

## 낮은 전송률 음성 부호화 방법

윤 대 희  
연세대학교 전자공학과

### Low Bit-Rate Speech Coder

Dae-Hee Youn  
Dept. of Electronics Eng. Yonsei Univ.

#### 요 약

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. 본 논문에서는 표준화되거나 표준안으로의 가능성이 높은 음성 부호화 알고리즘들에 대해 비교 설명하고, 최근에 이루어지고 있는 주요 연구 경향에 대해 서술한다. 또한 이로부터 향후 진행 방향에 대해 언급하고자 한다.

#### 1. 서 론

정보 및 통신문화가 급속히 발달해감에 따라 지금까지 사용되어 왔던 아날로그 셀룰라폰 시스템을 디지털 방식으로 대체하려는 연구가 활발해지고 있다. 디지털 방식은 데이터 압축을 통해 채널 용량면에서 효율적이며, 암호화가 가능하고 잡음에 강한 장점을 지니기 때문이다.

음성 신호 데이터 압축을 위해서는 RPE(Regular Pulse Excitation)[1], MPE(Multi Pulse Excitation)[2], CELP(Code Excited Linear Prediction)[3] 등의 복합 부호화(Hybrid Coding) 방법들이 주로 이용되고 있으며, 최근에는 성능 개선을 통해 실시간 시스템으로 확장되고 있다. 이를 바탕으로 각 국에서는 자국의 실정에 맞는 표준 음성 부호화 방식을 정하기 위한 작업도 활발히 진행되고 있다.

북미 및 일본 디지털 셀룰라 폰의 표준방식으로 채택된 VSELP(Vector Sum-Excited Linear Predictive)[4] 방법은 효율적인 코드북 구조와 탐색 방법을 통해 실시간 처리가 용이하며 gray code를 통한 부호화 기법에 따라 채널잡음에 강한 장점을 지니고 있다. CDMA(Code Division Multiple Access) 기술을 기반으로 TIA 북미 디지털 셀룰라 폰의 표준 방식으로 고려되고 있는 QCELP[5]는 음성의 존재 유무에 따라 전송률을 변화시키는 방식으로서 채널 용량면에서 효율적이며, 배경 잡음의 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 대표적인 예이다. 특히 MBE는 양자화 및 합성방법 개선을 통해 4kbps 내외에서 우수한 음질을 보이는 부호화 방법으로서 차세대 표준화 안으로 고려되고 있다.

본 논문에서는 표준화되거나 표준안으로의 가능성이 높은 음성 부호화 알고리즘들에 대해 비교 설명하고, 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 선형예측 부호화(Linear Predictive Coding; LPC)이다[9]. 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]. 그림 1은 일반적으로 많이 사용되는 CELP 부호화 방법의 블록도이다.

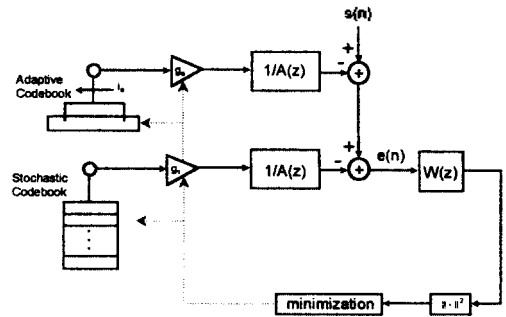


그림 1. CELP 부호화기의 블록도

낮은 전송을 음성부호화 방법

2.1 VSELP[4]

VSELP는 3개의 여기 신호원으로부터 여기 신호 벡터를 결정하고 이 벡터를 피치 합성 필터 및 포먼트 합성 필터를 구동하여 최종 합성음을 얻는 CELP형 부호화기이다. 전송율은 복음에서는 8kbps 혹은 4.8kbps이며, 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 양자화 : 반사 계수(reflection coefficient) 사용  
각 변수에 할당된 비트는 다음과 같다.

표 2-1. 8Kbps VSELP의 비트 할당

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

2.2 QCELP[5]

QCELP는 분석하고자 하는 음성 구간의 에너지와 배경 잡음(background noise)에 따라 전송 부호화율을 다르게 하는 CELP 계열의 부호화 방법으로서 주요 특징은 다음과 같다.

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

전송율에 따라 각 변수에 할당된 비트는 다음과 같다.

8 kbit/s	
LPC	40
Pitch	10 10 10 10
CB	10 10 10 10 10 10 10 10

4 kbit/s	
LPC	20
Pitch	10 10
CB	10 10 10 10

2 kbit/s	
LPC	10
Pitch	10
CB	10 10

1 kbit/s	
LPC	10
Pitch	0
CB	6

2.3 DoD 4.8kbps CELP[6]

DoD-CELP는 기존의 CELP 부호화 방법과 거의 유사하며, 주요 기능은 단구간 선형예측, 장시간 피치 코드북 검색, 단시간 코드북 검색으로 구성된다. 각 기능별 특징은 다음과 같다.

표 2-2. DoD CELP 부호화기의 기능별 특징

	Linear Predictor	Adaptive codebook	Stochastic codebook
Update	30ms	30/4=7.5ms	30/4=7.5ms
변수	10 LSPs	1 gain, 1 delay 256 codewords	1 gain, 1 index 512 codewords
분석방법	open loop no preemphasis 30ms Ham. window autocorrelation	closed loop delta search range : 20-147 noninteger delay	closed loop 77% sparsity ternary samples
Bit/frame	34 (3,4,4,4,4,3,3,3,3,3)	index:8+6+8+6 ± gain:5x4	index:9x4 ± gain:5x4
Rate	1133.33bps	1600bps	1866.67bps

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 이하로 낮추고자 진행되고 있다. 복음, 유럽 및 일본에서는 기존의 음성 부호화 시스템의 전송률을 절반(half-rate)으로 낮추기 위해 몇가지 부호화 알고리즘을 대상으로 표준화 작업이 진행중이다. half-rate 시스템이 이루어지려면 full-rate 시스템과의 호환성 및 성능은 그대로 유지되어야 하므로 보다 정교한 알고리즘이 필요할 것이다.

본 절에서는 half-rate 전송률을 위해 연구되고 있는 내용들을 세가지 주제로 나누어 설명하고, 앞으로의 연구 방향에 대해 고찰하고자 한다.

3.1 LPC 계수 양자화

음성 신호의 단구간 스펙트럼 포락선을 표현하기 위한 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 20 msec 프레임당 32-40 비트가 필요하다고 알려져 있다[11]. 보다 낮은 비트로도 원하는 조건을 충족시키기 위해 벡터 양자화(VQ) 방법이 도입되었으며[12], 특히 다단(multi-stage) VQ, 분할(split) VQ 등은 VQ 방법이 갖는 계산

항 문제를 다소 해결할 수 있는 방법으로서 자주 이용되는 방법이다. 다만 VQ는 벡터 양자화기를 직렬(cascade)로 연결하여 앞 단 양자화기를 거친 오차 신호를 다시 양자화하는 방법이며 [13], 분할 VQ는 입력 벡터를 두 개 이상의 작은 벡터로 나누어 양자화하는 방법[14]으로서 이러한 방법은 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 VQ 방법이 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 코드북에 따라 메모리 용량이 커지며, 탐색 시간이 길어진다는 것이었다. 이를 해결하기 위해 구조화된 코드북을 사용하는 방법이 제안되었다. 특히, 몇 개의 기저벡터로부터 다수의 코드를 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.

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, LPC 계열의 분석 합성 부호화기 경우에는 전체 사용 가능 비트의 70-80% 가 할당될 정도로 매우 중요한 변수이다.

낮은 전송률에서 여기 신호 구성을 위해 가장 중요한 부분은 유성을 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 커져야 한다는 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 코드를 후보로 정했다가 한 프레임이 끝난 후에 최적의 코드북을 결정(delayed decision) 하는 방법[19], 몇 개의 밴드로 나누어 모델링하는 방법[20], 피치 간격에 동기가 맞도록 코드북을 구성하는 방법[21] 등이 제안되었다. 4 kbps 정도를 위해 제안되고 있는 알고리즘은 대부분 구간 음성 신호의 특성에 따라 다른 코드북 혹은 다른 부호화 방법을 사용하는 다중 모드 부호화 방법[22,23]이 주류를 이루고 있다.

낮은 전송률에서도 성능을 유지하기 위한 방법은 결국 음성 신호의 특성에 따라 다른 방법으로 여기 신호를 모델링하는 방

법일 것이다. 이를 위해서는 [23]의 경우와 같이 시간축으로 일정 구간마다 모드를 결정할 수도 있고, MBE 방법과 같이 주파수 축을 분할하여 각 밴드의 특성을 반영할 수도 있을 것이다.

### 4. 결 론

본 논문에서는 표준화되거나 표준안으로의 가능성이 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 매우 중요한 역할을 한다는 것을 알 수 있으며, 제안된 알고리즘을 실시간 시스템으로 구현하기 위한 계산량 및 메모리 감소 문제 등도 많은 연구가 이루어져야 할 분야이다. 또한, 구현된 부호화기가 실제 상황에 적용되었을 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. Sluyter, "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. Jasiuk, "Vector-Sum Excited Linear Prediction(VSBLP) 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.

낮은 전송률 음성부호화 방법

- [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] J. Markel, A. Grey, *Linear Prediction of Speech*, Springer-Verlang, 1976
- [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.Saikaly, "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.