

# AMR부호화기와 MLT-VQ방법을 이용한 다전송률 광대역 음성부호화기

김은주, 이인성  
충북대학교 전자공학과

## Multi Rate Wideband Speech Coder with the AMR Speech Coder and MLT-VQ

Eunju Kim, Insung Lee  
Dept. of Radio Engineering, Chungbuk National Univ.  
E-Mail Address : ejkim424@korea.com

### 요약문

본 논문에서는 AMR(Adaptive Multi-Rate)과 MLT (Modulated Lapped Transform) 벡터 양자화 방법을 이용하여 광대역 음성부호화기를 설계하였다. 제한한 음성부호화 알고리즘은 split-band 구조를 가지고 있으며 16kHz로 샘플링된 신호를 입력받아 QMF 필터에 의해 두 개의 대역으로 나누어, 각각 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)부호화기와 MLT (Modulated Lapped Transform)벡터 양자화 방법을 사용하여 각각 부호화되어 전송된다. 수신단에서는 각 대역을 AMR과 IMLT(Inverse MLT) 벡터 양자화 방법으로 역부호화하여 음성신호를 합성한다. 제한한 음성부호화기는 20.2kbps에서 12.75kbps까지의 다전송률로 동작된다. 설계된 광대역 음성부호화기는 MOS시험 결과로부터 G.722의 56 kbps 음성이 설계된 코더의 20.2 kbps와 비슷한 음질을 갖음을 확인할 수 있었다.

### I. 서론

광대역 음성신호는 협대역(200Hz-3400Hz) 음성신호보다 넓은 대역폭(50Hz- 7000Hz)을 가지며 16 kHz 샘플링 주기를 요구하며 더욱 자연스러운 음질과 높은 인지도를 나타낸다. 최근 IMT-2000 이동통신 서비스가 넓은 대역폭으로 인해 영상에서 동영상까지 다양한 멀티미디어 데이터를 전송할 수 있게 되었다. 음성통신에 있어서도 기존의 협대역 음성뿐만 아니라 광대역 음성통신도 가능해진다. 또한 최근에 활발하게 보급되고 있는 고속 인터넷망에서는 전송 대역폭의 확대로 음질이 좋은 광대역 음성과 오디오 전송도 실시간 전송이 가능하게 되었다. 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. 고대역에서 사용될 MLT변환은 주파수영역의 신호를 처리하는데 있어 기존의 DCT방법과 같은 Unlapped Transform 대신 Lapped Transform을 사용하여 블록 경계에 있어서 합성된 신호의 불연속성이 감소하여 보다 선명한 음성을 합성하여 효과적이고, Dolby AC-3, MPEG-2 Layer III 등에 사용되고 있다.

본 논문에서는 II장에서는 제안한 광대역 음성부호

화기의 인코더 알고리즘을 설명하고, III장에서는 디코더 알고리즘에 대해서 설명을 한다. IV장에서는 제한한 음성부호화기의 성능을 평가하기 위해 G.722를 포함하여 주관적 음질 평가결과 및 시뮬레이션 결과를 다룬다.

### II. 광대역 음성부호화기의 인코더 알고리즘

본 연구에서 설계된 광대역 음성부호화기의 인코딩 과정이 그림1에 나타나 있다. 16 kHz로 샘플링된 입력 신호를 QMF 필터에 의해 두 개의 대역으로 나누어지고, 각각 downsampling하여 두 개의 8 kHz 샘플링 신호로 변환시킨 후 저대역(0 Hz-3400 Hz)의 신호와 고대역(3400 Hz-7000 Hz)의 신호로 나누어 각각 부호화되어 전송된다[6]. 저대역의 신호는 ETSI와 3GPP에서 차세대 이동통신 IMT-2000 서비스의 음성부호화기의 표준안으로 채택한 AMR를 이용하여 4.75에서 12.2 kbps까지 8개의 전송률로 부호화되어 전송된다. 고대역의 신호는 MLT방법으로 8 kbps의 전송률로 부호화한다[2][8]. 저대역과 고대역의 부호화된 정보는 전체적으로 20.2 kbps에서 12.75 kbps까지의 전송률을 갖는다.

#### 2.1 저대역을 위한 AMR 인코딩 알고리즘

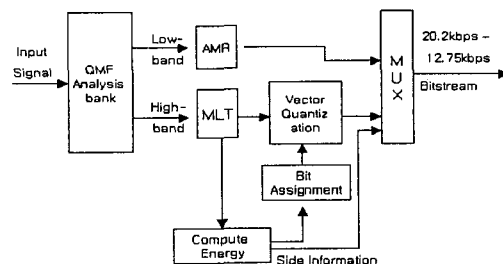


그림 1. 광대역 음성부호화기의 인코더

QMF필터에 의해 분리된 0 Hz -3400 Hz 대역의 음성신호는 AMR 인코더로 부호화된다. AMR은 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 4.75 kbit/s에서 12.2 kbit/s까지 8개로 나누고 하나의 저전송률 주변잡음 인코딩 모드를 갖는

하나의 통합된 음성부호화기를 구성한다. 프레임 사이 크는 8kHz로 샘플링된 5ms의 4개의 부프레임으로 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 *

표1. 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 고대역 여파기와 입력레벨 조절을 하여 전처리 과정을 한다. 선형예측 분석과 양자화 과정은 12.2 kbit/s 모드에서는 두 개의 선형예측 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 선형예측 필터 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 음성신호화 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 인자를 이용하여 양자화하게 된다.

### 2.2 고대역을 위한 MLT-VQ 인코딩 알고리즘

고대역에서는 QMF필터에 의해 분리된 3400 kHz-7000 kHz대역의 음성신호를 한 프레임 단위로 MLT 변환되어 주파수 영역의 값으로 변환된다. MLT는 인접 프레임간 50%의 오버랩이 적용된다. MLT(Modulated Lapped Transform)는 window, overlap과 DCT-IV로 이어지는 add로 구성된다.[5][6].

window, overlap 그리고 add는 다음과 같다.

$$v(n) = w\left(\frac{N}{2} - 1 - n\right)x\left(\frac{N}{2} - 1 - n\right) + w\left(\frac{N}{2} + n\right)x\left(\frac{N}{2} + n\right)$$

$$v\left(n + \frac{N}{2}\right) = w(N - 1 - n)x(N + n) - w(n)x(2N - 1 - n),$$

$$0 < n < \frac{N}{2}$$

$$w(n) = \sin\left(\frac{\pi}{2N}(n + 0.5)\right), \quad 0 < n < N$$

DCT-IV와  $v(n)$ 를 결합하여, MLT의 결과식은 다음과 같이 주어진다.

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

$$0 < m < N$$

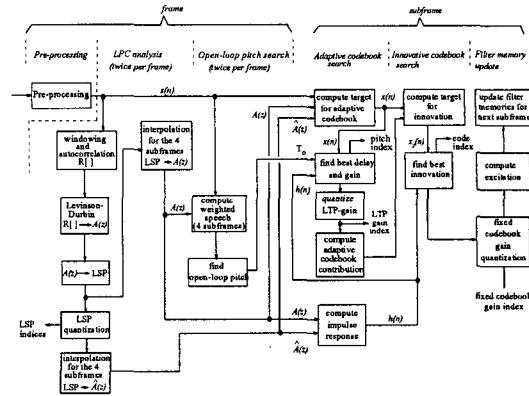


그림 2. AMR 인코더의 구조

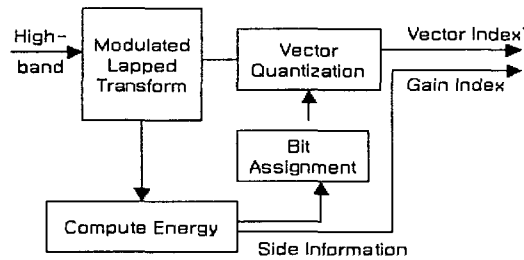


그림 3. 고대역 음성부호화기의 인코더

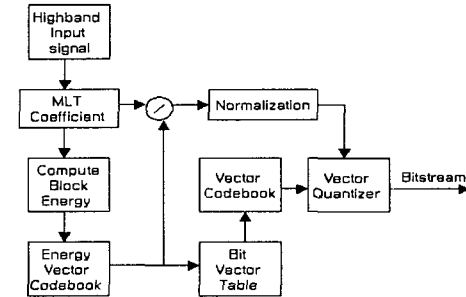


그림 4. 벡터 양자화의 구조

여기서  $v(n)$ : 입력신호,  $N$ : 입력샘플 수,  $M(m)$ : MLT 계수를 나타낸다.

그림3은 고대역 신호의 인코딩 과정을 보여준다. 한 프레임내에서 음성의 스펙트럼은 모든 대역에 걸쳐 균등하게 분포되어 있지 않으므로, 모든 MLT계수에 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. MLT계수는 주파수 영역의 값이므로 스펙트럼의 크기는 MLT계수와 연관하여 표현된다. 한 주파수 영역의 스펙트럼이 다른 영역보다 크게 나타난다면, 그 주파수 영역에 해당하는 MLT계수의 에너지는 다른 영역의 에너지 값보다 큰 값을 갖는다. 이러한 음성의 특성과 MLT계수의 관계를 이용하여 고대역을 부호화한다.

변환된 MLT계수는 몇 개의 서브블록으로 나누어지고, 각 서브블록마다 에너지가 계산된다. 계산된 각 서

브블록의 에너지값들은 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 MLT계수를 벡터 양자화할 때, 각 서브블록에 대하여 할당되는 비트 정보를 제공한다. 각 서브블록내의 MLT계수들은 에너지 벡터 코드북의 에너지 값으로 나누어짐으로써 정규화되고, 비트 벡터 코드북의 비트할당 정보를 이용하여 각 서브블록의 정규화된 MLT계수는 벡터 양자화된다.

그림4는 벡터 양자화 구조를 나타낸다. 이와 같은 구조로 고대역의 신호는 부호화되어 8 kbps의 전송률을 갖도록 설계한다[6][7][8].

### 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 역부호화되며, 고대역의 정보는 인덱스 값을 이용하여 신호를 구성하고 역양자화와 IMLT(Inverse Modulated Lapped Transform) 과정을 통해 출력 신호를 만들어 낸다. 역부호화된 두 신호는 QMF Synthesis Bank에 의해 합성되어 최종 출력 신호를 만들어낸다[3][4][5].

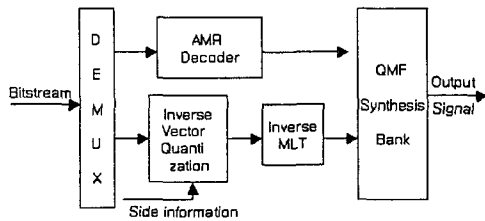


그림 5. 광대역 음성부호화기의 디코더

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 의해 구해진다. 그리고 음성은 LP 합성 필터를 통과하는 excitation의 필터링 과정을 통해 복원된다.

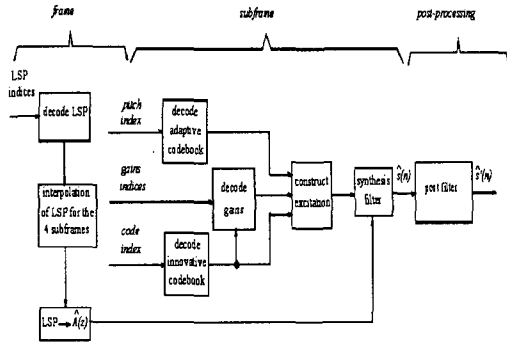


그림 6. 저대역 음성부호화기의 디코더

고대역의 정보는 역벡터 양자화 과정과 IMLT과정에 의해 역부호화된다. 고대역은 한 블록에 해당하는 이득과 비트 할당 정보는 부호정보로부터 얻어진다. 얻어진 이득, 비트 할당 정보는 MLT계수의 벡터양자화 인덱스와 결합하여 MLT계수를 복원한다.

Inverse MLT는 DCT-IV에 이어지는 window, overlap과 add로 이루어진다[2]. DCT-IV는 다음과 같다.

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

window, overlap/add는 현재 프레임과 과거 프레임의 DCT 출력의 각각 절반을 입력으로 사용하여 원래의 신호를 구성한다..

$$x(n) = w(n)u\left(\frac{N}{2} - 1 - n\right) + w(N-1-n)u_{old}(n)$$

$$x\left(n + \frac{N}{2}\right) = w\left(n + \frac{N}{2}\right)u(n) - w\left(\frac{N}{2} - 1 - n\right)u_{old}\left(\frac{N}{2} - 1 - n\right), \quad 0 < n < \frac{N}{2}$$

다음 프레임을 위해, 사용되지 않은  $u(n)$ 의 절반을  $u_{old}$  로 저장한다.

$$u_{old}[n] = u\left[n + \frac{N}{2}\right], \quad 0 < n < \frac{N}{2}$$

여기서  $x(n)$ : 입력신호,  $N$ : 입력샘플 수,  $M(n)$ : MLT 계수를 나타낸다.

### IV. 실험 및 고찰

본 논문에서 제안한 광대역 부호화기의 성능을 평가하기 위하여 ITU-T의 표준안인 G.722를 포함하여 MOS시험을 통하여 수행되었다. 실험에 사용된 음성은 16 kHz로 샘플링된 남자음성과 여자음성 각각 10개의 문장이다. 비전문가 9명이 음질 시험에 참가하였다. MOS시험에는 48, 56, 64 kbps 전송률의 G.722를 포함하였다. 음질시험의 결과는 표2와 그림 7, 8에 나타내었다. 표2는 채널환경이 없는 환경에서 MOS시험을 한 결과이고, 그림7은 원음, G.722의 56kbps, 그리고 본 논문에서 제안된 18.2 kbps부호화기의 파형이다.

MOS시험에서 원음, G.722의 64 kbps과 제안한 20.2 kbps의 음성은 거의 구별할 수 없을 정도로 음질이 좋았으며, 약간의 음질 선명도가 떨어지는 것을 느낄 수 있었다. 또한 AMR의 모든 전송모드 음성은 음질의 선명도의 차이만 느낄 수 있었다. 본 연구의 MOS시험 결과로부터 G.722의 64 kbps, 56 kbps, 48 kbps인 각각의 음성이 설계된 코더의 20.2 kbps, 18.2 kbps, 15.95 kbps와 비슷한 음질을 갖음을 확인할 수 있었다.

		남자	여자	전체
	원음	4.39	4.33	4.36
G.722	48kbps	3.89	3.79	3.84
	56kbps	4.09	3.95	4.02
	64kbps	4.33	4.28	4.305
설계된 코더	12.75kbps	3.42	3.12	3.26
	13.15kbps	3.47	3.21	3.34
	13.9kbps	3.55	3.35	3.425
	14.7kbps	3.52	3.4	3.46
	15.4kbps	3.6	3.57	3.585
	15.95kbps	3.77	3.64	3.705
	18.2kbps	4.1	3.9	4
20.2kbps	4.3	4.18	4.24	

표 2. MOS 시험 결과

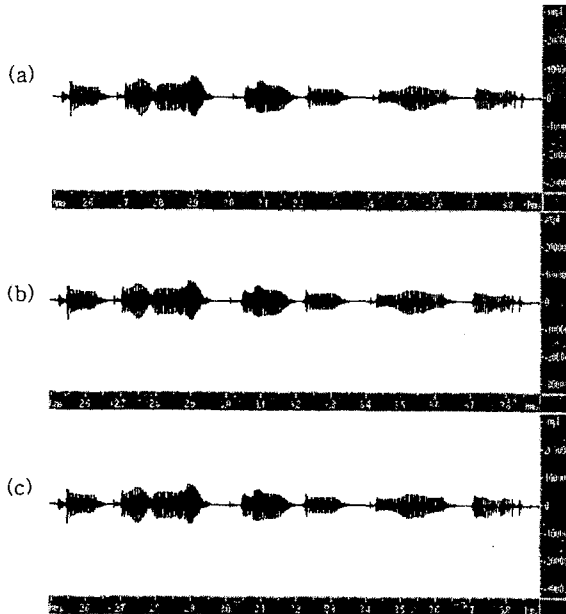


그림 7. 음성부호화기의 출력 음성 파형  
 (a) 원음 (b) 56 kbps G.722  
 (c) 설계된 음성부호화기의 18.2 kbps

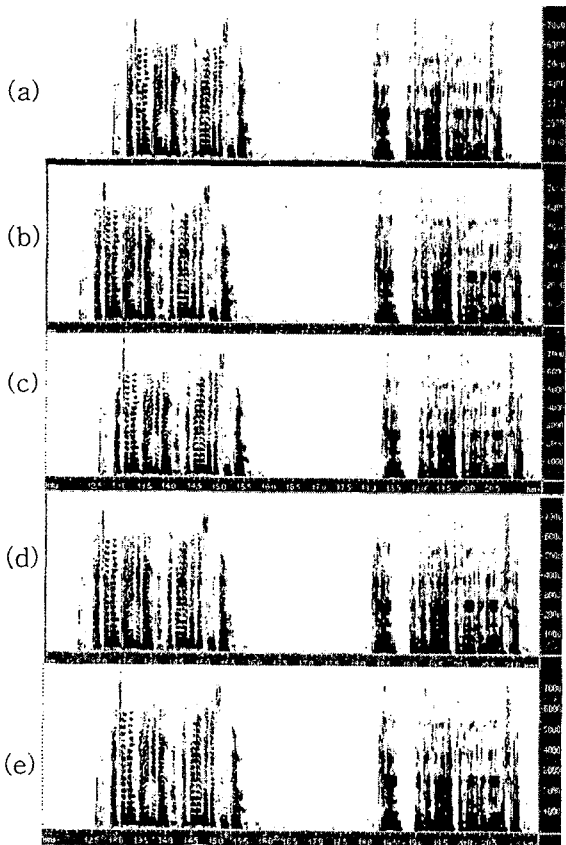


그림 8. 음성 부호화기의 출력 음성 스펙트럼  
 (a) 원음 (b) 64 kbps G.722  
 (c) 설계된 음성부호화기의 20.2 kbps  
 (d) 56 kbps G.722  
 (e) 설계된 음성부호화기의 18.2 kbps

## V. 결론

본 연구에서는 AMR과 MLT를 이용하여 채널의 환경에 따라 20.2 kbps에서 12.75 kbps까지의 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 호환성 유지하며, IMT-2000등의 무선 화상회의 시스템등에 적용될 수 있다. 32kbps이하의 전송률에서 동작되도록 설계되어, 화상회의 시스템, 고품질 디지털 AM 라디오방송, 고품질 전화 등에 사용 가능하다. 또한 다전송률 방식이므로 이동통신 시스템에서 트래픽이 높아질 때 전송률을 달리하는 방법으로 사용될 수 있다. 고대역은 기존의 DCT방법과 같은 Unlapped Transform 대신 Lapped Transform을 사용하여 블록 경계에 있어서 합성된 신호의 불연속성이 감소하여 보다 선명한 음성을 합성한다. 본 연구의 MOS시험 결과로부터 G.722의 64 kbps, 56 kbps, 48kbps인 음성이 설계된 코더의 20.2kbps, 18.2 kbps, 15.95 kbps와 비슷한 음질을 갖음을 확인할 수 있었다.

## VI 참고 문헌

- [1] 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
- [2] MALVAR, H. "Signal processing with Lapped Transform" Artech House, Boston, 1992 vol. 1, no 1, pp. 3-14, 1993.
- [3] CCITT Recommendation G.722, "7Khz audio coding with 64kbit/s," in Blue Book, vol. III Fascicle III, Oct.1988.
- [4] K.K Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24bits/frame", *IEEE Trans. Speech and Audio Processing*,
- [5] T.A .Ramstad, "Sub-band coder with a simple adaptive bit allocation algorithm." in Proc. ICASSP-82, p.203
- [6] 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
- [7] N. Ahmed, T. Natarajan and K.R. Rao, "Discrete Cosine Transform." *IEEE Trans. on Computer*, pp.90-93, January 1974
- [8] J. Makhoul et al. "Vector quantization in speech coding." *Pro of IEEE* 23, 1551-1588, November,1985
- [9] N.S. Jayynt-Peter Noll, "Digital coding of waveform," 1984

\* 본 연구는 한국과학재단의 기초 연구사업(No. 2000-1- 30200-003-3)의 지원으로 수행된 것임.