

PCS를 위한 음성 코딩 방식과 음질 비교

金 洪 國, 金 尙 龍

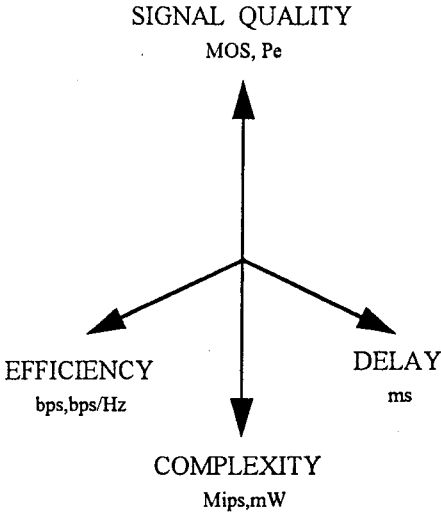
三星綜合技術院 基盤技術研究所
音聲研究室

I. 서 론

Personal Communication Services(PCS)는 지역에 제한없이 대용량으로 지상의 전기 통신 서비스에 무선으로 접근할 수 있어야 한다. 단말 이동성, 개인 이동성, 무선 접근, 그리고 사용자에 대한 연속적 이동성 등의 개념을 포함하는 PCS에서 사용자는 네트워크의 종류에 관계없이 개인에게 유일하게 할당된 번호를 통해 서비스를 제공받을 수 있어야 한다.^[1] 이러한 개념을 갖는 PCS는 기존의 공중전화망(PSTN: Public Switched Telephone Network), AMPS(American Mobile Phone System)와 같은 아날로그 (analog) 이동통신 시스템 (제1세대 PCS), cordless telephone, paging, 셀룰라 망 등의 디지털(digital) 이동통신 시스템 (제2세대 PCS), 그리고 위성통신 시스템 등의 각기 다른 기술을 모두 포함해야 한다.^[1] 따라서, PCS를 위한 음성 코딩 방식을 한마디로 정의하기는 매우 어려우며 PCS의 설계에 따라 음성 코더(speech coder)의 전송률, 전송 에러의 견고성, 전송을 위한 시간 지연(delay) 등이 정해져야 한다.

음성 coder의 성능은 크게 복원된 음성의 음질(signal quality), 전송률(bit rate), coder의 계산량(complexity), 그리고 통신지연(communication delay) 등의 4가지 요소로 평가될 수 있다. 그림 1은 이들 4가지 성능의 측정 단위와 그 관계를 보여준다.^[2,3]

Mean Opinion Score(MOS)는 표 1에 나타낸 바와 같이 음성의 자연성을 5단계로 분류한다.^[4] MOS가 4~4.5인 경우를 network quality, 3.5~4는 communication quality, 그리고 2.5~3.5인 coder는 synthetic quality를 갖는다고 말하여 진다.^[3] 예를 들면 32kbps ADPCM(Adaptive Differential Pulse Code Modulation)^[5]은 MOS 4.1로 network quality를 만족하며, 미국방성(U.S. Department of Defense) 표준의 하나인 2.4kbps LPC-10e^[6]는 MOS 2.3으로 synthetic quality 이하의 음성을 재생하기 때문에 일반 사용자를 위한



〈그림 1〉 음성 및 channel coder의 성능 척도

〈표 1〉 MOS Scale

MOS Scale	Speech Quality	Impairment Scale
1	bad	very annoying
2	poor	annoying
3	fair	slightly annoying
4	good	perceptible but not annoying
5	excellent	imperceptible

통신보다는 비화통신 등의 특수 목적에 주로 사용된다. 통신 시스템에서의 음질은 음성 coder 이외에 채널 코딩(channel coding), 변복조 방법 등에 영향을 받는다. 채널 코딩 방식은 bit error rate(BER)의 정도에 따라 설계된다. 오류정정 기술은 채널의 선택적 페이딩에 의한 burst error 대책으로 유효하며 음성 코딩의 중요 비트에 중점적으로 정정을 가한다. 유럽의 GSM의 경우 13kbps의 음성 신호에 Reed-Salomon 코드를 부가하여 22.8kbps의 신호를 만들어 interleaving 방식으로 전송한다.

Bit rate는 sampling rate와 sample당 bit수의 곱으로 표현된다. 표 2는 일반적으로 사용되는 오디오 신호의 sampling rate와 bandwidth를 보여

〈표 2〉 Digital Audio Formats

Format	Sampling Rate(kHz)	Bandwidth (kHz)	Frequency Band
Telephony	8	3.2	(200-3400Hz)
Teleconferencing	16	7	(50~7000Hz)
Compact Disk(CD)	44.1	20	(20-20000Hz)
Digital Audio Tape(DAT)	48	20	(20-20000Hz)

준다. 음성은 200~3400Hz의 대역에 음성의 에너지가 집중되어 있으며, 전형적인 sampling rate는 8kHz이다. 통신망에서의 spectral efficiency는 bandwidth당의 bit rate로 표현되며 단위는 bps/Hz가 된다.

Coder의 complexity는 음성을 부호화하고 복호화하는데 필요한 연산수(MIPS : Mega Instruction/Second)로 정의된다. 이 complexity는 codec의 가격 및 파워 소모량(mW)과 밀접한 관계가 있으며 PCS 단말가격 및 battery 용량을 결정하는 중요한 요소가 된다.

마지막으로, 음성 coder의 성능을 좌우하는 시간지연(delay)은 음성의 중복성을 제거하는데 소요된다. PSTN에서 echo없이 통화가 가능하기 위해서는 delay가 20~30ms 이내이어야 하며 위성 통신의 round trip delay는 280ms 이내이어야 한다. 그리고, 채널 코딩에서는 error protection bit를 소스 coder에 더하거나 burst error를 줄이기 위한 interleaving이 delay의 원인이 된다. 이들 4가지 성능 요소는 PCS 시스템의 형태, network 프로토콜, 사용자수, 다중접속 방식, 셀(cell) 크기, 그리고 변복조 방법들에 따른 trade-off를 갖는다.

PCS 설계시 가장 중요한 요소로는 통신을 위해 할당된 주파수 대역에 대한 사용자의 밀도로 정의되는 spectral efficiency이다. Spectral efficiency는 주로 셀의 크기와 할당된 대역폭, 그리고 다중접속(Multiple Access) 방법에 의해 정해진다. 또한 음성 coder의 전송률 감소는 보다 정교한 채널

〈표 3〉 음성 Coder의 국제 표준화 동향

Network	기 관	음성 coder		
PSTN	ITU-T	32kbps ADPCM	(G.721, G.726)	[5]
		16kbps LD-CELP	(G.728)	[8]
		8kbps ACS-CELP	(G.729)	[9]
		6.3 & 5.3kbps Coder	(G.723)	[10]
D-AMPS (IS-54)	TIA	8kbps VSELP		[11]
		4kbps VSELP		
Qualcomm CDMA (IS-95)	TIA	1~8kbps variable QCELP		[12]
GSM TDMA	ETSI/GSM	13kbps RPE-LTP		[13]
		5.6kbps VSELP		[14]
JDC	RCR	6.4kbps VSELP		[15]
		3.45kbps PSI-CELP		[16]
INMARSAT-M	INMARSAT	4.15kbps IMBE		[17]
Secure Communication	U.S. DoD	4.8kbps CELP		[18]
		2.4kbps LPC-10e		[6]

코딩을 요구하기도 하지만 일반적으로 spectral efficiency를 증가시킨다.

Microcell이나 picocell를 사용하는 CT2(Cordless Telephone 2)에서는 채널당 할당된 대역폭이 넓기 때문에 32kbps ADPCM coder를 사용한다. 반면 큰 셀(large cell)을 사용하는 유럽의 GSM(Global System for Mobile communications)은 13kbps의 RPE-LTP(Regular Pulse Excited with Long Term Prediction) coder를, 일본 디지털 셀룰라(JDC : Japan Digital Cellular) 망의 표준으로는 6.4kbps VSELP(Vector Sum Excited Linear Prediction)를 각각 채택하였다.^[7]

음성 coder의 국제 표준화는 PSTN의 경우는 ITU-T에서, 이동체 통신의 경우는 북미의 TIA, 일본의 RCR(Research and development Center for Radio system), 그리고 유럽의 ETSI(European Telecommunications Standard Institute)에서 각 이동통신 시스템 방식에 따른 음성 coder의 표준화를 진행한다. 또한, 위성통신을 위한 음성 coder의 표준화는 국제 콘소시엄인 INMARSAT(INternational MARitime Telecommunications SATellite organization)에서 진행되고 있다. 그밖

에 비화통신(secure communication)을 위해 미국 방성은 4.8kbps CELP와 2.4kbps LPC 보코더 등을 표준으로 사용하고 있다. 표 3은 각 주요기관에 의해 표준화된 coder의 일람을 보여준다.

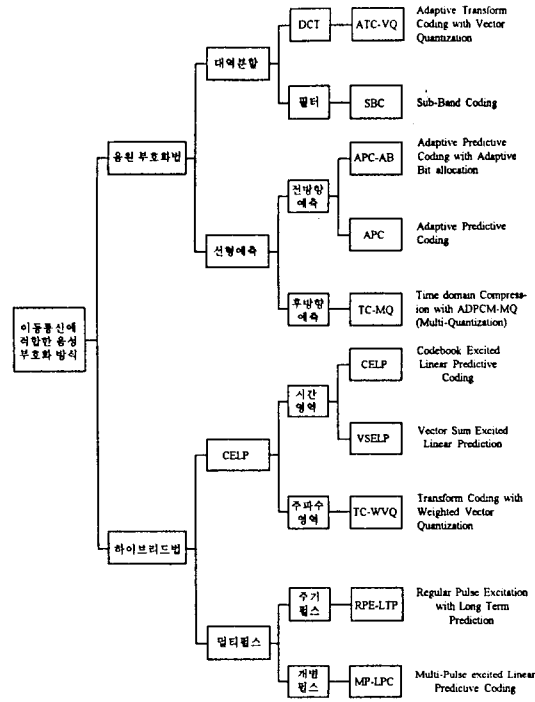
서론에 이어서 제2장에서는 이미 표준화로 선정된 제2세대 이동통신용 음성 코딩 방식을 기술하고, 제3장에서는 제3세대 PCS와 관련된 표준화 동향을 소개한다. 마지막으로 제4장에서는 음성 coder와 관련된 향후 연구방향을 제안하고 결론을 맺는다.

II. 제2세대 이동통신용 음성 코딩 방식

PCS의 제1세대라 불리는 아날로그 이동통신 시스템은 제2세대인 디지털 이동통신 시스템으로 바뀌어가고 있다. 디지털 시스템은 아날로그 시스템에 비해 음성 신호의 압축에 의한 spectral efficiency의 증가가 용이하고 잡음에 강하며, 특히 암호화 등 부가기능을 부여할 수 있다는 장점을 갖는다.

음성 coder는 크게 과형부호화 방식, 음원부호

화 방식, 그리고 이들의 장점을 취하는 hybrid 방식으로 나눌 수 있다. 주요한 음성 부호화 방식은 그림 2와 같다.



〈그림 2〉 이동 통신에 적합한 음성 coder의 분류

표 4는 CT2, D-AMPS, GSM, IS-95, 그리고 JDC의 음성 coder가 갖는 음질(MOS), 계산량(MIPS, mW), 전송률(bps)을 비교한다.

CT2는 무선으로 CT2 기저시스템과 PSTN에 접속이 가능하다. 음성 coder로 채택된 ADPCM의 음질은 network quality를 유지하며, 계산량이 적어 단말의 평균 파워는 5mW 정도에 지나지 않는다. 또한 채널 코딩이나 등화기가 없이도 고품질 합성음을 보장할 수 있으며 유럽의 이동통신 망인 DECT(Digital European Cordless Telecommunications) 시스템에서도 사용되고 있다.

한편, 경제적 실용성을 보장하기 위해 통신 시스템의 용량(capacity)을 증가시키기 위한 방법으로 더 낮은 bit rate를 갖는 음성 coder의 등장이 꾸준히 요구되어 왔다. 1980년대 GSM은 subband coder를 이동통신 시스템의 음성 coder 후보로 검토하기 시작했다. 그후 분석-합성 방식(analysis-by-synthesis)을 근간으로 하는 MOS 3.47의 RPE-LTP를 음성 coder로 확정하였으며 RPE-LTP의 채택으로 GSM은 25MHz의 대역내에 25kHz/channel의 1000채널을 구성할 수 있었다. 이는 ADPCM을 음성 coder로 사용하는 DECT에 비해 2.3배의 용량증가를 가져온 것이었다.

북미 디지털 셀룰라 시스템인 D-AMPS (혹은 IS-54)는 아날로그 시스템인 AMPS의 용량을 증가시키기 위해 제안되었다. 30kHz 내에 3개의 채널을 할당하는 방식으로 1개의 AMPS를 대체하여 결과적으로 3배의 용량 증가를 가져왔다. 10kHz 내에 한 사용자의 음성을 전송하기 위해 8kbps VSELP의 음성 coder(channel coder를 합

〈표 4〉 제2세대 이동통신 시스템의 음성 Coder의 성능비교

System \ ITEM	D-AMPS (IS-54)	GSM	IS-95	CT-2	JDC
Multiple Access	TDMA	TDMA	CDMA	FDMA	TDMA
Speech Coder	VSELP	RPE-LTP	QCELP	ADPCM	VSELP
voice rate(kbps)	13	22.8	19.2/28.8	72	11.2
uncoded voice rate (kbps)	8	13	variable 8	32	6.7
MOS	3.45	3.47	3.45	4.1	3.40
MIPS	13.5	6	11~23	2 이하	?
Mean power(mW)	200	125	600	5	?

칠 경우 13kbps가 된다)와 $\pi/4$ -DQPSK의 변복조 기법을 사용하였다. D-AMPS의 음성 coder는 13kbps RPE-LTP에 비해 계산량이 2배정도 증가하는 반면 음질은 거의 비슷한 수준을 유지하므로 25kHz/channel인 GSM에 비해 2.5배의 용량 증가를 가져 왔다.

IS-95에 채택된 QCELP는 voice activity detection(VAD) 기법을 적용하는 variable rate coder라는 점에서 지금까지 표준화된 다른 coder와 차이점을 갖는다. VAD는 silence 구간과 음성 구간을 분류하고 silence 구간에 대해서는 적은 bit로 음성을 표현하기 때문에 평균 전송률을 줄일 수 있다. QCELP에서의 VAD는 음성을 4가지 모드로 분류하여 각각 8kbps, 4kbps, 2kbps, 그리고 0.8kbps로 전송한다. 평균적으로는 4kbps 이하에서 coder가 동작하며 계산량은 각 모드에 따라 11~23MIPS가 된다. IS-54의 용량은 아날로그 셀룰라 시스템에 비해 3배 정도의 증가에 그치는 반면, IS-95는 CDMA와 QCELP의 결합으로 이론적으로 평균 15배의 용량 증가가 기대되나 실제 사용시 7~8배 용량 증가를 발휘한다.

일본에서는 RCR에 의해 JDC 시스템의 표준으로 채택된 6.7kbps VSELP를 통해 8kbps VSELP와 유사한 음질로 적은 계산량을 갖도록 하는 음성 coder를 채택함으로써 파워 소모를 줄이는데 성공하였다. 다음 장에서는 현재 연구 중이거나 표준화될 예정인 half-rate 음성 coder의 표준화 동향과 ITU-T의 표준화 작업에 대해 소개하기로 한다.

III. 차세대 PCS용 음성 coder의 표준화 동향

서론에서 언급한 바와 같이 제3세대 PCS는 PSTN, 디지털 셀룰라, satellites 통신 시스템 기술의 총체적 집합이다. 본 장에서는 PSTN용 음성 coder의 표준화를 주로 관장하는 ITU-T의 표준화 동향과 half-rate 디지털 셀룰라 시스템(2.5세대 PCS)을 위한 음성 coder의 표준화에 대해 기술하

고자 한다.

표 5는 1995년 이후 ITU-T의 음성 coder 표준화 동향을 보여준다. ITU-T G.728 LD-CELP는 16kbps에서 동작하고 MOS 4.0의 음질과 19MIPS 이하의 계산량을 갖는 음성 coder이다. 최근들어 ITU-R에서는 장래의 공중지상이동 통신 시스템인 FPLMTS(Future Public Land Mobile Telecommunication Systems)의 표준 규격 권고화에 대한 연구가 진행되고 있다. 이와 함께 ITU-T에서는 LD-CELP의 확장 버전으로 16kbps 이하와 16kbps 이상의 전송률로 동작하는 음성 coder에 대한 연구를 추진하고 있다. 12.8kbps와 9.6kbps에서 동작하도록 하는 LD-CELP는 1995년 11월 G.728의 Annex H로 표준 권고안이 마련될 예정이며, voice-band로 데이터를 전송하기 위해 16kbps 이상의 LD-CELP를 확장하려는 표준화 작업이 진행중이다.^[19]

음성과 데이터를 동시에 전송하기 위하여 11.2kbps 이하의 DSVD(Digital Simultaneous Voice and Data)용 음성 coder의 표준화가 진행중이며, 현재 University of Sherbrooke(9.6/13kbps dual rate CS-CELP), Rockwell(11.2kbps CELP), NTT(11.1kbps CELP), AT&T(SVD-CELP)등이 후보 coder들을 ITU-T에 제출해 놓고 있고, 1995년 9월에 이들 중 1개의 후보 coder를 선정

〈표 5〉 ITU-T의 음성 Coder 표준화 동향

음성 coder	study period	응 용
16kbps 이하 LD-CELP	1995.2~1995.11	FPLMTS
16kbps 이상 LD-CELP	1995.2~1996.3	Voice band data 전송
11.2kbps 이하의 DSVD용	1995.2~1996	Digital simultaneous voice and data modem
4kbps coder	1995.11~1998.12	Very low-rate PSTN PCS Mobile-telephony satellite system

할 예정이다.^[20]

또한, ITU-T G.729(8kbps CS-ACELP)의 차기 음성 coder로 4kbps 음성 coder의 표준화를 1995년 11월부터 시행하려 하고 있다.^[21] 이의 응용분야로는 PSTN용 visual telephone, PCS, mobile-telephony satellite 시스템 및 message retrieval 시스템 등이 있으며 1998년 말에 표준안으로 확정될 전망이다.

디지털 셀룰라 시스템에서 가입자의 수를 2배로 증가시키기 위한 half-rate 음성 coder의 표준화 작업이 북미, 유럽, 일본 등지에서 활발히 연구중이며 이는 표 6과 같다. 표 6에 나타난 5.6kbps VSELP는 유럽 GSM의 half-rate coder로서 유력시되며, 북미 TIA에서는 Motorola의 4kbps VSELP를 half-rate TDMA 표준으로 채택할 예정이다. 한편, 일본은 이미 half-rate JDC 시스템용으로 3.45kbps PSI-CELP(Pitch Synchronous Innovation CELP)를 개발, 표준으로 채택하였다.

(표 6) Half-rate Digital Cellular System 용 음성 Coder

기 관	시 스템	음성 coder
TIA	Half-rate TDMA	4kbps VSELP
ETSI/GSM	Half-rate GSM	5.6kbps VSELP
RRC(일본)	Half-rate JDC	3.45kbps PSI-CELP

IV. 결 론

음성 coder는 CELP 계열의 분석-합성 방식의 등장으로 8kbps 이하에서도 거의 toll quality를 나타내는 수준으로 발전되어 왔으며, 4~8kbps의 각종 음성 coder들이 PSTN이나 디지털 셀룰라 시스템의 표준으로 채택되었거나 채택될 예정이다.

음성 coder의 표준화 제정은 정해진 일정에 따라 공정한 경쟁을 통해 이루어진다. 외국의 우수한 대학, 기업, 연구소에서는 주어진 coder의 사양에

맞추어 독자적으로 개발된 coder 알고리즘을 제출하여 표준화에 참여하고 이를 통해 정보 교류 및 기술 발전을 도모한다. 따라서 국내의 대학이나 연구소에서도 국제 표준화 노력에 적극 동참할 필요가 있다. 특히 국내의 통신 시스템 구축시 음성 coder의 독자적 표준화를 위한 기구 설립이 시급하며 이는 국내 음성 coder 개발의 기술력 향상 및 저변확대에 일익을 담당하게 될 것이다.

참 고 문 헌

- [1] F. Ananasso and F. D. Priscoli, "The role of satellites in personal communication services," *IEEE Journal on Selected Areas in Commun.*, Vol. 13, No. 2, pp.180~196, Feb. 1995.
- [2] N. Jayant, "Signal compression : Technology targets and research directions," *IEEE Journal on Selected Areas in Commun.*, Vol. 10, No. 5, pp.796~818, June 1992.
- [3] A. S. Spanias, "Speech coding : A tutorial review," *Proc. of the IEEE*, Vol. 82, No. 10, pp.1541~1582, Oct. 1994.
- [4] ITU-T Recommendation P.80, "Telephone transmission quality subjective opinion tests," 1994.
- [5] ITU-T Recommendation G.721, "32kbit/s adaptive differential pulse code modulation(ADPCM)," 1988.
- [6] Federal Standard 1015, "Telecommunications : Analog to digital conversion of radio voice by 2400bit/second linear predictive coding," National communication system-Office of Technology and Standards, Nov. 1984.
- [7] R. Steele, "Speech codecs for personal communications," *IEEE Comm. Mag.*, Vol.

- 31, No. 1, pp.76~83, Nov. 1993.
- [8] ITU-T Recommendation G.728, "Coding of speech at 16kbit/s using low-delay code excited linear prediction(LD-CELP)," 1992.
- [9] ITU-T Draft Recommendation G.729, "Coding of speech at 8kbit/s using conjugate-structure algebraic-code-excited linear-predictive(CS-ACELP) coding," June 1995.
- [10] ITU-T Draft Recommendation G.723, "Dual rate speech Coder for multimedia telecommunication transmitting at 5.3 & 6.3 kbit/s," June 1995.
- [11] EIA/TIA-PN2398(IS-54), "The 8kbits/s VSELP algorithm," 1989.
- [12] EIA/TIA Interim Standard, "Wideband spread spectrum digital cellular system dual-mode mobile station," 1992.
- [13] ETSI/GSM, GSM 06.10, "GSM full-rate transcoding," Tech. Rep. Version 3.2, July 1989.
- [14] I. A. Gerson and M. A. Jasulic, "A 5600bps VSELP speech coder candidate for half-rate GSM," Proc. of Eurospeech' 93, pp.253~256, 1993.
- [15] Motorola, "Vector sum excited linear prediction(VSELP) 11200 bit per second voice coding algorithm including error control for Japan digital cellular," Draft text for specification, 1990.
- [16] K. Mano, T. Moriya, S. Miki, H. Ohmuro, K. Ikeda, and J. Ikedo, "Design of a pitch synchronous innovation CELP coder for mobile communication," *IEEE Journal on Selected Areas in Commun.*, Vol. 13, No. 1, pp.31~41, Jan. 1995.
- [17] Digital Voice Systems, Inc., "IMBE Vocoder description," July 1993.
- [18] Federal Standard 1016, "Telecommunications: Analog to digital conversion of radio voice by 4800bit/second code excited linear prediction(CELP)," National communication system-Office of Technology and Standards, Feb. 1991.
- [19] S. Hayashi, "Rapporteur's Report on question Q.5/15," ITU-T Temporary Document, Feb. 1995.
- [20] ITU-T Temporary Document No. DSVD-95-02, "Terms of reference for voice coder to support simultaneous voice and data transmission," Feb. 1995.
- [21] ITU-T Temporary Document No. COM15-R-E, Annex VI, "Terms of reference for 4kbit/s speech coding," 1995.

저자 소개



金 洪 國

1965年 10月 24日生

1988年 2月 서울대학교 제어계측공학과(학사)

1990年 2月 한국과학기술원 전기 및 전자공학과(석사)

1994年 8月 한국과학기술원 전기 및 전자공학과(박사)

1991年 9月~1993年 1月 삼성전자 정보통신연구소

1993年 2月~현재 삼성종합기술원 선임연구원

주관심분야 : 음성 압축, 음성 인식

金 尚 龍

1957年 8月 24日生

1980年 2月 한국항공대학(학사)

1982年 2月 한국과학기술원 전기 및 전자공학과(석사)

1989年 2月 한국과학기술원 전기 및 전자공학과(박사)

1980年 1月~1993年 1月 삼성전자 정보통신 연구소 근무

1993年 2月~현재 삼성종합기술원 음성연구실장

주관심분야 : 음성처리