AMR 부호화기와 결합된 다전송률 광대역 음성부호화기 설계

김은주* 이호창* 이인성*
충북대학교 전파공학과

Design of Multi Rate Wideband Speech Coder Using the AMR(Adaptive Multi-Rate) Coder

Eunju Kim* Hochang Lee* Insung Lee*
Dept. of Radio Engineering, Chungbuk National Univ.

요 약 문

본 논문에서는 AMR(Adaptive Multi-Rate)를 이용하여 광대역 음성부호화기를 설계하였다. 16kHz로 샘플링 된 입력 신호를 QMF 필터에 의해 두 개의 대역으로 나누어, 각각 decimation하여 두 개의 8kHz 샘플링 신호로 변환시킨 후 저대역(0Hz-3400Hz)의 신호와 고대역(3400Hz -7000Hz)의 신호로 나누어 각각 부호화한다. 나누어진 두 개의 협대역 음성신호는 AMR (Adaptive Multi-Rate)과 ATC(Adaptive Transform Coding)을 사용하여 각각 부호화되어 전송된다. 두 대역으로부터 부호화된 정보는 20.2kbps에서 12.75kbps까지의 전송률을 갖고, 수신단에서는 각 대역을 AMR과 ATC방법으로 역부호화하여 음성신호를 합성한다. 설계된 광대역 음성부호화기의 성능을 평가하기 위해 ITU-T의 표준안인 G.722를 포함하여 MOS 시험을 하였다.

Ⅰ. 서론

2001년부터 새롭게 선보일 예정인 IMT-2000 이동통신 서비스의 가장 큰 특징은 넓은 대역폭(2Mbps)을 사용하며 음성에서 동영상까지 다양한 멀티미디어 데이터를 전송할 수 있다는 것이다. 기존의 음성 위주의 이동통신 서비스에서 데이터, 영상데이터를 포함한 멀티미디어 통신이 가능하게 된다. 음성통신에 있어서도 기존의 협대역 음성 신호 뿐만 아니라 광대역 음성 통신도 가능하여진다. 또한 최근에 활발하게 보급되고 있는 고속 인터넷망에서는 전송 대역폭의 확대로 음질이 좋은 광대역 음성과 오디오 전송도 실시간 전송이 가능하게 되었다. 광대역 음성신호는 협대역(200Hz-3400Hz) 음성신호보다 넓은 대역폭(50Hz- 7000Hz)을 가지며 16 kHz 샘플링 주기를 요구하며 더욱 자연스러운 음질과 높은 인지도를 나타낸다. 이처럼 보다 자연스러운 음질을 멀티미디어 통신에서 느낄 수 있어야 한다. 따라서 본 논문에서는 IMT-2000 서비스의 음성부호화기의 표준으로 채택된 AMR부호화기를 저대역의 음성부호화기로 포함하여 광대역 음성부호화기를 설계한다. AMR 부호화기는 8개의 전송모드로 채널 환경에 적응적으로 대처하면서 toll quality를 갖는 부호화 방식이다. 본 논문에서 설계될 광대역 부호화 알고리즘은 32kbps이하의 전송률에서 동작되도록 설계하였으며, 화상회의 시스템, 고품질 디지털 AM 라디오 방송, 고품질 전화 등에 사용가능하다. 또한 다전송률 부호화 방식으로 이동통신 시스템에서 traffic높아질 때 전송률에 대해 유연성을 가질 수 있도록 설계되어진다. 저대역 부호화기에 AMR부호화기를 포함으로써 IMT-2000시스템과 호환성을 유지하며 그 응용분야는 더욱 확대되리라 예측되어진다.

본 논문에서는 Ⅱ장에서는 광대역 음성부호화기의 인코더 알고리즘 Ⅲ장에서는 디코더 알고리즘에 대해서 설명을 한다. Ⅳ장에서는 논문에 대한 시뮬레이션 결과 및 주관적 음질 평가결과를 다룬다.

Ⅱ. 광대역 음성부호화기의 인코더 알고리즘

본 연구에서 설계된 광대역 음성부호화기의 인코딩 과정이 그림1에 나타나 있다. 16kHz로 샘플링 된 입력 신호를 QMF 필터에 의해 두 개의 대역으로 나누어지고, 각각 decimation하여 두 개의 8kHz 샘플링 신호로 변환시킨 후 저대역(0Hz-3400Hz)의 신호와 고대역(3400Hz-7000Hz)의 신호로 나누어 각각 부호화되어 전송된다[4][5][6]. 저대역의 신호는 ETSI와 3GPP에서 차세대 이동통신 IMT-2000 서비스의 음성부호화기의 표준안으로 채택한 AMR를 이용하여 4.75kbps에서 12.2kbps까지 8개의 전송률로 부호화되어 전송된다. 고대역의 신호는 ATC(Adaptive Transform Coding)방법으로 8kbps의 전송률로 부호화되고, 변환은 DCT (Discrete Cosine Transform)알고리즘을 이용한다[7]. 저대역과 고대역의 부호화된 정보는 전체적으로 20.2kbps에서 12.75kbps까지의 전송률을 갖는다.

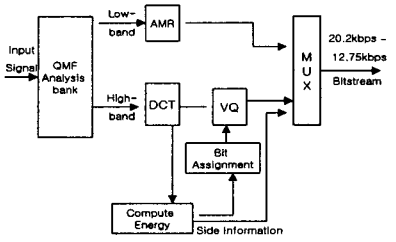


그림 1.광대역 음성부호화기의 인코더

2.1 저대역 인코더 알고리즘

QMF필터에 의해 분리된 0Hz -3400Hz 대역의 음성신호는 AMR 인코더로 부호화된다. ARM은 ACELP에 기반을 두고 있다. 또한 전송률이 여러개를 가짐으로써

MR-ACELP라고 하기도 한다. 기본 목적은 음성부호화의 무선 채널의 환경변화에 적응적으로 대처함에 있다. 또한 명료한 음질을 가진다는 장점이 있다. 이에 ETSI에서는 전송률을 4.75kbit/s에서 12.2kbit/s까지 8개로 나누고 하나의 저전송률 주변잡음 인코딩 모드를 갖는 하나의 통합된 음성부호화기를 구성한다. 프레임 사이즈는 8kHz로 샘플링된 5ms의 4개의 부프레임을 구성된 20ms크기를 갖는다. 모드에 따른 사용되는 전송률은 표1에 주어진다.

Codec-mode	Source-codec-bit-rate
AMR_12.20	12.20 kbit/s (GSM EFR)
AMR_10.20	10.20 kbit/s
AMR_7.95	7.95 kbit/s
AMR_7.40	7.40 kbit/s (IS-641)
AMR_6.70	6.70 kbit/s (PDC-EFR)
AMR_5.90	5.90 kbit/s
AMR_5.15	5.15 kbit/s
AMR_4.75	4.75 kbit/s
AMR_SID	1.80 kbit/s *

표.1 AMR코덱의 모드별 비트율
GSM-EFR : ETSI GSM06.90 Enhanced Full Rate
IS-641: TIA/EIA IS-641 TDMA Enhanced Full Rate
PDC_EFR : ARIB 6.7kbit/s Enhanced Full Rate
SID : 프레임이 연속적일 경우

AMR 인코더의 구조는 그림2에 나타나 있다. 처음에 모든 모드에 대해서 high-pass filtering과 down scaling을 하여 전처리 과정을 한다. LP 분석과 양자화 과정은 12.2kbit/s mode에서는 두 개의 LP 필터를 가지고 30ms 비대칭 윈도우를 이용해서 두 번 실행된다. 이때 lookahead는 사용되지 않고, Levinson-Durbin 알고리즘을 이용해 계수를 구한 후 양자화 한다. 나머지 모드에서는 비대칭 윈도우를 5ms lookahead를 사용한다. 계산된 LP filter 계수는 양자화와 보간을 위하여 LSP(Line Spectrum Pair)로 바뀐다. Algebraic 코드북 검색은 interleaved single-pulse permutation design (ISPP)에 기본을 두고 있다. 가중 필터를 통과한 음성신호와 가중필터와 합성필터를 통과한 음성신호 사이의 mean square error를 최소화 시키는 값을 갖는 코드북 인덱스를 검색한다. 모든 모드에 대하여 적응 코드북 이득을 위해 직접 양자화를 실행하고 algebraic 코드북 이득을 위하여 고정계수를 가진 MA-예측양자화를 수행한다. innovation 에너지와 예측된 에너지를 구하여 에너지차가 적은 값을 찾아 각 factor를 이용하여 양자화하게 된다.

2.2 고대역 인코더 알고리즘

고대역에서는 QMF필터에 의해 분리된 3400kHz-7000kHz대역의 음성신호를 한 프레임 단위로 DCT변환되어 주파수 영역의 값으로 변환된다. 고대역 인코더에서 사용한 DCT변환은 다음과 같이 주어진다[8][9].

$$S(k)=\sum_{n=0}^{N-1} s(n)\cos\left[\frac{(2n+1)\pi k}{2N}\right]$$

여기서 s(n) : 입력신호
S(k) : DCT계수

$$\lambda(k)=\begin{cases}1 & k=0\\ \sqrt{2} & otherwise\end{cases}$$

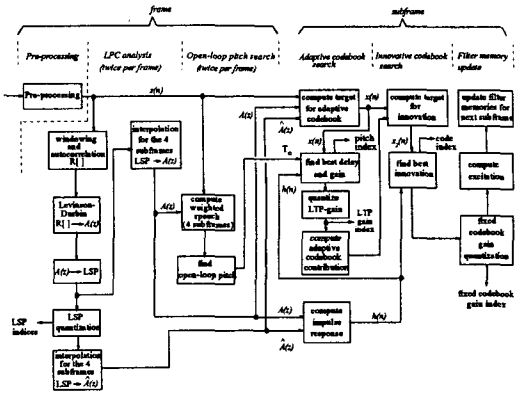


그림 2. AMR 인코더의 구조

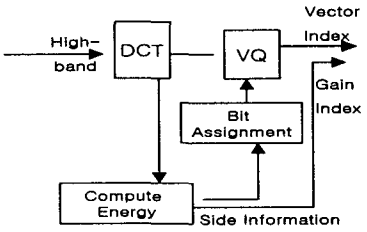


그림3. 고대역 음성부호화기의 인코더

그림 3은 고대역 신호의 인코딩 과정을 보여준다. 한프레임 내에서 음성의 스펙트럼은 모든 대역에 걸쳐 균등하게 분포되어 있지 않으므로, 모든 DCT계수에 같은 크기의 비트를 할당하여 양자화 하는 것은 비효율적이다. 즉, 한 부분의 스펙트럼이 다른 부분의 스펙트럼보다 크게 나타난다면, 스펙트럼이 큰 영역에 더 많은 비트를 할당하는 것이 효율적이다. DCT계수는 주파수 영역의 값이므로 스펙트럼의 크기는 DCT계수와 연관하여 표현된다. 한 주파수 영역의 스펙트럼이 다른 영역보다 크게 나타난다면, 그 주파수 영역에 해당하는 DCT계수의 에너지는 다른 영역의 에너지 값보다 큰 값을 갖는다. 이러한 음성의 특성과 DCT계수의 관계를 이용하여 고대역을 부호화한다. 변환된 DCT계수는 몇 개의 서브블록으로 나누어지고, 각 서브블록마다 에너지가 계산된다. 계산된 각 서브블록의 에너지값 들은 에너지 벡터 코드북을 이용하여 양자화되고, 코드북의 인덱스값을 얻는다. 에너지 코드북의 인덱스값은 각 서브블록에 할당되는 비트 정보를 포함하는 비트 벡터 코드북과 연결되어 설계된다. 그러므로 에너지 벡터코드북의 인덱스는 DCT계수를 벡터 양자화할 때, 각 서브블록에 대하여 할당되는 비트 정보를 제공한다 각 서브 블록내의 DCT계수들은 에너지 벡터 코드북의 에너지 값으로 나누어짐으로써 정규화되고, 비트 벡터 코드북의 비트할당 정보를 이용하여 각 서브블록의 정규화된 DCT계수는 벡터 양자화된다.

그림4는 에너지 벡터 코드북과 비트 벡터 코드북 그리고 벡터 코드북의 관계와 벡터 양자화 구조를 나타낸다. 이와 같은 구조로 고대역의 신호는 부호화되어 8kbps의 전송률을 갖도록 설계한다[10][11].

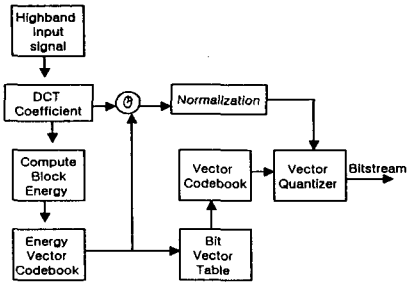


그림 4. 벡터 양자화기의 구조

Ⅲ. 광대역 음성 부호화기의 디코더 알고리즘

광대역 음성 부호화기의 전체적인 디코딩 과정은 그림5에 나타나 있다. 수신되는 비트스트림은 Demultiplexer에 의해 두 개의 정보로 분리된다. 저대역의 정보는 AMR로 역부호화되면, 고대역의 정보는 인덱스 값을 이용하여 신호를 구성하고 IDCT(Inverse Discrete Cosine Transform)과정을 통해 출력 신호를 만들어 낸다. 역부호화된 두 신호는 QMF Synthesis Bank에 의해 합성되어 최종 출력 신호를 만들어낸다 [4][5][6].

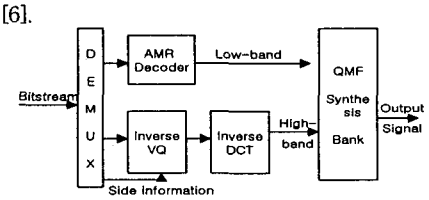


그림 5. 광대역 음성부호화기의 디코더

Demultiplexer에 의해 분리된 저대역의 신호는 AMR 디코더 알고리즘에 의하여 역부호화 된다. AMR의 디코더의 구조는 그림6에서 나타내고 있다. LP 파라미터 적응 코드북 벡터, 이득, 고정 코드북 벡터등을 수신하고 디코딩하여 합성음을 얻기 위해 과정을 수행한다. 여기신호는 적응과 innovative 코드벡터와 이득에 의 해 구해진다. 그리고 음성은 LP 합성 필터를 통과하는 excitation의 필터링 과정을 통해 복원된다.

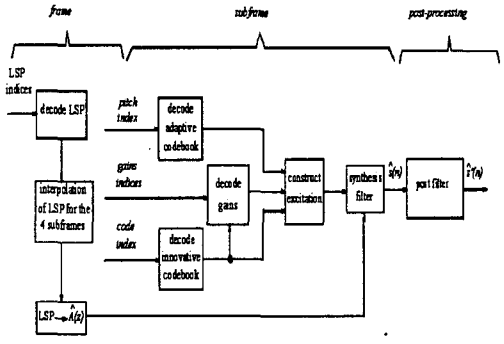


그림 6. 저대역 음성부호화기의 디코더

고대역의 정보는 역벡터 양자화 과정과 IDCT과정에 의해 역부호화된다. 고대역 역부호화기에서 사용한 IDCT의 식은 다음과 같이 주어진다[8].

$$s(n)=\frac{1}{N}\sum_{k=0}^{N-1}S(k)\lambda(k)\cos[\frac{(2n+1)\pi k}{2N}]$$

여기서, s(n) : 입력신호
S(K) : DCT 계수

$$\lambda(k)=\begin{cases}1 & k=0\\ \sqrt{2} & otherwise\end{cases}$$

고대역은 한 블록에 해당하는 이득과 비트 할당 정보는 부과정보로부터 얻어진다. 얻어진 이득, 비트 할당 정보는 DCT계수의 벡터양자화 인덱스와 결합하여 DCT계수를 복원한다. 복원된 DCT계수는 IDCT의 과정을 통하여 원래의 고대역 신호를 구성한다.

Ⅳ. 실험 및 고찰

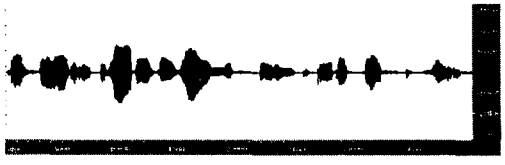
본 논문에서 제안한 광대역 부호화기의 성능을 평가하기 위하여 ITU-T의 표준안인 G.722를 포함하여 MOS(Mean Opinion Score)시험을 통하여 수행되었다. MOS시험에는 48kbps,56kbps, 64kbps 전송률의 G.722를 포함하였으며, 음질 시험에는 남자음성 10개와 여자음성10개의 문장을 사용하였다. 7명이 음질 시험에 참가하였다. 음질시험의 결과는 표2와 그림 7,8에 나타내었다. 표2는 채널환경이 없는 환경에서 MOS시험을 한 결과이고, 그림7은 원음, G.722의 64kbps, 그리고 본 논문에서 제안된 20.2kbps부호화기의 파형이다. 그림8은 원음, G .722의 64kbps, 56kbps와 설계된 음성부호화기의 20.2kbps 15.4kbps 등의 스펙트럼을 나타낸다.

MOS시험에서 G.722의 64kbps 음성은 원음과 거의 구별할 수 없을 정도로 음질이 좋았으며, G.722의 56kbps음성은 64kbps의 음성과 구별할 수는 없었지만, 음질의 선명도가 떨어지는 것을 느낄 수 있었다. 또한 AMR의모든 전송모드 음성은 음질의 선명도의 차이만 느낄 수 있었다. 본 연구의 MOS시험 결과로부터 G.722의 64kbps, 56kbps, 48kbps인 각각의 음성이 설계된 코더의 20.2kbps, 15.4kbps, 13.9kbps와 비슷한 음질을 갖음을 확인할 수 있었다.

		남 자	여자	전 체
	원 음	4.38	4.43	4.30
G.722	48kbps	3.31	3.2	3.25
	56kbps	3.48	3.66	3.57
	64kbps	4.2	4.3	4.25
설 계 된 코더	12.75kbps	3.16	3.1	3.13
	13.15kbps	3.22	3.17	3.19
	13.9kbps	3.29	3.3	3.29
	14.7kbps	3.42	3.49	3.45
	15.4kbps	3.5	3.6	3.55
	15.95kbps	3.66	3.7	3.68
	18.2kbps	4	3.9	3.95
	20.2kbps	4.2	4.13	4.16

표 2. MOS 시험 결과

(a)



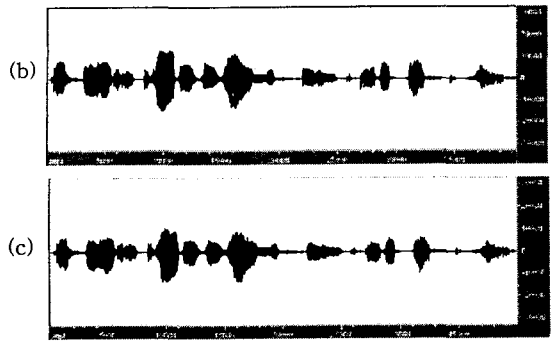


그림 7. (a) 원음의 파형 (b) G.722의 64kbps 음성파형
(c) 설계된 음성부호화기의 20.2Kbps 파형

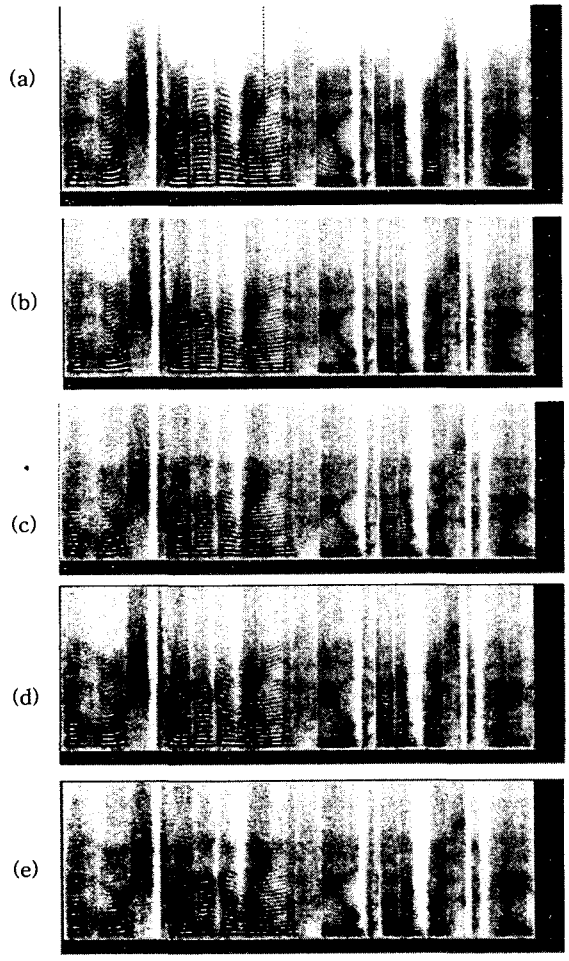


그림 8. (a) 원음의 스펙트럼
(b) G.722의 64kbps 음성스펙트럼
(c)설계된 음성부호화기의 20.2kbps 스펙트럼
(d) G.722의 56kbps 음성스펙트럼
(e) 설계된 음성부호화기의 15.4kbps 스펙트럼

V. 결론

본 연구에서는 채널의 환경에 따라 20.2kbps에서 12.75kbps까지의 전송률을 갖는 광대역 음성부호화기를 설계하였다. 제안된 광대역 음성부호화기는 멀티미디어 데이터 전송을 할 수 있는 IMT-2000 서비스의 음성부호화기의 표준으로 채택된 AMR을 포함하여 구성하고 있으므로 광대역 멀티미디어 서비스를 위한 IMT-2000과 호환성을 갖으며, IMT-2000등의 무선 화상회의 시스템등에 적용될 수 있다. 또한 다전송률 방식이므로 이동통신 시스템에서 Traffic이 높아질 때 전송률을 달리하는 방법으로 사용될 수 있다. 본 연구의 MOS시험 결과로부터 G.722의 64kbps, 56kbps인 음성이 설계된 코더의 20.2kbps, 15.4kbps와 비슷한 음질을 갖음을 확인할 수 있었다.

VI 참고 문헌

[1] F.K. Soong and B.H. Juang, "Line spectrum pair (LSP) and speech data compression", in *Proc. ICASSP'84*, pp. 1.10.1-1.10.4.

[2] K.K Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24bits/frame", *IEEE Trans. Speech and Audio Processing*, vol. 1, no 1, pp. 3-14, 1993.

[3] P. Kabal and R.P. Ramachandran, "The computation of line spectral frequencies using Chebyshev polynomials", *IEEE Trans. on ASSP*, vol. 34, no. 6, pp. 1419-1426, Dec. 1986.

[4] CCITT Recommendation G.722, "7Khz audio coding with 64kbit/s," inBlue Book, vol. III Fascicle III, Oct.1988.

[5] R. Crochiere, S. Webber, and J. Flanagan, "Digital coding of speech in sub-bands," Bell Syst.Tech. J., vol.55, no.8, p.1069, Oct.1976

[6] T.A .Ramstad, "Sub-band coder with a simple adaptive bit allocation algorithm." in Proc. ICASSP-82, p.203

[7] R. Zelinski and P. Noll, "Adaptive Transform Speech Coding of Speech Signal," IEEE Trans. on Acoustics Speech and Signal Proc. pp.299-309, August,1977

[8] N. Ahmed, T. Natarajan and K.R. Rao, "Discrete Cosine Transform." IEEE Trans. on Computer, pp.90-93, January 1974

[9] J. Makhoul et al."Vector quantization in speech coding." Pro of IEEE 23, 1551-1588, November,1985

[10] N.S. Jayynt-Peter Noll, "Digital coding of waveform," 1984

[11] A.K.Soman, P.P Vaidyanathan, "Coding Gain in Paraunitary Analysis/Synthesis System," IEEE Trans on Signal Proocessing, Vol.41, No.5, May,1993.