2000年度 第13回 信號處理合同學術大會 論文集 第13卷 1號

High-Band 신호에 웨이브렛 변환을 적용한 광대역 GSM-EFR 음성부호화 알고리즘 개발

이 승 원 , 배 건 성
경북대학교 전자·전기 공학부

Development of Wideband GSM-EFR Speech Coding Algorithm with Application of Wavelet Transform to High-Band Signal

Seung Won Lee , Keun Sung Bae
School of the Electronic & Electrical Engineering, Kyungpook National University

leesw@mmirl1.knu.ac.kr , ksbae@ee.knu.ac.kr

요 약

본 논문에서는 웨이브렛 변환을 적용한 광대역 음성부호화 알고리즘을 제안하였다. 제안한 음성부호화 알고리즘은 split-band 구조를 가지며, 16 kHz로 sampling된 입력신호를 QMF를 이용해서 동일한 대역폭을 갖는 두 개의 subband 신호로 나누고 이를 8 kHz의 sampling율을 갖도록 downsampling 한다. 그리고 저대역 신호는 GSM-EFR 음성부호화 알고리즘을 이용하여 부호화하고, 고대역 신호는 DWT(Discrete Wavelet Transform)을 적용하여 subband로 나누어 부호화하였다. 각 subband에서 양자화 된 파라미터는 IDWT(Inverse DWT)과정을 거쳐서 upsampling되고 합성 QMF를 통과시켜 최종 합성음을 구하였다. 제안한 음성부호화기는 저대역 신호의 GSM-EFR 부호화에 12.2 kbps, 웨이브렛 변환을 이용한 고대역 신호의 부호화에 7.8 kbps로 전체 20 kbps의 전송율을 가지면서 G.722 표준안의 56 kbps에서의 합성음과 비슷한 음질을 나타내었다.

1. 서 론

0 ~ 7 kHz의 대역폭을 갖는 광대역(Wideband) 음성부호화기는 300 ~ 3400 Hz의 대역폭을 이용하는 기존 전화망(PSTN)의 협대역(Narrowband) 음성부호화기와 비교할 때, 증가한 저주파성분이 음질의 자연스러움을 개선하고, 증가한 고주파성분이 마찰음(fricative)의 구분을 명확히 하여 전체적으로 음성의 명료도가 향상된다. 광대역 부호화기의 표준안으로는 1986년에 ITU-T에 의해 제정된 G.722[1,2]가 있지만 이 방식은 48 ~ 64 kbps의 비교적 높은 전송율을 가져서 널리 쓰이기에는 불리한 조건이 있다. 이러한 점을 보완하기 위해 최근까지 원음성과 거의 같은 음질을 유지하면서 낮은 전송율을 가지는 음성부호화 알고리즘을 개발하려는 연구가 계속 되고 있다[3,4]. 본 논문에서는 split-band 구조를 가지며 고대역 신호에 DWT를 적용한 광대역 음성부호화기를 제안하였다.

논문의 구성은 다음과 같다. 먼저 제안한 음성부호화 알고리즘을 2장에서 설명하고, 분석 및 합성과정에서의 실험 결과를 바탕으로 적절한 전송율을 구하는 과정을 3장에서 설명한다. 4장에서는 제안한 음성부호화기의 성능을 평가하기 위해 G.722 음성부호화기의 합성음과의 음질 비교를 수행한 결과를 설명하고, 마지막으로 5장에서 결론 및 앞으로의 연구 계획을 언급한다.

2. 광대역 음성부호화 알고리즘

제안한 음성부호화기는 그림 1과 같이 split-band 구조를 가진다. 16 kHz로 sampling된 입력 음성신호는 24-tap QMF를 통과해서 동일한 대역폭을 갖는 두 개의 subband 신호로 나누어진다. 그리고 downsampling을 통해 sampling율이 8 kHz로 된다. 이러한 두 개의 subband 신호는 각각의 subband 부호화기를 거친 후, 다시 upsampling을 거치고 분석단의 QMF와 동일한 QMF를 통과해서 합성음이 구해진다. 저대역 부호화에는 유럽의 이동통신 표준인 12.2 kbps의 전송율을 갖는 GSM-EFR 알고리즘[5]을 적용하였고, 고대역 부호화에는 DWT를 이용한 subband 부호화 방식[6,7]을 적용하였다.

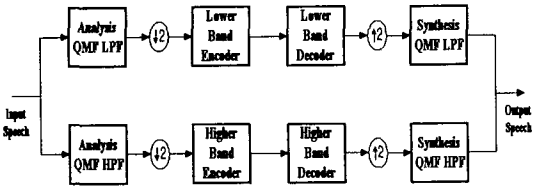


그림 1. 광대역 음성부호화기 구조

2.1 저대역 음성부호화 알고리즘

저대역 음성부호화기는 12.2 kbps의 전송율을 갖는 GSM-EFR 음성부호화기를 사용하였다. GSM-EFR 음성부호화기는 ACELP 음성부호화기에 기반을 둔 방식으로써, 20 ms의 길이를 갖는 프레임 단위로 동작하며, 한 프레임은 5 ms의 길이를 갖는 네 개의 subframe으로 나누어진다. 입력 음성신호로부터 CELP 음성부호화기의 합성모델에 필요한 파라미터들을 분석하고 부호화한다. 이때의 파라미터들은 LP(Linear Prediction) filter 계수(20 ms 프레임당 두 번)와 adaptive codebook delay와 gain, fixed codebook index와 gain(5 ms subframe 마다)들이다. GSM-EFR 음성부호화 알고리즘에서 프레임당 bit 할당은 표 1과 같다[5].

표 1. 저대역 음성부호화기의 bit 할당

Parameter	subframe (1st & 3rd)	subframe (2nd & 4th)	Total per frame
2 LSP sets			38
ACB index	9	6	30
ACB gain	4	4	16
FCB pulses	35	35	140
FCB gain	5	5	20
Total			244

디코더에서는 전송된 파라미터들을 이용하여 합성음을 만들고 이를 후처리하여 최종 합성음을 만든다.

2.2 고대역 음성부호화 알고리즘

고대역 부호화에 사용된 부호화기의 구조는 그림 2와 같다.

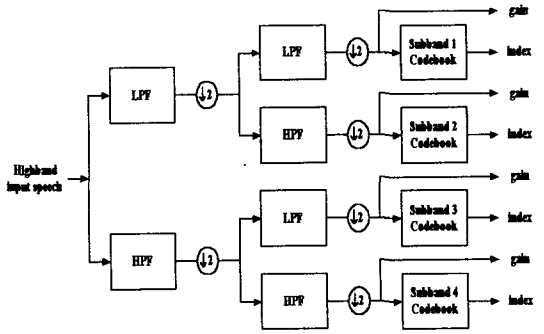


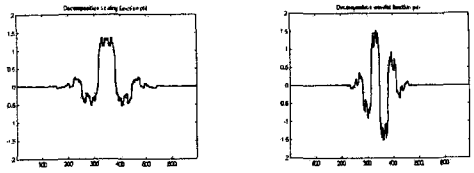
그림 2. 고대역 음성부호화기의 구조

QMF에 의해 분리된 고대역 신호는 저대역 음성부호화기와 동기를 맞추기 위해 20 ms길이의 프레임 단위로 50%의 overlap-and-add 방식으로 두 번의 DWT를 거쳐서 네 개의 subband로 분리된다. 이 과정을 거쳐서 구한 wavelet 변환계수들은 평균 전력을 이용하여 이득을 구하고, 일반적인 LBG 알고리즘[8]을 적용하여 미리 만들어둔 codebook과의 MSE (Mean Square Error)를 구하여 최적의 index를 찾는다. 이때 각 subband에 할당된 codebook의 size는 실험에 의해 lowlow, lowhigh, highlow, highhigh subband(이하 LL, LH, HL, HH라 칭한다)에 대해 각각 1024, 1024, 1024, 512로 결정하였다. 디코더에서는 전송된 codebook index와 gain 정보를 이용해서 네 개의 subband에서의 합성음을 구한 후 이들을 이용하여 최종 합성음을 구한다. DWT를 적용한 고대역 음성부호화 알고리즘에서의 프레임당 bit 할당은 표 2와 같다. 따라서, 고대역 신호에 7.8 kbps가 할당되어 전체적으로 20 kbps의 전송율을 갖는다.

표 2. 고대역 음성부호화기의 bit 할당

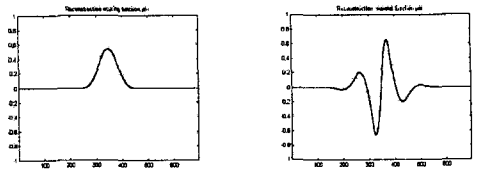
subband / parameter	LL	LH	HL	HH	Total
gain	10	10	10	9	78
index	10	10	10	9	

고대역 부호화의 DWT 및 IDWT에 사용된 wavelet은 biorthogonal wavelet이며, wavelet function 및 scaling function은 그림 3 및 그림 4와 같다[6,7].



(a) Decomposition scaling function　　(b) Decomposition wavelet function

그림 3. 분석과정에서의 함수



(a) Reconstruction scaling function　　(b) Reconstruction wavelet function

그림 4. 합성과정에서의 함수

3. 고대역 전송파라미터의 양자화 과정

본 연구에서 적용한 고대역 음성부호화 알고리즘은 프레임 단위의 음성신호를 biorthogonal wavelet을 이용해서 두 번의 DWT를 수행하고, 여기서 구한 wavelet 변환계수를 codebook으로 만들어서 codebook searching을 수행하는 것이다. Codebook 훈련과정에서는 16 kHz로 sampling된 뉴스 방송의 데이터 중 잡음이 없는 아나운서의 음성부분만을 사용하였다. 그리고 네 개의 subband마다 splitting vector quantization을 적용해서 codebook을 훈련시켰고, codebook size에 따른 훈련 데이터의 평균 오차를 구하면서 오차의 변화 정도를 고려하여 LL, LH, HL, HH subband의 codebook의 size를 각각 1024, 1024, 1024, 512로 결정하였다. 또한 이득을 양자화 하는 과정에서는 전체 훈련 데이터의 이득을 모두 구해보고 이득의 분포를 조사한 결과, 이득이 균일하게 분포하는 것이 아니라 특정 범위에 주로 분포한다는 점을 발견하여서 이 범위에는 작은 step size를 적용하고, 이득이 거의 분포하지 않는 범위에는 큰 step size를 적용하는 log-양자화 기법을 적용하였다. 네 개의 subband에서 이득의 양자화에 할당한 bit 수는 각각 10, 10, 10, 9 bit이다.

4. 실험 및 결과

제안한 음성부호화 알고리즘의 성능을 평가하기 위해서 ITU-T의 표준안인 G.722의 합성음과 파형 및 스펙트럼 비교, 그리고 MOS test 및 ITU-R 7의 grading point 비교(표 3의 분류방법)를 수행하였다. 실험에 사용한 음성은 16 kHz로 sampling되고 16 bit/sample로 양자화된 남자 음성 3문장(2개의 국어문장과 하나의 영어 문장)과 여자 음성 1문장이다.

표 3. ITU-R 7 point comparative scale

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

먼저 6명의 비전문가를 대상으로 제안한 음성부호화기를 이용한 합성음과 G.722 56 kbps 음성부호화기의 합성음을 헤드폰으로 들려주고 둘 중에서 어느 것이 나은가를 조사하여 그림 5와 같은 결과를 얻었다. 그림에서 나타나듯이 단순한 합성음의 우위를 비교하는 결과에서는 제안한 음성부호화기가 G.722 56 kbps 음성부호화기와 거의 비슷한 음질을 나타내었다.

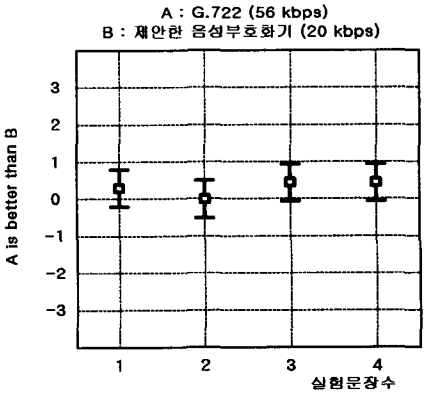


그림 5. 제안한 음성부호화기와 G.722 56 kbps 합성음의 음질 평가 비교결과

그리고 동일한 실험문장과 원음성을 이용한 MOS 테스트 결과는 전체 실험문장에 대해서 모두 "GOOD" 이상의 스코어를 얻었으며, 표 5와 같이 평균적으로 4.5 이상의 높은 점수를 얻었다.

표 4. MOS 테스트 결과

	문장1	문장2	문장3	문장4	평균
G.722 56 kbps	4.83	4.66	4.83	4.83	4.79
제안한 음성부호화기 20 kbps	4.66	4.66	4.50	4.50	4.58

그림 6과 그림 7은 원음성과 실험음성의 파형과 스펙트럼을 비교한 것이다. 파형을 비교해 볼 때, 저대역 부호화에 사용한 GSM-EFR이 파형부호화 방식이 아니므로 동일한 파형은 얻을 수 없지만 전체적으로 비슷한 형태를 보임을 알 수 있다. 그리고 스펙트럼을 비교해 보면 고대역에서의 스펙트럼 포락선의 형태가 원음성의 것과 비슷함을 알 수 있다.

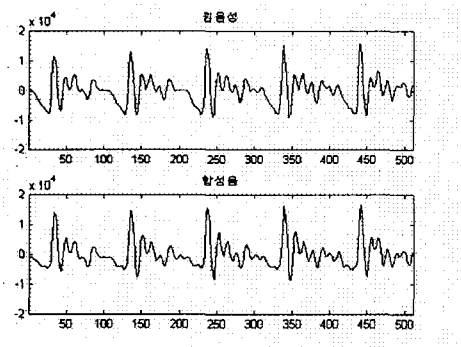


그림 6. 원음성과 합성음의 파형 비교

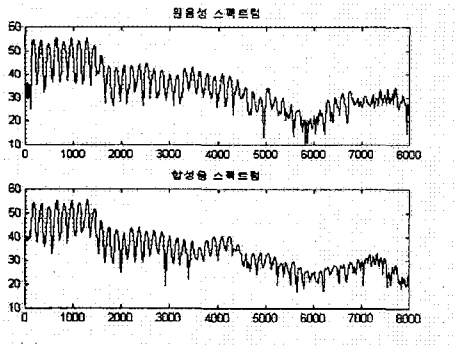


그림 7. 원음성과 합성음의 스펙트럼 비교

5. 결 론

본 논문에서는 기존의 음성부호화 알고리즘에 대해 wavelet 변환 기법을 추가 적용한 새로운 광대역 음성부호화기를 제안하였다. 제안한 음성부호화기는 기존의 이동통신 표준안인 GSM-EFR 방식을 내포하고 있어서 협대역 음성부호화기와의 호환성도 갖출 수 있어서 일반 이동통신에의 응용도 가능하리라 생각된다. 제안한 음성부호화기의 성능평가 결과, G.722 방식의 56 kbps 모드에서의 합성음과 비교할 때 낮은 전송율로 비슷한 음질을 얻을 수 있음을 확인하였다. 앞으로는 각 subband에 대한 codebook 훈련과정을 반복하면서 적절한 codebook size의 결정과, 양자화 기법의 연구 등을 통해 전송율을 낮추면서 음질을 개선할 수 있도록 연구할 계획이다.

참 고 문 헌

[1] Xavier 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.*, pp. 8-15, Jan. 1988

[3] G. Roy and P. Kabal, "Wideband CELP speech coding at 16 kbits/sec", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 17-20, 1991

[4] A. Ubale and A. Gersho, "A Multi-band CELP Wideband Speech Coder", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 1367-1370, 1997

[5] ETSI Draft ETSI 300 726, Enhanced Full Rate EFR) Speech Transcoding

[6] O. Rioul and M. Vetterli, "Wavelet and Signal Processing", *IEEE Signal Processing Magazine*, pp. 14-38, 1991.

[7] M. Vetterli, "Multi-dimensional Sub-band Coding," *IEEE Signal Processing, Vol. 6, No. 2*, pp. 97-112, 1984.

[8] Y. Linde, A. Buzo and R. M. Gray, "An Algorithm for Vector Quantization Design", *IEEE Trans. on Commun., vol. COM-28*, pp. 84-95, Jan. 1980