

낮은 전송률 음성 부호화 방법

강 홍 구*, 윤 대 회**

(*연세대 대학원 전자공학과 박사과정

**연세대 공대 전자공학과 교수)

1. 서 론

정보 및 통신문화가 급속히 발달해감에 따라 무선 및 네트워크 통신에도 디지털 방식을 적용하려는 연구가 활발해지고 있다. 디지털 방식은 데이터 압축을 통해 채널 용량 면에서 효율적이며, 암호화가 가능하고 잡음에 강한 장점을 지니기 때문이다 [25].

디지털로 변환된 음성 신호의 경우에도 효율적인 통신을 위해서는 데이터 압축 방법이 필수적이다. 특히 이동 통신 등에서는와 같이 9.6kbps 이하의 낮은 전송률을 요하는 응용 분야에는 RPE(Regular Pulse Excitation)[1], MPE(Multi Pulse Excitation)[2], CELP(Code Excited Linear Prediction)[3] 등의 복합 부호화(hybrid coding) 방법들이 주로 이용되고 있다. 이를 바탕으로 각 국에서는 자국의 실정에 맞는 표준 부호화 방식을 정하고, 실시간 시스템으로 확장하기 위한 작업이 활발히 진행되고 있다.

북미 및 일본 디지털 셀룰라 폰의 표준 방식으로 채택된 VSELP(Vector Sum-Excited Linear Prediction)[4][24] 방법은 8 kbps에서 우수한 성능을 보이는 CELP 계열의 방법으로서 코드북 구조와 탐색 방법 개선을 통해 실시간 처리가 용이하며, gray code를 통한 부호화 기법에 따라 채널잡음에 강한 장점을 지니고 있다. CDMA(Code Division Multiple Access) 기술을 기반으로 TIA 북

미 디지털 셀룰라의 표준 방식으로 고려되고 있는 QCELP(Qualcomm CELP)[5]는 음성의 존재 유무에 따라 전송률을 변화시키는 방식으로서 채널 용량면에서 효율적이며, 사용자 간의 상호 간섭(mutual interference) 영향에 강하다는 장점이 있다. 미국 국방성(Department of Defense)에서 기존 통신시스템 교환 목적으로 개발된 DoD-CELP [6]는 실수값(fractional) 피치 탐색 방법을 통해 음질향상을 시도했으며, ternary(+1, -1, 0)로 구성된 고정 코드북을 순환 적용하여 연산속도를 높이는 방법이다.

이와 같은 CELP계열 부호화 방법과는 달리 음성신호를 주파수 분석한 후 모델링 하는 Harmonic 부호화 방법도 활발히 연구되고 있다. 여기 신호를 주파수 분석한 후 각각의 합으로 모델링하는 정현파(sinusoidal) 부호화 방법[7]과 피치 간격으로 이루어지는 주파수 밴드의 유/무성음 특성에 따라 다른 여기신호를 사용하는 MBE(Multi-Band Excitation)[8] 방법이 그 대표적 예이다. 특히 MBE를 개선한 IMBE(Improved MBE)[9]는 양자화 및 합성방법 개선을 통해 4kbps 내외에서 우수한 음질을 보이는 부호화 방법으로서 차세대 표준화안으로 고려되고 있는 방법이다.

이러한 기존 표준안과 더불어 최근의 음성 부호화 알고리즘이 급속하게 발전하게 된 계기는 채널 용량을 확대하기 위한 half-rate 표준화 연구에서 비롯되었다. 즉, 기존 채널의 절반에서도 음질 저하

가 없는 부호화 방법 연구를 통해 다양한 접근이 이루어졌기 때문이다.

본 논문에서는 표준화되거나 표준안으로의 가능성이 높은 음성 부호화 알고리즘들에 대해 CELP 계열을 중심으로 비교 설명하고, half-rate 표준화를 위해 최근에 이루어지고 있는 주요 연구 경향에 대해 서술한다. 또한 이로부터 향후 진행 방향에 대해 간단히 언급하고자 한다.

2. 저전송률 음성 부호화 방법

음성 신호에 널리 사용되는 압축 방법은 선형 예측 부호화(Linear Predictive Coding ; LPC)이다. 특히, 높은 압축률에서 재생 신호의 왜곡을 줄이기 위해서는 원래 신호와 재생 신호 간의 오차가 최소가 되도록 잔차 신호(residual signal)를 모델링하여 부호화하는 분석 합성(analysis-by-synthesis) 방법이 더욱 효과적이다[10]. 백색 가우시안 잡음으로 구성된 벡터 코드북을 이용하여 잔차 신호를 모델링하는 CELP는 가장 활발히 연구되고 있는 방법이다. CELP는 그림 1에서와 같이 LPC계수, 장구간 적응 코드북(long-term adaptive codebook), 통계 코드북(stochastic codebook)으로 구성되어 있으며, 각각의 적용 방법에 따라 여러 알고리즘으로 구분된다.

며, 일본에서는 6.7kbps이다. 이 방법의 특징을 정리하면 다음과 같다.

- 벡터 합으로 고정 코드북 구성 : 계산량 및 메모리 감소
 - 여기 신호에 gray 코드 사용 : 채널 잡음에 강하다.
 - 계산량 감소
 - LPC 계수 : 반사 계수(reflection coefficient) 사용
- 부호화기의 블럭도 및 각 변수에 할당된 비트는 다음과 같다.

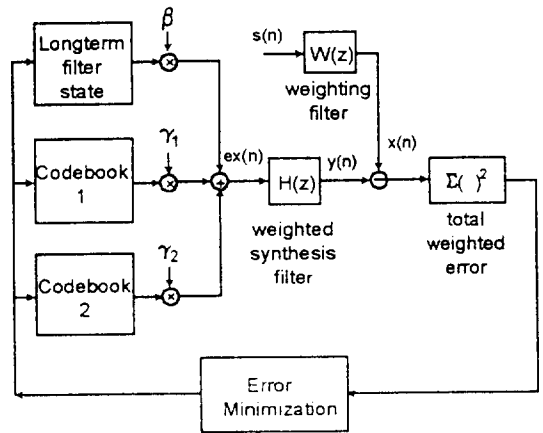


그림 2. VSELP 부호화기의 블럭도

표 1. 8Kbps VSELP 비트 할당

parameter	bit / 5ms subframe	bit / 20ms frame
LPC coefficient		38
frame energy		5
1st & 2nd codebook	7+7	56
LTP lag	7	28
gain (GS, P0, P1)	8	32
<unused>		1
total	29	160

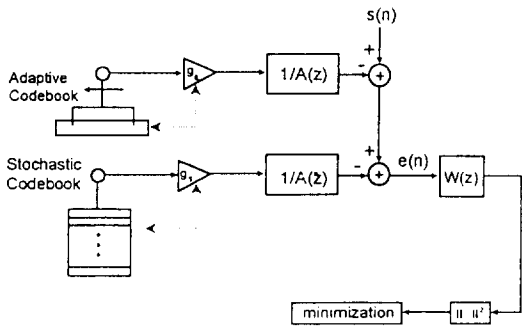


그림 1. CELP 부호화기의 블럭도

2.1 VSELP[4]

VSELP는 3개의 여기 신호원으로부터 여기 신호 벡터를 결정하고 이로부터 피치 합성 필터 및 포먼트 합성 필터를 구동하여 최종 합성음을 얻는 CELP형 부호화기이다. 전송률은 북미에서는 8kbps이

2.2 QCELP[5]

QCELP는 분석하고자 하는 음성 구간의 에너지와 배경 잡음(background noise)에 따라 전송 부

호화율을 다르게 하는 CELP 계열 부호화 방법으로서 주요 특징은 다음과 같다.

- 가변 전송률 : 8,4,2,1 kbps
- 고정 코드북 : center-clipped 가우시안
- LPC 계수 양자화 : LSP(Line Spectrum Pair) 사용

부호화기의 블럭도 및 전송률에 따라 각 변수에 할당된 비트는 다음과 같다.

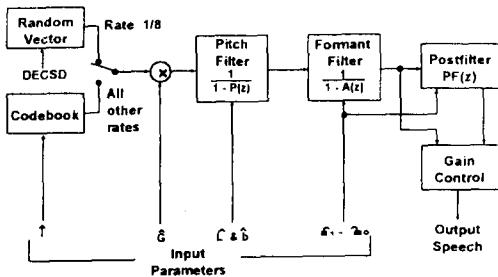


그림 3. QCELP 부호화기의 블럭도

	8 kbit/s	4 kbit/s
LPC	40	20
Pitch	10 10 10 10	10 10
Codebook	10 10 10 10 10 10	10 10 10 10
LPC	10	10
Pitch	10	0
Codebook	10 10	6

그림 4. 전송률에 따라 각 변수에 할당된 비트

2.3 DoD 4.8kbps CELP[6]

DoD-CELP는 기존 CELP와 동일한 구조를 가지며, 전송률을 낮추기 위해 프레임의 길이를 30msec, 즉 240샘플로 확장시켰다. 또한, 장구간

표 2. DoD CELP 부호화기의 기능별 특징

	선형 예측기	적용 코드북	통계 코드북
분석구간	30ms	30/4=7.5ms	30/4=7.5ms
변수	10LSP	1 gain, 1 delay 256 codewords	1 gain, 1 index 512 codewords
분석방법	·open loop ·no preemphasis ·30ms Hamm, window	·closed loop ·delta search range : 20-147 ·noninteger delay	·closed loop ·77% sparcity ·ternary samples
Bit / frame	34 (3,4,4, 4,4,3,3,3,3,3)	index:8+6+8 +6± gain:5×4	index:9×4 ± gain:5×4
Rate	1133.33bps	1600bps	1866.67bps

예측시 정확도를 높이기 위해 실수 간격으로 탐색하여 성능을 향상시킨 방법이다.

각 기능별 특징은 표2와 같다.

3. 최근의 주요 연구 방향

위에서 설명한 부호화 방법들은 대부분 8kbps 정도의 전송률에서는 우수한 성능을 보이지만 4kbps 이하로 낮춰질 경우에는 음질이 심각히 왜곡된다. 최근의 연구는 이러한 부호화 시스템의 개선 혹은 새로운 알고리즘 개발을 통해 전송률을 1/2 (half-rate) 이하로 낮추고자 진행되고 있다. half-rate를 위한 기본 조건은 다음과 같다.

- 전송률 : 북미(4kbps), 유럽(5~7kbps), 일본(4kbps 이하)
- 음 질 : full-rate 이상
- 복잡도 : full-rate의 4배 이하
- 지연시간 : 100 msec 이하

북미, 유럽 및 일본에서는 이러한 조건을 만족하는 몇가지 부호화 알고리즘을 대상으로 표준화 작업이 진행중이다.

본 절에서는 half-rate 전송률을 위해 연구되고 있는 내용들을 CELP 방법을 중심으로 LPC 계수 양자화, 코드북 구조 및 탐색 방법, 여기 신호 모델링 방법으로 나누어 설명하고, 앞으로의 연구 방향에 대해 고찰하고자 한다.

3.1 LPC 계수 양자화

음성 신호의 단구간 스펙트럼 포락선(envelope)을 표현하는 LPC 계수는 저장 혹은 전송을 위해 양자화 과정을 거쳐야만 한다. 그러나, LPC 계수 자체는 동적 범위(dynamic range)가 넓고, 미세한 오차에 의해서도 안정성을 유지하기 어려우므로 보통 LSP(Line Spectrum Pair), 반사계수(reflection coefficient), 또는 LAR(Log Area Ratio) 등으로 변환하여 양자화 된다[11]. VSELP에서와 같이 반사 계수를 이용하는 경우도 있지만, LSP를 사용하는 예가 대부분이다.

효율적인 음성 부호화를 위해 허용되는 평균 스펙트럼 오차는 1dB 이내로서 LSP 계수를 스칼라 양자화하는 경우에는 20msec 프레임당 32-40비트가 필요하다고 알려져 있다[11]. 보다 낮은 비트로

도 원하는 조건을 충족시키기 위해 벡터 양자화(VQ) 방법이 도입되었으며[12], 특히 다단(multi-stage) VQ, 분할(split) VQ 등은 VQ 방법이 갖는 계산량 문제를 다소 해결할 수 있는 방법으로서 자주 이용되는 방법이다.

다단 VQ는 벡터 양자화기를 직렬(cascade)로 연결하여 앞단 양자화기를 기진 오차 신호를 다시 양자화하는 방법이며[13], 분할 VQ는 입력 벡터를 두 개 이상의 작은 벡터로 나누어 양자화하는 방법[14]으로서 프레임당 24비트 내외에서 만족할만한 성능을 유지한다고 알려져 있다. 이외에도 LPC 계수가 갖는 시간상, 공간상의 상관 관계를 이용하여 2차원으로 양자화하는 방법[15], 스칼라 양자화와 벡터 양자화를 결합한 방법[16] 등도 꾸준히 연구되고 있는 분야이다.

전송률을 낮추기 위해서는 궁극적으로 VQ방법이 사용되어야 할 것이다. 그러나, 모든 음성 신호의 특성을 표현할 수 있는 코드북 발생 방법 문제와 얻어진 코드북을 저장하기 위한 메모리 용량 및 계산량 증가 문제는 심각히 고려해야 할 문제이다.

3.2 코드북 구조 및 탐색 방법

초창기 CELP 계열의 음성 부호화기가 갖는 가장 큰 단점은 코드북 저장을 위한 메모리가 많이 필요하며, 이에 따라 탐색 시간이 길어진다는 것이었다. 이를 해결하기 위해 구조화된 코드북을 사용하는 방법이 제안되었다. 특히, 몇 개의 기저벡터로부터 다수의 코드북을 만들어내는 방법[4], 하나의 큰 코드북을 몇 샘플씩 이동시키며 순환적으로 적용하는 방법[5,6], 나무(tree) 구조로 구성하여 단순화 시킨 방법 등은 자주 이용되는 방법이다. 또한, ternary로 구성된 기본 코드북 연산이 쉽도록 변환시켜 계산량을 감소시키는 방법도 실시간 시스템에 활발히 응용되고 있다.

이 분야는 많은 연구가 이루어져서 최근에는 독특한 구조가 제안되고 있지는 않지만, 기존 구조를 최적화하여 성능을 개선하고, 이를 실시간 시스템에 적용시키기 위한 방향으로 진행되고 있다.

3.3 여기 신호 모델링

여기 신호는 합성된 음성의 음질에 큰 영향을 미치는 변수로서 특히, LPC계열의 분석 합성 부호화

기 경우에는 전체 사용 가능 비트의 70-80%가 할당될 정도로 매우 중요한 변수이다.

낮은 전송률에서 여기 신호 구성을 위해 가장 중요한 부분은 유성음 구간에 존재하는 피치 성분을 모델링하는 것이다. 유성음 구간에서는 장시간 예측 후에도 펄스 형태의 신호가 존재하므로 랜덤 신호로 이루어진 코드북으로 이를 모델링 하기 위해서는 코드북의 크기가 커져야한다는 단점이 있다. 이를 해결하기 위해 샘플 사이의 값을 interpolation하여 피치 지연 신호를 실수 간격으로 사용하는 방법[17], 고정 코드북을 다단으로 구성하는 방법[18], 가능성이 높은 몇 개의 코드북을 후보로 정했다가한 프레임이 끝난 후에 최적의 코드북을 결정(delayed decision)하는 방법[19], 몇 개의 밴드로 나누어 모델링하는 방법[20][26], 피치 간격에 동기가 맞도록 코드북을 구성하는 방법[21] 등이 제안되었다. 특히 최근에는 구간 음성 신호의 특성에 따라 다른 코드북 혹은 다른 부호화 방법을 사용하는 다중 모드 부호화 방법[22,23]이 half-rate 부호화 방법의 주류를 이루고 있다.

낮은 전송률에서도 성능을 유지하기 위한 방법은 결국 음성 신호의 특성에 따라 다른 방법으로 여기 신호를 모델링하는 방법일 것이다. 이를 위해서는 [23]의 경우와 같이 시간축으로 일정 구간마다 모드를 결정할 수도 있고, MBE 방법과 같이 주파수축을 분할하여 각 밴드의 특성을 반영할 수도 있을 것이다.

4. Half-rate 음성 부호화 방법 예

4.1 PSI-CELP

(Pitch Synchronous Innovation CELP)[21]

PSI-CELP는 half-rate일본 디지털 셀룰라(JDC)에 사용되고 있는 방법으로서 모든 CELP 계열의 통계 코드북에 적용될 수 있다. 3.3절의 여기 신호 모델링에서도 언급했듯이 half-rate에서는 피치 성분이 음질에 심각한 영향을 주는 성분이다. PSI-CELP에서는 이를 위해 그림 5와 같이 유성음 구간에서는 통계 코드북도 피치 주기에 따라 적층적으로 변환시켜서 피치 성분을 강화하는 방법을 사용한다.

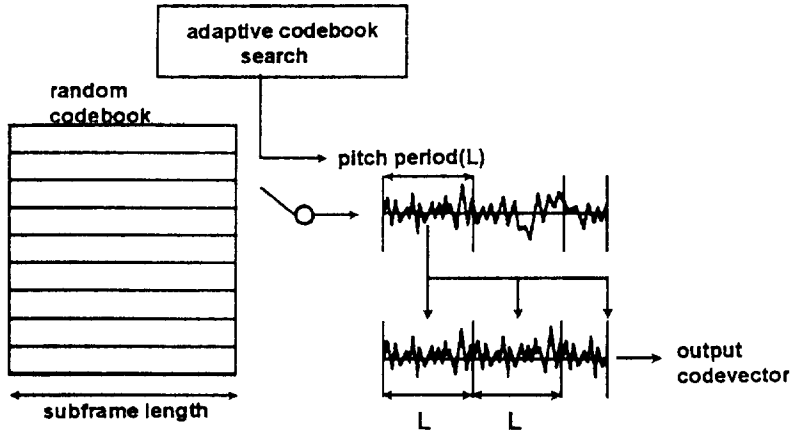


그림 5. PSI-CELP의 통계 코드북 구조

4.2 M-LCELP

(Multi-mode Learned CELP)[23]

M-LCELP는 복음 half-rate 디지털 셀룰라 표준 방법으로 제안된 방법으로서 주요 특징은 다음과 같다.

- 구간 특성(유성음/무성음)에 따라 다른 부호화 방법 사용
 - LPC 계수 : LSP 사용
 - 양자화 방법 : 다단 분할(multi-stage split) VQ
 - 유성음 구간에서는 comb필터 사용
- 부호화기의 블럭도는 그림 6과 같다.

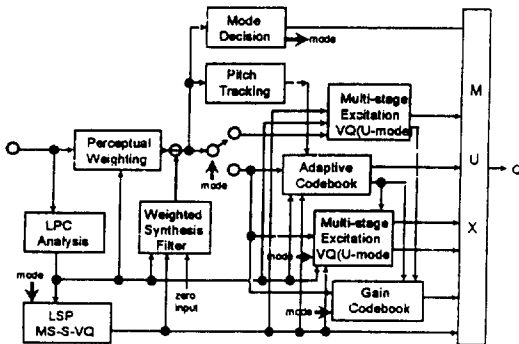


그림 6. M-LCELP 부호화기

5. 결 론

본 논문에서는 표준화되거나 표준안으로의 가능

성이 높은 음성 부호화 알고리즘들에 대해 비교 설명하였으며, 보다 낮은 전송률을 위한 최근의 주요 연구 경향에 대해 서술하였다.

정보 및 통신의 필요성이 증대되면서 음성 부호화 방법에 관한 연구는 꾸준히 진행되어 왔다. 특히, 이동 통신에 대한 수요가 증가함에 따라 선진국에서는 기본 표준안을 완성하고, 채널 용량을 확대하기 위한 half-rate 표준화 작업이 한창 진행되고 있다.

half-rate 전송률에서도 우수한 음질을 유지하기 위해서는 LPC 양자화 방법과 여기 신호 생성 방법이 중요한 역할을 하며, 이를 해결하기 위한 최근의 추세는 구간 특성에 따라 다른 부호화 방법(다중 모드 부호화)을 사용하는 것이다. 제안된 알고리즘을 실시간 시스템으로 구현하기 위한 계산량 및 메모리 감소 문제와 구현된 부호화기를 실제 상황에 적용하였을 때 발생할 수 있는 여러 문제점, 예를 들면 채널 에러에 대한 영향 등도 관심을 두고 연구해야 될 분야라고 생각된다.

무엇보다 중요한 것은 이러한 알고리즘들을 비교 평가하여 우리의 실정에 맞는 시스템으로 표준화하는 작업이며, 가능하면 순수한 우리의 기술로 이러한 연구가 이루어지도록 노력하는 것이다.

참 고 문 헌

- [1] P. Kroon, E. F. Deprettere, R. J. Slu-

- yster, "Regular Pulse Excitation : A Novel Approach to Effective And Efficient Multipulse Coding of Speech," IEEE Trans. ASSP, Vol. ASSP-34, pp. 1054-1063, Oct. 1986.
- [2] B. S. Atal, J. R. Remde, "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates," Proc. ICASSP, pp. 614-617, 1982.
- [3] M. R. Schroeder, B. S. Atal, "Code-Excited Linear Prediction(CELP) High-Quality Speech at Very Low Bit Rates," Proc. ICASSP, 25.1.1-25.1.4, 1985.
- [4] I. A. Gerson, M. A. Jasuik, "Vector-Sum Excited Linear Prediction(VSELP) Speech Coding at 8Kbps," Proc. ICASSP, pp. 641-644, 1990.
- [5] W. Gardner et al, "QCELP : A Variable Rate Speech Coder For CDMA Digital Cellular," Speech and Audio Coding for Wireless and Network Applications, Kluwer Academic Publisher, 1993.
- [6] J. P. Campbell, T. A. Tremain, V. C. Welch, "DoD 4.8Kbps Standard (Proposed Federal Standard 1016)," Advances in Speech Coding, Kluwer Academic Publisher, 1991.
- [7] R. J. McAulay and T. F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," IEEE Trans. Acoust., Speech and Signal Proc., vol. ASSP-34, No. 4, Aug. 1986.
- [8] D. W. Griffen and J. S. Lim, "Multi-Band Excitation Vocoder," IEEE Trans. Acoust., Speech and Signal Proc., vol. ASSP-36, Aug. 1988.
- [9] IMBE Vocoder Description, Digital Voice Systems, Inc., 1993.
- [10] P. Kroon, E. F. Deprettere, "A Class of Analysis-by-Synthesis Predictive Coders for High Quality Speech Coding at Rates between 4.8 And 16 Kbit /s," IEEE Journal on Selected Areas in Communications, Vol. 6, no. 2, pp. 353-363, Feb. 1988.
- [11] F. K. Soong, B. H. Juang, "Line Spectrum Pair(LSP) And Speech Data Compression," Proc. ICASSP, pp. 1.10.1-1.10.4, 1984.
- [12] K. Paliwal and B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 bits/frame," Proc. ICASSP, pp. 661-664, Mar. 1991.
- [13] B. Bhattacharya, W. P. LeBlanc, S. A. Mahmoud, and V. Cuperman, "Tree Searched Multi-Stage Vector Quantization of LPC Parameters for 4kb/s Speech Coding," Proc ICASSP, pp. 105-108, May 1992.
- [14] K. W. Law and C. F. Chan, "A Novel Split Residual Vector Quantization Scheme for Low Bit Rate Speech Coding," Proc ICASSP, pp. 493-496, May 1994.
- [15] C. C. Kuo, F. R. Jean, and H. C. Wang, "Low Bit-Rate Quantization of LSP Parameters Using two-Dimensional Differential Coding," Proc. ICASSP, pp. 97-100, May 1992.
- [16] J. Grass, and P. Kabal, "Methods of Improving Vector-Scalar Quantization of LPC Coefficients," Proc. ICASSP, pp. 657-660, Mar. 1991.
- [17] P. Kroon and B. S. Atal, "Pitch Predictors with High Temporal Resolution," Proc. ICASSP, pp. 661-664, 1990.
- [18] G. Davidson and A. Gersho, "Multiple-Stage Vector Excitation Coding of Speech Waveforms," Proc. ICASSP, pp. 163-166, 1988.
- [19] K. Mano, T. Moriya, "4.8 kbit /s Delayed Decision CELP Coder Using Tree Coding," Proc. ICASSP, pp. 21-24, 1990.
- [20] P. Mermelstein, P. Zheng and M. Sai-

kaly, "Multi-Band Residual Coding of CELP CODECs at 8 kb/s," Proc. ICASSP, pp. 117-120, May 1994.

- [21] S. Miki et al, "A Pitch Synchronous Innovation CELP(PSI-CELP) Coder for 2-4 kbit/s," Proc. ICASSP, pp. 113-116, May 1994.
- [22] S. Wang, A. Gersho, "Improved Phonetically-Segmented Vector Excitation Coding at 3.4 kb/s," Proc. ICASSP, pp. 349-352, May 1992.
- [23] K. Ozawa, M. serizawa, T. Miyano and T. Nomura, "M-LCELP Speech Coding at 4kbps," Proc. ICASSP, pp. 269-272, May 1994.
- [24] 낮은 비트율 음성부호화기(VSELP)에 대한 연구, 연세대학교 전자정보통신 연구소, 1994.
- [25] PSTN-Videophone용 저전송률 음성 압축 알고리즘 연구, 연세대학교, 1993.
- [26] 최용수, 서정태, 윤대회, 차일환, "대역 분할 장구간 예측을 이용한 CELP 부호화기의 성능 향상에 관한 연구," 제11회 음성 통신 및 신호처리워크샵 논문집, 1994.



강홍구(姜泓求)

1966년 11월 1일생. 1989년 2월 연세대 공대 전자공학과 졸업. 1991년 2월 동 대학원 전자공학과 졸업(석사). 1991년 3월~현재 연세대 대학원 전자공학과 박사과정. 주관심 분야: 음성 부호화, 오디오 부호화, 실시간 신호처리 시스템



윤대회(尹大熙)

1951년 5월 25일생. 1977년 2월 연세대 공대 전자공학과 졸업. 1979년 8월 Kansas State University, Electrical Eng., 공학석사 1982년, 8월 Kansas State University, Electrical Eng., Ph. D. 1978년 9월~79년 8월 Kansas State University, Research Assistant. 1979년 9월~19년 8월 Kansas State University, Research Associate. 1982년 8월~85년 6월 University of Iowa, Assistant Professor. 1985년 9월~현재 연세대 공대 전자공학과 교수. 주관심 분야: 음성 부호화, 오디오 부호화, 음성 인식, 적응 디지털 필터, 능동소음제어, 어레이 신호처리 및 응용, 실시간 신호처리 시스템, 음질 향상, 비선형 시스템