

웨이브렛 변환을 적용한 광대역 음성부호화 알고리즘

Wideband Speech Coding Algorithm with Application of Wavelet Transform

이 승 원*, 배 건 성*
(Seung Won Lee*, Keun Sung Bae*)

*경북대학교 전자·전기공학부

(접수일자: 2001년 8월 3일; 수정일자: 2002년 5월 3일; 채택일자: 2002년 6월 27일)

협대역 음성부호화기에 비해 훨씬 우수한 합성음의 음질을 보이는 광대역 음성부호화기는 상대적으로 높은 전송률을 가져서 협대역 음성부호화기에 비해 사용범위가 제한되었다. 광대역 음성부호화기에서 이러한 전송 속도를 협대역 음성부호화기와 비슷한 수준으로 낮출 수 있다면, 보다 나은 음질의 음성 통신 시스템을 구현할 수 있을 것이다. 본 논문에서는 16 kHz로 샘플링된 입력 음성신호를 동일한 대역폭을 갖는 두 부대역으로 분리하여, 저대역 부호화에는 유럽의 이동통신 표준안인 GSM-EFR 협대역 음성부호화기를 적용하고, 고대역 부호화에는 웨이브렛 변환을 이용하여 고안한 부대역 음성부호화기를 적용한 광대역 음성부호화기를 제안하였다. 제안한 음성부호화기는 저대역 신호와 고대역 신호의 부호화에 각각 12.2 kbps, 6.7 kbps의 전송 속도를 할당하여 18.9 kbps의 전송속도를 가지며, 합성음의 음질은 56 kbps의 전송속도를 갖는 G.722 음성부호화기의 합성음과 비슷한 음질을 유지하였다.

핵심용어: 광대역 음성부호화기, 협대역 음성부호화기, 웨이브렛 변환

주요분야: 음성처리 분야 (2,2)

Wideband speech, characterized by a bandwidth of 50 ~ 7000 Hz, sounds more natural and intelligible, and is less tiring to listen to when compared to narrowband speech characterized by a bandwidth of 300 ~ 3400 Hz. Wideband speech coders, however, have not been as successful as the narrowband speech coders because of their higher bit rate. In this paper, we propose a new wideband speech coder which combines the European standard of a narrowband speech coder, i.e., GSM-EFR, and a transform coder using the discrete wavelet transform. The proposed wideband speech coder operates as follows: input speech is first split into two subbands with equal bandwidth and the two subband signals are coded and decoded by each subband coder. A GSM-EFR is adopted as a lower subband coder and a subband coder with wavelet transformed speech is designed for an upper subband coder. The total bit rate of the proposed coder is 18.9 kbps (12.2 kbps for lower band coder and 6.7 kbps for upper band coder), and informal listening test results have shown that the proposed coder has comparable speech quality to that of G.722 with 56 kbps.

Keywords: Wideband speech coder, Narrowband speech coder, Wavelet transform

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

I. 서론

50 ~ 7000 Hz의 대역폭을 갖는 광대역 음성부호화

책임저자: 배건성 (ksbae@ee.knu.ac.kr)
702-701 대구광역시 북구 산격동 1370번지
경북대학교 전자·전기공학부
전화: 053-950-5527; 팩스: 053-950-5505)

방식은 300 ~ 3400 Hz의 대역폭을 갖는 협대역 음성부호화 방식과 비교할 때, 그림 1에서 볼 수 있는 것처럼 저주파 성분을 250 Hz, 고주파 성분을 3600 Hz 정도 더 고려하여, 협대역 음성부호화 방식에서 다루는 대역폭의 두 배 이상의 주파수 범위를 고려한다. 이와 같은 광대역 음성부호화 방식에서는 증가한 저주파 성분이 음질의 자

연스러움을 개선시키고, 증가한 고주파 성분이 마찰음의 구분을 명확히 하여 전체적으로 음성의 명료도가 향상된다. 이처럼 우수한 음질을 갖는 광대역 음성부호화기는 협대역 음성부호화기를 이용한 통신보다 원음에 충실한 음색을 나타내고, 장시간 사용해도 피로도를 적게 느끼게 된다.

광대역 음성부호화 방식에 대한 연구는 1983년에 고음질 음성부호화기의 필요성을 느껴서 연구를 시작한 CCITT (Consultative Committee on International Telephone and Telegraphy)의 활동을 시초로 들 수 있다. CCITT의 Study Group XVIII는 1988년에 G.722 방식[1,2]을 광대역 음성부호화기의 표준안으로 채택하였는데, 이 방식은 48, 56, 64 kbps 세 가지 전송속도를 갖는 방식으로 만들어졌다. 하지만 협대역 음성부호화 방식에 비해 상당히 높은 전송률을 가져서 그 응용 분야가 화상회의 등과 같은 유선통신에 제한될 수밖에 없었다. 하지만 전송속도를 적절하게 낮출 경우 기존의 유선통신뿐만 아니라 무선 이동통신 등에도 응용할 수 있으리라 기대된다. 이같은 응용 목적으로 광대역 음성부호화기와 관련한 최근의 연구로는, 8 kHz로 샘플링된 음성신호뿐만 아니라 16 kHz로 샘플링된 신호도 다루는 동영상 전문가 그룹(MPEG: Moving Picture Expert Group) IV의 광대역-코드북 여가 선형 예측 분석(Wideband-CELP: Code Excited Linear Prediction)[3]이 ISO/IEC의 표준안으로 채택되었고, 유럽의 이동통신 표준화 기관인 ETSI (European Telecommunications Standard Institute)에서는 차세대 이동통신의 표준안으로 채택된 협대역 음성부호화 방식인 적응 다중 비트율(AMR: Adaptive Multi-Rate)을 광대역 음성부호화에도 적용할 수 있도록 확장한 적응 다중 비트율-광대역(WB: WideBand) 음성부호화 방식을 개발하여 표준화 인증 작업을 수행 중이다[4].

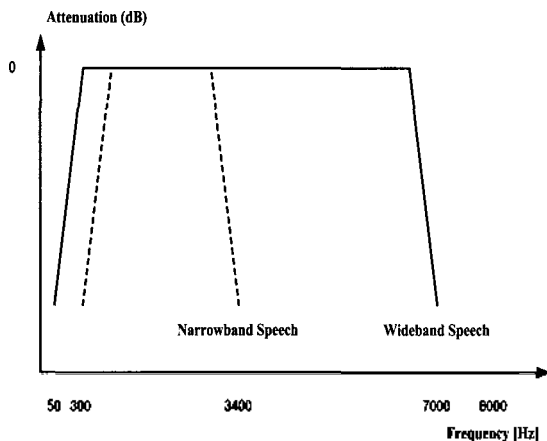


그림 1. 협대역 음성신호와 광대역 음성신호의 대역폭
Fig. 1. Narrowband and wideband speech bandwidth.

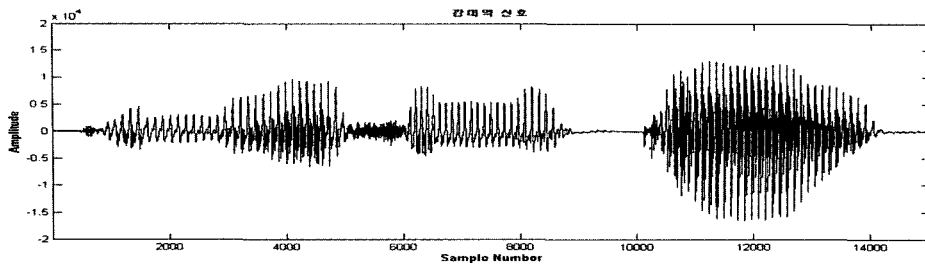
본 논문에서는 기존의 광대역 음성부호화 방식의 우수한 합성음을 갖는 장점을 유지하면서, 높은 전송률을 갖는 단점을 보완하여 유선통신뿐만 아니라 이동통신 등에도 응용 가능한 새로운 광대역 음성부호화 알고리즘을 제안하였다. 16 kHz로 샘플링된 음성신호를 동일한 대역폭을 갖는 저대역 및 고대역으로 나누어 저대역에는 GSM-EFR (Global Systems for Mobile communications - Enhanced Full Rate) 부호화기를 적용하였다. 고대역 신호의 부호화에는 음성신호의 고주파 성분이 협대역 음성신호에 비해 비정재적 특성을 보다 강하게 가지는 점을 고려하여, 비정재성을 갖는 신호의 특성을 표현하는데 적합하다고 알려진 웨이브렛 분석을 적용하였다. 고대역 부호화기는 웨이브렛 변환을 이용한 부대역 부호화기를 고안하여 사용하였으며, 전체적으로 18.9 kbps의 전송속도를 갖는 광대역 음성부호화기를 구현하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 대역 분리구조를 갖는 음성부호화기에 대해 설명하고 3장에서 저대역 음성부호화에 사용한 GSM-EFR 음성부호화기에 대해 간단히 설명한다. 그리고 4장에서는 고대역 부호화에 적용한 이산 웨이브렛 변환 이론을 언급한 후, 실제 음성부호화에 어떻게 적용하였는가를 설명한다. 5장에서는 제안한 광대역 음성부호화기와 기존의 광대역 음성부호화기의 표준안인 G.722 음성부호화기의 성능을 비교하고, 마지막으로 6장에서 결론을 맺는다.

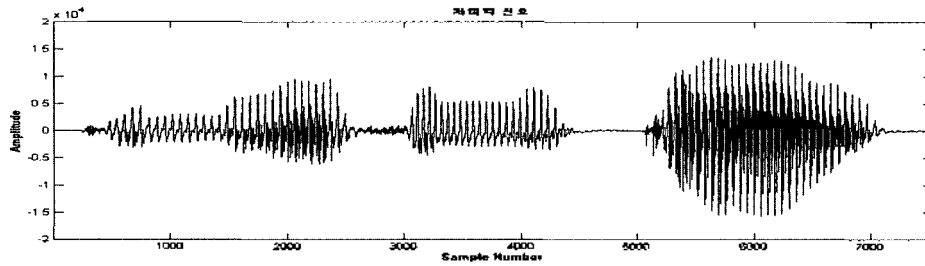
II. 대역 분리 구조의 광대역 음성부호화기

2.1. QMF를 이용한 부대역 신호 분리

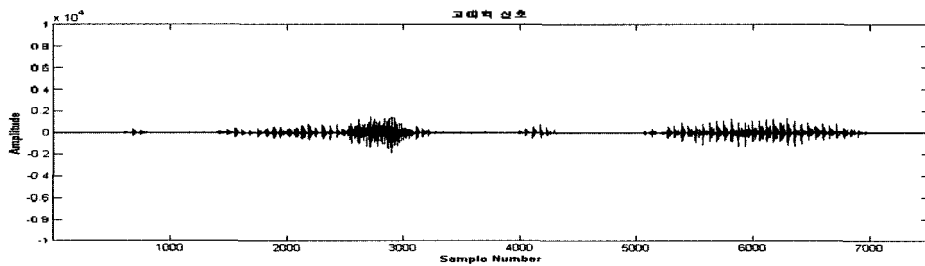
제안한 음성부호화기는 대역 분리구조를 가진다. 16 kHz로 샘플링된 입력 음성 신호는 분석 사분 대칭 필터(QMF: Quadrature Mirror Filters)를 통과해서 동일한 대역폭을 가지는 저대역 신호와 고대역 신호로 나누어지고 8 kHz의 샘플율을 갖도록 다운샘플링된다. 대역 분리에 사용한 QMF는 24-탭의 FIR 필터이며 위상의 왜곡이 없이 일정한 지연시간을 갖는 필터이다. 입력 신호의 다운샘플링으로 인해 발생하는 왜곡은 분석 필터와 같은 특성을 가지는 합성 QMF를 통과할 때 상쇄된다[5]. 24-탭의 QMF를 이용해서 분리한 저대역 신호와 고대역 신호와 원신호를 그림 2에 나타내었는데, 협대역 음성부호화에서는 고려되지 않는 고대역 신호가 상당히 큰 에너지를 가짐을 알 수 있고, 특히 무성음 부분에서는 상대적으로 에너지가 크다는 사실을 알 수 있다.



(a) 광대역 음성신호
(a) Wideband speech signal

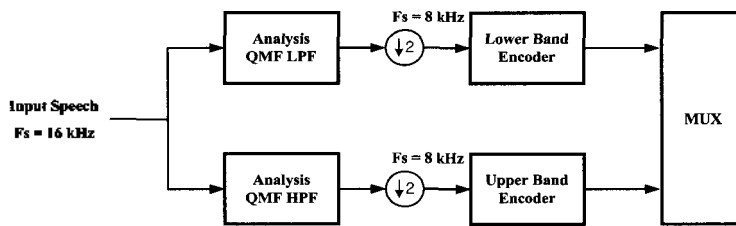


(b) 저대역 음성신호
(b) Lower subband speech signal

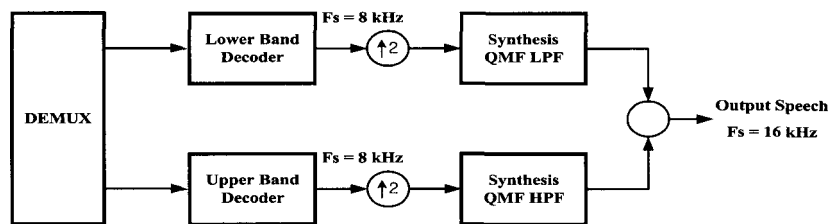


(c) 고대역 음성신호
(c) Upper subband speech signal

그림 2. 광대역 신호와 부대역 신호
Fig. 2. Wideband and subband speech signals.



(a) 대역 분리 부호화기의 구조
(a) Split-band encoder structure



(b) 대역 분리 복호화기의 구조
(b) Split-band decoder structure

그림 3. 대역 분리 부호화기 및 복호화기의 구조
Fig. 3. Split-band coder structure.

2.2. 대역 분리 음성부호화기

QMF를 통과한 두 가지 부대역 신호는 그림 3(a)와 같이 각각의 부대역 부호화기를 이용해서 부호화한다. 저대역 부호화에는 유럽의 이동통신 표준인 12.2 kbps의 전송률을 갖는 GSM-EFR 알고리즘[6]을 적용하였고, 고대역 부호화에는 이산 웨이브렛 변환을 적용한 부대역 부호화 방식을 적용하였다. 각각의 부대역 부호화기를 거친 신호는 다시 각각의 복호화기를 거쳐서 부대역 합성 신호로 변환된다. 두 가지 부대역 합성신호는 업샘플링 과정을 수행하고 합성 QMF를 통과한 후 더해져서 최종 합성음이 구해진다. 이 과정은 그림 3(b)에 나타내었다.

III. 저대역 음성부호화 알고리즘

저대역 신호의 부호화에 사용된 GSM-EFR 음성부호화 방식은 ACELP (Algebraic Code Excited Linear Prediction) 방식을 기반으로 하는 유럽 이동통신의 표준안이다. GSM-EFR 음성부호화 알고리즘에서는 네 개의 5 msec 길이의 부프레임으로 나누어지는 20 msec 길이를 갖는 프레임 단위로 동작하게 되는데, 전처리 과정을 거친 입력 음성신호는 CELP 음성부호화 모델에서 필요한 파라

표 1. 저대역 부호화기의 비트 할당
Table 1. Bit allocation of the lower subband coder.

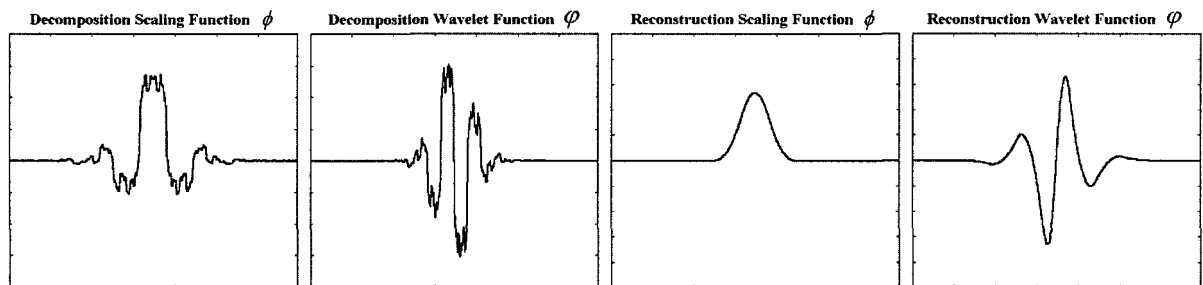
파라미터	첫 번째 세 번째 부프레임	두 번째 네 번째 부프레임	프레임 전체 비트수
2 LSP 집합	.	.	38
적용코드북 인덱스	9	6	30
적용코드북 아득	4	4	16
고정코드북 인덱스	35	35	140
고정코드북 아득	5	5	20
프레임 전체 비트수	.	.	244

미터, 즉 선형예측계수, 적용 코드북 지연과 아득, 고정 코드북 인덱스와 아득 등을 분석하고 이들을 양자화하여 전송한다. 복호화기에서는 이들 파라미터를 이용해서 합성음을 만들고 음질 향상을 위한 후처리 과정을 수행한다. GSM-EFR 음성부호화기는 12.2 kbps의 전송속도를 가지며, 프레임당 비트 할당은 표 1과 같다.

IV. 고대역 음성부호화 알고리즘

4.1. 웨이브렛 이론

웨이브렛 변환 이론[7]은 응용 수학에서 처음 소개된 이론으로 푸리에 변환과 같이 기저함수들의 집합을 이용하여 신호를 분해하는 신호해석 방법 중 하나이다. 이산 웨이브렛 변환에서는 입력신호가 저역통과 필터와 고역통과 필터를 거치고 2배의 간축 과정을 거치게 되면 한 번의 웨이브렛 변환이 수행되며, 이러한 과정을 원하는 스케일까지 반복적으로 수행하여 웨이브렛 변환된 신호를 얻을 수 있다. 그리고 일반적인 웨이브렛 변환에서 각 스케일은 상위 스케일에서 2배로 간축하여 구해지므로 각 스케일의 샘플수가 상위 스케일의 반이 된다. 신호의 주파수 해석에 널리 쓰이는 단구간 푸리에 변환이 고정된 길이를 갖는 창함수를 사용하여 시간-주파수 해상도가 고정되어 비정재성을 갖는 신호에 대해서는 신호의 특성을 표현하는데 적합하지 않다고 볼 수 있는 반면, 주파수 영역에서의 분해능을 조정할 수 있는 다중해상도 분석이 가능하다는 장점이 있다[8]. 본 논문에서는 이중직교 (biorthogonal) 웨이브렛 함수를 사용하여 QMF에 의해 분리된 고대역 신호를 두 번째 스케일까지 이산 웨이브렛 변환을 수행하였다. 분석 및 합성과정에서 사용된 웨이브렛 함수와 스케일링 함수는 그림 4와 같다.



(a) 분석 스케일링 함수 (b) 분석 웨이브렛 함수 (c) 합성 스케일링 함수 (d) 합성 웨이브렛 함수
(a) Decomposition scaling function (b) Decomposition wavelet function (c) Reconstruction scaling function (d) Reconstruction wavelet function

그림 4. 이중 직교 웨이브렛 함수와 스케일링 함수
Fig. 4. Biorthogonal wavelet and scaling functions.

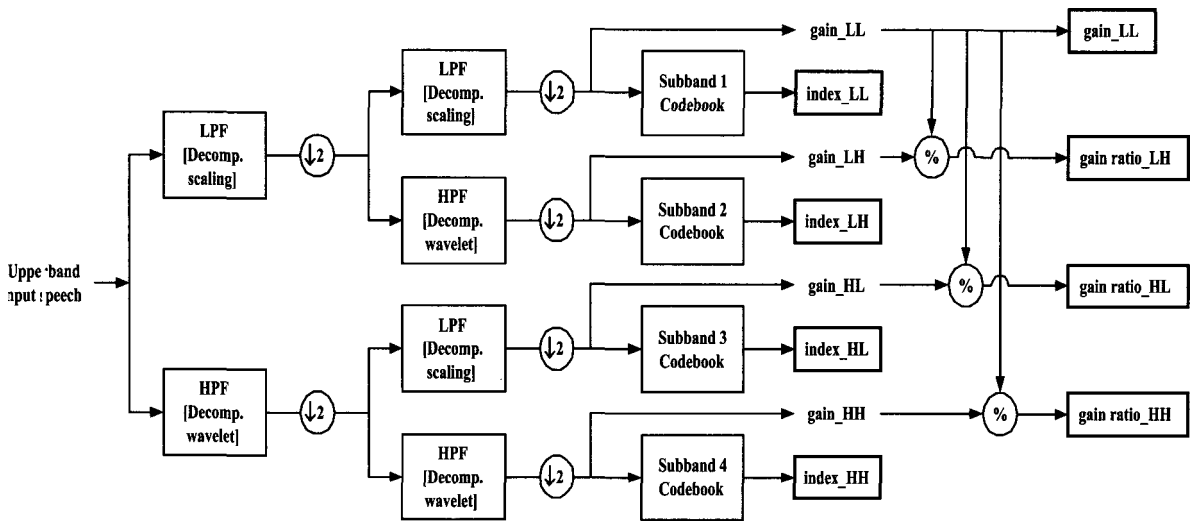


그림 5. 고대역 음성부호화기
Fig. 5. Upper subband encoder.

1.2. 전송 파라미터 양자화 기법

웨이브렛 변환을 적용한 고대역 음성부호화기의 구조는 그림 5와 같다. QMF에 의해 분리된 고대역 신호는 저대역 음성부호화기와 동기를 맞추기 위해 20 msec 길이의 프레임 단위로, 50%의 중첩 및 합산을 적용하여 처리된다. 고대역 신호는 두 번의 이산 웨이브렛 변환을 거친 후 네 개의 부대역 신호로 분리되고, 각각의 부대역(이하 LL, LH, HL, HH라 칭함) 신호에 대해서 미리 구해둔 코드북에서 가장 비슷한 형태를 갖는 코드벡터의 인덱스를 탐색한다. 그리고 웨이브렛 변환된 음성신호의 제곱평균(RMS: Root Mean Square)값을 이득으로 정의하고, 이를 구한후 양자화하게 된다.

4.2.1. 코드북 탐색 과정

두 번의 이산 웨이브렛 변환을 수행하여 분리한 네 가

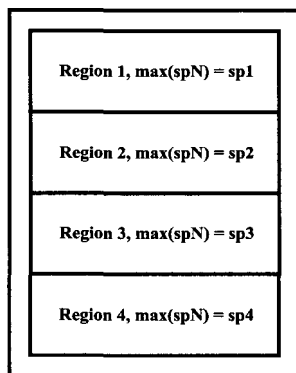


그림 6. 영역 선택 탐색법을 위한 코드북 정렬
Fig. 6. Arranged codebook structure for the "Region Selective Search" method.

지 부대역 신호를 먼저 크기가 ±1 사이의 값을 가지도록 정규화시킨다. 그리고 정규화된 부대역 신호를 미리 만들어진 코드북에서 평균 지승 오차를 최소화시키는 코드북 인덱스를 탐색한다. 실험에 사용한 코드북은 LBG (Linde-Buzo-Gray) 벡터 양자화 알고리즘[9]을 이용하여 네 가지 부대역에 대해 독립적으로 훈련시켰고, 훈련에 사용한 데이터는 16 kHz로 샘플링하고 한 샘플당 16 비트로 양자화한 뉴스 데이터 약 80000 프레임이며, 전체 뉴스 데이터 중 잡음이 섞인 리포터의 발성 부분은 제외하고 잡음이 섞이지 않은 아나운서의 실내 발성 부분만을 이용하였다. 네 가지의 부대역에 대해 코드북 사이즈에 따른 훈련 데이터의 평균 오차를 구하면서 오차의 변화 정도를 고려하여 LL, LH, HL, HH 부대역의 코드북 크기를 각각 1024, 1024, 1024, 512로 결정하였다. 훈련된 코드북은 다음에 설명될 영역 선택 탐색법에 적합하도록 다시 정렬된다. 정렬된 코드북은 그림 6과 같은 형태를 가지는데, 네 가지 서브벡터의 파워값 sp_N ($N = 1, 2, 3, 4$)을 계산하여 Region N에서 sp_N 이 최대값을 가지도록 정렬된다.

코드북 탐색에는 전체 코드 벡터를 모두 탐색하는 방법을 사용하지 않고 논문에서 제안한 영역 선택 탐색법을 이용하였다. 영역 선택 탐색법에서는 코드벡터를 서브벡터로 나누고, 각각의 파워 값을 계산하고 이를 참조하여

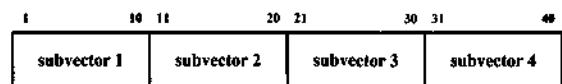


그림 7. 코드벡터의 분할
Fig. 7. Codevector division.

코드북의 일정 영역만을 탐색하는 방법이다. 전체 코드북을 모두 탐색하는 방식과 비교할 때, 제안한 방법이 코드북 탐색에 필요한 연산량이 감소하는 장점을 가진다. 영역 선택 탐색법은 다음과 같은 과정을 통해 수행된다.

- (1) 웨이브렛 변환을 통해 구한 40 샘플의 길이를 갖는 코드 벡터를 그림 7과 같이 동일한 길이를 갖는 네 가지 서브벡터로 분할한다.
- (2) 네 가지 서브벡터에서의 파워 $sp N$ 을 식 (1)~(4)와 같이 구한다.

$$sp 1 = \frac{1}{10} \sqrt{\sum_{i=1}^{10} codevector(i)^2} \quad (1)$$

$$sp 2 = \frac{1}{10} \sqrt{\sum_{i=11}^{20} codevector(i)^2} \quad (2)$$

$$sp 3 = \frac{1}{10} \sqrt{\sum_{i=21}^{30} codevector(i)^2} \quad (3)$$

$$sp 4 = \frac{1}{10} \sqrt{\sum_{i=31}^{40} codevector(i)^2} \quad (4)$$

이때, $codevector(i)$ 는 웨이브렛 변환을 통해 구한 변환계수 벡터이다.

- (3) 과정-(2)에서 계산된 서브벡터의 최대 파워값 $sp N$ 을 참조하여, 그림 6에 나타난 Region N을 탐색한다.

본 논문에서 제안하는 영역 선택적 탐색법은 코드북 탐색에 소요되는 시간을 감소시키는데 일차적인 목표를 두었으며, 전체 코드북을 탐색하는 방법과 비교했을 때 음질 저하 초래의 최소화를 고려하였다. 제안한 방법은 전체 코드북을 모두 탐색하여 최적의 코드 벡터를 결정하는 방법과 약 57%의 일치율을 나타내었으며, 두 가지 방법을 사용하여 선택한 각각의 코드 벡터를 사용해서 합성음의 음질 평가를 한 결과, 영역 선택적 탐색법이 합성음의 음질 저하를 거의 초래하지 않음을 확인하였으며 연산량 또한 이론적으로 25% 정도로 감소사킴을 확인하였다.

4.2.2. 이득 양자화 과정

이득의 양자화 과정에서는 먼저 LL 부대역의 이득을 독립적으로 양자화하고, 나머지 세 가지 부대역의 이득은 LL 부대역 이득과의 비율을 양자화하였다. 이것은 LL 부대역의 이득을 기준값으로 설정하고 나머지 부대역에서의 이득 정보를 LL 부대역의 이득과의 비율로 나타냄으로써, 양자화할 이득 정보의 동적 범위를 줄이고 양자

표 2. 고대역 부호화기의 비트 할당
Table 2. Bit allocation of the upper subband coder.

	LL 부대역	LH 부대역	HL 부대역	HH 부대역	전체
인덱스	10	10	10	9	67
이득	10	6	6	6	

화시 필요한 비트 수를 줄이기 위해서이다. LL 부대역의 이득을 양자화하는 과정에서 전체 혼련 데이터의 이득을 모두 구해보고 이득의 분포를 조사한 결과, 이득이 균일하게 분포하는 것이 아니라 특정 범위에 주로 분포한다는 점이 발견되었다. 이러한 특성을 이용하여 이득이 주로 분포하는 범위에는 작은 스텝 사이즈를 적용하고, 이득의 분포가 적은 범위에는 큰 스텝 사이즈를 적용하여 부분적으로 선형적으로 분포하는 이득 테이블을 만들고 실제 이득과 가장 가까운 이득 값을 탐색하고 그때의 인덱스를 10 비트로 양자화하였다. 나머지 세 가지 부대역에서의 이득은 LL 부대역 이득과의 비율을 식 (5)~(7)과 같이 정의하고, 이 비율을 양자화하였다.

$$LH \text{ subband gain ratio} = \frac{LH \text{ subband gain}}{LL \text{ subband gain}} \quad (5)$$

$$HL \text{ subband gain ratio} = \frac{HL \text{ subband gain}}{LL \text{ subband gain}} \quad (6)$$

$$HH \text{ subband gain ratio} = \frac{HH \text{ subband gain}}{LL \text{ subband gain}} \quad (7)$$

세 가지 부대역에서의 이득비의 양자화 과정에서도 코드북 혼련 데이터의 이득 비율값을 조사하여 그 결과를 고려해서, LL 부대역의 이득 테이블을 작성한 방법과 동일하게 이득비 값의 분포가 많은 범위의 값은 작은 스텝 사이즈를 적용하고 분포가 적은 범위의 값은 큰 스텝 사이즈를 적용한 이득비 테이블을 작성한 후 이를 탐색하는 방법으로 양자화하였다. 이 세 가지 부대역의 이득 비율을 양자화 하는데는 LL 부대역의 이득이 분포하는 범위에 비해 가변 범위가 상대적으로 작음을 고려하여 6 비트를 할당한 이득비 테이블을 이용하여 양자화하였다.

이산 웨이브렛 변환을 적용한 고대역 음성부호화 알고리즘에서의 프레임당 비트 할당은 네 가지 부대역에서의 코드북 인덱스와 이득 및 이득비 값을 양자화하기 위해 표 2와 같이 결정하였다. 따라서 고대역 신호에는 6.7 kbps가 할당되며 저대역 부호화기의 전송률과 함께 고려할 경우, 제안한 광대역 음성부호화기는 전체 18.9 kbps의 전송률을 갖는다.

V. 실험 및 결과

본 논문에서 제안한 음성부호화 알고리즘의 성능을 평가하기 위해서 ITU-T (International Telecommunications Union - Telecommunications standardization sector)의 권고안인 G.722 56 kbps 음성부호화기와 비교를 수행하였다. 먼저 전송률을 비교하면, 제안한 음성부호화기는 18.9 kbps의 전송률로써, 비교대상인 G.722 음성부호화기에 비해 절반 이하의 전송률을 갖는다. 그리고 두 가지 음성부호화기의 합성음의 음질을 평가하기 위해 G.722 56 kbps 음성부호화기의 합성음과 제안한 음성부호화기의 합성음 사이의 파형 및 스펙트럼 비교, 그리고 평균 청취 스코어 (MOS: Mean Opinion Score) 테스트 및 ITU-R 7 등급 단계 비교를 수행하였다. 실험에 사용한 음성은 16 kHz로 샘플링되고 16 bits/sample로 양자화된 남자 음성 세 문장 (두개의 국어문장과 하나의 영어문장)과 여자 음성 두 문장 (두개의 국어문장)이며, 표 3에 나타내었다.

먼저 10 여명의 비전문가를 대상으로 제안한 음성부호화기를 이용한 합성음과 G.722 56 kbps 음성부호화기의

합성음을 헤드폰으로 들려주고 두 가지 방식의 합성음 중에서 어느 것이 나은가를 표 4의 등급 비교를 이용해서 조사하였다. 이 방식은 먼저 실험 문장의 원음성을 청취자에게 들려주고, 비교 대상인 두 가지 음성부호화기의 합성음을 각각 들려준 후 단순하게 두 가지 합성음 중에서 무엇이 더 나은가를 일곱 가지 등급을 참고하여 평가하는 실험인데 그 결과는 그림 8과 같다.

그림 8에 나타나듯이 두 가지 음성부호화기의 합성음이 특별히 어느 하나가 다른 하나보다 월등히 나은 성능을 나타내지는 않았다. 청취 대상이 대부분 두 가지 음성부호화기를 비슷한 성능을 가진다고 평가한 것이다. 단순한 합성음의 음질 우위를 비교하는 이러한 결과로부터 제안한 음성부호화기가 G.722 56 kbps 음성부호화기와 거의 비슷한 성능을 가짐을 알 수 있다. 또한 동일한 실험 문장과 추가한 실험 문장을 동일한 청취자를 대상으로 MOS 테스트를 한 결과, 제안한 음성부호화기를 이용한 합성음은 전체 실험 문장에 대해서 모두 "GOOD" 이상의 스코어를 얻었으며, 표 5와 같이 평균적으로 4.5 이상의 높은 점수를 얻었다. 비록 비전문가를 대상으로 한 평가

표 3. 청취실험에 사용한 음성데이터
Table 3. Speech data for the listening test.

문장번호	내 용
남성화자	1 안녕하십니까 ?
	2 오늘 병원과 의원 상당수가 문을 닫아서 환자들이 헛걸음을 한 데가 많았습니다.
	3 A blockbuster movie in south Korea is breaking all box office records.
여성화자	4 IMF의 과거와 현재를 보셨는데, 이제는 ...
	5 전혀 나무가 뒤떨림이 없네요.

표 4. ITU-R 7 등급 비교 단계
Table 4. ITU-R 7 grading table.

Observation	Grading
A much better than B	3
A better than B	2
A slightly better than B	1
A same as B	0
A slightly worse than B	-1
A worse than B	-2
A much worse than B	-3

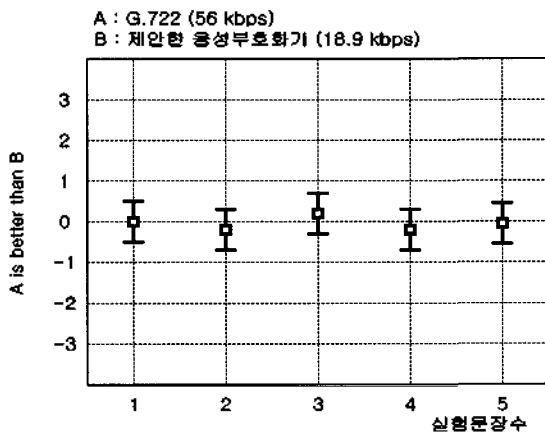


그림 8. 제안한 음성부호화기와 G.722 56 kbps 음성부호화기의 성능 비교
Fig. 8. Performance comparison between proposed and G.722 56 kbits/sec coder.

표 5. 평균 청취 스코어 테스트 결과
Table 5. Results of the MOS test.

	문장1	문장2	문장3	문장4	문장5	평균
G.722 56 kbps	4.78	4.67	4.78	4.67	4.67	4.71
제안한 음성부호화기 18.9 kbps	4.78	4.56	4.89	4.56	4.67	4.69

표 6. 로그 영역비 비교 결과
Table 6. Comparison results of the log area ratio.

문장번호	합성음의 결과값	비교값		
		저대역신호	G.722 56 kbps	
남성화자	1	2.172	2.356	1.757
	2	2.669		
	3	2.376		
여성화자	4	2.412		
	5	1.861		

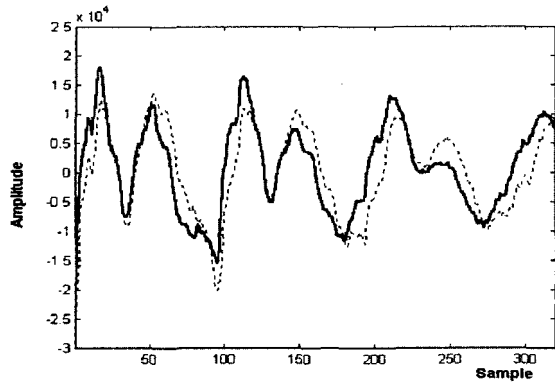


그림 9. 원음성, 합성음의 파형 비교 (원음성: 점선, 합성음: 실선)
 Fig. 9. Waveforms of the original synthesized speech (original: dash, synthesized: solid).

이지만 이상의 결과로부터 제안한 음성부호화기가 기존의 권고안보다 매우 낮은 전송률을 가지면서 거의 같은 수준의 음질을 유지함을 확인할 수 있다.

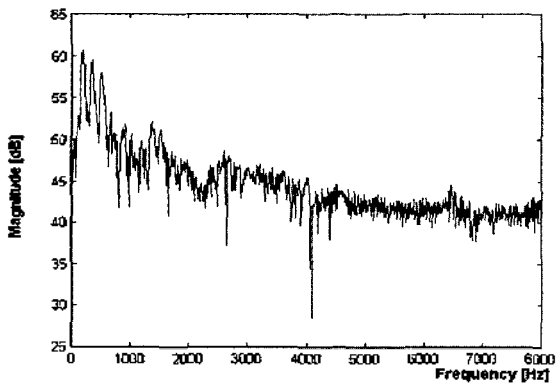
또한 제안한 음성부호화기의 객관적인 음질 평가[10]를 위해 로그 영역비 (Log-Area-Ratio)를 실험음성에 대해 구하고, 그 평균값을 표 6에 정리하였다. 로그 영역비는 합성음과 원음성의 선형예측 계수 차이에 기반하여 구해지는 값으로 원음성과 합성음의 주파수 특성이 얼마나 다른가를 반영하는 것으로 생각할 수 있으며, 스펙트럼 특성 차이가 작을수록 그 값이 작게 나타난다. 표 6의 비교값은 저대역 신호만을 GSM-EFR 방식으로 부호화한 합성음과 광대역 신호를 G.722 56 kbps 방식으로 부호화한 합성음에서 측정된 로그 영역비 값을 나타낸다. 이들 비교값으로부터 제안하는 음성부호화기가 원음성의 스펙트럼 특성을 충실히 반영하여 합성음을 복원하는 것으로 판단할 수 있다.

그림 9와 그림 10은 원음성과 합성음의 파형과 스펙트럼을 비교한 것이다. 이때 실험 문장 중에서 유성음의 구간 일부를 잘라서 비교하였다. 파형을 비교해 볼 때, 논문에서 제안한 음성부호화기는 저대역 부호화에 사용한 GSM-EFR이 파형부호화 방식이 아니어서 동일한 파형은 얻을 수 없지만 전체적인 형태는 거의 비슷하게 유지함을 알 수 있다. 스펙트럼을 비교해 보면 원음성과 제안한 음성부호화기의 합성음이 비슷한 형태의 스펙트럼 포락선을 가짐을 알 수 있다.

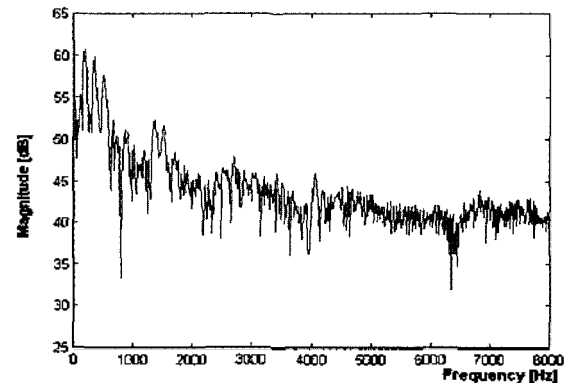
이상의 파형 및 스펙트럼 비교, MOS 테스트, 객관적인 수치 비교 등의 성능 평가에서 본 논문에서 제안한 음성부호화기는 합성음의 음질면에서 기존의 권고안보다 크게 떨어지지 않는 성능을 가지고, 전송속도 면에서는 크게 개선된 음성부호화기임을 확인할 수 있다.

VI. 결론

본 논문에서는 기존의 협대역 음성부호화 알고리즘에 이산 웨이브렛 변환 기법을 추가 적용한 새로운 광대역 음성부호화기를 제안하였다. 제안한 음성부호화기는 QMF를 이용한 부대역 분리 구조를 가지며, 저대역 부호화에는 유럽의 이동 통신 표준안인 GSM-EFR 음성부호화기를, 고대역 부호화에는 이산 웨이브렛 변환을 적용한 음성부호화 알고리즘을 제안하여 각각 사용하였다. 제안한 음성부호화기의 전송속도는 18.9 kbps로써 협대역 음성부호화기에 비해 크게 증가하지 않은 전송속도를 가지며, 합성음의 음질은 기존의 광대역 음성부호화 권고안의 합성음과 거의 비슷한 음질을 유지하였다. 구체



(a) 원음성의 스펙트럼
 (a) Original speech spectrum



(b) 제안한 음성부호화기 합성음의 스펙트럼
 (b) Synthetic speech spectrum of proposed coder

그림 10. 원음성, 합성음의 스펙트럼 비교
 Fig. 10. Spectrums of the original and synthesized speech.

적인 음질 평가 결과 56 kbps의 전송속도를 갖는 G.722 음성부호화기와 거의 비슷한 음질을 유지하였다. 또한 제한된 광대역 음성부호화기는 기존의 협대역 이동 통신 표준인 방식을 내포하기 때문에 적절한 호환성 유지 처리만 한다면 일반 이동 통신에의 응용도 가능하리라 생각된다.

앞으로는 고대역 부호화에 사용된 네 가지 부대역에 대한 코드북 훈련과정을 반복하면서 최적의 코드벡터 및 사이즈를 결정하고, 양자화 기법의 연구 등을 통해 전송률을 낮출 수 있는 방법에 대해 연구할 예정이다. 그리고 심리 음향 모델을 적용한 적절한 전/후처리 과정을 도입하여 합성음의 음질을 보다 개선할 계획이다.

참고 문헌

1. K. Maitre, "7 kHz Audio Coding within 64 kbits/s," *IEEE Journal on Selected Areas in Commun.*, 283-298, Feb. 1988.
2. P. Mermelstein, "G.722, a new ccitt coding standard for digital transmission of wideband audio signals," *IEEE Commun. Mag.*, 8-15, Jan. 1988.
3. SD/JTC draft : Information Technology-Coding of Audiovisual Objects, Part3 Audio, Subpart3 CELP.
4. ETSI Draft Tdoc S4/SMG11 (00)00167, AMR Wideband Development Overview (WB-1).
5. K. Sayood, "Introduction to Data Compression," 2nd edition, Morgan Kaufmann.
6. ETSI Draft ETSI 300 726, Enhanced Full Rate (EFR) Speech

Transcoding.

7. O. Rioul and M. Vetterli, "Wavelet and signal processing," *IEEE Signal Processing Mag.*, 14-38, Oct. 1991.
8. 한미경, 배건성, "웨이브렛 변환을 이용한 음성신호의 잡음 제거", *한국음향학회지*, 19 (5), 27-34, 2000.
9. Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantization design," *IEEE Trans. on Commun.*, COM-28, 84-95, Jan. 1980.
10. S. R. Quackenbush, T. P. Barnwell, M. A. Clements, *Objective Measures of Speech Quality*, Prentice-Hall, NJ, 1988.

저자 약력

● 이 승 원 (Seung Won Lee)



1999년 2월: 경북대학교 공과대학 전자·전기공학부 (공학사)
 2001년 2월: 경북대학교 대학원 전자공학과 (공학석사)
 2001년 3월~현재: 경북대학교 대학원 전자공학과 (박사과정)
 ※ 주관심분야: DSP, 음성코딩, VoIP

● 배 건 성 (Keun Sung Bae)



1977년 2월: 서울대학교 전자공학과 (공학사)
 1979년 2월: 한국과학기술원 전기 및 전자공학과 (공학석사)
 1989년 5월: University of Florida (공학박사)
 1979년 3월~현재: 경북대학교 전자·전기공학부 교수
 ※ 주관심분야: 음성분석 및 인식, 디지털 신호처리, 디지털 통신, 음성코딩, 웨이브렛이론 등