

MLT 여기신호를 이용한 광대역 음성 부호화기 설계

Design of Wideband Speech Coder Using the MLT Residual Signal

오연선*, 신재현*, 이인성*

(Yeon-Seon Oh, Jae-Hyun Shin, In-Sung Lee*)

*충북대학교 전자공학과

(접수일자: 2005년 4월 8일; 수정일자: 2005년 5월 24일; 채택일자: 2005년 6월 10일)

본 논문에서는 대역분할 광대역 음성 부호화기의 구조와 음질 향상을 위한 새로운 고대역 구조를 제안한다. 대역분할 방식에 의해 광대역 음성은 저대역 (0~4kHz) 음성과 고대역 (4~8kHz) 음성으로 나뉘어 각각 G.729E와 MLT(Modulated Lapped Transform) 여기모델을 적용하여 서로 독립된 방식으로 부호화한다. 4kbps의 낮은 전송률로 부호화되는 고대역에서는 MLT 여기모델을 효율적으로 이용하기 위하여 유·무성음을 구별하였고 유성음에 대해서는 저대역 피치주기를 이용한 MLT peak picking 방법을 적용하였다. 즉, MLT 변환된 여기신호는 주기적인 피크를 갖는 주기신호로 나타나며 이때의 피크값을 추출하여 양자화하여 전송한다. 무성음에 대해서는 에너지 값에 따라 비트를 달리 적용하고, 선형예측 스펙트럴 응답이 가중된 MLT 벡터 양자화 방법을 적용하였다. 제안된 15.8kbps 광대역 음성 부호화기의 성능평가는 주관적인 음질평가로 선호도 테스트를 수행하였다.

핵심용어: 광대역 음성 부호화기, 대역분할 부호화기, MLT, VQ

투고분야: 음성처리 분야 (2.2)

In this paper, the structure of a split bandwidth wideband speech coder and its highband coder for tone quality elevation are proposed. The lowband and highband by the split bandwidth method are encoded independently applying the G.729E and MLT (Modulated Lapped Transform) residual model. In the highband structure which is encoded by low bit rate of 4kbps, the MLT residual signals are distinguished to voice and unvoice signal. The voice signals are applied to MLT peak picking method by lowband pitch period. Because transformed MLT residual signals are represented by periodic signal that have periodic peak. The unvoice signals are applied to MLT which linear prediction spectral response is added and do vector quantization. Performance for proposed 15.8kbps wideband speech coder was verified through subjective listening test.

Keywords: MLT, VQ

ASK subject classification: Speech Signal Processing (2.2)

I. 서론

현재 이동통신 및 인터넷 통신의 가장 큰 특징은 넓은 대역폭을 사용하여 음성에서 동영상까지 다양한 멀티미디어 데이터를 전송할 수 있다는 것이다. 기존의 음성 위주의 통신 방법과 달리 데이터와 영상 전송이 가능하게 되었고 음성통신에 있어서도 자연스런 음질과 높은 인지도를 나타내는 50~7000Hz의 광대역 음성신호를 제공한다는 점이다. 이러한 광대역 음성의 품질은 이

동통신, VoIP 및 영상회의 등과 같은 분야에서도 절실히 요구되고 있다.

최근 이동통신 및 멀티미디어 분야에서 이러한 고품질 음성의 요구가 커짐에 따라 ITU-T에서는 패킷망, 고품질 오디오/비디오 회의, 3G wireless 등에 응용할 목적으로 embedded 개념의 새로운 가변 비트율 음성 코덱 개발이 활발하게 진행되고 있다. embedded 코덱은 채널 환경에 따라 최적의 비트율과 대역폭을 제공할 수 있어 패킷 교환망에 적합한 방식이다. 특히 코덱의 출력은 낮은 비트율의 저품질 음성에서 높은 비트율의 고품질 음성으로의 scalable 구조를 사용함으로써 넓은 범위의 서비스에 음성의 활용이 가능하게 된다. 또한 scalable 구

조를 갖춘 embedded 코더의 사용은 비트에러에 대한 강인성과 비트활용의 유연성을 제공함으로써 개선된 시스템을 설계할 수 있다[2]. 국내의 광대역 음성 코덱에 대한 연구는 유선 및 무선 음성통신 뿐만 아니라, 저장장치 등의 다양한 응용 분야에서 기존 협대역 음성 코덱과의 호환성을 지원하는 동시에 고품질의 음성 서비스를 제공할 수 있는 연구가 활발하게 이루어지고 있다[3].

본 논문에서는 ITU-T에서 최근 표준화가 진행 중인 Embedded Variable Bit Rate (EV) 음성 부호화 알고리즘의 설계를 목표로 15.8kbps의 전송률을 갖는 광대역 음성 부호화기를 설계하였다. 특히, 현재의 코덱들이 고정된 대역폭과 비트율만을 가짐으로서 네트워크의 트래픽 상태에 최적으로 적응할 수 없는 점을 이용하여 네트워크의 트래픽 상태와 단말의 처리속도에 따라 최적의 음질을 제공하는 비트율과 대역폭 관점에서 embedded 구조에 적합한 기본 구조의 코더를 설계하였다. 제안된 음성 부호화기는 저대역과 고대역 신호로 나누어 분석하는 대역분할 구조를 기반으로 하였다[4]. 대역분할에 의한 저대역 음성은 협대역 음성 통신과의 호환성을 고려하여 11.8kbps의 전송률을 갖는 G.729E를 사용하였다. 포맷트 특성과 주기적인 특성이 저대역 신호보다 더 미약한 고대역 신호는 transform 기반의 부호화 방법을 사용하되 신호의 특성을 고려하여 성능을 향상시킬 수 있는 방법에 대해 연구하였다.

본 논문은 다음과 같이 구성된다. II장에서는 제안된 광대역 음성 부호화기의 기본 구조에 대해 설명하고, III장에서는 고대역 음성 부호화기의 구조와 동작을 설명한다. IV장에서는 주관적인 음질평가에 따른 실험 결과를 설명하고 V장에서 결론을 맺는다.

II. 광대역 음성 부호화기 구조

광대역 음성 부호화기의 대역분할 구조는 일반적으로 그림 1과 같다. 16kHz로 샘플링된 광대역 음성신호는

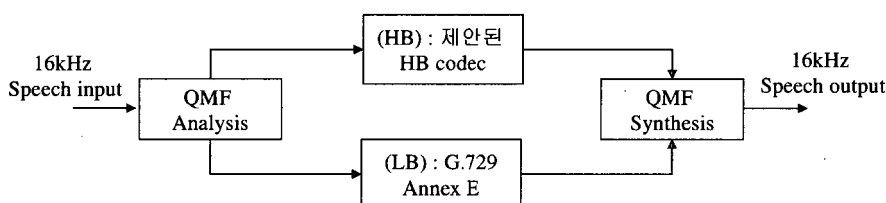


그림 1. 광대역 음성 부호화기 구조
Fig 1. Block diagram of the wideband speech coder.

96차 QMF (Quadrature Mirror Filter) 분석필터를 통과함으로써 8kHz의 저대역 신호와 고대역 신호로 나누어진다. 저대역 신호는 기존의 협대역 음성 부호화기 G.729E를 적용하여 11.8kbps의 전송률로 부호화한다 [5]. 이때 20ms 한 프레임을 처리하기 위하여 G.729E의 프레임 처리 단위인 10ms의 서브 프레임을 두 번 실행한다. 고대역 신호는 본 논문에서 제안한 방법을 적용하여 부호화한다. 입력신호는 20ms 프레임마다 MLT 전송 파라미터를 추출하고 벡터 양자화하여 전송한다. 복호화기에서 복원된 신호는 최종적으로 QMF 합성필터를 통과함으로써 광대역 출력신호를 얻을 수 있다. 이때 저대역과 고대역의 지연값이 서로 같지 않으므로 저대역에 지연값을 보상해 주어 합성필터를 통과한다. 고대역에서는 4kbps의 전송률을 가지고 부호화하며 전체 15.8kbps의 전송률을 갖는 광대역 음성 부호화기를 설계하였다.

III. 제안된 고대역 음성 부호화기

본 논문에서 제안한 고대역 음성 부호화기의 구조는 그림 2, 3과 같다. QMF에 의해 분리된 4000~8000Hz의 고대역 음성신호는 한정된 비트를 가지고 효율적으로 부호화하기 위하여 LSP 양자화 과정을 수행한다. 먼저 대역 분리된 고대역 입력신호는 20ms를 한 프레임으로 처리하며 자기상관방법에 의하여 6차 LPC 계수를 계산한다. 이 계수들은 양자화와 보간을 위해 LSP로 변환되고 1단계 벡터 양자화 과정을 수행한다. 전송 인덱스는 9bit를 할당하였다. 양자화되고 보간된 LPC 계수에 의하여 잔여신호를 구하게 되며 이 잔여신호는 주파수 변환방법에 있어 효율적인 틀을 제공하는 MLT를 이용한다 [6]. 이후 유성음과 무성음을 구별하여 각각의 특성에 맞게 양자화하여 부호화한다.

3.1. 유·무성음 검출

고대역 음성신호는 20ms의 매 프레임 마다 LPC 분석

을 실시하고, 유·무성음 모드를 판별한다. 모드 판별시에는 고대역 신호의 단독적인 모드 결정을 하지 않고, 저대역 신호의 모드판별 값을 고려하여 모드를 결정한다[7].

고대역 음성에 대한 모드 선택을 위해 다음과 같은 기본 가정을 한다.

1. 저대역 신호가 무성음 혹은 묵음이라면 고대역 신호에는 하모닉이 없다고 가정한다.
2. 저대역 신호가 유성음인 시간 축에서도 일부 하모닉이 없는 구간이 존재 한다고 가정한다.
3. 저대역 신호와 고대역 신호의 각각의 하모닉 존재 유무의 선택을 따로 하여 논리곱으로써 고대역의 하모닉 존재 유무를 알 수 있다고 가정한다.

이러한 가정과 각 대역에서의 하모닉 존재 유무를 파악하기 위한 잔여신호의 또렷함 비율 (Sharpness Rate : SR), 잔여신호의 에너지 비율 (Energy Rate : ER), 영교차율 (Zero-crossing Rate : ZR), 전처리된 음성신호의 1차 예측 계수 (First-order prediction Coefficient : FC)의 4가지 알고리즘을 사용하였다.

또렷함 비율, SR은 한 프레임 내에서 LPC 잔여신호의 가장 큰 크기를 갖는 신호와 평균 크기를 갖는 신호들의 비율을 나타낸 것이다. 이것은 하모닉 성분이 많을수록 더욱 큰 값을 갖는다.

$$SR = \frac{1}{HFRM} \frac{\sum_{n=0}^{HFRM-1} |r(n)|}{\max_{n=0,1,\dots,HFRM-1} |r(n)|} \quad (1)$$

여기서 HFRM은 한 프레임인 160샘플을 의미한다. 또한, $r(n)$ 은 한 프레임의 LPC 잔여신호를 의미한다.

LPC 잔여신호의 에너지 비율, ER은 한 프레임의 중간을 기준으로 잔여신호의 좌우 에너지 비율을 나타낸다. 식(2)에서 ER이 1에 가까울수록 프레임에서 좌우 에너지의 비율이 비슷하며, 0에 가까울수록 에너지 비율이 많은 차이를 나타낸다.

$$E(0) = \sum_{n=0}^{HFRM/2-1} r^2(n) \quad E(1) = \sum_{n=HFRM/2}^{HFRM-1} r^2(n)$$

$$ER = 1 - \frac{|E(0) - E(1)|}{E(0) + E(1)} \quad (2)$$

영 교차율, ZR은 전처리된 음성신호의 프레임당 부호의 변화율을 나타낸 것이다. 일반적으로 유성음의 ZR이 무성음의 ZR보다 작다고 알려져 있다.

전처리된 음성신호의 1차 예측 계수, FC는 식 (3)과 같이 음성신호의 상관도를 나타내는 척도로써 유성음일수록 상관도가 높아 FC 또한 1에 가깝다.

$$FP = \frac{\sum_{n=0}^{HFRM-2} \tilde{s}(n) \tilde{s}(n+1)}{\sum_{n=0}^{HFRM-1} \tilde{s}^2(n)} \quad (3)$$

여기서 $\tilde{s}(n)$ 은 전처리된 음성신호를 나타낸다.

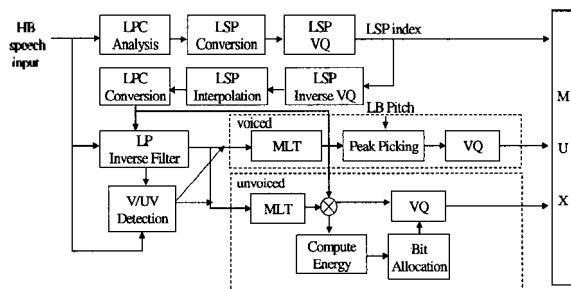


그림 2. 제안된 고대역 부호화기
Fig 2. Block diagram of the highband encoder.

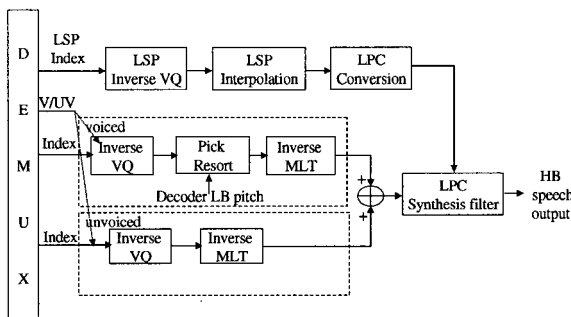


그림 3. 제안된 고대역 복호화기
Fig 3. Block diagram of the highband decoder.

3.2. MLT 변환방법

ITU-T의 권고안 G.722.1에 적용된 MLT 변환 방법은 블록킹 효과를 제거하기 위하여 블록 경계선의 범위를 넘는 확장된 기본 함수를 가지고 있다. 즉, 데이터 블록 간의 50%의 오버랩을 가지며 완전복원이 가능하고 변환 코딩 방법에 있어 거의 최상의 성능을 가진다[8]. 제안된 부호화기에 쓰인 MLT 구조는 그림 4에 나타내었다 [8-9].

MLT 변환식은 다음과 같다. DCT Type-IV와 $v(n)$ 이 결합하여 결과식 $M(m)$ 을 얻는다.

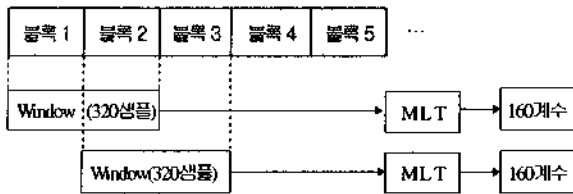


그림 4. MLT 구조
Fig 4. Structure of the MLT.

$$M(m) = \sum_{n=0}^{N-1} \sqrt{\frac{2}{N}} \cos\left(-\frac{\pi}{N}(n+0.5)(m+0.5)\right) v(n) \quad (4)$$

여기서 N 은 입력샘플 수를 의미하며, $v(n)$ 은 현재 프레임과 과거 프레임의 윈도우 값의 합으로 식(5)로 정의한다.

$$v(n) = u\left(\frac{N}{2}-1-n\right)x\left(\frac{n}{2}-1-n\right) + u\left(\frac{N}{2}+n\right)x\left(\frac{N}{2}+n\right) \quad \text{for } 0 \leq n \leq \frac{N}{2}-1$$

$$v\left(n + \frac{N}{2}\right) = u(N-1-n)x(N+n) - u(n)x(2N-1-n) \quad \text{for } 0 \leq n \leq \frac{N}{2}$$

$x(n)$ 은 MLT 변환을 위해 입력되는 신호를 의미하며 다음과 같은 윈도우가 사용된다.

$$u(n) = \sin\left(\frac{\pi}{2N}(n+0.5)\right) \quad \text{for } 0 \leq n \leq N \quad (6)$$

역변환 방법의 DCT Type-IV는 다음과 같다.

$$u(n) = \sum_{m=0}^{N-1} \sqrt{\frac{2}{N}} \cos\left(\frac{\pi}{N}(m+0.5)(n+0.5)\right) M(m) \quad (7)$$

현재 프레임과 과거 프레임의 MLT 출력의 1/2이 윈도우와의 중첩 합산방법에 사용된다. 역변환된 MLT의 출력은 식(8)과 같이 정의한다.

$$y(n) = u(n)u\left(\frac{N}{2}-1-n\right) + u(N-1-n)u_{old}(n) \quad \text{for } 0 \leq n \leq \frac{N}{2}-1$$

$$y\left(n + \frac{N}{2}\right) = u\left(\frac{N}{2}+n\right)u(n) - u\left(\frac{N}{2}-1-n\right)u_{old}\left(\frac{N}{2}-1-n\right) \quad \text{for } 0 \leq n \leq \frac{N}{2}-1 \quad (8)$$

사용되지 않은 $u(n)$ 의 1/2은 u_{old} 로 저장된다.

$$u_{old}(n) = u_{old}\left(n + \frac{N}{2}\right) \quad \text{for } 0 \leq n \leq \frac{N}{2}-1 \quad (9)$$

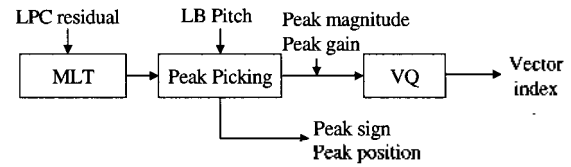


그림 5. 유성음 부호화기
Fig 5. Block diagram of the voice encoder.

위의 MLT 변환된 계수는 유·무성음 구별에 따라 각각의 신호 특성에 적합한 독립된 부호화 방법을 따른다.

3.3. 유성음 모드의 부호화

그림 5는 유성음 모드의 부호화기를 나타낸 것이다.

그림 5. 유성음 부호화기

유성음으로 판별된 LPC 잔여신호는 한 프레임마다 MLT 변환되고 이 신호는 주기적인 피크를 갖는 주기신호로 나타난다. 이에 주기적인 피크 및 주변피크를 검출하고 양자화하여 전송하는 부호화 방식을 적용하였다. 여기서 MLT 신호의 큰 피크 값을 많이 찾을수록 스펙트럼의 하모닉 표현이 잘 나타나므로 좋은 성능을 가지게 된다. 그러나 검출된 피크 값들은 피크 크기, 부호, 위치 등 많은 비트를 필요로 하므로 적절한 비트할당이 필요하다. 따라서 4kbps의 전송률을 고려하여 검색되는 피크 수를 제한하였고 그것의 위치정보는 주변피크의 위치 전송만으로 최소화 하였다. 중심피크의 위치정보는 저대역의 피치주기를 이용하여 주기적인 피크 위치를 찾을 수 있다. 이러한 저대역 신호의 기본 주파수에서 고대역 신호의 피크 위치를 찾는 이유는 저대역 신호의 기본 주파수의 정수배 되는 지점이 고대역의 시작점과 일치하지 않기 때문이다. 고대역 신호의 피크 위치와 저대역 신호의 주파수 관계를 식(10)으로 정의한다.

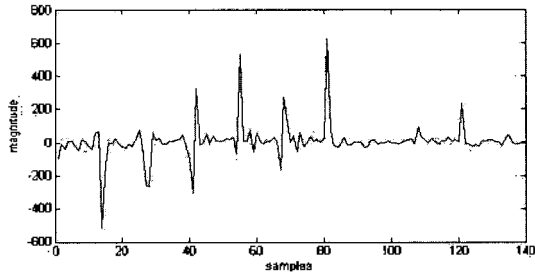
$$NUMHAM = LBpitch/2$$

$$P(k) = f(k + NUMHAM) - 4000Hz \quad (10)$$

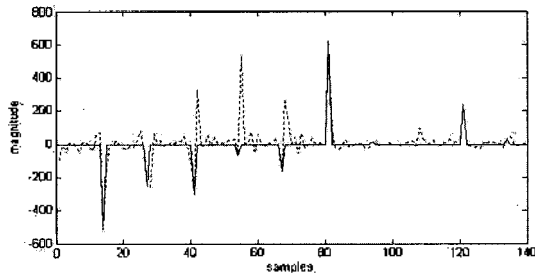
저대역의 하모닉 개수 $NUMHAM$ 에 의하여 고대역의 k 번째 피크 위치 $P(k)$ 를 구할 수 있다. 저대역의 피치 주기를 주파수로 변환 시킨 것은 f 로 정의한다. 또한 이러한 피크를 중심피크라 부른다. 먼저 저대역의 피치주기를 이용하여 기본 주파수의 정수배 되는 지점에서 10개의 중심피크 위치를 찾는다. 그러나 양과 음 동시에

나타나는 피크 값이나 피치주기의 배수와는 맞지 않은 큰 피크 값을 검색하기 위하여 중심피크 10개의 주변을 검색하여 19개의 추가적인 주변피크를 찾는다. 즉 10개의 중심피크 위치의 좌우 피크를 검색하여 더 큰 값의 피크를 찾아내는 것이다.

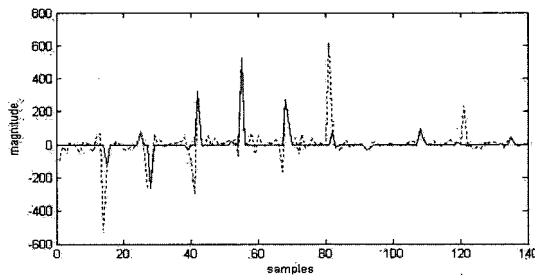
그림 6에 검출된 중심피크와 주변피크를 나타내었다.



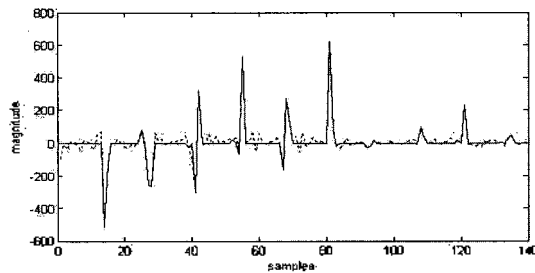
(a) MLT 변환된 음성 신호
(a) The voice signal transformed MLT.



(b) 저대역 피치에 의해 10개의 중심피크 검출
(b) Detection of 10 center peak by lowband pitch.



(c) 중심피크 주변 검색에 의한 19개의 주변피크 검출
(c) Detection of 19 surrounding peak by surrounding search of center peak.



(d) 최종 선택된 29개 피크
(d) Finally selected 29peak.

그림 6. 29개의 피크검출 과정
Fig 6. Detection process of 29peak.

주변피크를 검색시에는 이에 대한 위치 전송이 필요하게 되며 주변피크 한 개당 1bit의 위치 정보를 할당한다. 이에 추가적으로 19bit를 할당하였다. 이렇게 구한 총 29개에 대한 중심 및 주변피크 값은 에너지 값으로 나누어져 정규화 시키며 크기와 부호, 위치, 이득 값을 벡터 양자화하여 전송한다. MLT 신호의 크기 이득과 정규화된 크기는 각각 11bit를 사용하여 양자화하고 피크 부호는 피크당 1bit를 할당하여 양 또는 음 값을 표시하여 전송한다. 주변피크의 위치 정보는 중심피크의 좌 혹은 우 값을 표시한다. 이때 부호화기에 사용된 저대역 피치주기는 저대역의 복호화기로 전송되므로 따로 전송하지 않는다. 고대역의 복호화기에서는 저대역의 복호화기에서 구한 피치 값을 사용하여 10개의 중심피크 위치를 복원해내고 전송된 피크 정보를 가지고 19개의 주변피크 위치를 정렬한다. 이로서 MLT의 여기신호를 복원한다. 선택되지 못한 위치 값에 대해서는 잠음을 추가한다.

3.4. 무성음 모드의 부호화

유·무성음 구별에 의해 무성음으로 판별된 잔여신호는 20ms마다 MLT 변환되어 주파수 영역 값으로 나타나고 음성 정보를 포함하는 7kHz 대역 즉, 120샘플만을 선형예측 효과를 가중시켜 벡터 양자화하여 전송한다. 그림 7은 벡터 양자화기의 구조를 나타내고 있다.

무성음 부호화를 위해 입력되는 잔여신호는 백색잡음과 같이 스펙트럼 정보를 포함하고 있지 않다. 따라서 스펙트럼 정보를 얻기 위하여 역 양자화된 LPC 계수를 이용하여 선형예측 스펙트럼 응답을 계산하고 여기에 MLT 변환신호를 가중시킨 $P(m)$ 값을 이용하여 부호화 한다.

$$P(m) = |H(m)|M(m) \tag{11}$$

선형예측 스펙트럼 응답을 가중시킨 MLT 잔여신호는 10차원의 코드북을 구성되고 12개의 부 블록으로 나뉘어

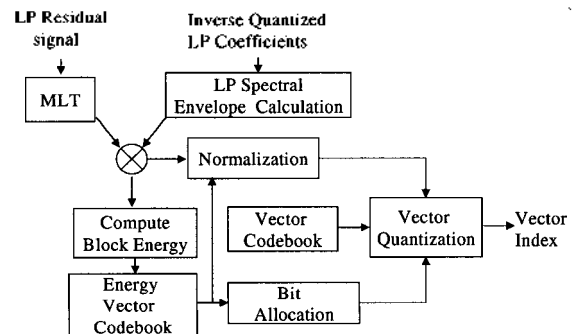


그림 7. 벡터 양자화기 구조
Fig 7. Structure of the vector quantization.

저 각 블록마다 에너지가 계산된다. 에너지 값은 에너지에 대한 벡터 코드북을 이용하여 양자화하고 코드북의 인덱스를 부가 정보로 전송한다. 에너지 인덱스 값은 비트할당 함수와 연결되어 설계되므로 각 부 블록에 할당되는 비트 정보를 제공한다. 한 프레임 내에서 음성 스펙트럼은 모든 대역에 균등하게 분포되어 있지 않으므로 각 대역에 동일한 비트를 할당하지 않고 에너지가 큰 중요대역에 더 많은 비트를 할당한다. 식 (12)는 블록 에너지와 블록마다 할당되는 비트와의 관계를 나타낸 것이다 [10].

$$R_i = R_{avr} + \frac{1}{2} \log_2 \frac{\sigma_i^2}{\prod_{k=1}^N \sigma_k^2} \quad (12)$$

식 (12)로 부터 i 번째 부 블록에 할당되는 비트는 평균 비트율 R_{avr} 과 σ_i^2 에 관계됨을 알 수 있다. 각 부 블록 내의 선형예측 효과가 가중된 MLT 계수들은 에너지 벡터 코드북의 에너지 값으로 나뉘어 짐으로서 정규화되고 비트할당 정보를 이용하여 벡터 양자화 한다. 이득 양자화에는 11bit를 사용하였고 크기에는 59bit를 사용하였다. 복호화기에서는 역 양자화를 통해 한 블록에 해당하는 비트정보와 이득과 크기값을 복원하고, 역MLT과정에 의하여 무성음 신호를 복원한다.

IV. 실험 결과

제안된 고대역 부호화기는 20ms 프레임 길이를 사용하며 look-ahead 5ms와 MLT delay 20ms를 가지고 있어 총 45ms의 지연값을 갖는다. 이는 저대역의 look-ahead 5ms에 대하여 지연 차이가 발생하므로 복호화기에서 두 대역을 합성하기 전에 지연 값을 고려해야만 한다. 고대역 음성 부호화기에 대한 비트할당이 표 1에 나타나 있다. ITU-T에서는 16kbps의 새로운 광대역 음성 코덱의 음질을 G.722.2[11]의 12.65kbps보다 나쁘지 않은 수준을 요구하고 있다. 이를 참고하여 제안된 광대역 음성 부호화기의 성능은 G.722.2와 비교실험을 하였고 주관적인 음질평가로 a-b선호도 테스트를 수행하였다. 실험에 사용된 광대역 입력음성은 16kHz로 샘플링된 NTT 표준음성을 사용하였다. 여자20, 남자20, 총 40분 장을 사용하였으며 일반인 8명으로 이루어진 청취자를 대상으로 선호도 측정을 하였다. 표 2는 주관적인 음질

표 1. 4kbps 고대역 부호화기의 비트할당

Table 1. Bit allocation of the 4kbps highbandcoder.

	유성음	무성음
LSP	9	9
V/UV	1	1
MLT magnitude gain	11	11
MLT magnitude	11	59
MLT sign	29	
MLT position	19	
Total	80bit/20ms	

표 2. 광대역 음성 부호화기의 주관적 성능평가

Table 2. Performance evaluation of the wideband speech coder.

	설계된 부호화기의 15.8kbit/s	G.722.2의 12.65kbit/s
여자음성	51.2%	48.8%
남자음성	49.4%	50.6%
전체	50.3%	49.7%

평가의 결과이며 신뢰도 95%에 표본오차는 $\pm 9.8\%$ 이다.

저대역 음성의 11.8kbps와 고대역 음성의 4kbps로 이루어진 15.8kbps 광대역 음성을 G.722.2의 12.65kbps와 비교한 결과 비슷한 음질을 가지는 것으로 확인되었다. 또한, 제안된 부호화기를 통과한 출력신호는 고대역의 하모닉이 잘 표현되고 있음을 확인하였다. 추가적인 실험으로 16kHz로 샘플링된 음악신호를 광대역 입력신호로 사용하였다. 이 경우 코더의 성능은 그다지 나쁘지 않음을 확인하였다. 특히 저주파 부분의 CELP 부호화 방법에 의하여 대부분의 중요한 신호가 잘 복원되고 있어 낮은 전송률에 대한 고대역 신호의 단점을 감추는 효과를 얻을 수 있었다.

V. 결론

본 논문에서는 ITU-T에서 최근 표준화가 진행 중인 EV 음성 부호화기의 기본 코더 설계를 목표로 15.8kbps의 전송률을 갖는 광대역 음성 부호화기를 설계하였다. 제안된 대역분할 광대역 음성 부호화기는 저대역에 G.729E를 사용함으로써 다양한 응용분야에서 기존 협대역 통신 시스템과의 호환성을 지원하도록 하였다. 고대역에는 MLT 여기신호를 이용하여 유성음과 무성음을 구별하였고 각각을 다르게 부호화함으로써 데이터의 압축율을 높이는 방법을 제안하였다. 제안된 고대역 구조는 LPC 잔여신호를 주파수 영역으로 변환하고 유성음과 무성음을 구별한다. 신호의 주기성을 갖는 유성음일 경우

에는 저대역 피치주기를 이용하여 주기적인 피크와 그 주변피크 값을 검색하는 MLT peak picking 방법을 적용하였다. 저대역의 피치주기를 이용함으로써 중심피크에 대한 위치 정보는 전송하지 않으며 MLT의 크기와 부호, 주변위치 값만을 양자화하여 전송한다. 무성음일 경우에는 주파수 응답이 가중된 MLT 신호를 몇 개의 블록으로 나누어 블록마다 벡터 양자화 하는 방법을 적용하였다. 이와 같은 방법은 유성음과 무성음 구별 없이 몇 개의 블록으로 나누어 벡터 양자화하는 방법보다 유성음에서의 하모닉 표현을 많이 살릴 수 있었다. 또한 G.722.2 와 비교하여 비슷한 성능을 가지는 것으로 확인되었다.

제안된 광대역 부호화기의 고대역 구조는 MLT 파라미터를 양자화하여 전송하는 방식으로 파라미터 수가 적으며 추가적인 비트할당에 따른 embedded 구조에 쉽게 적용할 수 있다. 이는 소스코더의 수정 없이도 비트율의 변화를 쉽게 만들 수 있어 기존 코덱의 제한된 비트율과 대역폭의 단점을 보완할 수 있으며 scalable 구조에 따른 유연성과 강인성을 갖는 시스템으로 개선시킬 수 있다는 장점을 가지고 있다. 향후 설계된 기본 코덱을 바탕으로 추가적인 비트율 및 대역폭 조정으로 음성 뿐만 아니라 음악신호의 전송이 가능한 embedded 구조의 멀티 코덱을 설계하는 것이 가능하다.

감사의 글

본 연구는 2004년도 충북대학교 학술연구 지원사업의 연구비 지원으로 수행되었습니다.

참고 문헌

1. ITU Radiocommunication Assembly, Document 8 1045-E, "Detailed specifications of the radio interfaces of IMT-2000," Document 8 126(Rev.1), 2000.
2. B. Tang, A. Alwan, "A perceptually based embedded subband speech coder," IEEE transactions on speech and audio processing, 5 (2), Mar. 1997.
3. 이우석, 박호종, 손창용, 이양범, "대역폭 계층 구조의 광대역 음성 부호화기 개발," 음향학회지 23 (6), 481~487, 2004.
4. P. P Vaidyanathan, *Multirate Systems and Filter Banks*, (Prentice Hall), 189~285, 1993.
5. ITU-T Recommendation. G.729 Annex E, "11.8kbit/s CS-ACELP speech coding algorithm," Sep. 1998.
6. H. Malvar, "Fast algorithms for orthogonal and bio-

- orthogonal modulated lapped transforms," 1998 IEEE symposium on, 159~163, Jun. 1998.
7. 3GPP C.S0030-0, "Selectable mode vocoder service option for wideband spread spectrum communication system," Dec. 2001.
8. S. Shlien, "The modulated lapped transform, its time-varying form, and its application to audio coding stands," IEEE Transactions on speech and audio processing, 4, July 1997.
9. ITU-T Recommendation. G.722.1, "Coding at 24 and 32kbit/s for hands-free operation in systems with low frame loss," Sep. 1999.
10. A. K. Soman, P. P. Vaidyanathan, "Coding gain in paraunitary analysis/synthesis system," IEEE Transactions on signal processing, 41, 1824~1835, May 1993.
11. ITU-T Recommendation. G.722.2, "Wideband coding of speech at around 16kbit/s using Adaptive Multi-Rate Wide-Band," 2001.

저자 약력

• 오연선 (Yeon-Seon Oh)



2003년 2월: 홍익대학교 컴퓨터 정보통신공학과 학사
 2003년 9월~현재: 충북대학교 전자공학과 석사
 *관심분야: 음성 신호처리

• 신재현 (Jae-Hyun Shin)



1997년 2월: 충북대학교 전자공학과 학사
 2003년 8월: 충북대학교 전자공학과 석사
 2003년 9월~현재: 충북대학교 전자공학과 박사
 *관심분야: 음성 신호처리, 적응필터

• 이인성 (In-Sung Lee)



1983년 2월: 연세대학교 전자공학과 학사
 1985년 2월: 연세대학교 전자공학과 석사
 1992년 12월: Texas A&M University 전기공학과 박사
 1986년 5월~1987년 7월: 한국 통신 연구 개발단 전임 연구원
 1993년 3월~1995년 8월: 한국 전자통신 연구원 이동통신 기술연구단 선임 연구원
 1995년~현재: 충북대 전기전자컴퓨터 공학부 교수
 *관심분야: 음성 신호처리, 적응필터, 이동통신