

낮은 전송률 음성 부호화 연구 동향

강 용 구, 윤 대 희
연세대학교 전자공학과

Trends of Low Bit-Rate Speech Coding

H. G. Kang, D. H. Youn
Department of Electronic Engineering, Yonsei University

요 약

음성 신호를 압축한 후, 주관적 왜곡 없이 복원하기 위한 부호화 연구는 대역 제한된 채널에 적용하기 위해 전송률을 더욱 낮추기 위한 연구 분야와 전송률은 높더라도 복잡도 및 성능 면에서 효과적인 연구 분야로 나눌 수 있다.

본 논문에서는 기존 시스템의 절반에 해당하는 전송률(half rate)에서도 비슷한 성능을 유지함으로써 디지털 셀룰라 표준화안으로의 가능성이 가장 높은 음성 부호화 방법들에 대해 비교 설명하고, 전송률을 좀 더 낮추기 위해 진행되고 있는 새로운 연구 경향에 대해 고찰한다. 또한, 전송률은 높더라도 기존 방법에 비해 복잡도는 매우 낮은 부호화에 대한 설명을 통해 향후 연구 방향에 대해 언급하고자 한다.

1. 서 론

정보 통신 문화가 발달함에 따라 시간 및 장소에 제약받지 않는 이동 통신 시스템에 대한 필요성이 증대되고 있다. 디지털 방식은 데이터 압축을 통해 채널 용량 면에서 효율적이며 암호화가 가능하고 감음에 강한 장점을 지니므로 기존의 아날로그 방식을 대체하고 있다[1].

신호를 압축한 후 왜곡없이 복원하기 위한 부호화 연구는 디지털 셀룰라, 음성 저장 및 전송 시스템에 적용되고 있으며, 일반적으로 선형 예측(linear prediction)을 기본으로 하는 분석-합성(analysis-by-synthesis) 부호화 방법이 가장 활발히 연구되고 있다[2]. 분석-합성이란 부호화 단계에서 선택 가능한 모든 여기 신호로부터 합성음을 얻고 이를 원음과 비교하여 오차가 최소가 되는 여기 신호를 선택하는 방법으로서 여기 신호 발생 방법에 따라 RPE(Regular Pulse Excitation)[3], MPE(Multi-Pulse Excitation)[4], CELP(Code Excited Linear Prediction)[5] 등으로 나눌 수 있다. CELP는 4.8 kbps 내외에서 만족할 만한 성능을 유지하며

구조화된 코드북과 고속 탐색 알고리즘을 통해 실시간 시스템 구현이 가능하므로 가장 활발히 연구되고 있는 방법이다.

최근에는 4kbps 내외의 전송률을 갖는 half-rate에서도 기존 시스템과 비슷한 성능을 유지하기 위한 연구를 통해 낮은 전송률 음성 부호화 기술은 급격한 발전을 이루고 있으며, 이를 바탕으로 각 국에서는 표준안을 정하기 위한 작업이 진행 중이다[6]. 그러나, CELP를 이용하는 방법은 전송률이 4 kbps 내외로 낮춰지면 유성음에 존재하는 주기 성분을 정확히 모델링하지 못하게 됨에 따라 성능이 급격히 저하된다[6]. 이를 극복하기 위해 여러 알고리즘이 제안되고 있으며, 대표적인 것이 다중 모드(multi-mode)[7][8] 방법과 피치 동기화(pitch synchronous)[9] 방법이다.

이러한 CELP 계열 음성 부호화의 성능을 개선하기 위한 방법과 아울러 최근에는 기존과는 다른 방향의 접근 방식을 통해 전송률을 좀 더 낮추려는 연구도 진행되고 있는데 그 중 가능성이 가장 높은 방법이 PWI(Prototype Waveform Innovation)이다[10]. 이것은 음성 신호의 유성음 구간에서는 피치 간격마다 비슷한 형태의 신호가 반복된다는 특징을 이용하여 일정 간격(20-30 msec)마다 원형(prototype)이 되는 신호를 전송하고, 복호화단에서는 보간(interpolation)을 이용하여 합성음을 얻는 방법이다.

이러한 연구는 전송률을 낮추기 위한 방법이 주요 관심사이므로 기존 시스템에 비해 복잡도가 어느 정도 증가되는 것은 허용하고 있다. 그러나, 전송률을 무작정 낮추는 데에는 한계가 있으며, 일정 전송률에서도 상용 시스템에 적용하기에는 성능 면에서 문제가 있다[11].

음성 무편이나 음성 안내, 자동 응답기, 멀티미디어와 같은 음성 저장 시스템에서는 이동 통신 시스템의 경우와는 달리 대역폭 보다는 복잡도를 줄이고, 성능을 향상시키는 것이 주요 관심사가 된다. 따라서, 이러한 연구도 음성 부호화 방법의 한 흐름이 되고 있다.

본 논문에서는 전송률을 half rate로 낮추기 위해 연구되고 있는 대표적인 음성 부호화 방법과 복잡도 및 성능에 중점을 두는 부호화 방법을 설명함으로써 앞으로의 연구 경향에 대해 고찰하고자 한다. 논문 구성은 다음과 같다. 2장에서는 복미, 유립, 일본에서 half rate 표준안을 위해 제안된 방법으로서 표준화 가능성이 가장 높은 세 방법을 설명하고, 3장에서는 4 kbps 이하의 보다 낮은 전송률을 위해 연구되고 있는 PWI 방법에 대해 설명한다. 4장에서는 복잡도가 낮은 부호화 방법의 예를 설명하며, 5장에서 결론을 맺는다.

II. Half-rate 음성 부호화 방법 예

음성 부호화 방법에서 전송률을 낮추기 위한 가장 간단한 방법은 분석 구간의 길이를 길게 하는 것이다. 같은 비트 수를 사용하더라도 분석 구간의 길이가 길어지면 그만큼 전송률도 낮출 수 있기 때문이다. 그러나, 프레임의 길이가 길어지면 장구간 예측 성능도 저하되므로 잔차신호(residual)가 갖는 백색 잡음 특성이 사라지게 되어 음질에 심각한 영향을 초래한다[12][13]. 이를 극복하기 위해 장구간 예측 시에 보간(interpolation)을 통해 지연 시간을 실수 간격으로 구하는 방법[12], 장구간 예측 후의 신호를 여러 대역으로 나눈 후 한번 더 모델링하는 방법[13], 다중 모드 방법[7][8], 백색 가우시안 잡음으로 구성된 통계 코드북 자체를 피치에 의존하도록 적응적으로 구성하여 해결하는 방법 등이 제안되었다[9]. 본 논문에서는 이러한 방법 중 각 지역의 half-rate 표준안 가능성이 가장 높은 세 방법에 대해 알아본다.

2.1 M-LCELP

M-LCELP(Multi-mode Learned CELP)[7]는 복미 half-rate 디지털 셀룰라 표준안의 검토 대상이 되고 있는 방법으로서 그림 1과 같은 부호화 구조를 갖는다.

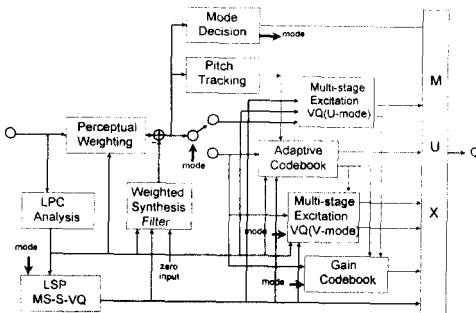


그림 1. M-LCELP 음성 부호화기

입력된 음성 신호는 선형 예측 계수를 이용해 구성된 심리 가중 필터(perceptual weighting filter)에 통과되어 청각 특성이 반영되고, 선형 예측 계수는 양자화 어득을 얻기 위해 LSP(Line Spectrum Pair)로 변환되어 전송된다. 모드 결정 블록에서는 3 개의 임계값으로 구성된 피치 예측 어득을 이용하여 모드-0, 모드-1, 모드-2, 모드-3의 네 가지로 구분하며, 결정된 모드 정보에 따라 각 구간의 여기 신호 생성 방법이 달라진다.

모드 1, 모드-2, 모드-3에서는 간혹비율 2로 갖는 일정 간격의 펄스(regular pulse) 코드북을 2단으로 구성하여 여기 신호로 사용하며, 무성음으로 이루어지는 모드-0에서는 장구간 예측을 하지 않고, 3단으로 구성된 비정기적 펄스 코드북을 사용하여 여기 신호를 생성한다. 코드북을 다단으로 구성함에 따른 계산량과 성능과의 trade-off 관계에 의해 첫번째 단계에서는 몇 개의 후보를 미리 선택하고, 두번째 단계에서는 선택된 코드와의 조합을 통해 최적의 코드북 찾는 방법을 사용한다.

2.2 5.6 kbps VSELP 부호화기

GSM(Global Systems for Mobile Communications)에서는 22.8 kbps로 구성된 기존 채널의 절반(half-rate)에 해당하는 11.4 kbps에서 기존 시스템과 유사한 성능을 유지하는 부호화 알고리즘 표준화 작업이 진행 중에 있다. 11.4 kbps는 여러 결정 방법이 포함된 전송률로서 실제 음성 신호에는 5~7 kbps 정도의 여유만을 갖게 된다.

half rate GSM 표준안으로의 가능성이 가장 높은 방법은 VSELP(Vector Sum Excited Linear Prediction)[14]이다. 이 방법은 그림 2와 같이 미리 정해진 3개의 여기 신호원(L, I, H)으로부터 최적의 여기 신호 벡터와 이득을 결정하고, 구해진 벡터를 피치 합성 필터 및 포먼트 합성 필터에 통과시켜 출력 신호를 발생시킨 후 스펙트럼을 보상하여 최종 합성음을 얻는 방법이다. 복미 full rate 디지털 셀룰라의 표준 전송률은 8 Kbps인데 반해, GSM 표준위원회에 제출된 부호화기는 5.6 Kbps를 채택하고 있다[15]. 5.6 kbps VSELP의 주요 특징은 다음과 같다.

1) 다중 모드 프레임 구분

프레임의 특성에 따라 모드를 0, 1, 2, 3으로 나누고, 각 모드에 따라 장구간 예측 여부와 여기 신호에 할당되는 비트 할당 방법이 다르다.

2) 반사 계수 벡터 양자화

기존 방법에서는 스칼라 양자화를 사용하는 데 반해 비트율을 낮추기 위해 벡터 양자화 방법을 적용한다.

3) 서브샘플(subsample) 지연시간 탐색과 펄스 부호화

유성음 구간으로 판정되는 구간에서는 장구간 예측시에 탐색 영역에 따라 서브 샘플 간격에 대해서도 탐색을 시도한다.

4) 하모닉 스펙트럼 가중 잡음

피치에 해당하는 하모닉 간격에 대해서도 가중치를 반영한다.

5) 적응 프리 필터(prefilter) 및 포스트 필터(postfilter)

주관적 성능을 향상시키기 위해 pole-zero 형태의 필터를 사용한다.

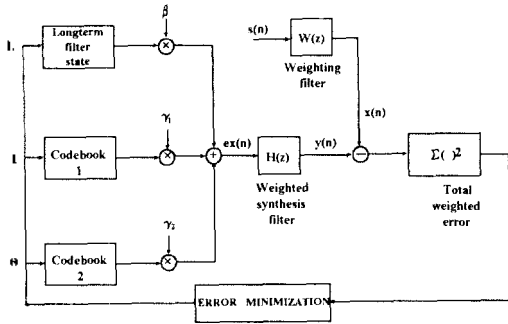


그림 2. 5.6 kbps VSELP 부호화기

2.3 PSI-CELP

PSI-CELP(Pitch Synchronous Innovation CELP)[9]는 일본 half-rate 디지털 셀룰라 표준안으로의 가능성이 가장 큰 부호화 방법으로서 4 kbps 이하에서도 우수한 성능을 보인다고 알려져 있다. 이 방법의 기존 구조는 기존 CELP 부호화기와 같으며, 유성음 구간의 잔차 신호에 존재하는 피치 주기성 신호를 모델링하기 위해 그림 2와 같이 랜덤 코드북을 피치에 따라 적응적으로 변환시켜 음질을 향상시킨 점에만 차이가 있다. 따라서, 코드북을 사용하는 기존 부호화 방법에도 커다란 수정없이 효과적으로 적용될 수 있는 장점을 지닌다.

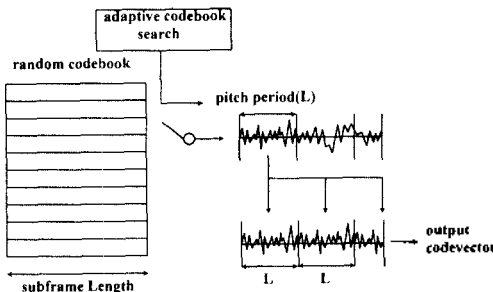


그림 3 PSI-CELP 코드북 생성 방법

III. half-rate 이하의 전송률을 갖는 음성 부호화 방법

이동 통신 시스템에서는 대역폭이 제한되어 있으므로 사용자가 늘어감에 따라 채널 부족 현상은 더욱 심해질 것이다. 이를 해결할 수 있는 한 방법은 음성 부호화를 더욱 낮추는 것이다. 그러나, 4 kbps 이하의 전송률에서는 기존 방법으로는 일정한 성능을 얻기는 어려우므로 다른 접근 방법을 필요로 한다.

PWI(Prototype Waveform Innovation)[10]는 새로운 접근 방식의 부호화 방법으로서 3kbps 이하의 아주 낮은 비트율에서 뛰어난 음질을 갖는다고 알려져 있다. 음성 신호의 대부분 구간을 차지하는 유성음은 서서히 변하는 피치 주기와 파형들의 연결된 형태로 볼 수 있다는 가정하에 그림 4와 같이 음성 신호를 20~30ms 간격에 대해 피치 주기 만큼의 길이로 갖는 원형 파형(prototype waveform)을 구하여 전송하고, 복호화 단계에서는 원형 파형간의 신호를 보간하여 합성음을 얻는다. 이러한 원형 파형은 일반적으로 포먼트(formant) 성분이 제거된 원형 여기 파형(prototype excitation waveform)으로 표현된다. 한 피치 간격의 원형 여기 파형은 벡터 양자화 방법이나 푸리에 급수로 모델링한 후 구한 계수를 양자화하여 전송한다. 여기서 음질에 중요한 영향을 미치는 요소는 피치 주기와 파형들간에 존재하는 단구간 및 장구간 상관 관계 정도에 따른 정확한 주기성 정도를 측정하는 것이다.

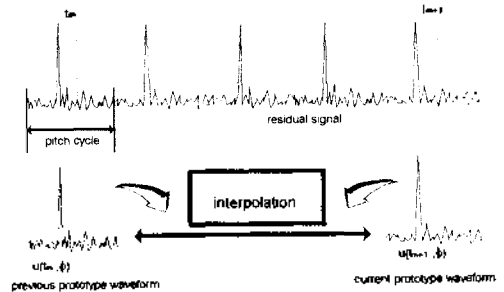


그림 4. PWI 관련 신호

IV 낮은 복잡도를 갖는 음성 부호화기

디지털 셀룰라 경우에는 채널 대역의 한계로 인해 부호화율을 낮추는 것이 필요하지만 음성 무편이나 안내 시스템, 멀티미디어와 같은 음성 저장 시스템에서는 대역폭 보다는 복잡도를 줄이고, 성능을 향상시키는 것이 중요한 관심사이다. 이러한 응용 분야에 적용되는 방법을 보통 저복잡도(low complexity) 부호화기라고 부른다[11].

그림 5는 저 복잡도 부호화기의 하나인 LC-CELP(Low Complexity-CELP)의 복잡도를 나타낸다. 이는 16 kbps 전송률에서 복잡도는 기존 시스템의 1/3~1/6 정도를 유지하며, 주요 특징을 다음과 같다.

1) 고속 LPC 양자화

양자화 시간을 줄이기 위해 일반적으로 스칼라 양자화 방법을 사용한다.

2) 3차 개-회로(open loop) 피치 예측

기존 방법에서는 피치 예측 시에 폐회로(closed loop) 방법을 사용했지만, 시간 상의 이득을 얻기 위해 개회로를 사용한다.

3) 역방향 적응 여기 신호 이득

4) 벡터 차원과 코드북 크기가 작다.

차원이 줄어들면 전송률은 높아지지만, 그만큼 코드북의 크기를 줄일 수 있으므로 연산량에서의 이득을 볼 수 있다.

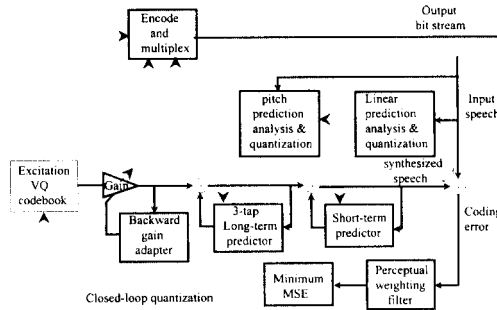


그림 5. LC-CELP 부호화기의 블럭도

V. 결 론

본 논문에서는 half-rate 표준화 가능성이 가장 높은 세 가지 부호화 방법과 복잡도가 아주 낮은 부호화 방법에 대한 설명을 통해 앞으로의 연구 방향을 살펴보았다.

음성 부호화는 전송률, 복잡도, 성능이라는 세 변수가 복잡하게 얽혀 있는 분야로서 모든 조건을 최상으로 유지하기는 어렵다. 낮은 전송률이 요구되는 디지털 셀룰라 분야에서는 기존의 음성 부호화 방법이 효과를 거두기는 어려우므로 음성 합성 연구에서 얻어진 결과를 좀 더 확장함으로써 성능을 개선하거나 PWI와 같이 다소 다른 접근 방식이 필요할 것으로 생각된다.

음성 저장 시스템 등과 같이 비교적 높은 전송률에서 시스템 복잡도를 최소로 하기 위한 분야에서는 주로 16 kbps 내외의 전송률에서 연구가 이루어졌으나 예전의 음성 부호화 연구와 마찬가지로 이러한 분야도 전송률을 더욱 낮추면서도 복잡도는 기존 시스템과 비슷하도록 유지하는 방향으로 연구가 지속될 것으로 생각된다.

참 고 문 헌

[1] L. R. Rabiner, R. W. Shafer, *Digital Processing of Speech Signals*, Prentice-Hall Inc., 1978.
 [2] 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.

[3] P. Kroon, E. F. Deprettere, R. J. Sluyter, "Regular Pulse Excitation : A Novel Approach to Effective And Efficient Multipulse Coding of Speech," *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-34, pp. 1054-1063, Oct. 1986.

[4] B. S. Atal, J. R. Remde, "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp. 614-617, 1982.

[5] M. R. Schroeder, B. S. Atal, "Code Excited Linear Prediction(CELP) High Quality Speech at Very Low Bit Rates," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 25.1.1 25.1.4, 1985.

[6] *Speech and Audio Coding for Wireless and Network Applications*, Kluwer Academic Publisher, 1993.

[7] K. Ozawa, M. Senzawa, T. Miyano and T. Nomura, "M LCELP Speech Coding at 4kbps," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp.269-272, May 1994.

[8] 강 홍구, 서 경태, 이 인성, 차 일환, 윤 대희, "단순화된 다중 모드 방법을 이용한 음성 부호화기" 한국 통신학회 계출, 1996. 3.

[9] S. Miki et al, "A Pitch Synchronous Innovation CELP(PSI CELP) Coder for 2.4 kbit/s," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp.113-116, May 1994.

[10] W. Bastiaan Kleijn, W. Granzow, "Waveform Interpolation in Speech Coding", *Speech and Audio Coding for Wireless and Network Applications*, Kluwer Academic Publisher, 1993.

[11] J. H. Chen, "Toll-Quality 16 kb/s CELP Speech Coding with Very Low Complexity", *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp. 9-12, May 1995.

[12] P. Kroon, B. S. Atal, "On the Use of Pitch Predictors with High Temporal Resolution", *IEEE Trans. on Signal Processing*, Vol. 39, No. 3, pp. 733-735, Mar. 1991.

[13] G. Yang, H. Leich and R. Boite, "Multiband Code Excited Linear Prediction(MBCELP) for Speech Coding," *Signal Processing*, vol. 31, pp.215-227, 1993.

[14] I. A. Gerson, M. A. Jasiuk, "Vector-Sum Excited Linear Prediction(VSELP) Speech Coding at 8Kbps," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp. 641-644, 1990.

[15] I. A. Gerson and M. A. Jasiuk, "A 5600 BPS VSELP Speech Coder Candidate for Half-Rate GSM," EURO-SPEECH '93, 1993.