

디지털 음성코우딩

李 商 郁 *

■ 차 례 ■

- 1. 서 론
- 2. 음 질
- 3. 음성코우딩 방식
- 4. DSP의 응용

1. 서론

우리 인간의 음성에 관한 연구는 오래전부터 시작되어 왔으나, 컴퓨터 및 집적회로 특히 VLSI의 눈부신 발달에 힘입어 최근 음성처리(speech processing)에 대한 일반 관심도가 높아지고 있다. 음성처리에 관한 연구분야는 매우 광범위하나, 보통 음성코우딩(speech coding) 그리고 음성 및 화자인식(speech/speaker recognition)으로 나눌 수 있다. 그러나 본 글에서는 음성코우딩에 대하여 기술하려고 한다. 우리 일상 생활에서 음성통신의 중요성은 재삼 재론할 필요가 없다. 통신에 있어 전송신호를 애널로그 형태보다는 디지털로 보내면 많은 잇점이 있다. 그러나 디지털 신호로 보내면 전송 주파수대역이 증가되므로 여기에 음성코우딩의 중요성이 있다. 음성통신에서 광범위하게 이용되고 있는 mobile radio를 예를 들면, 현재 거의 대부분의 mobile radio가 애널로그 방식을 사용하고 있다. 그러나 mobile radio 사용자 간의 통화내용보장(privacy 혹은 security) 및 다중통신(multiplex communication)을 위하여 최근 디지털 방식을 사용한 mobile radio의 개발이 진행중에 있다.¹⁾

그러나 우리에게 잘 알려진 PCM(pulse-code modulation) 방식을 사용하여 원음과 거의 같게 전

송하려면, 입력음을 8000 Hz로 샘플링 해서 8-bit로 양자화(quantization)해야 하므로²⁾, bit-rate는 64 Kbps가 된다. 보통 1-bit의 데이터를 보낼때 1-Hz의 주파수대역이 필요하므로, PCM 방식으로는 64 KHz의 주파수대역이 요구된다. 이것은 현재 사용되고 있는 FM의 16 KHz, AM의 6 KHz, 그리고 SSB(single side band)의 3 KHz 주파수대역과 비교할 때 큰 차이가 있다. 또 사회가 복잡함에 따라 자연히 요구되는 것이 통신설비의 증대이다. 그러므로 제한된 채널에서 더욱 더 많은 양의 정보를 보내려고 하는 문제가 대두 되었다. 따라서 음성코우딩의 목적은 다음과 같이 말할 수 있다. 어떻게 하면 많은 양의 음성을, 거의 원음과 같게, 경제적으로 보낼 수 있느냐 하는 것이다. 많은 양의 음성을 보낸다는 것은 결국 전송 bit-rate를 줄이는 것을 의미한다. 이와 같은 이유로 음성코우딩을 한편에서는 음성데이터 compression 이라고 부르고 있다.

본 글에서는 음성코우딩 할 때 제일 고려되어야 할 음질(speech quality)에 대해 2장에서 설명하고, 현재 많이 사용되고 있는 대표적인 음성코우딩 방식에 관해 3장에 설명 하려고 한다. 4 장에서는 3장에서 기술한 각 음성코우딩 방식을 실시간(real-time)에서 동작시키기 위해서 현재 많이 사용하고 있는 DSP(digital-signal-processor)에 대해 설명하려고 한다.

*正 會 員 : 서울대 工大 制御計測工学科 助教授 · 工博

2. 음질

음성코우딩에서 제일 중요한 문제가 코우딩 되어서 전송된 음성의 음질이다. 일반적으로 음질을 설명하는데 많이 인용되는 몇가지를 표1에 보였다.

표1. 대표적으로 인용되는 음질

음 질	주파수 대역	S N R	음 용 예
High - Fidelity	20~40 KHz	60 dB	HI -FI stereo 시스템
Program	5KHz	50 dB	AM 방송
Toll	3 KHz	30 dB	양호한 상태의 장거리 통화
Communication	3 KHz	20 dB	CB Radio
Synthetic	-	-	장난감 혹은 음성음 담장치

고감도음질(high-fidelity quality)은 고급가전 Hi-Fi 음향기에서 요구되는 음질인데, 가청주파수 상계선인 20KHz 또는 그 이상의 주파수대역을 필요로 한다. Program quality는 일반적으로 사용되고 있는 AM방송의 음질을 의미한다. 참고적으로 FM방송음질은 high-fidelity와 program 음질의 중간이라고 알려져 있다. Toll-quality는 양호한 상태의 장거리전화에 사용되는 음질로써, 보통 음성코우딩에서는 toll-quality를 목표로 한다. 좀더 부연하면, toll-quality는 주파수대역 폭이 20~3200Hz, 고주파 찌그러짐(harmonic distortion)이 2~3%미만, 그리고 SNR이 30dB 이상인 애널로그 음성에 대응하는 음질이다.³⁾ CB나 mobile radio의 음질은 communication-quality라고 부르는데, toll-quality와 비교하면 통화자가 여러가지 형태의 잡음 및 찌그러짐을 들을 수 있다. Synthetic-quality는 교육용 장난감(Texas Instrument의 Speak and Spell이 좋은 예)이나 공상과학영화에서 로봇이나 컴퓨터가 말할 때 흔히 들을 수 있는 음질을 일컫는데 억양의 고저가 없고 화자인식도가 매우 낮다.

그러나 인간이 음질을 판단할 때에는 청각기간인 귀에 의존할 수 밖에 없으므로, 청각에 의한 음질

판단(subjective listening test)은 주관적인 의미를 가지고 있다. 왜냐하면 개인의 선호도에 따라 음질에 대한 의견이 각각 틀리기 때문이다. 많은 사람들이 음성이나 영상(image)을 주관적으로 판단하는 대신, 객관적인 입장에서 사용할 수 있는 판단계수(figure of merit)를 정의하려고 많은 노력을 해왔으나 아직도 해결되지 않고있는 실정이다.^{3),4)} 통신시스템 해석에서 널리 사용되고 있는 SNR은 청각에 의한 음질판단시험에서는 그렇게 큰 의미를 가지지 못한다. 실제 많은 사람들이 낮은 SNR의 음성을 높은 SNR의 음성보다 청각적으로 더 우수하다는 판단을 내리는 경우를 많이 보고 있다. 그러나 음질을 비교판단할 때에는 다음과 같은 사항은 반드시 고려할 필요가 있다.

1) 이해도

이해도는 전송된 단어나 문장을 통화자(listner)가 이해하는 능력을 나타내는 척도이다. 이해도는 음성통신에서 제일 중요한 요소이다. 이해도를 측정하는 데는 여러가지 방법이 있으나 널리 사용되고 있는 것이 DRT(diagnostic rhyme test)이다.⁵⁾ DRT는 언어학적으로 발음들이 비슷한 단어들을 전송시켜 통화자들로 하여금 구별하게 해서 이해도를 측정한다. 보통 DRT에서는 3~6명의 통화자들로 하여금 같은 단어들을 듣게해서, 얻은 데이터를 통계학적으로 처리하므로써 이해도를 측정한다.

2) 인식도

인식에는 화자(speaker) 및 음성(speech) 인식이 있으나, 음성코우딩에서 요구하는 것은 대부분 화자인식이다. 인식도란 우리가 누구와 통화할 때, 통화가자가 만일 아는 사람이면, 그가 누구인지를 음성을 듣고 알아내는 능력을 말한다. 인식도는 특히 특수목적용통신에서는 빼 놓을 수 없는 사항이다. 표1에 보인 각 음질중 synthetic-quality를 제외한 다른 음질들은 모두 높은 화자인식도를 가지고 있다.

3) 자연성

우리 인간의 음성은 감정 및 억양의 고저에 의해 듣기에 자연스럽다. 그러므로 음질을 판단할 때에는 얼마나 우리에게 자연스럽게 들리느냐, 즉 음성의 자연성을 고려할 필요가 있다. 자연성은 인식도와 밀접한 관계가 있다. 자연성을 측정하는데는 DAM(diagnostic acceptability measure)방식이 널리 사용되고 있다.⁶⁾ 음성에서 자연성이 결여되어 있

면 마치 기계가 말하는 것처럼 들려 거북스럽다. Synthetic-quality가 그 좋은 예이다.

4) 잡음 및 찌그러짐

좋은 음질이란 말할 것도 없이 잡음 및 찌그러짐이 적어야 한다. 그러나 잡음 및 찌그러짐을 수량화하여 음질을 판단하려고 하는 것은 많은 문제점이 있다.

5) 강인성

강인성이란 코우딩된 음성이 채널을 통해 전송될 때, 여러가지 원하지 않는 영향, 즉 간섭(interference), 잡음 또는 찌그러짐이 어떻게 디코우딩된 음성에 나타나느냐 하는 특성을 의미한다. 강인도가 좋다는 말은 위에서 말한 여러가지 영향을 적게 받는다는 뜻이다. 보통 강인성은 전송 bit-rate에 비례한다고 믿어진다.

음성코우딩은 위의 모든 사항을 고려하는 것이 좋으나, 때때로 응용목적에 따라 몇가지는 생략될 수 있다. 예를들면 음성응답장치(voice answering system)에서는 이해도는 필수적이거나 인식도는 그렇게 요구되지 않는다.

3. 음성코우딩 방식

음성코우딩에는 음성파형을 A/D 변환하여 디지털 형태로 코우딩하는 파형코우딩(waveform coding)과 인간의 음성발생 과정을 모델링하여 그모델의 파라미터를 코우딩하는 source 코우딩 방법이 있다. 파형코우딩에는 PCM⁷⁾, ADPCM(adaptive differential PCM)⁸⁾, ADM(adaptive delta modulation)⁹⁾, SBC(sub-band-coding)¹⁰⁾, TDHS(time-domain harmonic-scaling)¹¹⁾, tree-coding¹²⁾ 등이 있고 source 코우딩에는 여러가지 형태의 vocoder가 이에 해당되는데 LPC(linear predictive coding)¹³⁾가 대표적인 예다. 또 코우딩하는 영역에 따라 주파수영역 또는 시간영역 코우딩으로 나눌 수 있는데, SBC나 TC(transform coding)가 주파수영역코우딩 예이고 대부분의 파형코우딩은 시간영역 코우딩이다. 또한 파형과 source 코우딩을 합친 hybrid 코우딩이 있는데, APC(adaptive predictive coding)¹⁴⁾ 또는 RELP(residual excited linear prediction)¹⁵⁾ 코우딩 방식이 좋은 예이다. 한편으로는 코우딩된 신호의 전송 bit-rate에 따라 광대역(wide-band), 중대역(medium

band), 또는 협대역(narrow-band) 코우딩으로 나누는데 그림 1에 이 관계를 보였다. 표 2는 대표적인 코우딩 방식에 의해 toll 및 communication quality를 얻을 수 있는 최소 bit-rate를 도시하였다.

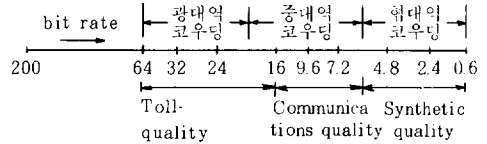


그림 1. 음성코우딩에 있어 bit-rate 및 음질의 관계

표 2. 대표적인 코우딩의 bit-rate

코우딩 방식	요구되는 최소 bit-rate (Kbps)		
	Toll	Communication	Synthetic
μ -Law PCM	60	36	N/A
ADM	40	24	N/A
ADPCM	32	16	N/A
SBC	24	9.6	N/A
APC	16	9.6	N/A
RELP	N/A	7.2	N/A
LPC Vocoder	N/A	N/A	0.8~2.4

3.1. 파형코우딩

3.1.1 시간영역 코우딩

3.1.1.1 DPCM(Differential Pulse Code Modulation)

음성을 Nyquist 샘플링 주파수 8000 Hz로 샘플링하면 샘플된 신호, x_i 들은 불규칙신호(random process)적인 면에서 보면 인접한 샘플간에 상관계수(correlation coefficient)가 대단히 높다. 그러므로 x_i 는 바로전 샘플값 x_{i-1} 로 예측할 수 있다. 또 x_i 와 x_{i-1} 의 차(differential)신호의 variance는 원신호 x_i 의 variance보다 적으므로, DPCM에서는 x_i 를 양자화 하는 대신, 차신호($x_i - x_{i-1}$)를 양자화 해서 전송한다. 그러나 실질적으로는 양자기(quantizer)의 step-size를 입력신호 크기에 비례해서 변환시키는 adaptive DPCM(ADPCM) 방식을 많이 사용한다. ADPCM의 구성도를 그림 2에 보였다.

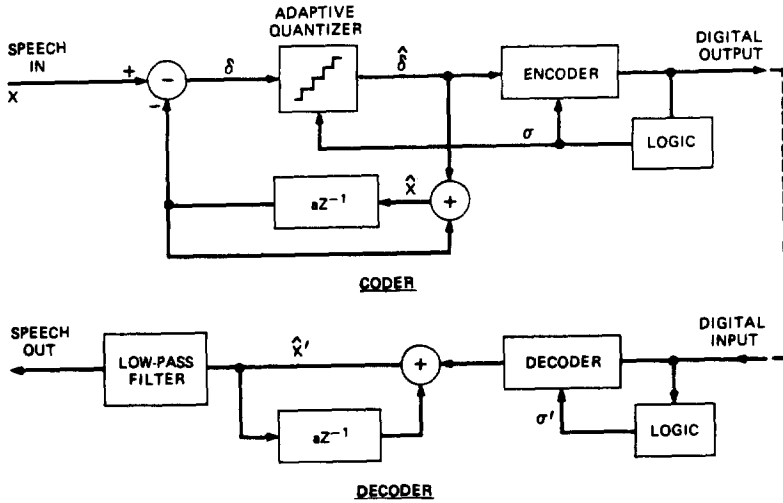


그림 2. ADPCM 구성도

DPCM은 하아드웨어의 구성이 간단하고, 강인성이 좋아 음성 및 영상 코우딩에 넓게 응용되고 있으나, 32Kbps의 비교적 높은 bit-rate에서 toll-quality가 가능하다.

3.1.1.2 DM (Delta Modulation)

DM은 지금까지 알려진 코우딩 방식중 가장 간단할 뿐만 아니라 잡음 및 채널 error에 대한 강인성이 아주 우수하다. DM은 1-bit DPCM이므로 샘플링 주파수가 곧 bit-rate이다. 그러므로 DM은 Nyquist 주파수보다 훨씬 높은 샘플링 주파수가 필요하다. DM에서도 ADPCM과 같이 양자기의 step-size가 입력신호의 크기에 따라 비례하는 adaptive DM (ADM)이 사용되고 있다.⁹⁾ Adaption 하는 방식에 따라 여러가지 방법의 ADM이 소개되어 있다. 그림 3에 ADM 구성도의 한예를 보였다.

ADM중 가장 널리 사용되고 있는 것이 CVSD (continuous variable slope delta) 방식이다. 그 특징은 step-size가 입력크기에 따라 순간순간 변하는 것이 아니라, 음절의 변화 (syllabic-rate)에 따라 천천히 적응한다. CVSD 방식은 Motorola나 Harris 같은 반도체회사에서 채택, 1-chip CVSD 집적회로를 개발 판매 중에 있다. 그러나 DM도 역시 32~40Kbps의 높은 bit-rate에서 toll-quality가 가능하므로 낮은 bit-rate를 요구하는 곳에서는 사용이 용이하지 못하다.

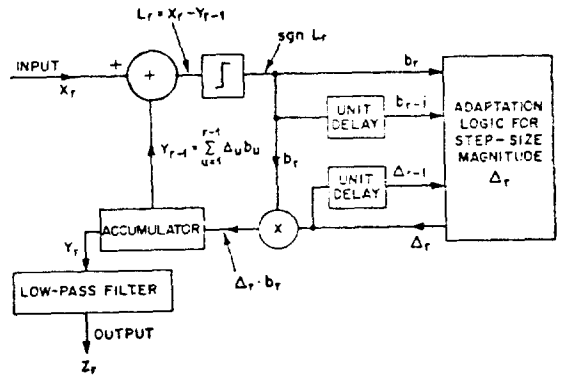


그림 3. ADM 구성도의 예

3.1.2 주파수영역 코우딩

주파수영역 코우딩에는 음성의 주파수성분을 여러개의 대역필터로 분리하는 SBC와 적절한 변환 (transformation) 을 이용하여 시간영역의 신호를 주파수영역의 신호로 변환하여 코우딩하는 TC가 있다¹⁶⁾.

3.1.2.1 SBC (Sub-Band-Coding)

SBC는 그림 4에 그 구성도를 도시한바와 같이 4~5개의 대역필터로 주파수성분을 분리한 후, 각각 분리된 신호의 대역폭에 맞게 다시 샘플링 (down-sampling 또는 decimation) 해서 코우딩한다. 대역필터를 통과한 각 band들의 신호들은 각샘플간의 상관관계수가 거의 없는 백색잡음적인 성질을 가지므

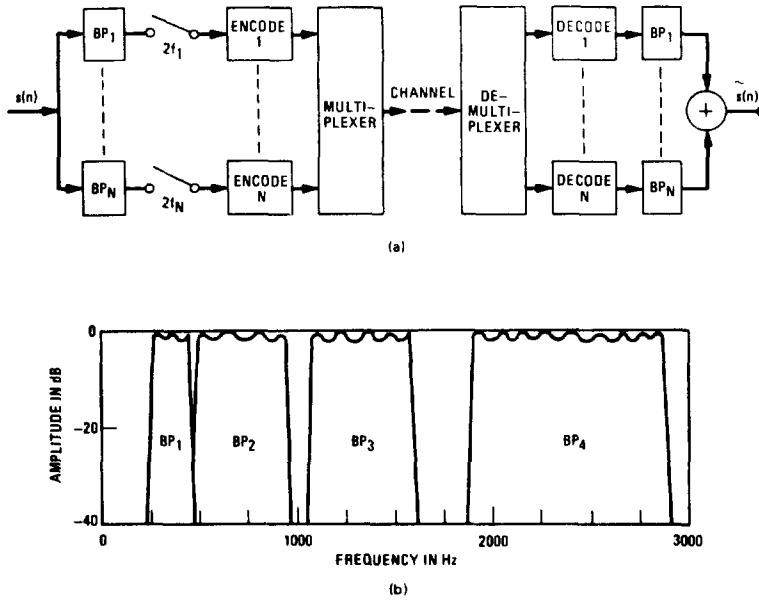


그림 4. a) SBC 구성도 b) SBC에서 사용한 필터의 주파수 특성

로 ADPCM 보다 APCM 이 효율적인 코우딩 방식이다. 음성에 있어서 이해도나 인식에 있어 중요한 역할을 하는 주파수 성분 들은 대개 1000 HZ 보다 낮은 성분들이다. 그러므로 각 band 들을 코우딩 할때는 저역주파수 성분들은 고역주파수 성분들 보다 좀더 치중하여 세밀하게 코우딩 한다. 이렇게 하므로써 저역주파수 신호들의 양자잡음을 줄이므로 같은 bit - rate 로써 ADPCM, DM 보다는 전반적인 음질향상을 얻을 수 있다. SBC에서 가장 중요한 부분은 필터이다. 만일 각 인접한 필터간의 겹치는 부분이 넓어지면, 양자잡음 이외에 aliasing 잡음이 생겨, 음질이 좋지 못하게 되므로, 필터 설계시 세심한 주의가 필요하다. 또 한가지 고려되어야 할 문제는 band 분할 방식인데, 최적의 방법은 각 band 의 AI (articulation index) 가 서로 균등하게 되도록 한다.¹⁰⁾ 그러나 AI 에 의한 분할방법은 복잡한 변조가 필요하므로, 보통 정수배샘플링 (integer sampling) 에 의한 분할방법이 널리 이용된다. SBC 는 ADPCM과 비교하여 적어도 10 kbps 의 코우딩 이득이 있어, 12~16 kbps 로써 쉽게 communication-quality를 얻을 수 있다.

최근 switch capacitir 필터의 소개로 저렴한 필터의 설계 및 제작이 용이함에 따라 SBC 를 사용한 mobile radio 의 개발도 고려되고 있다.¹⁷⁾

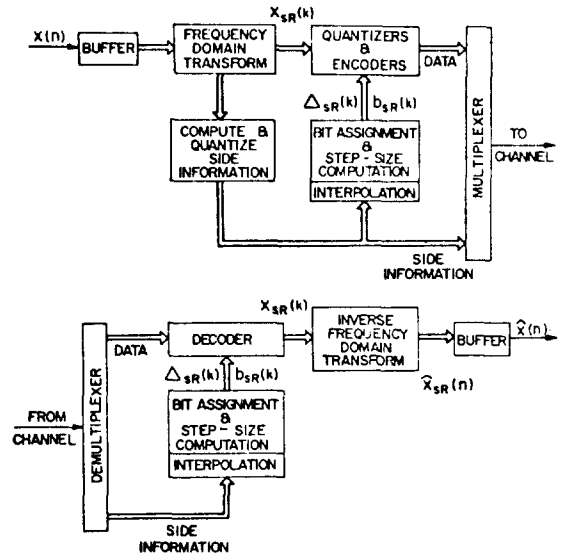


그림 5. ATC 구성도

3.1.2.2 TC(Transformation Coding)

영상에서는 여러가지 orthogonal 변환, 즉 Walsh-Hadamard, Slant, Harr 또는 DCT (discrete cosine transform) 를 이용한 TC 가 널리 사용되어 왔으나,⁴⁾ 음성에서는 DCT를 제외 하고는 널리 이용되고 있지 않다. DCT를 이용한 adaptive

TC (ATC) 의 구성도를 그림 5에 보였다. ATC에서는 먼저 입력음성을 짧은 간격으로 분할하여 DCT변환 한 후, 변환된 계수 (coefficient) 들을 각계수들의 Variance에 비례하는 bit 수를 양자기에 할당하여, 양자화 해서 보낸다. 아울러 어떤 기준치에 미달하는 적은 크기의 계수들은 전송시키지 않고, 수신측에서 적당한 방법으로 보상시키고 있다. ATC는 APC와 동작특성이 비슷하나 하아드웨어의 구성도가 복잡한 것이 문제로 지적 되고 있다.

3.2 Source 코우딩

음성은 인간의 음성발생기관의 구동신호(excitation signal)에 따라 두 가지로 구분 할 수 있다. 즉 성문 (glottis)에 의해 발생하는 반주기 (quasi-period) 적인 구형파형 공기에 의해 성도(vocal tract) 를 진동시켜 발생하는 voiced 음원과, 구강내의 난류성 공기로 성도를 구동시키는 unvoiced 음원이다. 그러면 조음기관인 성도의 형태에 따른 음향전달특성에 의하여 음원의 주파수성분이 선택적으로 공명하여 음성이 생성된다. 공명주파수를 Formant 라고 부르며 성대의 모양 및 개인에 따라 각각 그값이 틀린다. 그러므로 음성은 시간에 따라 특성이 변해가는 복잡한 파형이다. 이러한 음성발생 과정을 성

도는 시간적으로 그 특성이 변하는 필터, 반주기형 구형파는 어떤 주기를 가진 임펄스, 그리고 난류성 공기는 백색잡음 발생기 (보통 random number 발생기) 로 모델링 할 수 있다. Soura 쿠우딩의 대표적인 예인 LPC에서는 입력음성에서 voiced 음원의 주기 (pitch) 그리고 필터의 파라미터를 계산하여, 파형대신 전송하고 있다. 수신 측에서는 전송된 파라미터를 이용하여 음성을 발생시키고 있다.

음성주기 및 필터 파라미터를 구하는 과정을 음성의 해석 (speech analysis), 음성발생장치를 이용하여 음성을 발생시키는 과정을 음성의 합성 (speech synthesis) 이라고 부른다. 그림 6은 LPC Vocoder 의 구성도를 보였다.

필터는 일반적으로 all-pole 필터를 사용하며 필터 파라미터는 10~40 msec 동안의 짧은 시간동안에는 그 특성이 변하지 않는다는 가정하에 여러가지 방법으로 계산할 수 있다.³⁾ All-pole 필터 모델을 사용하면 식 (1)에 보인것과 같이 시간 n 에서의 값은 p 개의 과거값 $x_{n-i}, i=1,2,\dots,p$, 로 예측할 수 있다.

$$x_n = \sum_{i=1}^p a_i x_{n-i} \quad (1)$$

여기서 p 는 필터의 pole 수, 그리고 a_i 는 필터

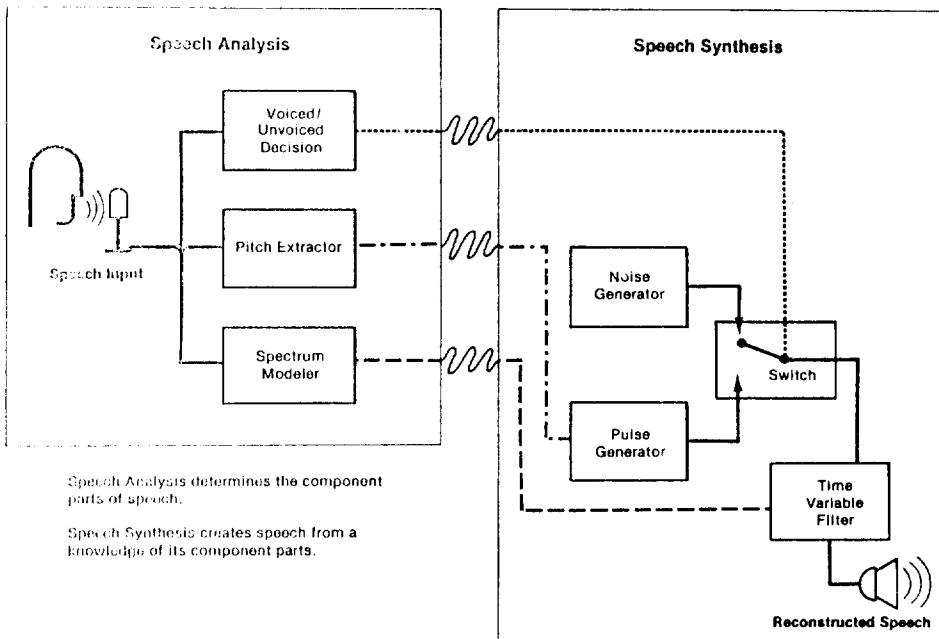


그림 6. LPC Vocoder 구성도

의 계수 (LPC계수) 이다.

그러므로 실제값 x_n 과 예측값 \hat{x}_n 의 차이 e_n 은

$$e_n = x_n - \hat{x}_n = \sum_{i=0}^p a_i x_{n-i} \quad (2)$$

로 주어진다. 만일 voiced 음원일때는 e_n 은 식 (3) 으로 표시된다.

$$e_n = \alpha e_{n-T} \quad (3)$$

여기서 T 는 pitch, α 는 이득계수이다. LPC 에서는 파형코딩과는 달리, 음성발생에 필요한 파라미터를 전송하므로, 낮은 bit-rate (보통 0.8 ~ 2.4 Kbps)에서의 음성통신은 가능하나, 음질자체가 이해도는 아주 높으나 인식도 및 자연성이 결여되어있다. 특히 필터계수는 입력음성에 포함된 잡음에 대한 민감도가 아주높아 LPC를 이용한 음성통신은 제한된 응용을 가지고 있다. 그러나 음성합성모델을 이용한 여러가지의 음성응답장치 응용에는 전망이 밝다.

3.3 Hybrid 코딩

DPCM 또는 DM 방식은 비교적 높은 bit-rate 에서 toll-quality 가 가능하고, source 코딩을 이용한 여러가지 형태의 vocoder 는 합성된 음질에 문제가 있다. 또한 LPC에서는 시간적으로 변하는 pitch 를 산출해야하는데, 이 문제가 그렇게 쉽지는 않다.

Pitch 산출은 현재 음성처리에서 가장 복잡한 문제 중의 하나이고, 또한 100% 정확한 pitch 산출방법은 아직도 알려져 있지않다. 그러므로 전송 bit-rate는 다소 증가시키더라도, 복잡한 pitch 산출을 피하고 synthetic 이 아닌 toll 혹은 communication-

quality 의 음질향상에 목적을 둔 코우딩이 바로 hybrid 코우딩 이다. Hybrid 코우딩은 9.6 ~ 16 kbps bit-rate 에서 널리 응용되고 있다. 왜냐하면 9.6 kbps 정도의 bit-rate 는 상용전화선으로 전송이 가능하기 때문에 전화통신분야의 광범위한 응용이 밝기 때문이다. 그러나 아직까지도 12 kbps 이하의 toll-quality 코우딩은 해결되고 있지 않다. 음성코우딩의 활발한 응용을 위해서는 꼭 해결 되어야 되겠다. Hybrid 코우딩방식중 앞으로 가장 음질 향상 가능성이 있는 것이 APC 이다.^{14), 18)} APC에서는 DPCM 과 LPC 의 장점을 이용하고 있다. LPC 에서 가장 문제점이 구동신호의 정확성 이므로, APC는 임펄스 대신 residual 신호로 구동신호를 대치한다. Residual 신호는 식 (2)에서 표시한 것 같이 실제값 x_n 과 예측값 \hat{x}_n 의 차신호를 말한다. Residual신호는 합성필터의 역적달함수 (inverse filter) 를 가진 필터에 입력음성을 통과하여 얻을 수 있는데, 이것을 inverse filtering 이라 부른다. 그림 7은 APC의 구성도이다.

Residual 신호를 양자화 하는데 가장 효과적인 방법은 center-clipping 을 이용한 양자기이다.¹⁹⁾ Center-clipping 양자기는 residual신호의 크기가 어떤기준치, θ , 보다 적으면 양자화 하는 대신 "0" 를 보낸다. 이렇게 하므로써 전체entropy 를 낮추므로써, entropy 코우딩의 효율을 얻게된다. 그림 8은 entropy 대 θ 와의 관계를 보였다. 많은 청각실험에서 center-clipping 양자기의 효율성은 입증되고 있다.

또한가지 빼놓을 수 없는 것은 noise-shaping 필

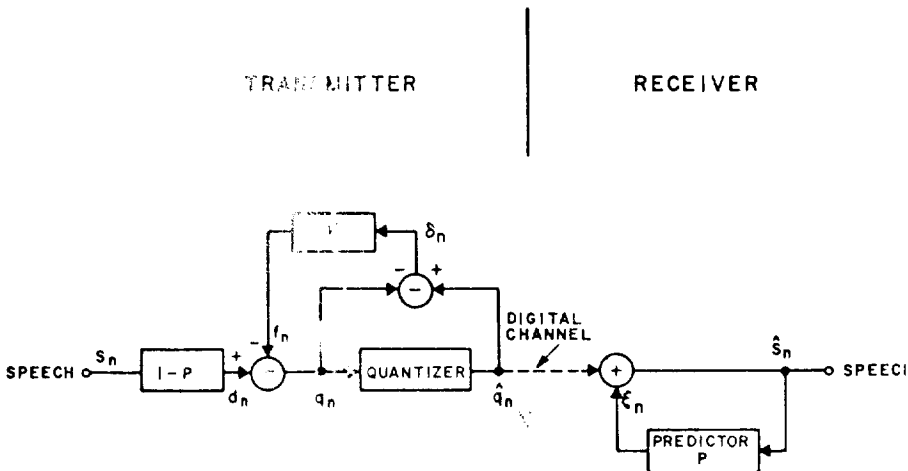


그림 7. APC 구성도의 한 예

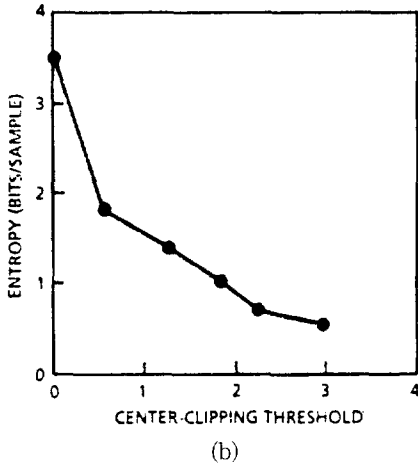
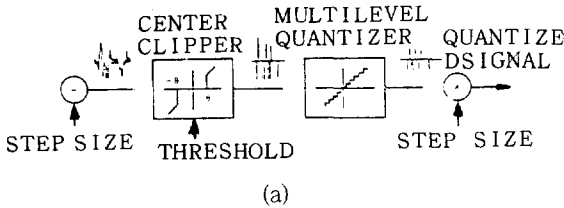


그림 8. a) Center - Clipping 양자기 구성도
b) 기준치(θ) 대 Entropy 의 관계

터의 역할이다¹⁴⁾. 일반적으로 양자기에서 발생하는 양자잡음은 백색잡음적인 성질을 가지므로 주파수 상에서 보면 그분포는 flat 하다. 그림 9에서 점선 “—”로 표시하였다. 그림에서 음성 스펙트럼의 peak 부분은 formant 를 나타내는데, peak 와 peak 사이의 valley 에서는 때때로 백색잡음적인 양자잡음보다 그 크기가 적을 수 있다. 이런 현상이 생기면 SNR 이 낮게 되고 전반적인 음질손실을 가져오게 된다. 그러므로 noise-shaping 필터를 사용해서 백색잡음적인 양자잡음의 분포를 음성의 스펙트럼분포와 비슷하게 만들면 음질향상을 얻을 수 있다. 이것을 그림 9에서 점선 (·····)으로 표시 하였다. 실제 noise-shaping 필터는 전체 SNR은 증가시키나 청각에 의한 음질향상에 그 목적이 있다고 하겠다. SBC도 이와 비슷한 원리를 이용한 코우딩이라 할 수 있다. APC 구성도 그림 7에서 보면 noise-shaping 필터는 블럭선도 F로 표시하였다. APC와 함께 중대역코우딩에서 널리 이용되고 있는 것이 RELP이다. RELP는 보통 4.8~9.6 Kbps 정도의 bit-rate에서 응용되고 있으며, 그림 10에 구

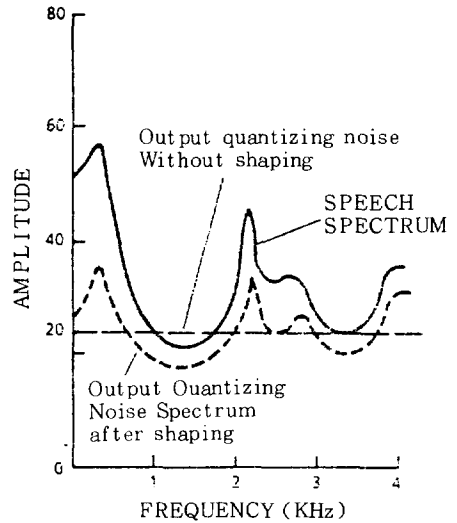


그림 9. Noise-shaping 필터의 특성

성도를 보였다.

REL P의 특징은 APC처럼 전대역 (full-band width)의 residual 신호를 보내는 대신, 1,000Hz 미만의 base-band 만 코우딩 해서 보낸다. 이렇게 되면 수신기에서는 전송된 base-band 신호를 가지고, 전송안된 고주파성분을 재생시켜야 한다. RELP에서는 고주파성분재생 (high-frequency-regeneration)의 역할이 제일 중요하다. 여러 가지의 고주파성분재생 방법이 소개되어 있으며 현재도 연구중에 있다. 고주파성분재생이 완전하지 못하면 음질은 금속성 (metallic sound)을 가지는게 특징이다. 특히 여성이나 어린이 음성에서 이 영향이 크다. RELP는 full-band 대신 base-band 만 코우딩 하므로 toll-quality는 불가능하나 7.2~9.6 kbps에서 아주 우수한 communication-quality를 얻을 수 있다.

4. DSP(Digital Signal Processor)의 응용

지금까지 몇 가지 대표적인 음성코우딩 방식을 설명하였다. 그러나 음성코우딩의 광범위한 활용을 위해서는 실시간동작이 가능한 codex (coder/decoder)의 개발이 필요하게 되었다. ADPCM이나 ADM은 하드웨어 구성도가 간단하므로 실제 실시간 codex가 사용되고 있으나, APC, ATC, SBC 또는 LPC 같은 코우딩방식은 해석이 복잡하여 대량 또는 중형

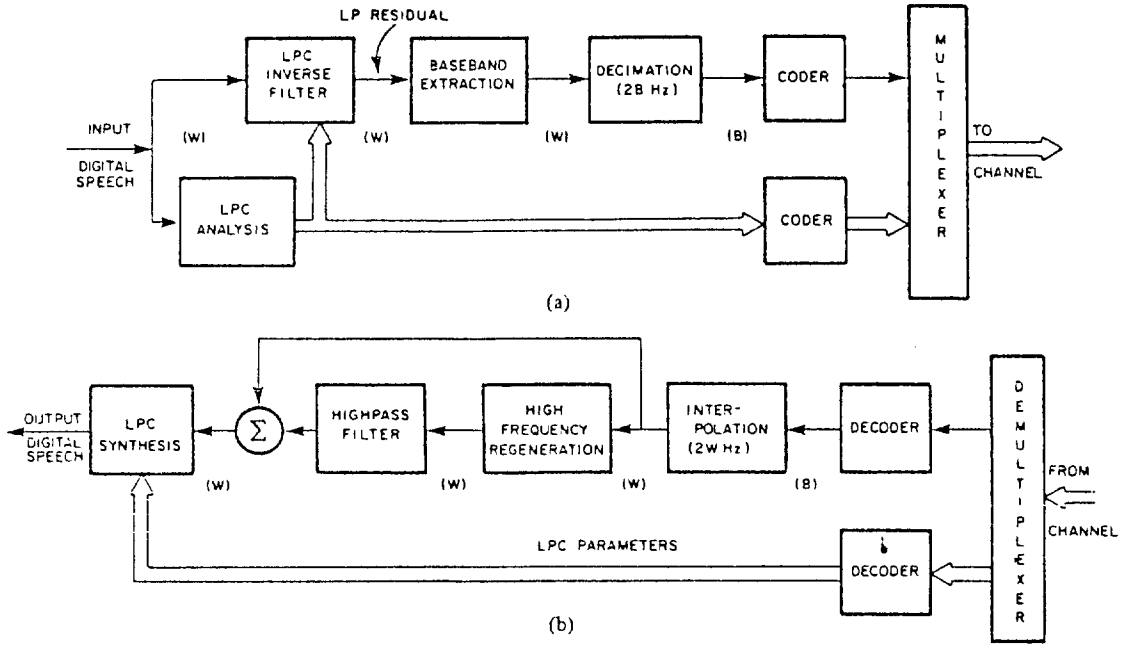


그림 10. RELP구성도

컴퓨터에 의한 실시간동작이 아닌 Simulation을 제외하고는 널리 응용되어 있지않은 실정이었다. 물론 floating-point array processor를 사용하면 실시간동작은 가능하나 경제성이 없다. 그러나 최근 VLSI의 눈부신 발전에 힘입어 고속배율기 (high-speed multiplier) 및 DSP가 소개되어 협, 중 대역 실시간 codex 개발에 많은 관심을 기울이고 있다²⁰⁾. DSP는 디지털신호처리를 위해서 고속배율기와 microprocessor를 배합한 일종의 processor이다. DSP는 Texas Instrument 회사가 음성합성을 위해 TMS 5200을 소개하면서 박차를 가하게 되었다. 현재 다목적용 DSP는 Intel 2920, AMI 2811, NEC 7720 그리고 Bell Lab DSP 등 네가지가 있다. 현재 TI에서 NEC 7720와 동작특성이 비슷한 TMS 320을 개발중에 있으나 아직 시판은 되지않고 있다. 표 3에 각 DSP의 중요한 특성을 보였다.

그중 가장 관심을 얻고있는것이 NEC 7720이다. NEC 7720는 16 × 16 bit 곱하기를 250 nsec에 행하는데, 이 속도면 실시간 음성해석을 하는데 충분하다. 그러나 NEC 7720는 off-chip addressing 능력이 없고, on-chip program ROM이 EPROM이 아닌 mask-programmable인 단점은 있다. 그러나 최근 MIT Lincoln 연구소에서 NEC7720를 사용하여 LPC vocoder를 개발성공한 예가 있다²¹⁾.

표 3. DSP들의 특성

Audio Frequency Signal Processors				
	INTEL 2920	AMI 2811	BELL LABS	NEC 7720
Chip Size	29mm ²	25mm ²	68mm ²	28mm ²
No. of Transistors	≈ 30,000	30,000	45,000	40,000
RAM (Bits)	1k	2k	2.5k	2k
ROM (Bits)	4.5k	6k	16k	18k
Clock Rate	≈ 2.5 MHz	≈ 3MHz	5MHz	8 MHz
Multiplier-Accumulator	Software	12 × 12-16	16 × 20-40	16 × 16-31
Power Consumption	0.8W	1W	1.5W	0.9W
Process	NMOS L = 4.5 μm	VMOS	NMOS L = 4.5 μm	NMOS L = 2.5 μm

TRW 배율기를 사용한 LPC vocoder는 Time and Space Processing (San Jose, Calif)에서 판매중에 있으나 대당가격이 \$ 10,000 이상이고 부피가 매우크다. 그러나 MIT의 LPC vocoder는 모두 16개의 I.C.로 구성하였으니, DSP의 효율성을 짐작할 수 있다.

Bell Labs.의 DSP는 16 × 16 bit 곱하기를 행하는데 800 nsec가 필요하므로 NEC 7720와 비교할때 속도가 느리는게 큰 결점이나, 최근 BSTJ

(Bell System Technical Journal)에 여러가지 재미있는 응용사례가 발표되어 있으니 참고하기 바란다²²⁾. 지금까지 소개한 DSP들은 모두 fixed-point 연산을 사용하고 있어, 실제 사용할적에는 finite-word length의 영향을 해석할 필요가 있다. 또 fixed-point 연산에서 자주 일어나는 over- 및 underflow를 피하기 위해서는 scaling도 아울러 해결하여야 한다. 최근 Hitachi 에서 floating-point 연산을 사용한 DSP 개발을 보고중에 있으나 아직 그 결과는 두고봐야 되겠다.

参考文献

- 1) M. Laughlin; "2400 bps digital speech for land mobile radio," 1982 Global Comm. Conf. pp. 66~70.
- 2) L.R. Rabiner and R.W. Schafer; "Digital processing of speech signals," Prentice-Hall, 1978
- 3) J.F. Flanagan and et. al; "Speech coding", IEEE. Trans. Comm., vol. COMM-27, pp. 710~736, April 1979.
- 4) W.K. Pratt; "Digital image processing," Wiley-Interscience, 1977.
- 5) W.D. Voiers and et. al; "Research on diagnostic evaluation of speech intelligibility"(internal paper, Tracor, Inc., Austin, Texas, 1973).
- 6) W.D. Voiers; "Diagnostic acceptability measure for speech communication system,"IEEE ASSP International conf. Proc., 1977.
- 7) B. N. Oliver, J.R. Pierce and C.E. Shannon; "The philosophy of PCM," Proc. IRE, 36, pp. 1324~1331, Nov. 1948.
- 8) N.S. Jayant; "Digital coding of speech waveforms: PCM, DPCM and DM quantizers," Proc. IEEE, vol. 62, pp. 611~632, May 1974.
- 9) R. Steele; "Delta modulation system," John. Wiley and Sons, 1975.
- 10) R.E. Crochiere and et. al; "Digital coding of speech in sub-bands," BSTJ, 55, pp. 1069~1085, October 1976.
- 11) D. Malah; "Time-Domain algorithms for harmonic bandwidth reduction and time scaling of speech signals," IEEE Trans. ASSP, vol. ASSP-26, pp.121~133, April 1978.
- 12) J.B. Anderson and J.B. Boding; "Tree encoding of speech" IEEE Trans. Infor. Theory, vol. IT-20, pp.389~379~389, July 1975.
- 13) J.D. Markel and A.H. Gray; "Linear prediction of speech," Springer-Verlag, 1976.
- 14) B.S. Atal; "Predictive coding of speech at low bit rates," IEEE Trans. Commo vol. COMM-30, pp.600~614, April 1982.
- 15) C.J. Un and D.T. Magill; "The residual excited linear prediction vocoder with transmission Rate below 9.6kbps," IEEE Trans. Comm. vol. COMM-23, pp. 1466~1474, December 1975.
- 16) J.M. Tribolet and R.E. Crochiere; "Frequency domain coding of speech," IEEE Trans. ASSP vol. ASSP-27, pp.512~530, Oct. 1979.
- 17) S.U. LEE and R. Stroh; "Study of SBC for the mobile radio application," M/A-COM Research center, Technical Report, Rockville, Maryland, 1983.
- 18) B.S. Atal and M.R. Schroeder; "Adaptive predictive coding of speech signals," BSTJ, 49, pp. 1973~1986, October 1970.
- 19) B.S. Atah and M.R. Schroeder; "Improved quantizer For adaptive predictive coding of speech signals at low bit rates," ASSP Inter'l Conference Proc., pp. 535~538, April 1980.
- 20) "Joint special issue on Integrated circuits for speech," IEEE Trans. ASSP vol. ASSP-31 (part II), whole issue, Feb. 1983.
- 21) J.A. Feldman and et. al; "A compact, flexible LPC vocoder based on commercial signal processing microcomquter," IEEETrans. ASSP vol. ASSP-31, pp.252~258, Feb. 1983.
- 22) BSTJ, vol. 60, No. 7(part 2), September 1981