무성음 판별 과정은 두 단계로 구성된다. 우선 입력되는 음성 신호의 에너지에 따라 몇 가지 등급 (rate)[18]으로 나눈 후 이 등급에 따라 식 (1)과 같이 대략적인 유/무성음을 판별한다. 그 다음으로 개회로 피치 예측 이득과 영교차율을 사용하여 1 단계에서의 유/무성음 판별을 수정한다.

$$voicing = \begin{cases} 1, & rate \geq 2 \\ 0, & otherwise \end{cases} \tag{1}$$

최종적인 유/무성음 판별은 개회로 피치 예측 이득 β와 영 교차율 (Zero-Crossing Rate: ZCR), PKN (PeaKi-Ness) [19], 등을 고려하여 다음과 같이 극단적인 경우에 대한 교정 과정을 거친다.

$$voicing = \begin{cases} 1, & \beta > \zeta_1 \ or \ PKN > \zeta_2 \\ 0, & \beta < \zeta_3 \ or \ ZCR > \zeta_4 \end{cases}, \tag{2}$$

여기서, ζ_1, ζ_2, ζ_3, ζ_4는 실험적으로 결정된다.

다중 모드는 식 (3)과 같이 유성음(V)/무성음(UV)의 순차적인 관계에 의해서 결정되며, 이 정보를 이용하여 여기 신호의 시간적인 변화 특성에 맞게 서로 다른 방식의 부호화 및 복호화를 수행한다. 기본적으로 현재 프레임이 무성음인 경우는 CELP 부호화를 수행하고 유성음인 경우는 하모닉 부호화를 수행하지만 모드에 따라 양자화 코드북과 여기 신호 합성이 다르다. 이러한 이유는 프레임 길이가 큰 경우에 신호의 천이 특성을 반영하기 위한 것이다.

$$mode = \begin{cases} 0, & UV \ to \ UV \\ 1, & V \ to \ UV \\ 2, & UV \ to \ V \\ 3, & V \ to \ V \end{cases} \tag{3}$$

3.4. 무성음 여기 신호 부호화

일반적으로 인간의 청각은 무성음 여기 신호는 에너지 컨투어에 민감하고, 유성음 여기 신호는 파형에 민감하다[20]. 다시 말해서, 무성음 신호는 단구간 에너지와 스펙트럼 크기에 충실해야 하고, 유성음 신호는 스펙트럼 크기와 위상에 충실해야 양호한 음질을 얻을 수 있다. [20]에서 무성음 LPC 여기 신호는 5 ms 간격으로 에너지 값을 갱신하고 원래 파형 대신 백색 잡음을 사용하여 합성하여도 원음과 구별할 수 없다는 것을 청취 실험을 통하여 확인하였다. 그러나 이러한 결과는 백색 잡음 특성을 갖는 무성음 여기 신호에 국한되며 파찰음이나 파열음과 같이 백색 잡음 특성이 아닌 여기 신호의 경우에는 음질 왜곡이 발생한다. 따라서 다양한 잡음성 신호에 대해서 양호한 음질을 얻기 위해서는 파형 양자화가 수행되어야 한다.

본 논문에서는 위에서 언급한 문제점을 고려하여 무성음 여기 신호를 그림 8과 같이 부프레임 단위 고속 CELP 방식인 RP-VSELP를 사용하여 AbS 벡터 양자화한다. 본 저자가 이미 제안한 바 있는 RP-VSELP는 기존의 VSELP [21]가 갖는 채널 에러에 대한 강인함, 우수한 음질 등의 장점들을 유지하면서 코드북 검색 과정의 계산량을 크게 감소시킨다. 또한, RP-VSELP는 반복적인 폐회로 훈련 과정을 통해서 코드북 최적화를 수행할 수 있다는 잠재적인 장점도 가지고 있다[7]. 또 ±1의 크기의 몇 개의 펄스로 구성된 ACELP (Algebraic CELP)[11]는 무성음 구간에서 잡음 특성을 잘 표현하지 못하지만[22], RP-VSELP는 임의의 크기를 갖는 일정 간격의 펄스 열로 구성되어 있

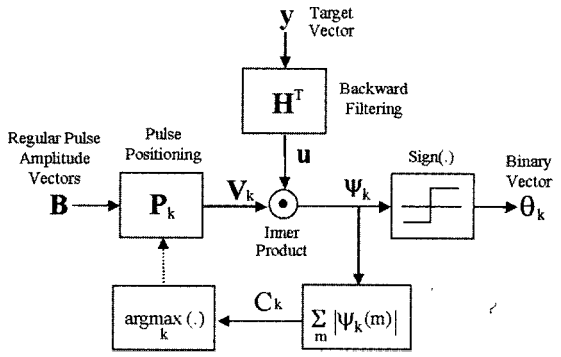


그림 8. RP-VSELP 코드북 검색 과정
Fig. 8. RP-VSELP codebook search process.

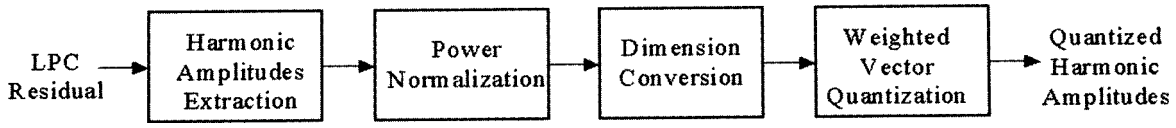


그림 9. 하모닉 크기 벡터 양자화 과정
Fig. 9. Harmonic amplitude vector quantization process.

기 때문에 무성음 구간에서도 만족할 만한 음질을 얻을 수 있는 고속 CELP 알고리듬이다.

일반적인 CELP 부호화기는 인간의 청각적인 특성을 반영하는 인지 가중 함수를 사용하지만 기본적으로는 시간축 파형을 충실하게 복원한다. 이것은 시간 영역에서 신호의 스펙트럼 크기와 위상을 동시에 부호화하는 방식으로 위상 양자화에 많은 비트가 할당된다고 해석될 수 있다. 그러나 낮은 비트율에서는 할당된 비트가 충분하지 않으므로 위상을 충실히 복원하지 못하므로 파형 정합(matching)이 부정확해져서 코드북 이득은 실제 신호의 전력에 비해 작아지게 된다. 이러한 문제점을 극복하기 위해서 여기 신호의 평균 전력을 개회로 방법으로 구하여 코드북 검색 전에 여기 신호를 평균 전력으로 정규화하고 파형만을 부호화한다. 복호화기에서는 코드북 이득을 1로 놓고 여기 신호를 재생한다.

3.5. 유성음 여기 신호 부호화

유성음 부호화를 위한 개선된 하모닉 부호화는 정수 피치를 이용하는 고속 하모닉 추정, 가변 차원 하모닉 벡터 양자화, 청각 특성을 반영한 인지 가중치, 고속 하모닉 합성, 대역별 유성음 정도에 따른 자연성 제어, 유/무성음 판별에 따른 다중 모드 등을 주요 특징으로 한다.

3.5.1. 하모닉 추정 및 양자화

대개의 하모닉 부호화기에서는 MBE[5]에서와 같은 FFT-기반 방법을 사용하여 하모닉을 추정한다. 이 방법은 하모닉 추정의 정확도를 높이고자 실수 단위의 피치 해상도를 높이면 상당한 연산량의 증가를 가져오며, 고주파수 영역으로 갈수록 원래 하모닉 주파수와 추정된 하모닉 주파수의 차이가 커져서 원래 스펙트럼과 합성 스펙트럼과의 오차가 커지게 된다. 또한 이러한 문제는 실제 음성 스펙트럼의 하모닉 주파수 간격이 주파수 대역에 따라 약간의 편차가 있는 경우에 대해서도 발생하며 피치 해상도가 낮으면 문제는 더욱 심각해진다. 이러한 문제점을 극복하기 위해서 하모닉 간격을 조정하여 하모닉 대역에서 스펙트럼 오차 에너지가 최소가 되도록 하는 하모닉 크기를 구하는 방법을 사용한다. 이 델타 조정법은 두 단계로 나누어질 수 있다. 우선 식 (4)로 주어진 오차 에너지 $E_l(\Delta_l)$을 최소화하는 최적의 하모닉 조정값 Δ_l^*을 선택한다. 다음, Δ_l^*를 이용하여 식 (5)로 주어진 하모닉 크기 $A_l(\Delta_l^*)$을 구한다. 이 과정은 하모닉 대역별로 순차적으로 이루어진다.

$$E_l(\Delta_l) = \sum_{m=a_l}^{b_l} [\,|X_w(m+\Delta_l)| - |\widehat{X_w}(m,\omega_0)|\,]^2 \tag{4}$$

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

$$-d_l \le \Delta_l \le d_l,\quad d_l = \left\lfloor \frac{a\omega_0}{L-1}(l-1) \right\rfloor, \tag{5}$$

이때, 하모닉 주파수의 누적 오차를 고려하여 Δ_l는 고주파수 대역으로 갈수록 크게 한다.

정수 피치만을 사용하는 이 방법은 실수 피치 분석 합성 방법을 사용하는 기존이 방법[5]에 비해서 평균 1/13의 적은 계산량으로 스펙트럼 왜곡 0.6 dB 이하의 양호한 결과를 보였다.

추정된 하모닉 크기는 HVXC에서 사용된 방식을 사용하여 가변 차원 벡터 양자화된다. 그림 9는 본 논문에서 사용된 하모닉 양자화 과정을 보여 준다. 그 과정을 살펴보면, 먼저 LPC 잔차 신호를 푸리어 변환한 스펙트럼으로부터 하모닉 크기 벡터를 추출하고 이를 평균 전력이

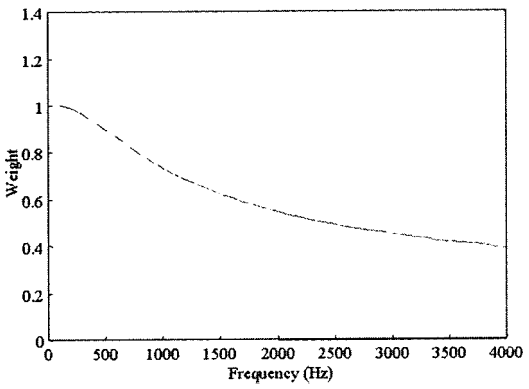


그림 10. 바크 가중치
Fig. 10. Bark weighting.

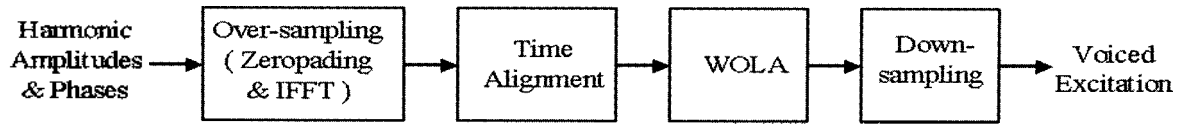


그림 11. 고속 하모닉 합성 과정
Fig. 11. Fast harmonic synthesis process.

1이 되게 정규화한다. 평균 전력은 정규화 과정 후에 개회로 형식으로 스칼라 양자화된다. 정규화된 하모닉 벡터는 가변 차원이므로, 보간/간축 과정을 거쳐 고정 차원으로 변환된 후 가중 벡터 양자화된다.

가중 벡터 양자화에 사용되는 가중치 $w_{pb}(j)$ $(1 \le j \le D)$는 인지 가중치 $w_p(j)$와 바크 가중치 $w_b(j)$의 곱으로 얻어진다. $\{w_p\}$는 일반적인 인지 가중 합성 필터의 주파수 응답으로부터 구해지고, 바크 가중치 $\{w_b\}$는 식 (6)과 같이 주어진 바크 주파수[23]를 미분하여, 식 (7)과 같이 근사화하여 사용하는 방법을 제안한다.

$$\Omega(f) = 6\log\left(\frac{f}{600} + \left(\left(\frac{f}{600}\right)^2 + 1\right)^{0.5}\right), \quad 0 \le f \le \frac{f_s}{2}, \quad (6)$$

$$w_b(j) = \left[\frac{\partial \Omega(f)}{\partial f}\Big|_{f=f_j}\right]^c \approx [\Omega(f_j) - \Omega(f_{j-1})]^c, \quad 1 \le j \le D, \quad (7)$$

여기서, $f_j = \frac{jf_s}{2D}$으로 D차원에서 j번째 하모닉에 해당하는 주파수이며, c는 가중치 벡터의 기울기를 결정한다. 그림 10은 $c = 0.5$ 일 때의 식 (7)로 주어진 바크 가중치 w_b를 보여 준다. 그림에서 주파수 해상도를 반영하여 저 주파수를 강조함을 알 수 있다. 실제로 하모닉 스펙트럼 크기 벡터의 양자화 계산에 사용되는 바크 가중치는 미리 계산하여 표로 저장함으로써 계산상의 이득을 얻을 수 있다.

3.5.2. 하모닉 합성

직접적인 정현파 합성 방법[5]은 성능은 우수하지만 계산량이 많이 요구되는 문제점이 있다. 이러한 문제점을 해결하기 위해서는 IFFT를 이용한 WOLA 방법을 사용할 수 있다[4,14]. 이중에서 HVXC에서의 고속 하모닉 합성 방법은 계산량이 적고 성능이 우수하다. 본 논문에서는 HVXC에서의 IFFT와 표본화율 변환 과정으로 구성된 고속 하모닉 합성 방법을 도입한다. 그림 11은 고속 하모닉 합성 과정을 보여 주며, 이 과정을 통해서 하모닉 크기와 기본 주파수가 선형 보간된 유성음 여기 신호를 얻게된다.

3.5.3. 자연성 제어

음성 신호를 관찰해보면 주기적인 신호라도 저 주파수 대역에서는 하모닉 성분이 강하고 고 주파수 대역에서는 하모닉 (유성음) 성분과 잡음 (무성음) 성분이 혼합되어 있으며 고 주파수 대역으로 갈수록 잡음 성분이 많다[14]. 만약 주파수 대역에 따라 유/무성음이 혼합되어 있는 신호를 하모닉으로만 표현하면 기계적인 (buzzy) 음질을 갖게 되며 잡음으로만 표현하면 거칠게 (harsh) 되어 자연성이 결여된다. 이러한 문제점은 LPC 보코더에서 잘 나타난다. 자연성 향상을 위한 방법으로는 하모닉 위상 분산 (phase dispersion) 방법[14]과 백색 잡음 혼합 방법[5, 6]이 있다. 위상 분산 방법은 합성 프레임의 길이가 충분히 짧지 않으면 순음 효과 (tonal effect)가 발생하는 문제점이 있다[24]. 백색 잡음 혼합 방법은 부가적인 계산이 필요하지만 합성 프레임의 길이와는 무관하게 자연성을 제어할 수 있다.

MBE[6]에서는 하모닉 간격의 다중 대역에서의 이진 유/무성음 판별을 하여, 유성음 대역에서는 위상 분산 기법을 도입하여 하모닉 합성을 하고, 무성음 대역에서는 백색 잡음을 합성하여 시간 영역에서 두 신호를 합한다. 따라서 MBE는 유/무성음이 혼합되어 있는 신호를 잘 표현할 수 있지만 다중 대역 유/무성음 정보에 상당한 비트가 필요하다. 2.4 kbps 이하의 낮은 전송률에서는 차단 주파수를 기준으로 두 개의 대역으로 분할하여, 저주파 대역은 유성음 대역으로 하고, 고주파 대역은 무성음 대역으로 하는 방법도 제안되었다[14].

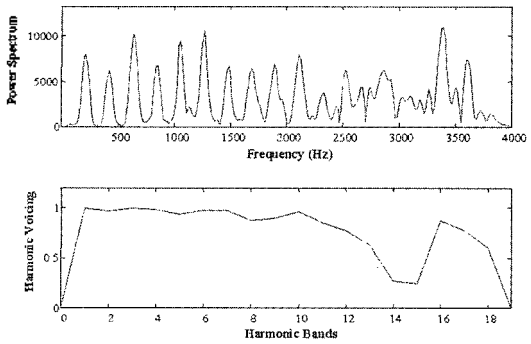


그림 12. LPC 잔차 신호의 스펙트럼과 대역 유성음정도
Fig. 12. LPC residual spectrum and band voicing.

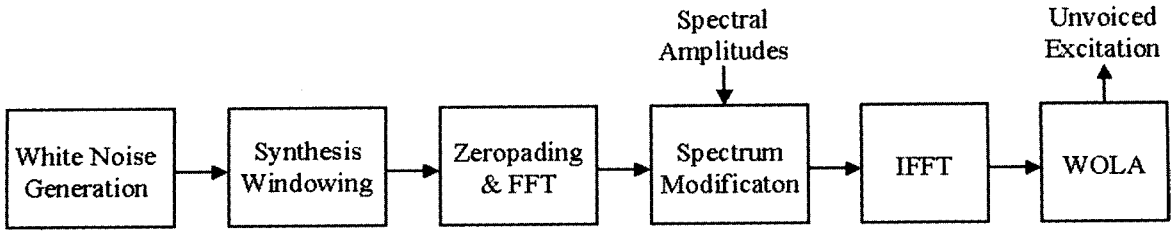


그림 13. 무성음 여기 신호 합성 과정
Fig. 13. Unvoiced excitation synthesis process.

본 논문에서는 하모닉 대역의 유성음 정도에 따라 하모닉과 잡음을 혼합하여 자연성을 제어하는 방법을 제안한다. 이 방법은 MBE의 다중 대역에서의 이진 유/무성음 판별과는 달리 0과 1 사이의 실수 값을 사용하므로 보다 자연스런 음질을 얻을 수 있다. 대역 유성음 정도는 벡터 양자화를 통해서 효과적으로 부호화되므로 낮은 전송률에서도 구현이 가능하며 전송률에 비례하는 성능을 얻을 수 있다.

하모닉 대역에서의 유성음 정도 BV_l은 원 스펙트럼과 합성 스펙트럼의 정규화된 오차 에너지 E_l'를 사용하여 다음과 같이 정의된다.

$$BV_l = 1 - E_l' = 1 - \frac{\sum_{m=a_l}^{b_l} \left| \, |X_w(m+\Delta_l)| - |\hat{X}_w(m)| \, \right|^2}{\sum_{m=a_l}^{b_l} |X_w(m+\Delta_l)|^2}, \quad (8)$$

여기서, Δ_l는 5.1 절에서 언급한 하모닉 간격의 조정 값이며, 가변 차원의 BV_l은 고정 차원 변환을 거쳐 벡터 양자화된다.

그림 12는 LPC 잔차 신호의 전력 스펙트럼과 그로부터 추정한 하모닉 대역의 유성음 정도 BV_l을 보여 준다. 그림 12에서 미리 고정된 비율의 잡음을 혼합하는 HVXC와는 달리 본 논문에서 제안된 부호화기에서는 하모닉 성분이 강한 주파수 대역에서는 BV_l이 높은 값을 갖고 잡음성 성분이 존재하는 주파수 대역에서는 BV_l이 낮은 값을 갖는 등 주파수 대역에 따라 다양하게 변하는 유/무성음 혼합 비율을 잘 표현하고 있음을 알 수 있다. 또한 비공식적인 선호도 평가를 통해서 제안된 방법에 의해 추정된 유성음 정도에 따라 잡음을 혼합하는 것이 HVXC에서의 고정된 비율에 의해 잡음을 합성하는 것 보다 우수한 음질을 보임을 확인하였다.

복호화기에서는 전송된 피치, 하모닉 크기, 그리고 대역 유성음 정도에 따라 주파수 영역에서 잡음 (무성음) 성분 신호를 합성한다. 전체 과정을 간략하게 살펴보면 백색 잡음을 주파수 영역에서 주어진 스펙트럼 크기를 갖도록 수정한 후, 이것을 다시 시간 영역으로 변환하여 최종적인 무성음 성분 신호로 사용한다. 이 때 주어진 스펙트럼 크기는 대역 유성음 정도에 따라 하모닉 크기로부터 구해진다. 그림 13은 본 논문에서 사용된 무성음 성분 여기 신호의 합성 과정이다.

IV. 비트 할당 및 성능 평가

제안된 EHC 부호화기의 성능을 평가하기 위하여 2.4 kbps 비트 할당을 하여 컴퓨터 모의 실험을 하였다. 입력 신호는 8 kHz로 샘플링되고 16 비트 A/D (Analog-to-Digital) 변환된 음성 신호를 사용하며, 프레임은 25 ms이고, 한 프레임은 다시 4개의 부프레임으로 분할된다. 제안된 혼성 부호화기의 2.4 kbps 비트 할당은 표 2와 같다.

부호화 및 복호화는 25 ms 길이의 프레임 단위로 수행된다. 파라미터 분석을 위한 미리보기 (lookahead)는 10 ms로서 전체 지연은 35 ms이다. 우선 음성 신호가 입력되면 고역 통과 필터를 거쳐 DC 성분을 제거한 다음, 프레임 단위로 10차 LPC 분석을 수행한다. 10차의 LPC는 LSP로 변환되고, 프레임 간 상관 관계를 제거하고 2 단계 벡터 양자화를 한 후, 5차씩 묶어 분리 벡터 양자화를 한다. 이때 LSP 양자화에 할당된 비트는 유성음일 경우는 24 비트이며 무성음일 경우는 23 비트이다. LSP 양자화 코드북은 유/무성음에 따라 다르게 설계하였으며, 유/무성음 정보에 따라 스위칭된다. 다음으로 프레임 당 한 개의 개회로 방식으로 20~120 샘플의 범위에서 피치를 구하고 이를 6 비트 양자화한다. 유/무성음 정보는 1 비트로 양자화하고 유/무성음 정보에 따라 나머지 파라미터의 비트 할당을 다르게 한다. 무성음에서는 부프레임 단위로 개회로 방식으로 평균 전력을 구하고 이를 이용하여 목적 잔차 신호를 정규화한다. 전력은 5 비트를 할당하여 스칼

표 2. 제안된 2.4 kbps EHC 부호화기의 비트 할당
Table 2. Bit allocation of the proposed 2.4 kbps EHC coder.

	Unvoiced	Voiced
LPC	23	24
V/UV	1	1
Pitch		6
Band Voicing		3
Power	4×4	5×2
Shape	5×4	8×2
Total/25ms	60	60
kbps	2.4	2.4

라 양자화하고, 정규화된 잔차 신호는 고속 RP- VSELP를 사용하여 5 비트로 벡터 양자화된다. 유성음 구간에서는 두 개의 부프레임 단위로 하모닉을 추정하고 평균 전력으로 정규화한 후, 전력은 5 비트 스칼라 양자화한다. 정규화된 하모닉 크기 벡터는 차원 변환을 거쳐 코드북 검색을 통해 8 비트 벡터 양자화된다. 하모닉 합성 과정에서 혼합할 무성음 (잡음) 성분의 양을 나타내는 대역 유성음 정도는 프레임 단위로 3 비트 벡터 양자화된다.

본 논문에서는 제안된 2.4 kbps EHC 부호화기의 계산량을 평가하기 위해서 2 kbps HVXC와 개인용 컴퓨터에서 비교 모의 실험을 수행하였다. 실험에서 2.4 kbps EHC와 2 kbps HVXC는 32 비트 부동 소수점 연산을 갖는 C 코드로 구성되어 있으며, 2 kbps HVXC는 MPEG-4에서 제공하는 코드[25]를 사용하였다. 계산량 평가 실험 환경은 다음과 같다.

- 음성 파일: 8 초, 8 kHz 샘플링, 여성화자.
  "여동생에게 책을 사주었습니다. 친구가 전화를 걸어주었습니다."
- 컴퓨터 기종: 펜티엄-III 866 MHz, 128MB RAM
- 플랫폼: Visual C++ 6.0
- 컴파일 옵션: Debug, Enable profiling, Generate mapfile
- 프로파일 옵션: function timing

표 3은 위의 실험 환경에서 10회 반복하여 프로파일링을 통해 구한 수행 시간을 평균한 것이다. 이 때 수행 시간은 실험 환경에 따라서 달라질 수 있지만 수행 시간의 상대적인 비율을 통해 볼 때 제안된 부호화기는 개선된 모듈의 도입과 제안으로 인해 전체적으로 HVXC의 20 %의 계산량만을 필요로 함을 알 수 있다. 이러한 비율은 프레임 단위로 계산량을 측정하면 달라질 수 있으며, PC의 CPU와는 구조가 다른 DSP (Digital Signal Processor)에서는 차이가 발생할 수 있다. 그러나 대략적으로는 큰 차이는 없을 것으로 기대된다.

표 3. 계산량 비교
Table 3. Complexity comparison.

	Encoder (ms)	Decoder (ms)	Total (ms)
2 kbps HVXC	4228.386	1230.645	5459.032
2.4 kbps EHC	615.502	469.321	1084.823
Ratio (EHC/HVXC)	14.55 %	38.11 %	19.87 %

제안된 2.4 kbps EHC 부호화기의 음질 평가를 위해서는 부호화기의 특성상 segSNR (segmental Signal-to-Noise Ratio)과 같이 객관적인 음질 평가는 부적절하므로 주관적이고 비공식적인 평가 방법인 MOS (Mean Opinion Score) 평가를 수행하였다. 비교 기준 부호화기로 G.723.1 5.3 kbps ACELP[26]와 2.4 kbps MELP[19], 4 kbps / 2 kbps HVXC[4,25]를 사용하였으며, 양자화하지 않은 제안된 부호화기의 합성음과 원음도 비교 대상에 포함시켰다. 청취 집단으로는 훈련되지 않은 연구실의 남녀 15명을 대상으로 하였고, 시험에 사용된 음성 샘플은 화자가 서로 다른 남녀 각각 5 문장씩 총 10 개의 문장을 사용하였다. MOS 평가 결과는 표 4에 나타내었으며 절대적인 값보다는 원음과의 상대적인 값만이 의미가 있다. 평가 결과에서 CELP 계열 부호화기는 여성 화자보다 남성 화자에서 성능이 우수하였으며 하모닉 계열 부호화기에서는 남녀 화자 모두 비슷한 성능을 보이는 등 각 부호화기의 특성을 나타내었다. 제안된 2.4 kbps EHC 부호화기는 5.3 kbps ACELP와 4 kbps HVXC 보다는 낮은 음질을 보여 주었지만, 2 kbps HVXC와는 유사한 음질을, 그리고 2.4 kbps MELP 보다는 우수한 음질을 보여 주었다.

표 4. MOS 평가 결과
Table 4. MOS results.

	Female	Male	Total
Original Speech	4.31	4.65	4.48
5.3 kbps ACELP	2.88	3.46	3.17
4 kbps HVXC	2.91	3.38	3.15
2 kbps HVXC	2.65	2.84	2.75
2.4 kbps MELP	1.94	2.47	2.21
2.4 kbps EHC	2.52	2.83	2.68

V. 결 론

본 논문에서는 자연과 계산량 면에서 효율적인 저 전송률 하모닉-CELP 부호화기를 제안하였다. 제안된 부호화기에서는 프레임 단위 유/무성음 판별에 따라 무성음 구간에서는 고속 CELP 방식인 RP-VSELP로 부호화하고 유성음 구간에서는 개선된 하모닉 부호화를 수행한다. 이러한 개선된 부호화 알고리듬을 통해서 주관적인 음질 평가 결과, 제안된 2.4 kbps EHC 부호화기는 적은 계산량으로 양호한 음질을 얻을 수 있음을 확인하였다. 현재 제안된 EHC 부호화기의 음질 향상을 위해서 청각 특성을 반영한 대역별 유/무성음 합성 방법에 관한 연구를 진행 중에 있다.

감사의 글

논문이 완성될 수 있도록 HVXC에 관한 자료 수집에 큰 도움을 주신 충북대학교 대학원 전파공학과 김종학 박사과정께 깊은 감사를 드리며, 음질 평가에 참여해주신 연세대학교 대학원 전기전자공학과 음향 음성 및 신호처리 연구실 여러분께 감사드립니다.

참 고 문 헌

1. M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp. 25.1.1-25.1.4, 1985.
2. W. B. Kleijn, "Encoding Speech Using Prototype Waveforms", *IEEE Trans. on Acoust. Speech and Signal Proc.*, Vol. **1**, No. 4, pp. 386-399, 1993.
3. J. C. De Martin and A. Gersho, "Mixed-Domain Coding of Speech At 3 kbps," *IEEE Proc. Int. Conf. Acoust., Speech, Signal Proc.*, pp. 216-219, 1996.
4. ISO/IEC FCD 0.1 Subpart 2, Information Technology Very Low Bit Rate Audio-Visual Coding, 1998.
5. D. W. Griffin and J. S. Lim, "Multi-Band Excitation Vocoder," *IEEE Trans. on Acoust. Speech and Signal Proc.*, Vol. **36**, No. 8, pp. 1223-11235, 1988.
6. Digital Voice Systems, Inc., APCO Project 25 Vocoder Description Version 1.3, Jul. 1993.
7. Yong-Soo Choi, Hong-Goo Kang, and Dae-Hee Youn, "A Fast VSELP Speech Coder Based on Mutually Orthonormal Regular Pulse Vectors," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp. 554-557, 1996.
8. Yong-Soo Choi, Sang-Wook Park, and Dae-Hee Youn, "Fast Vector-Sum Codebook Search Method for Low Bit Rate Speech Coding," *IEE Electronic Letters*, Vol. **33**, No. 6, 1997.
9. Yong-Soo Choi, Hong-Goo Kang, Sang-Wook Park, Jae-Ha Yoo, and Dae-Hee Youn, "Improved Regular Pulse VSELP Coding of Speech at Low Bit-rate," *EUROSPEECH'97*, pp. 1267-1270, 1997.
10. I. M. Transco, L. B. Almedia and J. M. Tribolet, "A Study on the Relationships between Stochastic and Harmonic Coding," *IEEE Proc. Int. Conf. Acoust., Speech, Signal Proc.*, pp. 216-219, 1986.
11. ITU-T Recommendation G.729, Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic Code Excited Linear Prediction (CS-ACELP), 1995.
12. Y. Linde A. Buzo and R. M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. on Comm.*, Vol. **COM-28**, pp. 84-95, 1980.
13. K. K. Paliwal and B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/frame," *IEEE Trans. on Acoust. Speech and Signal Proc.*, Vol. **1**, No. 1, pp. 3-14, 1993.
14. W. B. Kleijn and K. K. Paliwal, *Speech Coding and Synthesis*, Elsevier, 1995.
15. 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.
16. Ki-Seung Lee and R. V. Cox, "TTS Based Very Low Bit Rate Speech Coder," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp. 181-184, 1999.
17. T. Eriksson and H. G. Kang, "Pitch Quantization in Low Bit-Rate Speech Coding," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp. 489-192, 1999.
18. EIA/TIA/IS-96A Interim Standard: Speech Service Option Standard for Wideband Spread Spectrum Digital Cellular System, Qualcomm Inc., May 1995.
19. 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.
20. G. Cubin, B. S. Atal, and W. B. Kleijn, "Performance of Noise Excitation for Unvoiced Speech," Proc. *IEEE Workshop on Speech Coding for Telecomm.*, pp. 35-36, 1993.
21. I. A. Gerson and M. A. Jasuik, "Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8 kbps," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp. 461-464, 1990.
22. R. Hagen, E. Ekudden, B. Johansson, and W. B. Kleijn, "Removal of Sparse-Excitation Artifacts in CELP," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp. 145-148, 1998.
23. E. Zwicker and H. Fastle, Psycoacoustics: Facts and Models, Springer-Verlag, 1990.
24. M. W. Macon and Mark A. Clements, "Sinusoidal Modeling and Modification of Unvoiced Speech," *IEEE Trans. on Speech and Audio Proc.*, Vol. **5**, No. 6, 1997.
25. ISO/IEC JTC1/SC29/WG11, Report on the MPEG-4 Speech Codec Verification Tests, N2424, October 1998.
26. ITU-T Recommendation G.723.1, General Aspects of Digital Transmission Systems: Dual Rate Speech Coder For Multimedia Communications Transmission at 5.3 and 6.3 kbit/s, 1996.

저자 약력

● 최 용 수 (Yong-Soo Choi)

1993년 2월: 연세대학교 전자공학과 (공학사)
1995년 2월: 연세대학교 전자공학과 (공학석사)
1999년 8월: 연세대학교 전기컴퓨터공학과 (공학박사)
1999년 9월~2000년 8월: LG정보통신 네트워크연구소 선임연구원
2000년 9월~현재: LG전자 네트워크연구소 선임연구원
※ 주관심분야: 음성신호처리, 음성부호화, 음질평가, 실시간 DSP 응용

● 김 경 민 (Kyong-Min Kim)

1999년 2월: 연세대학교 전자공학과 (공학사)
1999년 9월~현재: 연세대학교 전기전자공학과 (공학석사)
※ 주관심분야: 음성신호처리, 음성부호화

● 윤 대 희 (Dae-Hee Youn)

1977년 2월: 연세대학교 전자공학과 (공학사)
1979년 8월: Kansas State University 전기공학과 (공학석사)
1982년 8월: Kansas State University 전기공학과 (공학박사)
1982년 8월~1985년 6월: University of Iowa 전기공학과 조교수
1985년 9월~1990년 2월: 연세대학교 전자공학과 조교수
1990년 3월~1995년 2월: 연세대학교 전자공학과 부교수
1995년 3월~1996년 2월: 연세대학교 전자공학과 교수
1996년 3월~현재: 연세대학교 기계전자공학부 전기공학전공 교수
1999년 3월~현재: 연세대학교 대학원 전기, 전자공학과 교수
1998년 2월~현재: 연세대학교 신호처리연구센터 소장
1998년 8월~2000년 6월: 연세대학교 정보통신처장
※ 주관심분야: 디지털신호처리, 음성/오디오신호처리, 적응신호처리, 레이다/소나 신호처리, DSP 응용, ASIC