

이동통신을 위한 음성 부호화 방식

이 황 수
SK 텔레콤 중앙연구원

Speech Coding Algorithms for Mobile Communication

Hwang-Soo Lee
SK Telecom Central R&D Center
hslee@sktelecom.re.kr

요 약

정보통신 문화가 발달함에 따라 디지털 이동통신이나, 멀티미디어, 음성우편 시스템 등 음성을 이용하여 여러 가지 새로운 산업들이 급속히 성장하고 있다. 이 중에서도 특히 디지털 이동통신분야에 대한 연구가 활발한데, 이는 디지털 시스템에서는 부호화기를 사용하여 음성신호를 압축하기 때문에 아날로그 시스템에 비해 채널 증대를 가져올 수 있기 때문이다. 이처럼 음성 부호화기는 실질적인 상업화와 매우 밀접한 관계를 갖고 있기 때문에 그에 대한 연구가 활발히 진행되고 있다.

본 논문에서는 먼저 일반적인 음성 부호화 방법들에 대해 살펴본 다음에, 현재 디지털 셀룰라 시스템에서 사용하고 있는 full-rate 음성 부호화기 및 half-rate 음성 부호화기의 표준화 동향과 최근에 여러 응용분야에서 널리 사용되고 있는 음성 부호화기에 대해서 설명하기로 한다. 또한 ITU-T의 표준화 동향 및 4 kbps 이하의 전송률을 갖는 음성 부호화기의 연구추세에 대해서 살펴보기로 한다.

제 1 장 서론

정보통신 문화가 발달함에 따라 현재 여러 나라에서는 아날로그 이동통신 시스템에서 디지털 이동통신 시스템으로 서비스를 대체하고 있다. 특히 디지털 이동통신 시스템에서는 음성부호화기를 사용하여 spectral efficiency가 증가된다. 이에 따라 음성 부호화 알고리즘에 대한 연구도 매우 활발히 진행되고 있다. 특히 많은 연구가 되고 있는 부호화기는 선형예측을 기본으로 하는 음성 분석-합성(analysis by synthesis) 방식이다. 분석-합성이란 부호화 단에서 선택 가능한 모든 여기신호로부터 합성음을 얻고 이를 원음과 비교하여 오차를 최소로 하는 여기신호를 선택하는 방법이다.

최근 들어서는 현재 디지털 셀룰라에서 사용되고 있는 부호화기에 비해 전송률이 반으로 감소된 음성 부호화기에 대한 표준화 작업을 시행함에 따라 약 4 kbps에서 동작하는 음성 부호화기에 대한 연구도 활발하다. 특히, CELP나 MBE 알고리즘을 기본으로 한 부호화기에 대한 연구결과 등이 많이 발표되고 있다.

일반적으로 음성 부호화의 목적은 고품질의 재생 음질을 유지하면서 가능한 한 낮은 비트율로 음성 신호를 표현하는 데 있다. 이론상으로는 약 2 kbps 정도의

적은 전송률로도 자연스러운 재생음을 만들 수 있다고 하지만, 현재까지 제안된 음성 부호화 방식들은 그 정도의 전송률에서 고음질의 재생이 불가능하다.

음성 부호화는 전송률, 계산량 그리고 음질 사이에 trade-off가 존재 한다. 적은 계산량을 가지고 높은 음질을 얻기 위해서는 64 kbps에서 32 kbps 정도의 전송률이 요구된다. 16 kbps 이하의 전송률로 전송하려면 계산량이 많이 증가하고 음질이 떨어지며, 전송률이 감소함에 따라 부호화 지연이 늘어난다. 그러므로 음성 부호화기를 선택할 때는 어떤 분야에 사용할 지에 따라 음질, 전송률, 지연시간, 계산량 등을 잘 고려하여 선정하여야 한다. 음성 부호화기의 주된 응용분야로는 음성통신, 음성저장, 암호화 등을 들 수 있다.

본 논문에서는 디지털 이동통신에서 사용되고 있는 부호화기에 대한 연구와 표준화 추세에 대하여 살펴보기로 한다. 논문의 구성은 다음과 같다. 2장에서는 파형 부호화, 보코딩, 혼성 부호화법에 대하여 설명하고, 3장에서는 디지털 셀룰라와 ITU-T 표준으로 채택된 음성 부호화기에 대해 살펴보기로 한다. 제 4장에서는 half-rate 음성 부호화기의 표준화 및 연구 동향에 대해 살펴보고, 5장에서는 4 kbps 이하의 전송률에서 동작하는 부호화기에 대한 연구추세에 대해 알아보고 6장에서 결론을 맺기로 한다[1, 3, 6].

제 2 장 음성 부호화 방식의 분류

음성 부호화란 음성 신호를 디지털 부호로 변환시키는 것을 말하며 음성의 전송 또는 저장 등에 주로 사용된다. 음성 부호화의 방법은 크게 파형 부호화(waveform coding), 보코딩(vocoding : voice coding) 그리고 혼성 부호화(hybrid coding)로 분류할 수 있다. 이러한 음성 부호화의 주된 기능은 음성 신호를 분석하여 잉여성분(redundancy)을 제거한 후에 제거되지 않는 나머지 성분들을 심리적으로 타당한 방법을 이용하여 부호화하는 것이다. 다음 세부 절에서는 이 세 가지 부호화 방식에 대하여 간략히 살펴보기로 한다[2, 6].

제 1 절 파형 부호화

파형 부호화의 목적은 수신측에서 재생된 신호가 원래 신호의 파형 모양을 그대로 보존하도록 하는데 있다. 따라서 음성 신호뿐만 아니라 크기가 제한된 다른 입력 신호에 대해서도 동작 가능하다. 이와 같은 이유

초청감연: 이동통신을 위한 음성부호화 방식

때문에 파형 부호화기는 보통 샘플단위로 동작하며, 신호대 잡음비(signal to noise ratio : SNR)와 같은 객관적인 척도로 부호화기의 성능을 측정할 수 있다. 일반적으로 신호 파형의 형태를 유지하기 위한 샘플링율은 나이퀴스트(Nyquist) 샘플링 이론으로 정해져 있기 때문에 파형 부호화법에서는 주로 샘플당 비트의 수를 줄이는 데 연구의 초점을 맞추고 있다. 파형 부호화법에는 pulse code modulation(PCM), delta modulation(DM), adaptive PCM(APCM), differential PCM(DPCM), adaptive DPCM(ADPCM) 등이 있다.

처음으로 널리 사용되기 시작한 표준 음성 부호화기는 1972년에 ITU-T recommendation G.711로 채택된 압축 알고리즘(유럽의 경우 A-law, 북미의 경우 μ -law)을 갖는 64 kbps PCM 부호화기이다. 이 알고리즘은 지금까지도 많은 디지털 시스템에서 사용되고 있다. 12년 뒤인 1984년에는 32 kbps ADPCM이 ITU-T recommendation G.721로 채택되었다. 이 부호화기는 적응 예측기와 양자화기를 이용하여 음성 신호의 잉여성분을 효과적으로 제거함으로써 log-PCM (G.711)에 비해 전송률을 반으로 낮춘 부호화기로 저전송률 음성 부호화기의 음질 평가 시 기준으로 많이 사용된다. 32 kbps 전송률을 갖는 다른 부호화기로 continuous variable slope delta modulation (CVSDM)도 여러 응용분야에서 사용되고 있다.

이러한 파형 부호화는 16 kbps 이상에서 양질의 음성을 보장하지만 그 이하의 전송률에서는 음질의 저하가 심하다는 문제점이 있다. 그러나 계산량이 작기 때문에 아직도 여러 응용분야에서 이용되고 있다. 특히, ADPCM은 digital European cordless telecommunication (DECT), 미국의 personal access communication system (PACS), 일본의 personal handyphone system (PHS)와 같은 통신 시스템등에서 사용되고 있다[3, 4, 6].

제 2 절 보코딩

보코딩은 음성 신호의 생성모델에 근거하여 음성 신호로부터 여기원(excitation source)과 성도(vocal tract) 특성 파라미터를 추출하고, 합성 시에는 이 두 특성 파라미터를 이용하여 다시 원래의 음성을 복원하는 방법이다. 보코딩에서는 신호 파형을 재생해 내는 것이 아니라 사람의 귀로 듣기에 원래 신호와 차이가 없도록 소리(sound)만을 재생한다. 따라서 보코더의 성능 평가에는 SNR이 적합하지 않으므로 주관 척도인 mean opinion score(MOS), diagnostic rhyme test(DRT) 그리고 diagnostic acceptability measure(DAM)를 이용한다. 보코더는 저 전송률 음성 부호화기로 주로 4.8 kbps 이하에서 동작하는 보코더에 대한 연구가 주도적이다. 보코더의 종류로는 linear predictive coding(LPC), homomorphic, channel, formant, phase 보코더 등이 있다.

대표적인 표준 보코더는 1977년에 채택된 U.S. Government LPC 알고리즘(LPC-10)으로 이 보코더는 2.4 kbps에서 동작한다. 이 표준은 상업적인 용용을 목적으로 하고 있는 것이 아니므로 음질은 그다지 좋지않다. 왜냐하면 이 LPC 부호화기의 전달함수는 전극형(all-pole) 모델이어서 실제 음성과 같이 제로 성분이 있는 경우에는 부적당하고 음원이 펄스열(유성음) 또는 백색 불규칙 잡음(무성음)으로 어원화되어 있기 때문에 다양

한 음성의 변화를 표현하지 못하며, 유성음과 무성음의 구별 및 피치의 추출이 정확해야 한다는 어려움이 있다. 따라서 때로는 신뢰성 있는 메시지 전송을 위해서는 혼련된 오퍼레이터(operator)를 필요로 한다.

주파수영역에서 동작하는 대표적인 부호화기로는 포먼트(formant) 보코더가 있다. 포먼트 보코더는 음성 스펙트럼 포락선(spectrum envelope)에서 음성학적 개념의 포먼트 주파수와 대역폭을 이용하여 합성하는 방법으로 여러 공진파로의 조합으로 구성된다. 포먼트 합성에는 직렬합성기와 병렬합성기가 있으며, 각각의 포먼트 필터는 2차 공진 회로로 구성되어 있다. 포먼트 합성기는 음성학적으로 다루기 쉽고 음원 모델이 다양하며 제로의 영향을 고려할 수 있으므로 파라미터들을 잘 조정함으로써 2 kbps에서도 좋은 음질을 얻을 수 있다. 따라서 제한된 단어를 합성하는 비 실시간 시스템으로 응용이 가능하다.

이상의 보코딩 방법들은 대부분 분석 시에 여기원과 성도 필터의 특성이 독립적이라는 가정하에서 선형적으로 처리된 모델이며, 이를 독립적으로 부호화하는 동안 제한된 비트율로 양자화 및 샘플링하기 때문에 이로 인한 음질의 저하는 합성 시에 완전히 복구하기가 어렵다[2, 3, 4, 6].

제 3 절 혼성 부호화

혼성 부호화법은 파형 부호화법과 보코딩법의 장점을 결합시킨 부호화법으로 분석시에는 보코딩법에서 사용하는 음성 생성모델을 그대로 적용하여 성도의 특성을 양자화 및 샘플링하지만 성도의 특성이 제거된 예측 오차신호(여기원 신호라고도 함)를 최대한 그대로 전송 또는 보관하는 방법이다. 따라서 8 kbps 정도의 전송률에서도 우수한 음질을 유지한다. 이러한 혼성 부호화법에는 예측 오차신호를 어떻게 모델링하느냐에 따라 residual excited linear prediction(RELP), multi-pulse LPC(MPLPC), code excited linear prediction(CELP), vector sum excited linear prediction(VSELP), regular pulse excited-long term prediction(RPE-LTP), improved multi-band excitation(IMBE) 등이 있다 [2, 3, 4, 5, 6, 15].

RELP 부호화기에서는 성도 파라미터는 선형예측법으로 구하지만, 예측 오차신호는 파형 부호화한다. 그러나 예측 오차신호를 그대로 파형 부호화하면 최소한 20 kbps의 전송률이 소요되므로 부호화하기 전에 예측 오차신호의 대역폭을 800 Hz로 제한해야 전송률을 낮출 수 있다. 일반적으로는 예측 오차신호를 혼성 압축 DM (hybrid companding DM)법을 이용하여 부호화하지만, 합성음의 음질을 향상시키기 위해서는 2-비트 APCM을 사용한다. 이 부호화기는 9.6 kbps 이하에서는 음질 저하가 심하기 때문에 표준으로 채택된 음성 부호화기 중에는 순수하게 RELP 알고리즘에 기본을 둔 부호화기는 없다.

MPLPC 법은 성도의 특성이 선형예측 분석필터에서 걸러진 예측 오차신호를 여러 개의 펄스열로 모델링하여 부호화하는 방법이다. 부호화기는 기본적으로 합성필터의 분석 블럭과 여기펄스 조사 블럭으로 구성되고, 합성필터의 분석단은 기존의 LPC 법과 같이 성도의 특성 파라미터를 추출한다. 여기펄스를 찾는 블럭에서는 선형 예측계수와 음성 신호로부터 여기 펄스의 위

초청강연: 이동통신을 위한 음성부호화 방식

사용하고 있는데 8 kbps VSELP 인코더의 블록도는 그림 2와 같다. 이 부호화기에서는 한 프레임의 신호에 대하여 covariance matrix를 구한 후에 binomial window를 이용하여 SST (spectral smoothing technique) 처리를 하고 fixed point covariance lattice (FLAT) 알고리즘을 사용한 LPC 분석을 통하여 반사계수를 구한다. 그런 다음 한 프레임을 4개의 서브 프레임으로 분할하여 선형예측 오차신호를 부호화한다. 이 예측오차 신호를 부호화하기 위해서 피치지연 검색을 위한 적응 코드북과 2개의 여기 코드북을 사용한다. 먼저 적응 코드북을 이용하여 피치 지연과 이득을 구한 후에 피치성분이 제거된 잔여신호를 M개의 basis vector를 갖는 2개의 코드북을 이용하여 부호화한다. 이때 두번째 코드북을 검색하기 전에는 첫번째 코드북 검색에 의해 얻어진 이기 벡터의 영향을 분리시키기 위하여 orthogonalization을 하며, 코드북 검색시간을 줄이기 위해서 그레이(gray) 코드를 적용하고 있다[5, 6, 17].

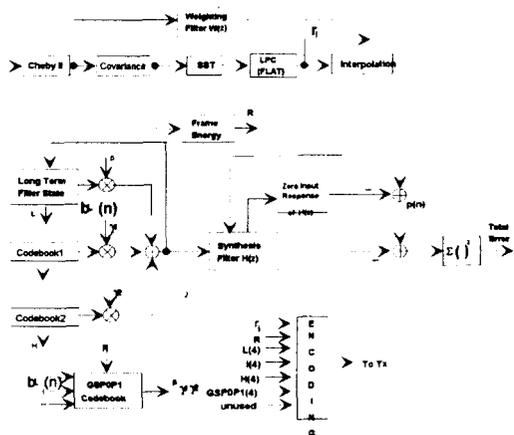


그림 2. VSELP 인코더의 블록도

CDMA를 사용한 디지털 셀룰라에서는 가변전송률 부호화기인 qualcomm CELP (QCELP)를 표준으로 채택하였다. QCELP 부호화기는 기본적으로 AT&T 사의 CELP 알고리즘을 사용하고 있으며, 그림 3은 QCELP 부호화기의 디코더 블록도이다. 즉, 이 부호화기에서는 예측오차 신호를 벡터 양자화하며, 최적의 코드북 값을 찾기 위해서 분석-합성 방법을 이용한다. 기존의 CELP 계열 부호화기와의 차이는 입력 음성 신호의 활성도에 따라 출력 데이터를 조정하는 가변전송률 부호화기라는 점이다. 즉, 활성도에 따라 전송률이 8 kbps, 4 kbps, 2 kbps, 1 kbps 중 하나로 변한다. 이때 전송률이 낮아질수록 피치와 코드북 파라미터의 갱신을 낮추며 1 kbps로 전송할 때는 피치값을 전송하지 않는다. 이 부호화기의 평균 전송률은 약 4 kbps 정도인데 음질은 8 kbps VSELP와 같다고 보고되어 있다[7]. 이 부호화기에서는 LPC 계수를 Durbin 알고리즘으로 구하고 다시 line spectrum pair (LSP) 계수로 변환하여 전송한다. 그리고 한 프레임을 4개의 서브 프레임으로 나누어서 VSELP에서와 마찬가지로 피치 지연과 이득을 구한다. 그런

다음에 피치성분이 제거된 잔여신호를 코드북을 이용하여 부호화한다[1, 6, 7].

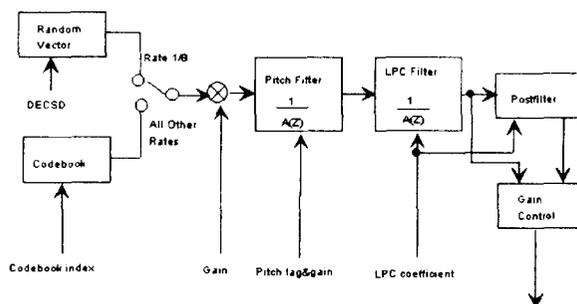


그림 3. QCELP 디코더의 블록도

국내 CDMA 표준으로 채택되었던 8 kbps QCELP는 음질이 그다지 좋지 않아, 대안으로 enhanced variable rate codec(EVRC)가 또 하나의 표준으로 채택되었다. EVRC는 Lucent, Nokia, Motorola, Qualcomm 등 4개 회사가 모여 제안한 보코딩 알고리즘으로, 최대 8 kbps의 전송률을 가지면서도 toll-quality의 음질을 보이는 알고리즘이다. 특히, 잡음 감쇄 기능이 있어, 프레임 에러 및 잡음 환경에서도 좋은 음질을 나타낸다고 보고되어 있다. 기본적인 구조는 Lucent의 relaxed CELP(RCELP) 방식을 따르며, 블록도는 그림 4와 같다. EVRC는 북미의 TIA 표준인 IS-127로 채택되었다.

이름에서 의미하듯이 가변 전송률을 사용하며, QCELP와는 달리 rate 1/4(2 kbps)은 사용하지 않는다. 다른 CELP 계열 부호화기와 마찬가지로, 분석-합성 방법을 사용한다. QCELP처럼 Durbin 알고리즘으로 LPC 계수를 구하고, 이를 LSP 계수로 변환하여 전송한다. 피치는 한 프레임에 한 번만 검색하고, 이를 각 샘플별로 interpolation하여 사용한다. 대수적인 구조의 고정 코드북을 사용하여 저장 용량, 검색의 복잡도, 강인성 등에서 장점을 갖는다[24].

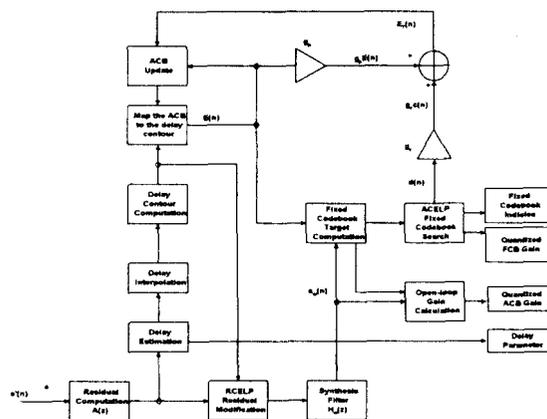


그림 4. EVRC 인코더의 블록도

AT&T사에서 제안한 LD-CELP는 1992년에 ITU-T recommendation G.728로 채택되었다. 이 표준의 최대 응용분야는 public switched telephone network (PSTN) 망이다. 그림 5는 LD-CELP의 인코더 블럭도를 나타내고 있는데, 이 부호화기의 가장 큰 특징은 one-way 지연이 다른 16 kbps 부호화기 보다 매우 짧다는 것이다. 이 부호화기의 알고리즘 지연은 0.625 ms(5 샘플)이며 one-way 지연은 2ms 정도이다. 이 부호화기는 기존의 CELP 부호화기에서 파라미터를 찾기 위해 사용하던 분석-합성 방법은 사용하지만 low delay에서 좋은 음질을 얻기 위해 CELP 부호화기의 다른 부분들은 대부분 수정되었다. 즉, 이 부호화기에서는 역방향 적응 예측기 (backward adaptive predictor)를 사용함으로써 LPC 계수를 전송하지 않아도 되고 LPC 분석 필터의 차수를 50차로 높이고 피치예측 필터를 사용하지 않는다. 이 부호화기는 postfilter를 사용하지 않은 상태에서 32 kbps ADPCM 과 같은 정도의 음질을 보장하고 있다. postfilter를 사용하지 않는 이유는 이 필터가 tandem coding 동안 음질의 저하를 초래할 수 있고, 또한 differential phase shift keying (DPSK) 모뎀 신호와 같은 비음성 신호를 디코딩 할 때 여러 가지 문제를 초래할 수 있기 때문이다 [6, 12].

최근에 ITU-R에서는 차세대 통신망인 IMT-2000의 표준 규격 권고화에 대한 연구가 진행되고 있다. 이와 함께 ITU-T에서는 LD-CELP의 확장 버전으로 16 kbps 이상과 16 kbps 이하의 전송률로 동작하는 음성 부호화기에 대한 연구를 추진하고 있다. 또한 음성대역으로 데이터를 전송하기 위해 16 kbps 이상의 LD-CELP를 확장하려는 표준화 작업이 진행중이다[12, 22].

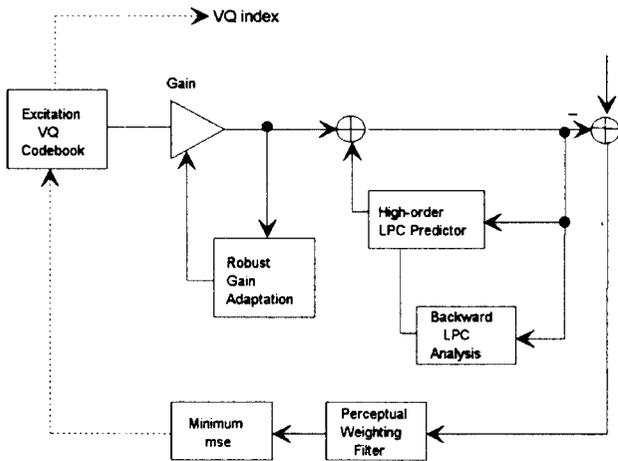


그림 5. LD-CELP 인코더의 블럭도

ITU-T에서는 또한 personal communication system (PCS), IMT-2000, maritime, digital circuit multiplication equipment (DCME) 등에서 사용할 수 있는 8 kbps 음성 부호화기에 대한 표준화 작업을 하였으며 conjugated structure-algebraic CELP (CS-ACELP)를 G.729 draft recommendation으로 선정하였다. CS-ACELP는 일본의 NTT사에서 제안한 CS-CELP와 캐나다의 Sherbrooke 대학과 France telecom.에서 제안한 A-CELP를 결합하여 만

든 것으로 인코더의 블럭도는 그림 6과 같다. 이 부호화기에서는 10 msec를 한 프레임으로 하여 LPC 분석을 한 다음 계수를 다시 LSP 계수로 변환하여 2-단계 벡터 양자화한다. 그리고 일단 open-loop로 피치값을 구한 다음 두 개의 서브 프레임에 대해 앞에서 구한 피치값을 중심으로 closed-loop로 피치를 구한다. 이때 피치값의 resolution은 1/3으로 한다. 첫번째 서브 프레임의 피치값은 그대로 양자화하고 두번째 프레임의 피치값은 첫번째 피치값과의 차이만을 전송한다. 그리고 피치가 제거된 오차신호를 모델링하기 위해서 algebraic 코드북을 사용한다. 표 1에서는 현재 디지털 이동통신에서 사용되고 있는 full rate(8 kbps 이상) 음성 부호화기를 비교하였다.

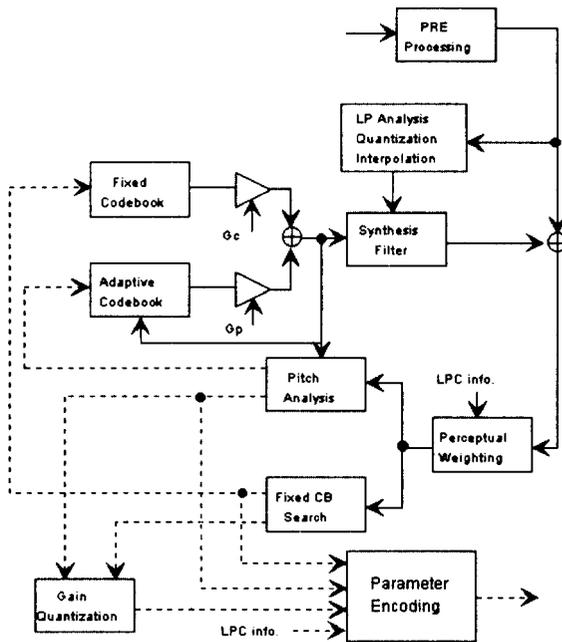


그림 6. CS-ACELP 인코더의 블럭도

표 1. 디지털 이동통신용 음성 부호화기의 비교

알고리즘	RPE-LTP	VSELP	QCELP	EVRC	CS-ACELP
프레임 길이	20 ms	20 ms	20 ms	20 ms	10 ms
전송률	8 kbps	8 kbps	8, 4, 2, 1 kbps	8, 4, 1 kbps	8 kbps
응용 분야	Pan-European digital mobile radio cellular system	North American digital mobile radio system	CDMA	CDMA	PCS IMT-2000 DCME

제 4 장 Half-rate 음성 부호화기의 표준화 및 연구동향

디지털 셀룰라 서비스를 하고 있는 여러 나라에서는 더 많은 spectral efficiency 증가를 위해 기존의 디지털 셀룰라 시스템에 비해 전송률이 반으로 줄어든 (half-rate) 음성 부호화기에 대한 연구가 활발히 진행되고 있다. 전송률을 낮추기 위한 가장 간단한 방법은 분석 프레임의 길이를 길게 하는 것이다. 그러나 프레임 길이가 길어지면 선형 예측기의 성능도 저하되어 예측 오차신호가 더 이상 백색잡음 특성을 가지지 못하므로 음질에 심각한 영향을 초래한다. 따라서 낮은 전송률을 갖는 부호화기에서는 이러한 문제점들을 해결해야 한다.

이 장에서는 half-rate 디지털 셀룰라 음성 부호화기의 표준화 및 연구동향, 그리고 INMARSAT standard M land mobile 통신 시스템(4.15 kbps)과 북미 digital land mobile radio(4.4 kbps) 시스템 등 여러 응용분야에서 표준 부호화기로 채택된 IMBE 부호화기에 대해 간단히 설명하기로 한다.

유럽에서는 디지털 셀룰라 표준 부호화기로 Motorola에서 제안한 5.6 kbps VSELP 부호화기를 선정하였다. 이 시스템의 블록도는 그림 7과 같다. LPC 계수는 보다 적은 비트로 양자화하고 양자화 왜곡을 최소화하기 위해 반사계수를 벡터 양자화 한다. 이때 코드북은 검색에 필요한 계산량과 저장에 필요한 메모리를 줄이기 위하여 세 개의 분할 코드북으로 구성되었으며, 2단계 검색방법을 사용하였다. 이 부호화기는 네 가지 모드로 동작하는 데 모드 0일 경우에는 입력 신호를 무싱음으로 간주하여 7 개의 basis vector로 된 2개의 코드북을 사용하고 모드 1,2,3인 경우에는 적응 코드북과 9 개의 basis vector로 된 코드북을 사용한다. 또한 심리 가중필터(perceptual weighting filter)의 성능 향상을 위하여 10차의 pole-zero 모델로 된 새로운 flexible spectral noise weighting filter를 사용한다. 또한 adaptive spectral postfilter와 adaptive pitch prefilter에 의해 성능을 향상시킨다[10,15,18].

일본에서는 1993년 4월 RCR에 의해 디지털 셀룰라 시스템의 half-rate 표준으로 NTT사에서 제안한 pitch synchronous innovation (PSI)-CELP를 채택하였다. 이 부호화기는 3.45 kbps로 동작하며, 성능이 우수한 것으로 알려져 있으며 블록도는 그림 8과 같다. 이 부호화기는 기본적으로 CELP 알고리즘을 바탕으로 하고 있으며, 여기 벡터를 피치지연과 동기시킴으로써 성능을 향상시킨 것이 특징이다. 한 프레임 길이는 40 ms이며, 네 개의 서브 프레임(서브 프레임당 10ms)으로 나누어진다. 피치, 예측오차 에너지, 이득은 매 서브 프레임마다 양자화된다[15, 23].

북미 디지털 셀룰라 표준으로 검토대상이 되고 있는 부호화기 중 하나는 NEC사에서 제안한 multi-mode learned CELP (M-LCELP)이다. 이 부호기는 그림 9와 같은 구조로 이루어져 있다. 입력 음성 신호는 선형 예측 계수를 이용해서 구성된 심리 가중필터(perceptual weighting filter)에 통과되어 청각특성이 반영되고, 선형 예측 계수는 양자화 이득을 얻기 위해 LSP(line spectrum pair)로 변환되어 전송된다. 모드 결정 블록에서는 3개

의 무게값으로 구성된 피치예측이득을 이용하여 모드 0, 모드 1, 모드 2, 모드 3의 네가지 구조로 구분하며, 결정된 모드 정보에 따라 각 구간의 여기신호 생성방법이 달라진다.

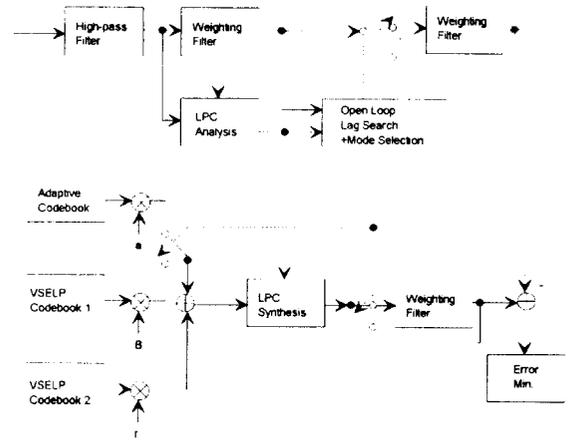


그림 7. 5.6 kbps VSELP 인코더 블록도

모드 1, 모드 2, 모드3, 에서는 간속비를 2로 갖는 일정간격의 펄스 코드북을 2단으로 구성하여 여기신호로 사용하며, 무싱음으로 이어지는 모드 0에서는 장구간 예측을 하지 않고 3단으로 구성된 비정기적 펄스 코드북을 사용하여 여기신호를 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 코드를 찾는 방법을 이용한다[11, 15].

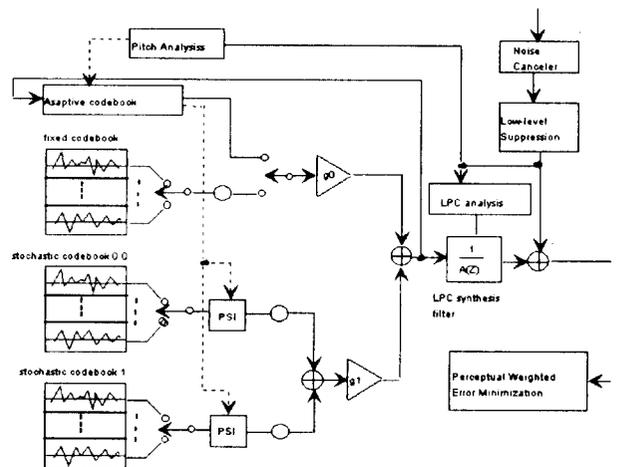


그림 8. PSI-CELP 인코더의 블록도

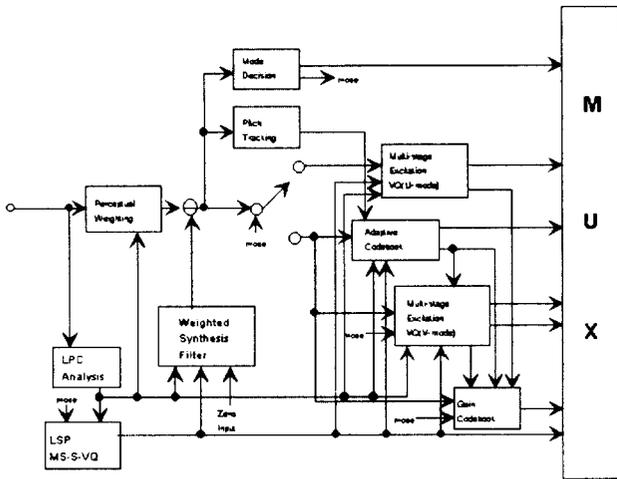


그림 9. M-LCELP 인코더의 블럭도

여러 응용분야에서 연구가 활발한 MBE 부호화기는 처음 제안되었을 당시에는 전송률이 8 kbps였는데, 그 이후 기본적으로는 MBE 알고리즘을 따르면서 전송률을 낮춘 부호화기들이 많이 제안되었으며 이들을 improved MBE(IMBE)라고 부른다. MBE와 IMBE라는 용어가 특별한 구분 없이 혼용해서 사용되기도 한다. 이 부호화기의 특징은 기존의 부호화기와 달리 여기 신호를 모델링하는 데 융통성이 있기 때문에 특히 잡음이 섞인 음성 신호의 음질을 향상시킬 수 있다는 점이다. MBE 알고리즘은 그 이름에서 알 수 있듯이 한 프레임의 음성 스펙트럼을 여러 주파수 대역으로 나누어 각 대역에 대하여 유/무성음 판정을 한 후 그 결과에 따라 여기 신호를 모델링한다. 이때 유성음의 여기신호는 음성 신호에서 구한 기본 주파수(즉, 피치의 역수)를 갖는 임펄스 신호로, 무성음의 여기신호는 랜덤 잡음으로 모델링한다. 그림 10은 IMBE의 인코더를 나타내고 있다. 먼저 입력 신호에 윈도우를 씌운 후에 초기 피치값을 구한다. 그런 다음 fast Fourier transform (FFT) 하여 주파수 영역에서 피치 refinement 과정을 통하여 피치값의 정확성을 높인다. 그리고 원래 음성 스펙트럼과 합성된 스펙트럼과의 정합 정도를 비교하여 유/무성음 판별을 하고 스펙트럼 진폭을 구한다. 이때 스펙트럼의 진폭은 LPC, cepstrum, 또는 스펙트럼의 포락선 정보를 이용하여 구할 수 있다[6, 16].

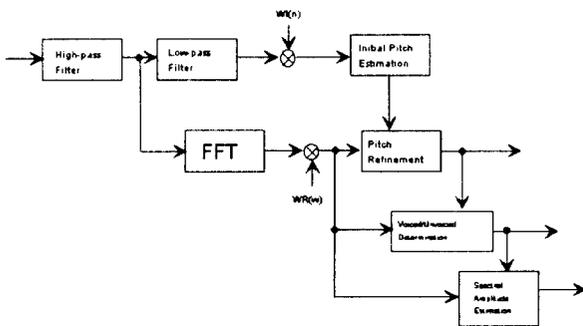


그림 10. IMBE 인코더 블럭도

이상으로 half-rate 음성 부호화기의 연구동향을 살펴 보았다. 현재 ITU-T에서는 8 kbps 전송률을 갖는 부호화기의 표준화 작업에 이어 1993년부터 4 kbps 음성 부호화 알고리즘에 대한 표준화 작업을 진행하고 있다.

제 5 장 4 kbps 이하의 음성 부호화기 표준화 및 연구동향

세계 각국에서는 half-rate 이하의 전송률에서도 어느 정도의 음질을 유지하는 저 전송률 부호화기에 대한 연구를 꾸준히 진행 중에 있으며, 특히 2.4 kbps에서 동작하는 부호화기에 대한 연구에 관심이 모아지고 있다. 또한 1 kbps 이하에서 동작하는 부호화 알고리즘에 대한 연구 결과들도 국제 학술대회를 통해 발표되고 있다.

1994년 3월에 NATO에서는 HF-electronic counter counter measure (HF-ECCM)에서 사용할 수 있는 800 bps 음성 부호화기를 표준으로 채택하였다. 이 부호화 알고리즘은 기존의 NATO 표준 알고리즘인 2.4 kbps LPC-10c를 그대로 수용하면서 양자와 무성음을 다르게 구성하였다. 즉, 이 부호화기에서는 연속적인 세 프레임으로 이루어진 하나의 슈퍼 프레임(superframe)에 대하여 파라미터들을 jointly optimize한다. 이 LPC 계열의 부호화기 중에서 가장 많은 전송률을 차지하는 LPC 계수는 두가지 모드로 부호화한다. 즉, 슈퍼프레임동안 음성 스펙트럼의 변화가 적으면 평균 스펙트럼을 양자화하고, 변화가 심한 경우에는 정확한 필터 계수보다는 스펙트럼의 변화를 강조하여 부호화한다. 이 부호화기의 DRT 점수는 약 87점 정도이다[14].

미국의 defense digital voice processing consortium 부(DDVPC)에서는 2.4 kbps에서 동작하는 새로운 부호화기에 대한 표준화 작업을 하고 있는데, 이 작업은 1993년 3월부터 시작하여 1994년에는 8개의 후보 부호화기들이 테스트를 받았다. 이들 중에서 mixed excitation linear prediction(MELP) 알고리즘이 표준으로 채택되었다[13].

4 kbps이하의 전송률을 갖는 음성 부호화기에 대한 연구에서는 LPC와 MBE 알고리즘을 기반으로 하는 부호화기에 대한 연구가 활발하다. 기본적으로 LPC 알고리즘을 사용한 부호화기는 전송률을 상당히 낮출 수 있지만 음질이 좋지 못하다. 따라서 전송률을 약간 높이면서 선형예측 오차신호의 모델링을 좀 더 융통성 있게 하는 방법들이 많이 제안되고 있다. MBE 알고리즘을 이용한 음성 부호화에서는 음성 신호의 스펙트럼 정보를 벡터 양자화 하거나 LPC 모델링을 하여 전송률을 낮추고 또한 유/무성음 정보를 주파수 대역마다 보내는 것이 아니라 유/무성 혼합함수를 사용하여 유성음에서 무성음으로 혹은 그 반대로 천이되는 주파수만을 전송한다든지 하는 여러 가지 방법으로 전송률을 낮춘다. 현재는 1.2 kbps의 전송률을 가지면서 좋은 성능(DRT 점수가 91.5)을 보이는 부호화기도 제안되었다. 참고로 코딩되지 않은 음성 신호의 경우에는 DRT 점수가 97.0이다.

최근에는 또한 prototype waveform interpolation(PWI) 방법을 사용한 음성 부호화에 대한 연구가 활발한데, 이 방법은 3 kbps 이하에서도 우수한 음질을 갖는 것으로 알려져 있다. 이 방법은 음성 신호의 대부분을 차지

하는 유성음은 서서히 변하는 피치주기 파형들의 연결된 형태로 볼 수 있다는 가정하에서 음성 신호를 20-30 msec 간격에 대해 피치 주기만큼의 길이를 갖는 원형 파형을 구성하여 전송하고 수신단에서는 원형 파형간의 신호를 보간하여 합성음을 얻는 방식이다[15].

제 6 장 결론

정보통신 문화가 발달함에 따라 디지털 이동통신이나, 멀티미디어, 음성우편 시스템 등 음성을 이용한 여러 가지 새로운 산업들이 급속히 성장하고 있다. 이 중에서도 특히 디지털 이동통신분야에 대한 연구가 활발한데, 이는 계속되는 이동통신 가입자를 충분히 수용하기 위해서는 무엇보다도 spectral efficiency를 높여야 한다는 필요성에 의한 것이다. 즉, 디지털 이동통신 시스템에서는 음성 부호화기를 사용하여 아날로그 이동통신 시스템에 비해 TDMA 전송방식을 사용할 경우 약 3배, CDMA 전송방식의 경우에는 약 15배의 채널 용량을 가져올 수 있기 때문이다. 이렇게 음성 부호화기는 실제적으로 사용되는 시스템과 밀접한 관계를 갖고 세계 여러 나라에서 연구가 매우 활발히 진행되고 있으며 우수한 결과들이 계속 발표되고 있다. 본 논문에서는 현재 디지털 셀룰라 시스템에서 사용되고 있는 8 kbps 정도의 full-rate 표준 부호화기 및 4 kbps 정도의 half-rate 음성 부호화기의 표준화 동향에 대하여 알아보았다. 또한 ITU-T에서의 표준화 및 4 kbps 이하의 전송률을 갖는 음성 부호화기에 대한 연구 동향에 대하여 살펴본 것이다.

결론적으로 음성 부호화기에 대한 최근의 연구방향은 크게 두 가지로 나누어 진다. 즉, 어느 정도의 음질을 보장하는 한도 내에서 전송률을 최대한 낮추는 저 전송률 부호화기와, 낮은 전송률보다는 전송률을 조금 더 높이면서 적은 계산량으로 실시간 구현이 용이하도록 하는 음성 부호화기에 대한 연구분야로 분류된다.

현재 여러 나라에서 이미 디지털 이동통신 서비스를 실시하고 있는데, 이동통신 가입자의 계속되는 증가로 인해 현재 사용하고 있는 부호화기보다 전송률이 절반인 half-rate 음성 부호화기에 대해서도 표준안을 결정하였거나 한창 진행 중에 있다. 특히, 군용으로 사용하기 위하여 4 kbps 이하의 전송률을 갖는 음성 부호화기에 대한 연구도 계속 진행되고 있다. 더불어 차세대 이동통신인 IMT-2000에 대한 연구도 활발히 진행되고 있다. 저 전송률로 내려 갈수록 음질의 저하나 계산량의 증가 혹은 지연시간 증가 등의 문제들이 발생할 수 있는데, 이러한 문제점들을 해결하기 위해 다양한 각도에서 많은 연구들이 필요하다.

이렇게 같은 음질을 유지하면서 전송률을 낮추다 보면 복잡도가 높아지게 되는데, 그렇게 되면 실시간 구현이 어려워진다. 음성 우편, 멀티미디어, 인터넷 폰 등과 같이 고품질의 합성음을 필요로 하는 응용분야에서는 전송률을 약간 더 높더라도 계산량을 낮추어 실시간 구현이 가능하도록 하는 것이 더 중요한데, 이러한 응용분야에 적용되는 음성 부호화기를 보통 저 복잡도(low complexity) 부호화기라 한다. 최근에는 13 kbps나 16 kbps에서 동작하는 저 복잡도 고품질 CELP 부호화기에 대한 연구결과들도 발표되고 있다. 특히, 음질 향상을 위해서는 인간의 청각특성에 중점을 두어

perceptual waiting filter나 noise masking 효과 등에 대한 많은 연구도 진행되어야 할 것이다.

현재 우리 나라에서는 1996년도 상반기에 세계 최초로 CDMA 전송방식을 채택하여 디지털 이동통신을 서비스 하고 있다. CDMA 방식에서는 가변전송률 음성 부호화기를 사용하고 있으므로 저 전송률, 저 복잡도 음성 부호화기와 더불어 가변전송률 음성 부호화기에 대한 많은 연구가 있어야 할 것이다. 또한 차세대 통신망인 IMT-2000을 비롯한 이동통신망에 사용하기 위한 음성 부호화기로 ITU-T에서는 CS-ACELP를 draft recommendation으로 정하였는데, 이 부호화기는 8 kbps fixed rate로 동작하는 부호화기이다. 따라서 IMT-2000의 효율적인 서비스를 위해 저 복잡도, 가변전송률 음성 부호화기에 대하여 지속적인 연구가 있어야 할 것이다.

[참고문헌]

- [1] 차명희, "디지털 음성 지장의 응용을 위한 CELP 음성 부호화기의 효율적 구현에 관한 연구," KAIST, Feb. 1991.
- [2] C. Wheddon and R. Linggard, *Speech and Language Processing*, Chapman and Hall, 1990.
- [3] Panos E. Papamichalis, *Practical Approaches to Speech Coding*, Prentice-Hall, Inc. Englewood Cliffs, New Jersey.
- [4] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
- [5] Bishnu S. Atal and Vladimir Cuperman, & Allen Gersho, *Advances in Speech Coding*, Kluwer Academic Publishers, 1991.
- [6] A. M. Kondoz, *Digital Speech (coding for low bit rate communication systems)*, John Wiley & Sons Ltd., 1994.
- [7] Bishnu S. Atal, Vladimir Cuperman, Allen Gersho, *Speech and Audio Coding for Wireless and Network Applications*, Kluwer Academic Publishers, 1993.
- [8] P. Kroon, E. F. Deprettere and R. J. Sluyter, "Regular-Pulse Excitation-A Novel Approach to Effective and Efficient Multipulse Coding of Speech," *IEEE, Trans. On Acoust. Speech and Signal Proc.*, pp.1054-1063, Oct., 1986.
- [9] ED F. Deprettere and Peter Kroon, "Regular Excitation Reduction for Effective and Efficient LP-Coding of Speech," *ICASSP'85*, pp. 25.8.1-25.8.4, 1985.
- [10] Tim Fingscheidt, Thomas Wicchers and Eckhard Delfs, "Implementation Aspects of the GSM Half-rate Speech Codec," *EUROSPEECH'95*, pp.723-726, 1995.
- [11] Kazunori Ozawa, Mashiho Scrizawa, Toshiaki Miyano and Toshiyuki Nomura, "M-LCELP Speech Coding at 4 kbps," *ICASSP'94*, pp. 269-272, 1994.
- [12] Craig R. Watkins and Juin-Hwey Chen, "Improving 16KB/S LD-CELP Speech Coder for Frame Erasure Channels," *ICASSP'95*, pp.241-244, 1995.
- [13] M.A.Kohler, L.M.Supplee and T.E.Tremain, "Progress Toward a New Government Standard 2400 BPS Voice Coder," *ICASSP'95*, pp.488-491, 1995.
- [14] B.Mouy, P.de La Noue, G.Goudezeune, "NATO STANAG 4479: A Standard for an 800 BPS Vocoder and Channel Coding in HF-ECCM System," *ICASSP'95*, pp.480-483, 1995.
- [15] 강홍구, 윤대화, "낮은 전송률 음성 부호화 연구 동향," *음성통신 및 신호처리 워크샵 논문집*, pp.27-30, 1995.
- [16] Daniel W. Griffin and JAE S. LIM, "Multiband Excitation Vocoder," *IEEE Trans. on Acoustics, Speech and Signal Proc.*

VOL.36, NO.8, Aug., 1988.

[17] "Principles of Vector-Sum Excited Linear Predictive (VSELP) Speech Coder and its Implementation on the DSP56156," Motorola, Inc., Digital Processor Operations, Austin, Texas.

[18] I.A.Gerson and M.A.Jasiuk, "A 5600 bps VSELP Speech Coder Candidate for Half rate GSM," EUROSPEECH'93, 1993.

[19] R.Salami, C.Laflamme and J.P.Adoul, "8kbit/s ACELP coding of speech with 10ms speech frame: A candidate for CCITT standardization", ICASSP'94, pp. II-97-II-100.

[20] Akitoshi Kataoka, Takehiro Moriya and Shinji Hayashi, "Implementation and Performance of an 8-kbit/s Conjugate Structure CELP Speech Coder", ICASSP'94., pp.II-93-96.

[21] Akitoshi Kataoka, Takehiro Moriya and Shinji Hayashi, "An 8-kbit/s Speech Coder Based on Conjugate Structure CELP", ICASSP'93, pp. II-592-595.

[22] 김 홍국, 김 삼룡, "PCS를 위한 음성코딩방식과 음질비교", 전자공학회지 22권 9호, pp.1060-1066., 1995.

[23] T. Ohya, H. Suda and T.Miki, "5.6 kbits/s PSI-CELP of the Half-rate PDC Speech Coding Standard", IEEE VTC, pp.1680-1684, 1994.

[24] TIA/EIA/IS-127 Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems.