

주 제

차세대 음성/음향 부호화기 기술 동향

한국전자통신연구원 윤병식, 최송인 한양대학교 강상원

차 례

I. 서론

II. 음성 부호화기 기술 동향

III. 음향 부호화기 기술 동향

IV. 음성/음향 부호화 통합 기술 동향

V. 결론

요 약

디지털 음성 및 음향 부호화기는 디지털 통신 시스템 및 방송 시스템의 발전과 더불어 양방향 음성 통신 서비스, 멀티미디어 방송 서비스 등과 같은 응용분야에 각광을 받고 있다. 특히 음성 부호화기의 경우 심한 채널에러와 한정된 주파수 자원과 같은 디지털 무선 통신 시스템 환경에 적합하게 더욱 발전해왔으며 디지털 음향 부호화기의 경우 다채널 디지털 방송, 무선 멀티미디어 서비스에 적합하도록 전송 속도와 성능이 향상되어 왔다. 최근 들어 통신 및 방송 기술의 비약적인 발전으로 네트워크와 전송 채널 상태에 따라 전송 속도가 유연하게 변화함과 동시에 높은 품질의 멀티미디어 서비스가 가능한 가변 전송 속도(Scalability)의 기술과 음성과 음향 부호화기를 하나로 통합시키는 기술들이 각광을 받고 있다. 본 논문에서는 유선 및 무선 디지털 망과 방송망에서 사용되는

음성 / 음향 부호화기의 기술에 대하여 살펴보고 이와 관련된 표준화 활동 및 기술동향에 대하여 알아본다. 또한 향후 통신 및 방송의 융합에 따른 디지털 음성 / 음향 부호화 통합 기술의 방향에 대하여 언급한다.

I. 서론

최근 들어 새로운 디지털 무선 통신 시스템의 출현과 디지털 멀티미디어 서비스의 비약적인 발전과 더불어 높은 수준의 디지털 음성 부호화기와 디지털 음향 부호화기의 개발이 진행되어 왔다. 과거의 음성 부호화기는 단순한 대화 위주의 의사 전달이 가능한 협대역(0.2 ~ 3.4 kHz) 영역에서 사용되어 왔고, 초창기 음향 부호화기 또한 높은 대역폭(15 ~ 20 kHz)을 가지는 방송 서비스 영역에서만 독자적으로 발전해왔으나, 최근 들어 낮은 수준의 음향 서비스와

높은 수준의 음성 서비스가 특징인 광대역 (0.05 ~ 7 kHz) 음성 부호화기들이 등장했을 뿐 아니라 기존의 방송 영역에서나 가능한 다양한 멀티미디어 콘텐츠들을 수용할 수 있는 음향 부호화 영역의 주파수 대역까지 처리할 수 있는 음성/음향 부호화기들이 디지털 통신 및 방송 시스템에 적용되고 있는 실정이다.

디지털 통신 및 방송 시스템이 각기 발전함에 따라 두 시스템간의 구분이 모호해질 뿐 아니라 두 가지 시스템이 하나로 융합되는 방향으로 기술 진화가 이루어지고 있다. 따라서 디지털 음성/음향 부호화기는 화상통신, 서킷 통신망, 패킷통신망, 멀티미디어 전송, 디지털 방송 등의 기술적 발전과 높은 품질의 음성/음향 서비스 등의 사용자의 요구사항 변화에 맞추어 발전하고 있다.

음성/음향 부호화기 개발에서 중요한 요소는 전송 속도, 다양한 환경 하에서의 품질, 부호화 지연시간, 복잡도로 나누어질 수 있다. 전송속도 관점에서 음성 부호화기를 본다면 협대역 음성 부호화기는 평균 전송속도 4 kbit/s 에서 Toll 품질 (Toll quality) 수준의 음성부호화기들이 개발되었으며, 광대역 음성부호화기는 24 kbit/s 급에서 개발이 되었고 최근의 음향 부호화 기술은 약 48 kbit/s 에서 CD 수준의 스테레오 음향 품질을 나타내는 부호화기들이 개발을 완료하였다. 부호화 품질 관점에서 본다면 다양한 무선 채널 환경 변화와 배경 잡음 하에 음성/음향 품질의 현격한 성능 저하를 보이지 않아야 하며 이는 실제 응용분야에 중요한 요소로 작용한다. 부호화 지연 시간은 디지털 부호화를 위하여 입력 데이터를 수집 및 처리하는데 요구되는 시간을 의미한다. 일반적으로 낮은 전송 속도의 부호화기는 높은 지연 시간을 가지며 이러한 지연 시간은 디지털 통신 시스템의 지연 시간과 맞물려 양방향 통화 서비스에서는 라인 에코와 같은 형태로 양방향 통화 품질을 저하 시킨다. 복잡도는 음성/음향 부호화기를 실제 DSP (Digital Signal

Processor) 등으로 구현할 때 DSP가 요구하는 계산량과 메모리 크기를 의미하며 다양한 기능을 하나의 단말로 융합하려는 추세에 따라 적은 용량의 복잡도를 가지는 음성/음향 부호화기가 바람직하다[1].

본 논문에서는 다양한 네트워크 및 방송 시스템에서 사용중인 디지털 음성/음향 부호화 기술의 동향, 표준화 동향, 향후 발전 방향에 대하여 언급한다. 먼저 현재까지 발전해온 디지털 음성 및 음향 부호화 기술에 대하여 언급하고 각 기술 별 특성과 응용분야에 대하여 설명한다. 먼저 ITU-T, 3GPP, 3GPP2 등과 같은 통신 표준화 단체에서 제정된 음성부호화 기술들에 대하여 알아본다. 그리고 MPEG에서 다루어지는 음향 부호화기의 기술 개발 현황과 표준화 상황에 대하여 설명하고, 마지막으로 현재 개발중인 디지털 음성/음향 부호화기의 통합 추세와 기술동향에 대하여 언급한다.

II. 음성 부호화기 기술 동향

통신 시스템이 발전함에 따라 디지털 음성 부호화기도 더불어 발전해왔다. 이러한 발전은 각종 통신 표준화 단체의 통신 시스템의 일부로 채택되어 왔는데, 특히 근래 들어 디지털 이동통신 시스템의 비약적인 발전에 따라 각 이동 통신 시스템에 적합한 디지털 음성 부호화기가 개발되어왔고 진화되어 왔다.

이와 관련한 대표적인 표준화 단체는 UN산하 각종 통신 표준화를 담당하는 ITU (International Telecommunication Union) 가 있으며, 유럽지역의 각종 통신 표준화를 담당하는 ETSI (European Telecommunications Standards Institute) 및 3GPP (3rd Generation Partnership Project) 가 있고, 북미지역의 각종 통신 표준화 업무를 수행하는 TIA (Telecommunication Industry Association)

및 3세대 이동통신 시스템 표준화 기구인 3GPP2(3rd Generation Partnership Project 2)가 있다. 또한 그 밖의 관련 표준화 기구로 ISO/IEC(International Organization for Standardization / International Electrotechnical Commission), IETF(Internet Engineering Task Force) 등이 있다. 본 단락에서는 음성부호화 모델링 기술에 대하여 간단히 언급하고 각 표준화 단체별 디지털 음성 부호화기의 종류와 특성에 대하여 알아보고 실제 응용 분야에 대하여 살펴본다.

1. 음성 부호화 모델링 기술

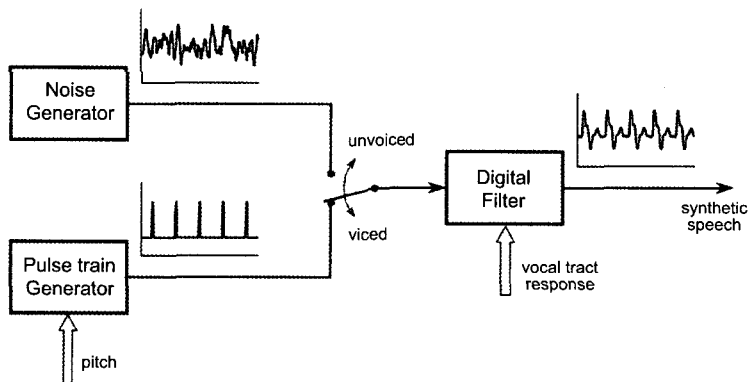
초창기 음성 부호화 방식은 아날로그 신호를 일정한 시간 간격으로 샘플링한 후 샘플링된 신호의 크기를 디지털 값으로 변환하는 PCM(Pulse Coded Modulation) 방식으로 사용되었다. 이러한 PCM 방식은 부호화 방식이 간단하고 음성 왜곡이 작은 장점이 있으나 전송속도가 높아 주파수 자원이 한정적인 디지털 무선 통신시스템에서의 사용에 적합하지 않다. 따라서 인간의 발성 원리를 모델링한 LPC(Linear Prediction Coefficient) 부호화방식이 등

장하게 되었다. 인간의 음성은 성도에서(Vocal cord) 발생하는 불규칙한 공기의 흐름이 성대를(Vocal tract) 통과하면서 발생된다. 이는(그림 1)에서와 같이 인간의 성대에 해당하는 부분을 LPC 디지털 필터로 모델링하고 인간의 성도에서 발생하는 신호를 유성음과 무성음으로 구분하여 여기신호(Exciter)신호로 표현하게 된다. 따라서 LPC 필터의 계수와 여기신호를 실제 음성 대신 부호화하게 되고 이 전송 계수를 복호화함에 따라 음성 신호를 재생하게 된다[2].

최근의 음성 부호화 연구에서 가장 큰 이슈는 여기 신호를 효율적으로 표현하는 방법과 적은 비트수로 LPC 필터 계수를 양자화 방법이다. 효율적인 여기 신호를 모델링하기 위해서 대부분의 분석/합성 방법(Analysis by Synthesis)이 널리 이용되고 있으며 다양한 채널에러와 배경 잡음에 강인한 음성부호화기를 설계하는 기법들이 연구되고 있다.

2. 이동통신 시스템용 음성 부호화기 표준화

2세대 디지털 이동통신 시스템의 개발로 디지털



(그림 1) LPC 음성 부호화 모델링 방식

음성 부호화기가 표준화 단체로부터 주목을 받기 시작하였고 이동통신 시스템에 적합한 이동통신용 부호화기의 표준을 위하여 각 통신 시스템 표준화 단체 산하 분과에서 음성 부호화기 표준화를 담당하였다.

3GPP 표준화 단체에서 3세대 이동통신 시스템 개발이 되면서 이 시스템에 적합한 음성 부호화기가 개발되었다. 3GPP 산하 SA(System Aspect) WG4(Working Group 4)에서는 1999년 유럽형 3세대 이동통신 시스템인 WCDMA(Wideband Code Division Multiple Access)에서 사용될 협대역 음성호화기로 AMR 음성 부호화기를 표준화하였다[3]. AMR 음성부호화기는 최소 4.75 kbit/s의 전송속도를 갖는 모드부터 시작하여 최고 12.2 kbit/s의 전송속도를 갖는 모드까지 총 8개 모드 중 하나로 음성을 부호화한다. 각 부호화 모드는 무선 채널의 상태나 네트워크의 상태에 따라 변경된다. 예를 들어 무선 채널 환경이 나빠지거나 혹은 많은 통화 접속이 이루어지는 환경에서는 낮은 전송 속도로 부호화가 이루어지고 반대의 경우는 높은 전송 속도로 음성을 부호화하여 전반적인 음성 품질을 높이고 이동통신 시스템 효율을 높일 수 있는 장점이 있다.

2001년 3GPP에서는 WCDMA Release 4 규격

의 일환으로 AMR-WB를 표준화 하였다[4]. 이 음성 부호화기는 7kHz 음성 대역폭의 신호를 부호화가 가능하여 이동통신 시스템에서 획기적인 통화 품질을 개선할 수 있을 뿐 아니라 음성 이외의 다양한 멀티미디어 서비스를 제공할 수 있다. 이 음성 부호화기는 ITU-T에서도 IP 텔레포니 응용 분야로 표준화가 되었기 때문에 향후 이동통신 시스템과 IP 텔레포니 시스템간의 호환성도 보장된다.

북미 통신 산업의 표준화를 담당하는 TIA에서는 1989년 북미 TDMA(Time Division Multiple Access) 이동통신 시스템의 음성부호화기로 VSELP(Vector Sum Excited Linear Prediction)을 표준화하였다. 이 음성부호화기는 8 kbit/s 전송속도로 동작하며 여기신호는 128개 벡터들의 합으로 표시하는 특징이 있다. 기술 발전이 이루어짐에 따라 1998년 VSELP 음성부호화기를 대체하는 8kbit/s 전송속도의 ACELP 기반의 IS-641 음성부호화기를 표준화 하였다. 1994년 TIA는 북미 CDMA 이동통신 시스템 개발을 위하여 QCELP(Qualcomm Code Excited Linear Prediction)라는 음성부호화기를 IS-96으로 표준화 하였다[5].

QCELP는 CDMA 특성을 살리기 위해 입력 음성

〈표 1〉 이동통신용 음성 부호화기 특징

음성부호화기	표준화 연도	전송속도	코딩알고리즘	알고리즘 지연	응용분야
GSM FR 코덱	1988	13 kbit/s	RPE-LTP	20 msec	GSM
GSM HR 코덱	1995	3.45 kbit/s	PSI-CELP	24.4 msec	GSM
EFR GSM 코덱	1996	12.2 kbit/s	ACELP	20 msec	GSM, TDMA
AMR	1999	4.75 ~ 12.2 kbit/s	ACELP	25 msec	GSM, WCDMA
AMR-WB	2001	6.6 ~ 23.85 kbit/s	ACELP	25 msec	GSM, WCDMA
IS-54	1989	8 kbit/s	VSELP	20 msec	TDMA
IS-96	1994	8 kbit/s (VBR)	QCELP	24 msec	CDMA
IS-733	1995	13 kbit/s (VBR)	QCELP	24 msec	CDMA
IS-127	1996	8 kbit/s (VBR)	RCELP, ACELP	30 msec	CDMA
SMV	2000	8.5 kbit/s (VBR)	eX-CELP	25 msec	CDMA 2000
VMR-WB	2004	12.65 kbit/s (VBR)	ACELP	35 msec	CDMA 2000
PDC FR 코덱	1990	6.7 kbit/s	VSELP	20 msec	PDC
PDC HR 코덱	1995	5.6 kbit/s	PSI-CELP	40 msec	PDC
PDC EFR 코덱	1999	6.7 kbit/s	ACELP	20 msec	PDC, GSM

에 따라 전송 속도를 변화하는 특성을 가진다. 목음이 50% 이상 차지하는 양방향 통화 특성에 따라 목음 구간의 전송속도를 줄임으로써 무선 주파수 자원을 더욱 효율적으로 이용할 수 있다. 그러나 낮은 음성 품질로 인한 문제가 제기되어 개선된 음성 품질을 보여주는 13kbit/s 전송속도의 QCELP 음성 부호화기가 1995년 개발되어 IS-733으로 표준화 되었다. 또한 8 kbit/s의 전송 속도로 유선 전화 품질의 서비스가 가능한 EVRC 음성부호화기가 CDMA 이동통신에서 IS-127로 채택되었는데 EVRC 음성 부호화기는 여기신호 발생을 위하여 RCELP(Relaxed Code Excited Linear Prediction) 방식과 ACELP 방식을 혼합하여 채택하였다[6].

북미 3세대 이동통신 시스템용 음성 부호화기를 개발하기 위하여 3GPP2 TSG-C(Technical Study Group - C) WG1.1(Working Group 1.1)에서 SMV(Selectable Mode Vocoder)를 2000년 표준화 하였다[7]. 주파수 채널 환경과 네트워크 환경에 따라 CDMA 이동통신 Rate-set I에서 프리미엄 모드, 표준 모드, 이코노미 모드로 나누어지게 되며 SMV 음성 부호화기는 각 모드에 맞게 최적의 음성 부호화 전송속도로 부호화하게 된다. 2004년에는 북미 3세대 이동통신 시스템에 채용할 VMR-WB(Variable-rate Multi-mode Wideband) 광대역 음성부호화기가 표준화 되었다[8]. 이 음성부호화기는 CDMA 이동통신 Rate-set II에서 광대역 음성부호화가 가능하도록 설계 되었다. 또한 유럽의 3세대 이동통신에서 채택한 AMR-WB와 상호 호환이 가능하다.

일본의 경우, PDC(Personal Digital Cellular) 시스템에 적용한 PDC-FR 음성부호화기는 1990년에는 6.7 kbit/s 전송속도의 VSELP를 사용하였으나 1999년부터 ACELP 계열의 PDC-EFR 음성부호화기로 바꾸어 사용하고 있으며, PDC-HR 음성부

호화기로는 3.45 kbit/s PSI-CELP(Pitch Synchronous Innovation CELP)를 사용하고 있다. <표 1>은 이러한 이동통신용 음성부호화기들의 특징을 나타내었다.

3. 기타 표준화 단체 음성 부호화기

ITU-T는 1972년 64 kbit/s 전송 속도의 G.711 음성 부호화기를 표준화하였다. G. 711은 8kHz의 표본 주파수로 음성을 샘플링하고, 한 샘플당 8bit를 할당하여 PCM 코딩을 수행한다. 이 음성 부호화기는 다시 2가지 방식으로 나누어지는데, 미국과 일본 지역은 mu-law PCM 방식을 사용하고 유럽은 A-law PCM 방식을 사용한다. 이 두 가지 방식은 상용 전자 교환기간 음성 통화에서 가장 기본적으로 사용하는 부호화 방식이다.

그 후, ADPCM(Adaptive Differential PCM) 방식을 이용한 G.726이 표준화되었다. 이 음성부호화기는 전송속도에 따라 16, 24, 32, 40 kbit/s로 나뉘어진다. 1992년에는 16 kbit/s 전송속도의 G.728 LD-CELP(Low Delay - Code Excited Linear Prediction) 음성부호화기를 표준화하였고, 1996년에는 두 개의 전송 속도를 갖는 G.723.1 음성부호화기를 표준화하였는데[9], Algebraic 코드북의 여기신호 구조의 5.3 kbit/s 와 MP-MLQ(Multi - Pulse Maximum Likelihood Quantization) 구조의 6.3 kbit/s 전송 속도를 갖는다.

또한 같은 해에 8 kbit/s 전송 속도를 가지는 CS-ACELP (Conjugated Structure - Algebraic CELP) 음성 부호화기를 표준화 하였다[10]. 이 방식은 15 msec 의 짧은 알고리즘 지연 시간을 가지고 있으며 최초로 8 kbit/s 급의 전송속도에서 Toll 음질을 보여준 음성부호화기이다. G.723.1과 G.729 음성 부호화기는 특히 IP 텔레포니 시스템에 널리 사용

되어 왔다. 앞서 언급한 음성 부호화기들은 협대역 음성 부호화기의 표준화 상황이다.

1988년 음성의 주파수 대역폭이 7kHz이고 전송 속도가 64 kbit/s 인 SB-ADPCM(Sub Band - ADPCM) 음성부호화기가 G.722로 표준화가 되었다. 1999년에는 24 kbit/s 및 32 kbit/s 두개의 전송 속도를 가지고 MLT(Modulated Lapped Transform) 방식을 사용하는 G.722.1 음성부호화기를 표준화하였으며, 2002년에는 3GPP에서 표준화가 완료된 AMR-WB(Adaptive Multi Rate - Wideband) 음성부호화기의 일부 모드를 G.722.2로 표준화 하였다. 광대역 음성부호화기는 상호 통화시 음성의 자연성(Naturalness) 및 인지성(Intelligibility)을 높여 자연스러운 대화가 가능할 뿐만 아니라 낮은 수준의 음향레벨의 응용분야까지 영역을 확장 시킬 수 있어 미래 통신 시스템의 음성 통화에 높은 이용이 기대된다.

국제 해상 위성(INMARSAT)은 위성을 통한 양방향 통화 서비스를 제공하는데, Inmarsat-B 시스템으로는 16kbit/s APC-MLQ(Adaptive Predictive coder-MLQ), Inmarsat-Aero(Skyphone) 시스템에서는 9.6kbit/s MP-LPC(Multi-pulse-LPC), Inmarsat-M 시스템에서는 4.15 kbit/s IMBE(Improved Multi Band

Excitation) 음성부호화기를 사용하고 최근 성능이 향상된 AMBE(Advanced MBE)를 사용하고 있다 [11].

미국 국방성 DOD(Department of Defense)는 군사용 목적의 표준 음성부호화기로 전송 속도는 낮지만 음성의 명료성이 높은 2.4 kbit/s LPC-10e 음성부호화기(FS-1015)를 1984년에 표준화 하였고 1990년 4.8 kbit/s 전송속도의 DOD CELP(FS-1016)를 표준화하였다. 이후 2.4 kbit/s MELPe(Enhanced Mixed Excitation Linear Prediction)를 표준화 하였다[12].

2004년 IETF(Internet Engineer Taskforce Team) AVT(Audio Video Transport) WG에서 iLBC(Internet Lowbit Rate Codec)을 표준화 하였다[13]. iLBC는 기존의 CELP 음성부호화기가 이전 프레임의 메모리들을 사용함으로써 생기는 에러의 영향을 최소화하기 위하여 이전 프레임의 메모리를 사용하지 않고 현재 프레임에 대해서만 부호화함으로써 인터넷상에서 빈번히 일어나는 패킷손실이나 패킷지연에 매우 강인하게 설계되어 향후 VoIP등의 인터넷 멀티미디어 응용에 사용될 예정이다. <표 2>는 기타 표준화 음성부호화기들의 특징을 나타내었다.

<표 2> 기타 표준화 음성 부호화기 특징

음성부호화기	표준화 연도	전송속도	코딩알고리즘	알고리즘 지연	응용분야
G.711	1972	64 kbit/s	PCM	-	유선통신
G.726	1990	16, 24, 32, 40 kbit/s	ADPCM	-	유선통신
G.728	1992	16 kbit/s	LD-CELP	0.625 msec	유선통신
G.723.1	1996	5.3, 6.3 kbit/s	ACELP, MP-MLQ	37.5 msec	VoIP
G.729	1996	8 kbit/s	CS-ACELP	15 msec	VoIP
G.722(WB)	1988	64 kbit/s	SB-ADPCM	-	유선통신
G.722.1(WB)	1999	24, 32 kbit/s	ATC	40 msec	VoIP
G.722.2(WB)	2003	6.6 ~ 23.85 kbit/s	AMR-WB	20 msec	IMT-2000, VoIP
IMBE	1993	6.4 kbit/s	MBE	78.75	INMARSAT
MIL-STD-3005	1996	1.2, 2.4 kbit/s	MELPe	240 msec	Military
RFC-3951	2004	13.3, 15.2 kbit/s	iLBC	30 msec	VoIP

III. 음향 부호화기 기술 동향

1. 음향 부호화 모델링 기술

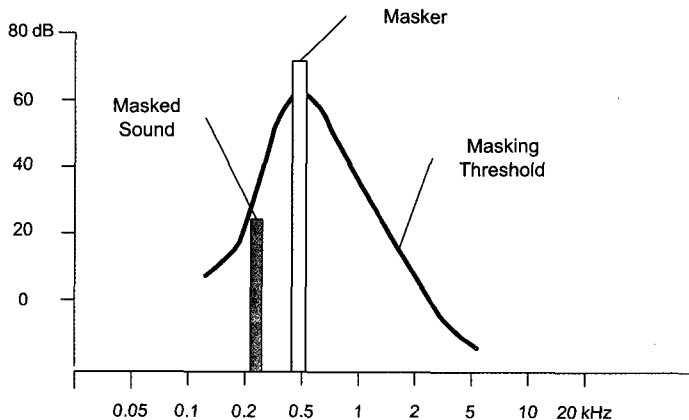
최근의 음향 부호화 방식은 음성부호화기에서 인간의 음원 발생 모델과는 달리 마스킹 현상을 이용하는 심리 음향 모델 (Psychoacoustic Modeling) [14] 혹은 양이 청취 모델링 (Binaural Hearing Modeling) 기법 [15] 등 인간의 청각 특성을 이용한 부호화 방식이다. 심리 음향 모델은 (그림 2)와 같이 마스크에 의하여 마스크된 신호들은 마스크 문턱치 아래에 위치함으로써 인간의 청각으로는 들리지 않게 되는 특성을 이용하여 주파수 영역에서 신호의 크기에 대한 정보를 부호화하게 된다. 양이 청취 모델링 기법은 하나의 음원이 양쪽 귀에 도달하는 시간과 크기가 다름을 이용하여 가상의 멀티 채널 음향신호를 부호화하는 방식에 이용된다.

2. MPEG 음향 부호화 기술

ISO/IEC(International Organization for

Standardization / International Electrotechnical Commission) 산하 MPEG-Audio 분과에서는 통신 방송용 음향부호화기 표준화를 수행 하여 왔다. 1992년 저장 매체를 위한 음향부호화 방식으로 3개의 계층(Layer)을 갖는 MPEG-1 음향부호화기를 표준화 하였다. MPEG-1은 복잡도와 음향품질에 따라 3개의 계층(layer) 구조를 가지는 음향부호화기이다. 계층 구조 3 (MPEG-Layer3, MP3)의 경우 192kbps에서 CD 음향 품질을 가진다[16].

1994년 디지털 방송용 음향 부호화기로 확장 개발된 MPEG-2 는 MPEG-1과 호환이 가능한 MPEG-1 BC (Backward Compatible)와[17] MPEG-2 AAC (Advance Audio Coding) [18]으로 나눌 수 있다. MPEG-1 BC의 경우 MPEG-1과 호환이 가능할 뿐 아니라 5개의 음향채널과 1개의 저음 영역 채널을 지원하는 멀티 채널 기능이 포함되었다. MPEG-2 AAC 는 MPEG-1과 호환은 되지 않지만 채널당 64 kbit/s 전송속도만으로도 높은 품질을 얻을 수 있다. 특히 고해상도 필터뱅크, 시간영역에서 노이즈 셰이핑, 백워드 예측 기법 등을 사용하여 부호화 효율을 높였다. MPEG-2 AAC의 경우 3가



(그림 2) 마스킹 효과

지의 확장 표준이 있는데 메모리나 구현 프로세서에 제약이 없을 경우 사용하는 메인 프로파일, 구현 제약이 있을 경우 사용하는 LC (Low complexity Profile) 프로파일, 비트 전송 속도의 변화에 적용 가능한 SSR (Scalable Sampling Rate) 프로파일이다.

MPEG-4 음향 부호화기는 기존의 단순히 압축 효율을 높이기 위한 부호화 관점에서 통신, 데이터베이스 접근, 양방향 멀티미디어 서비스와 같은 어플리케이션에 쉽게 사용 가능한 부호화 방식을 목표로 만들어 졌다[19]. 특히 음성/음향 부호화, 음성 합성 등을 위한 다양한 툴을 표준화하는 것을 목표로 한다. 음성의 경우 HVXC (Harmonic Vector eXcitation Coding), NB/WB CELP 등을 이용한 부호화 방식이 표준화 되었으며, 음향의 경우 MPEG-2 AAC의 부호화 방식을 기본으로 한 MPEG-4 AAC 가 표준화 되었다. 특히 낮은 전송 속도에서 성능을 최적화하기 위하여 Twin VQ 혹은 PNS(Perceptual Noise Substitution) 같은 툴들이 제공되며, 채널 에러 환경을 위한 에러복원기능, LTP (Long Term Prediction) 과 같은 툴들을 제공하여 성능을 더욱 향상시킨 것이 MPEG-2 AAC와 차이점이다. MPEG-4 BSAC (Bit Slice Arithmetic Coding)은 MPEG-4 음향부호화기에서 16kbps~64kbps 영

역을 1kbps 단위로 Scalability를 주기 위한 압축방식이다.

MPEG aacPlus라고도 불리는 MPEG-4 HE (High Efficient) AAC 는 높은 주파수 대역의 신호 성분과 낮은 주파수 대역의 신호성분과의 연관성을 이용하여 고주파 영역의 성분을 부가적인 파라미터로 표현하는 SBR (Spectral Band Replication) 기술을 사용하였다[20]. 따라서 48 kbit/s 의 전송속도로 128 kbit/s MP3 품질을 보여준다고 알려져 있다. 이 음향부호화기는 높은 부호화 효율로 인하여 주파수 자원이 한정된 디지털 방송, 이동통신 시스템에서 크게 각광을 받고 있다. 특히 DVB-H (Digital Video Broadcasting - Handheld), 3G UMTS (Universal Mobile Telecommunications System) 에서 방송 및 멀티미디어 콘텐츠용 표준 음향 코덱으로 선정되었다.

MPEG-4 HE AAC v2는 aacPlus에 PS (Parametric Stereo) 기법이 추가된 음향 부호화기이다[21]. 이는 모노의 음향신호를 가상의 스테레오 신호로 만들어내어 낮은 전송 속도에서 압축효율을 더욱 높일 수 있다.

MP3 혹은 AAC와 같은 손실 음향부호화기와는 달리 원음의 음향 품질을 보장하는 비손실 음향 부호화 기법인 MPEG-4 ALS (Audio Lossless

〈표 3〉 MPEG 음향 부호화기 특징

음향부호화기	표준화 연도	CD Quality 전송속도	확장 표준	주요특징
MPEG-1	1992	192 kbit/s (2ch)	Layer 1 ~ 3	심리음향모델, 2 channel 디지털 저장매체로 사용
MPEG-2 BC	1994	160 kbit/s (2ch)	Layer 1 ~ 3	5.1 channel, Backward compatible with MPEG-1
MPEG-2 AAC	1997	64 kbit/s (per 1 ch)	Main, LC, Scalable Sampling Rate	Filter bank, Backward Prediction, Noise shaping 기법 사용
MPEG-4 AAC	1998	64 kbit/s (per 1 ch)	-	Twin VQ, LTP, PNS
MPEG-4 BSAC	1999	64 kbit/s (per 1 ch)	-	Bit rate control
MPEG-4 HE AAC	2003	24 kbit/s (per 1ch)	-	SBR
MPEG-4 HE AAC v2	2004	-	-	Parametric Stereo Coding
MPEG-4 ALS	2006	-	-	LPC, Entropy coding

Coding) 기법이 2006년 표준화가 완료되었다. 이 방식을 통하여 원음의 음향품질로 사운드의 저장 및 편집이 가능하다. 또한 AAC와 같은 GA (General Audio codec)와 결합되어 최고의 품질을 나타내는 비손실 음향에서부터 손실 음향 품질까지 Scalable 하게 부호화하는 MPEG-SLS (Scalable Lossless Coding) 방식이 현재 활발히 연구되고 있다. <표 3>은 이러한 MPEG 음향부호화기들의 특징을 나타내었다.

멀티 채널 음향을 모노 혹은 스테레오 신호를 이용하여 효율적으로 압축하는 기법인 MPEG Surround 기술 개발이 현재 진행 중에 있으며 Fraunhofer IIS/Agere 와 Coding Technologies/Philips 결합 모델인 RMO (Reference Model 0)에 대한 표준화가 진행 중이다[22].

IV. 음성/음향 부호화 통합 기술 동향

통신 기술의 발전과 통신/방송 융합 추세에 발맞추어 최근의 음향 부호화기들의 개발 방향은 고품질의 멀티미디어 응용 서비스 혹은 고품질의 양방향 통화 서비스가 가능한 음성/음향 통합 부호화기들이 주목 받고 있다.

현재 ITU-T SG16 (Study Group 16) Q.9 (Question Group 9), Q.10에서는 통신망에서 적용적으로 전송 속도 변화가 가능하고 협대역 음성, 광대역 음성 그리고 음향 신호를 효율적으로 부호화 가능한 음성/음향 부호화기의 표준화를 진행 중에 있다.

현재 SG 16에서 진행중인 표준화 활동을 살펴보면 G.722.1의 확장 표준으로 14kHz의 대역폭을 가지는 음향신호 (Super-Wideband)에 대한 부호화

방식을 표준화 중에 있다. 또한 G.729를 코어 코덱으로 사용하고 비트율 및 주파수 대역이 가변적으로 변화하는 G.729EV에 대한 표준화가 진행 중에 있으며 이 표준화에 참여한 업체들로부터 제안된 음성부호화기의 성능을 검증하는 작업이 (Characterization Process) Q.10에서 현재 수행 중에 있다. Q.9에서는 EV-VBR 코덱에 대한 요구사항에 대한 협의를 수행 중에 있으며 14kHz의 대역폭을 가지는 Super-Wide band 모드를 EV-VBR 코덱에 포함 시키기로 최근 결정하였다.

3GPP에서는 SA WG4에서 PSS (Packet Switched Streaming), MMS (Multimedia Messaging Service), MBMS (Multimedia Broadcast Multicast Service)용 음성 및 음향 부호화기에 대한 표준화를 수행하였다.

최근 두개의 음성/음향 부호화기에 대한 표준화가 완료되었는데, AMR-WB+ 음성부호화기는 ACELP와 TCX (Transform Coded Excitation) 기술을 이용하여 6 ~ 48 kbit/s의 전송속도와 20 kHz의 주파수 대역을 갖는 음성 혹은 음향에 대한 부호화를 수행할 수 있으며 기존의 AMR-WB와 상호 호환이 가능하다. 또한 높은 수준의 음향부호화를 위하여 MPEG HE AAC v2를 표준화 하였으며 이는 SBR기술과 BCC 방식을 이용하여 3G 이동통신 시스템에서 다양한 멀티미디어 콘텐츠를 부호화하는 데 이용될 예정이다[23] [24].

OMA (Open Mobile Alliance) BCAST (Broadcast) STI (Standard Transcoding Interface) 분과에서는 다양한 네트워크에서 멀티미디어 데이터들의 호환을 위하여 표준화된 멀티미디어 변환 인터페이스 및 변환 엔진에 대한 연구가 진행 중에 있다. 현재 서로 다른 기기종류 간에 영상, 음향, 음성의 효율적인 전송 및 변환을 위한 표준화 작업이 진행 중에 있다.

작년 MPEG-Audio 분과에서 Scalable Speech / Audio 부호화기에 대한 표준화 요구 움직임이 있었다. 이는 지금까지 음성 및 음향으로 구분된 부호화 방식을 하나의 부호화기로 통합하는 추세에 발맞추어 다양한 멀티미디어 사운드 콘텐츠들을 하나의 코덱으로 수용하자는 움직임이다. 또한 이 부호화기는 콘텐츠의 종류 및 네트워크의 상황에 맞게 비트스트림 Scalability를 갖추어야 한다. 현재 MPEG에서는 Scalable Speech / Audio Codec에 대한 연구 방향을 협의 중에 있다.

V. 결 론

본 논문에서는 지금까지 표준화되거나 현재 표준화 진행중인 음성 및 음향 부호화기에 대하여 그 특징과 응용분야에 대하여 설명하였다. 디지털 음성 및 음향 부호화기는 고품질의 양방향 통화 서비스, 디지털 멀티미디어 서비스, 디지털 방송 서비스등과 같은 어플리케이션에서 그 수요가 최근 들어 더욱 높아지고 있다. 음성부호화기는 그 성능과 전송 속도 그리고 통신 시스템의 특성에 맞게 최적으로 설계 되었으며 성공적으로 사용되어 왔다.

또한 음향 부호화기의 경우 디지털 음향 저장 장치에서 DAB(Digital Audio Broadcasting), HDTV(High-Definition TV), DMB(Digital Multimedia Broadcasting)과 같은 방송 응용분야에서 큰 각광을 받아 왔다. 그러나 앞으로 도래할 차세대 유무선 통합망 그리고 통신 방송 융합 시스템에서는 고품질 통화서비스, 패킷 통신용 VoIP 서비스, 멀티미디어 스트리밍 서비스, 휴대용 방송 서비스 같은 응용 서비스 분야에서 음성 및 음향서비스에 대한 기대치와 요구 사항이 더욱 높아지고 있다.

예를 들어, 현재 IMT-2000 서비스를 위해 할당

된 주파수는 송수신 합쳐 230MHz로서 많은 대역이 할당되어 있으나, IMT-2000 시스템을 사용하게 될 가입자 수가 매우 많을 것이고 제공될 서비스도 상당히 다양할 것이다. 그리고 패킷 네트워크를 통한 전화 서비스로서 현재 각광을 받고 있는 VoIP 시스템의 다음 버전인 차세대 VoIP 시스템 혹은 MoIP (Multimedia over IP network)의 가장 중요한 특징 중 하나는 고품질 음성 및 음향이므로, face-to-face 음질에 해당하는 음성 서비스와 CD 품질의 음향 서비스 제공은 필연적이다.

유무선 통합 네트워크를 이용한 디지털 방송 서비스에서는 scalability 기능을 가지면서 높은 품질과 낮은 전송 속도를 가지는 음향 부호화기가 필요하다. 또한 휴대 단말기의 통신 방송 융합화가 급격히 진행되고 있는 가운데 하나의 프로세서에서 음성과 음향을 모두 처리할 수 있는 음성/음향 통합 코덱의 등장도 멀지 않았다.

그러나 네트워크의 트래픽 상태와 단말의 처리속도에 따라 최적의 음질을 제공하는 embedded 개념의 음성 및 음향 부호화 기술에 대한 국내 연구는 아직까지 미미한 실정이므로, 관련 핵심 기술 개발을 통한 지적 재산권 확보와 음성/음향 부호화기 분야에서 기술적 우위를 선점하여 현재 우리가 이제껏 쌓아온 통신 방송 기술의 산업 경쟁력을 더욱 강화 시켜야 할 것이다.

[참 고 문 헌]

- [1] W.B. Kleijn and K.K. Paliwal, Speech Coding and synthesis, Elsevier, 1995.
- [2] N.S. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception", Proc. IEEE, Vol. 81,

- pp. 1385–1422, Oct. 1993
- [3] 3GPP Rec. TS 26.090 “AMR speech coder; Transcoding functions,” Jun. 2002.
- [4] 3GPP Rec. TS 26.190. “AMR Wideband speech codec; Transcoding function,” Jan. 2005.
- [5] TIA/EIA IS-96 “Speech Service Option Standard for Wideband Spread Spectrum Digital Cellular System,” Apr. 1994.
- [6] TIA/EIA/IS-127, “Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems,” Jan. 1997.
- [7] TIA/EIA/IS-893, “Selectable mode vocoder service option for wideband spread spectrum communication system,” Jun. 2005.
- [8] TIA-1016. “CDMA2000 wideband codec algorithm description,” Oct. 2004.
- [9] ITU-T Rec. G.723.1, “Dual-rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s,” May 1996.
- [10] G.729, “Coding of speech at 8 kbit/s using conjugate-structure algebraic code-excited linear prediction (CS-ACELP),” May 1996.
- [11] S. Dimolitsas et al. “Evaluation of Voice Codec Performance for the Inmarsat Mini-M System,” Proceedings, 10th Int. Digital Satellite, Brighton, England, May, 1995.
- [12] L. Supplee, et al. “MELP: The new federal standard at 2400 bps,” ICASSP-97, Munich, Germany
- [13] S. V. Andersen et al. “iLBC – a linear predictive coder with robustness to packet losses,” Speech Coding, IEEE Workshop Proceedings, pp. 23–25, Oct. 2002.
- [14] E. Zwicker and H. Fastl, Psychoacoustics, Springer-Varlag, 1990
- [15] J. Blauert, Spatial Hearing, The MIT Press, 1983
- [16] ISO/IEC JTC1/SC29/WG11 No. 71 “Coding of Moving Picture and Associated Audio for Digital Storage Media at up to about 1.5Mbit/s – CD 11172-3 (Part 3. MPEG-Audio)”
- [17] ISO/IEC 13818-3:1998, Information technology – Generic coding of moving pictures and associated audio information – Part 3 : Audio.
- [18] ISO/IEC 13818-7:1997, Information technology – Generic coding of moving picture and associated audio information – Part 7 : Advanced Audio Coding (AAC)
- [19] ISO/IEC 14496-3:2001, Information technology – Coding of audio-visual objects – Part 3 : Audio
- [20] M. Dietz, L. Lijeryd, K. Kjorling, O. Kunz, “Spectral Band Replication, a Novel Approach in Audio Coding,” 112nd AES Convention, Munich 2002, Preprint 5553.
- [21] ISO/IEC SC29/WG11 N6675, “Report on the Verification Tests of MPEG-4 Parametric Coding for High Quality Audio”, Redmond, USA, July 2004.
- [22] ISO/IEC SC29/WG11 N7138, “Report on

MPEG Spatial Audio Coding RM0 Listening Tests”, Busan, Korea, April 2005

- [23] 3GPP Rec. TS 26.290, “Audio codec processing functions; Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions,” 2004.
- [24] 3GPP Rec. TS. 26.401, “General audio codec audio processing functions; Enhanced aacPlus general audio codec; General description,” 2004.



윤병식

1990년 경북대학교 공과대학 전자공학과 졸업 (학사)

1992년 경북대학교 공과대학 전자공학과 졸업 (석사)

1992년 ~ 현재 한국전자통신연구원 이동통신연구단 선임연구원

2004년 ~ 현재 한양대학교 전자컴퓨터 공학부 박사과정

관심분야 : 음성 코딩, 멀티미디어 신호처리, 이동통신 시스템



최송인

1982년 광운대학교 응용전자공학과 졸업 (학사)

1987년 광운대학교 전자계산기공학과 졸업 (석사)

1982년 ~ 현재 한국전자통신연구원 이동통신연구단 책임연구원 (현 융합단말연구팀장)

관심분야 : 음성 신호처리, 영상 신호처리, 이동통신 시스템



강상원

1980년 한양대학교 전자공학과 졸업 (학사)

1982년 서울대학교 전자공학과 졸업 (석사)

1982년 ~ 1994년 한국전자통신연구소 근무

1990년 Texas A&M대 전기공학과 졸업 (박사)

1994년 ~ 현재 한양대학교 전자컴퓨터공학부 교수

관심분야 : 음성코딩, 신호처리 및 디지털통신