

음성 및 오디오 통합 부호화에 관한 MPEG 표준 기술 동향

*김민제, **이태진, ***서정일, ****강경옥

한국전자통신연구원

*mkim@etri.re.kr, **tjlee@etri.re.kr, ***seoji@etri.re.kr, ****kokang@etri.re.kr

On Current MPEG Standardization Process about Universal Speech and Audio Coding

*Minje Kim, **Taejin Lee, ***Jeongil Seo, ****Kyeongok Kang

Electronics and Telecommunications Research Institute

요 약

본 논문은 MPEG audio group 에서 진행중인 음성/오디오 통합 부호화 기술 관련 표준화 동향을 기술한다. 먼저 현재까지의 디지털 오디오 신호 압축 기술 표준화 진행 현황에 대한 분석을 하며, 이를 위해 MPEG-4 AAC 오디오 부호화 기술의 잡음 및 순음 코딩 모듈, 고주파수 대역확장을 통한 압축 성능 향상 모듈, 채널간 정보 파라미터를 이용한 스테레오 채널 신호 압축 모듈이 추가된 MPEG-4 HE-AAC v2 코덱을 기술한다. 또한 이러한 성능 향상에도 불구하고, 주로 음성 신호 입력과 낮은 전송률 환경에서 성능이 낮아지는 HE-AAC v2 코덱의 성능 향상을 위해 진행중인 음성/오디오 통합 부호화 표준화 동향을 기술하며, 관련해서 목표로 선정된 입력 신호에 대한 분석 및 비교 코덱인 AMR-WB+와의 성능 비교를 수행한다.

1. 서론

최초의 상용 디지털 오디오 매체인 CD(Compact Disk)는 1976년 상용화 이후 약 20년 이상 상용 오디오 시장에서 고품질의 음질을 손실 없이 제공하며 시장을 주도해 왔다. 그러나 CD는 transparent한 음질을 제공하기 위해 기본적으로 압축 없이 PCM 신호를 저장하고 있으며, 이에 따라 1.4Mbps라는 높은 비트 전송률을 가지고 있다. 이는 유통되는 광학 기록 매체에 저장할 수 있는 콘텐츠 양에 한계를 가지고 있으며, 방송/통신 등의 응용에서 사용되기에 어려움이 있다는 단점으로 작용하였다. 따라서 CD와 같은 고품질의 음질을 유지하면서도 데이터 전송률을 낮추기 위한 디지털 오디오 신호 압축 기술이 발전해 왔다.

국제적으로 표준화된 오디오 부호화 기술 개발의 필요성은 다양한 기술 연구 기관과 ITU, ISO/IEC 등의 국제 표준화 단체의 공동 표준화 요구를 이끌어 내었으며, 이에 따라 1992년 MPEG-1 오디오 부호화 기술이 표준화 완료되었다. MPEG-1 오디오 부호화 기술은 품질, 지연시간, 복잡도, 전송률 등의 여러 가지 요소를 고려한 3개의 계층으로 구성되어 있으며, 특히 192kbps에서 원음과 차이가 없는 음질을 제공하는 계층 3은 MP3(MPEG-1 Layer 3)라는 확장자 명칭으로 더 많이 알려져 디지털 오디오 음원의 새로운 시장 영역을 확보하여 기존의 CD 매체를 대체하기에 이르렀다 [1][2].

또한 1994년 표준화된 MPEG-2 BC(Backward Compatible)가 MPEG-1과의 호환성을 제공하기 위해 [3] 압축률 향상에 제약이 있었던 점을 해소하기 위해 1997년에 표준화된 MPEG-2 AAC(Advanced Audio Coding)은 TNS(Temporal Noise Shaping), 예측(prediction) 등 호환성을 고려하지 않은 다양한 새로운 틀을 적용하였고, 결과적으로 채널당 64kbps의 낮은 전송률로도 transparent한 음질을 제공할 수 있었다 [4]. MPEG-2 AAC는 일본 디지털 위성 방송의 표준 등 다양한 방송환경의 오디오 표준으로 사용되고 있

다 [5].

2001년에 표준화된 MPEG-4 오디오 기술은 MPEG-2 AAC를 더욱 확장하여 PNS(Perceptual Noise Substitution), LTP(Long-term Prediction) 등의 새로운 부호화 도구를 추가하였으며 [6], 또한 오류에 대한 강인성을 높이고, 계층 부호화를 위해 BSAC(Bit-Sliced Arithmetic Coding) 등의 도구를 추가하였고, 특히 고대역 신호의 효율적 부호화 도구인 SBR(Spectral Band Replication), 채널 신호의 파라미터 표현 방식인 PS(Parametric Stereo) 기법을 추가하여 저 비트율에서도 우수한 음질을 제공한다 [7][8]. AAC 기술에 SBR과 PS 기법을 추가한 기술을 각각 HE-AAC(High Efficiency AAC)와 HE-AAC v2(HE-AAC version 2)라 한다. AAC 기반 코딩 방식은 최근 Apple의 온라인 음악 다운로드 서비스인 iTunes에서의 부호화 방식으로 채택되면서 온라인 음악시장 활성화에 기여하였다.

최근 MPEG에서는, 24kbps 이하의 낮은 전송률에서 특히 음성 신호의 압축시 음질이 열화되는 HE-AAC v2의 특성을 보완하기 위해, 음성/오디오 통합 부호화 기술인 USAC(Universal Speech and Audio Coding)에 관한 부호화 기술이 진행중이며, 이는 12kbps ~ 64kbps까지의 다양한 전송률에서 음성 및 오디오 신호에 대해 공히 고른 음질과 압축 성능을 보이는 것을 목표로 하고 있다.

본 논문에서는 먼저 기존 AAC를 기반으로 하는 오디오 코딩 방식에 대해 기술하고, 현재 MPEG에서의 USAC 표준화 동향에 대해 설명한다.

II. AAC 기반 오디오 부호화 기술[9]

1. MPEG-4 오디오 부호화 기술

MPEG-4 오디오는 채널당 16kbps ~ 64kbps의 비트율에서, AM보다 우수한 음질에서 CD와 동등한 음질까지를 지원한다. MPEG-4 일반 오디오는 양자화와 부호화 방식에 따라 MPEG-4 AAC, Twin-VQ [10] 및 BSAC [11]으로 나눌 수 있다.

이중 MPEG-4 AAC는 그림 1과 같이 일반적인 지각 오디오 부호화(perceptual audio coding) 방식으로 구성된다. 먼저 시간-주파수 변환된 각 프레임은 마스킹 현상과 같은 인간의 청각특성(human auditory system)을 기반으로 지각에 불필요한 정보를 제거하게 된다. 추가적으로, 양자화 과정을 통해 통계적인 신호 중복성을 제거한 다음, 최종 부호화된 오디오 데이터를 생성한다.

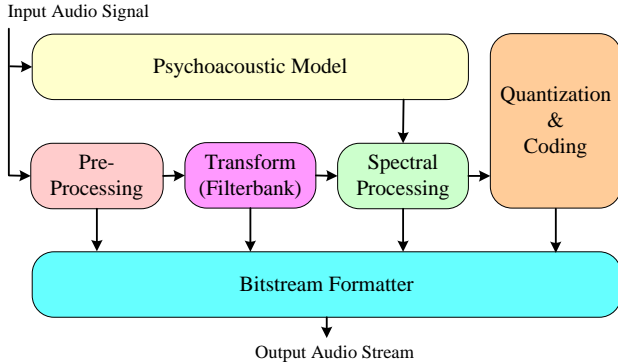


그림 1. MPEG-4 AAC 구성도

MPEG-4에 추가된 대표적인 부호화 도구는 PNS와 LTP이다. PNS는 청각적으로 잡음 특성을 가지는 성분을 파라미터로 표현하여 부호화 효율을 높이는 것을 목적으로 하며, transient가 많은 음성 신호 및 낮은 전송률에서 효과적인 성능을 발휘한다. LTP는 잡음 성분에 비해 순음 성분의 정밀도가 높은 것이 청각적으로 더 높은 품질의 부호화를 가능하게 한다는 특성을 기반으로 하여, 순음 성분을 효율적으로 부호화함으로써, 피치 특성이 있는 신호에 대해 좀더 효과적인 압축을 가능하게 하는 방식이다.

HE-AAC에서는 AAC 기술에 파라미터 기반으로 고대역을 표현하는 SBR 모듈이 추가되었으며, 이에 파라미터 기반의 채널 부호화를 통해 스테레오 신호를 효율적으로 부호화하는 기술인 PS를 추가한 것이 HE-AAC v2이다.

2. SBR(Spectral Band Replication)

SBR 기술은 고대역 성분신호를 파라미터로 표현하고 이를 복호화 단계에서 합성하는 방식으로 표현함으로써 부호화 효율을 향상시킨다. 이는 인간의 청각 특성이 고대역 신호에 상대적으로 낮은 해상력을 가진다는 사실에 기반한 것이다. SBR 기술은 core 대역의 부호화에 대해 독립적인 동작이 가능하므로, 기존 오디오 부호화 기술을 낮은 대역의 스펙트럼을 부호화 하고, 고대역은 SBR을 통해 파라미터화 하는 형태로 기존 오디오 부호화 기술과 융합이 가능하다. 이는 MP3와 SBR 기술을 융합한 MP3Pro, AAC와 SBR 기술을 융합한 HE-AAC(또는 aacPlus)로 표준화되었다 [12][13].

그림 2는 SBR을 이용한 HE-AAC 부호화/복호화 과정을 보여준다. SBR 부호화기는 광대역(wideband) 오디오 입력 신호를 QMF(Quadrature Mirror Filter) 분석을 통해 고대역 신호를 포락선으로 표현하는 제어 파라미터와 저대역으로 제한된 오디오 신호를 생성한다. SBR 부호화기에서 대역 제한된 오디오 신호는 AAC의 core 부호화기를 통해 부호화 되고, SBR을 위한 부가 데이터가 이에 포함되어 HE-AAC 비트스트림을 생성한다. 복호화 단계에서는 상기 비트스트림에서 core 대역에 해당하는 스펙트럼을 먼저 생성하고, 고대역 스펙트럼의 포락선 정보를 이용하여 광대역 오디오 신호를 생성한다.

SBR 기술은 절반 이상에 해당하는 고대역 스펙트럼 정보를 부가 정보로 표현하고, 이것의 기술을 위해 1~3kbps 정도의 미미한 데이터를 사용하므로 32kbps 이하에서도

우수한 음질을 제공한다. SBR은 2005년 MPEG-4 오디오의 표준으로 포함되었으며, HE-AAC 프로파일은 AAC LC 객체와 SBR 객체로 구성되어, 기존 AAC 복호화기에 대한 역호환성을 제공한다 [14].

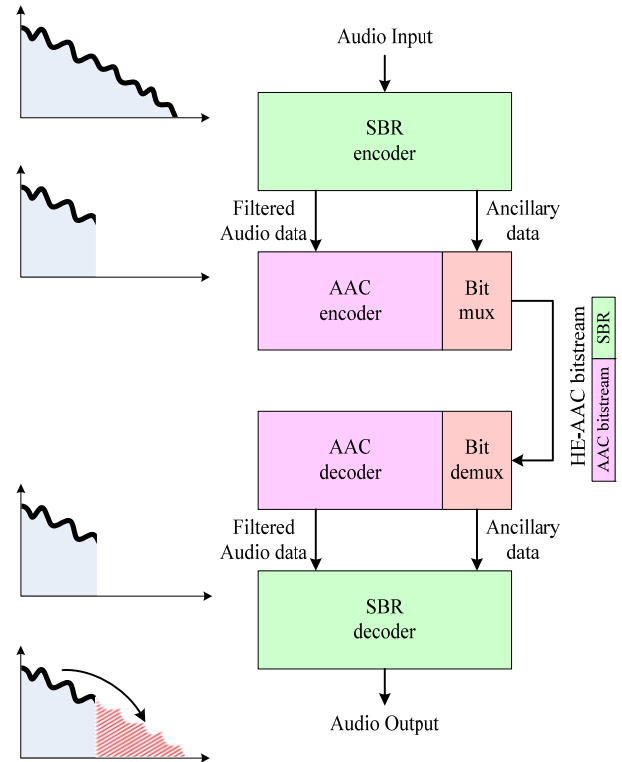


그림 2. HE-AAC 부호화/복호화 과정

3. PS(Parametric Stereo)

PS 기술은 입력 신호의 채널간 관계 정보를 파라미터로 표현하여, 다운믹스된 모노 신호에서 가상의 스테레오 채널을 생성함으로써 원 신호를 복원하는 기술이다 [15]. 그림 3은 PS 기술을 개념적으로 설명하고 있다. PS 부호화기는 스테레오 입력 신호를 분석하여 스테레오 음상을 제어할 수 있는 파라미터를 추출하고 이를 다운믹스된 모노 신호와 함께 전송함으로써, 추후 복호화 단계에서 원본 2채널 신호를 복원한다.

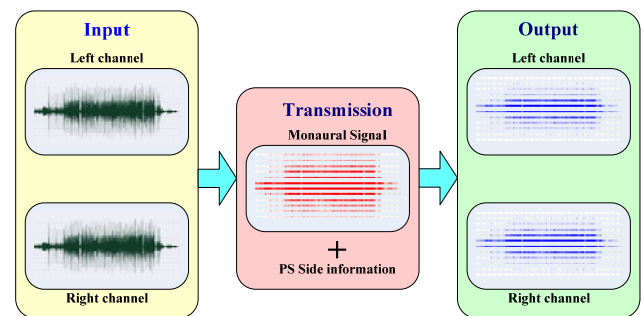


그림 3. Parametric Stereo 개념

이를 위해 사용되는 채널 간 관계 파라미터에는, 채널 간 신호 세기 차이(IID:Inter-channel Intensity Difference), 채널 간 상호 상관관계(ICC:Inter-channel Cross Correlation), 채널 사이의 위상 차이(IPD:Inter-channel Phase Difference) 및 채널 사이의 전체적인 위상 분포에 대한 정보(OPD:Overall Phase Difference) 등이 있다 [16].

PS 기술은 채널당 24kbps 이하의 낮은 전송률에서도 기존 HE-AAC의 성능 저하를 보완하여 고품질의 음질을

제공하고 있으며, HE-AAC 기술에 PS 기술이 추가된 HE-AAC v2는 2005년에 표준화 되었다. PS는 기존 AAC 계열 기술과 역호환이 가능하며, HE-AAC v2 기술은 일본의 지상파 ISDB(Integrated Services Digital Broadcasting), 3GPP(3rd Generation Partnership Project)/3GPP2에서 오디오 표준으로 채택 되었다.

4. HE-AAC 및 HE-AAC v2의 성능

MPEG에서는 성능에 대한 검증을 기본적으로 블라인드 형태의 주관적 음질 평가를 통해 수행하고 있다 [17]. 그림 4은 7kHz로 대역이 제한된 anchor 신호와 18~48kbps의 HE-AAC, HE-AAC V2 및 AAC-LC의 주관적 음질평가 점수의 차이를 나타낸다 [18]. 그림에서 세로축은 7kHz anchor 신호와 HE-AAC, HE-AAC v2, AAC-LC의 점수 차이를 의미하며, AAC-LC에 비해 HE-AAC나 HE-AAC v2는 전송률이 낮아짐에도 음질을 유지함을 알 수 있다. HE-AAC v2 기술은 기존 기술에 비해 낮은 전송률 환경에서도 기존 기술에 비해 상대적으로 우수한 성능을 보여준다.

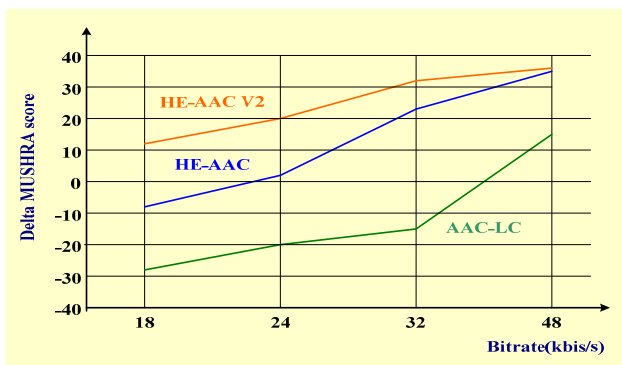


그림 4. HE-AAC/HE-AAC V2 음질평가 결과

II. USAC 표준화 동향

2007년부터 MPEG audio group에서 시작된 USAC 표준

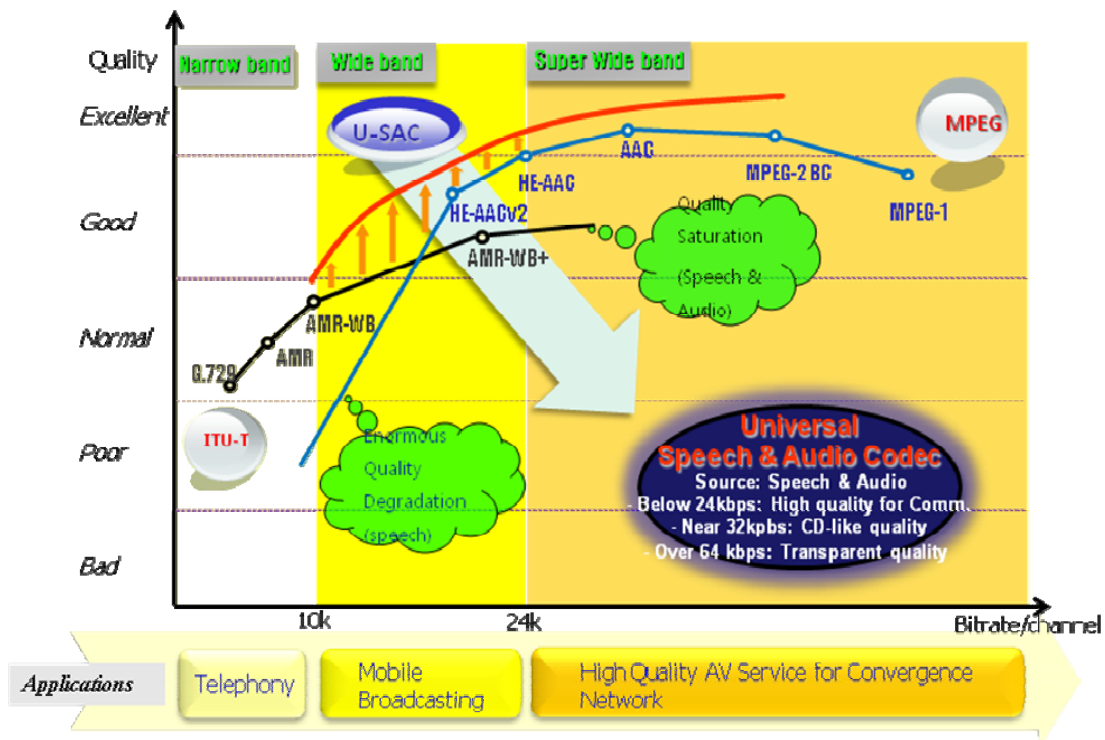


그림 5. VT 성능 현황 및 USAC 성능 요구 사항

화 기술은 그림 5에서와 같은 VC(Virtual Codec)의 요구조건을 만족시키도록 하고 있다. VC는 현재 3GPP에서 표준화된 AMR-WB+ 계열 부호화 방식들과, MPEG의 AAC 계열 부호화 방식들 중 전송률 별로 성능이 가장 우수한 코덱을 동적으로 선택하여 부호화할 수 있는 가상의 상황을 가정한 것이다. USAC의 RM(Reference Model) 채택의 기본적인 조건은 이러한 VC 보다 우수한 성능을 내는 것이며, 이는 연구 기관 별로 특히 두 계열의 코덱이 성능이 병목을 이루는 10kbps ~ 24kbps에서의 성능 향상에 집중하게 하고 있다.

1. 선정된 Test Sequence 및 VT 성능

제 84차 MPEG 회의에서는 향후 제출될 USAC 코덱들의 성능 평가 시 사용될 12개의 test sequence가 선정되었다. Test sequence들은 음성, 음악, 음성/음악 혼합 신호로 구분되어 광범위한 응용에서의 코덱 성능을 측정할 수 있는 것을 목표로 하고 있다. 그림 6, 7은 각각, 20kbps와 24bps mono 환경에서의 AMR-WB+와 HE-AAC v2 코덱의 test sequence 별 본 연구기관 내부에서의 주관적 음질 평가 비교이다.

그림 6, 7에서 실선 위쪽의 x 마크는 각 코덱 별 주관 청취 평가 평균 점수가 AMR-WB+가 높은 쪽임을 나타내며, 실선 아래쪽일수록 HE-AAC v2의 성능이 좋음을 의미한다. 평가 결과에서 알 수 있듯이, PS가 적용되지 않은 mono 신호에 대해서는 전반적으로 AMR-WB+의 성능이 HE-AAC v2보다 우수함을 알 수 있다. 그러나 특정한 음악 신호에 대해서는 HE-AAC v2의 성능이 우수하고, 이러한 경향성은 신호의 음성 신호의 특징이 줄어들수록, 비트 전송률이 높아질수록, 모노 신호에서 스테레오 신호로 바뀔수록 사라져서 HE-AAC v2의 성능이 향상된다.

Test sequence 별 코덱 성능은 표 1에 정리된 바와 같다.

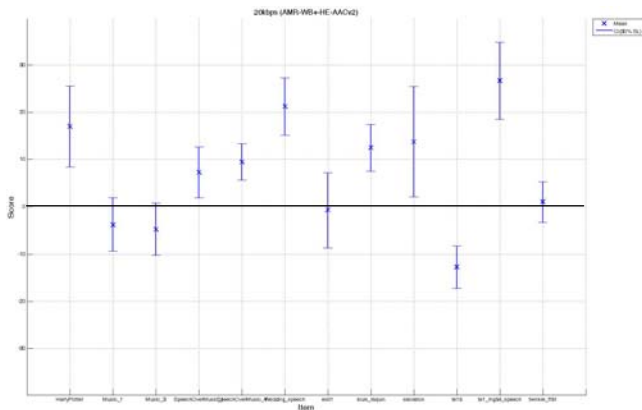


그림 6. 20kbps mono 에서 AMR-WB+와 HE-AAC v2의 test sequence 별 주관 평가 점수 차이

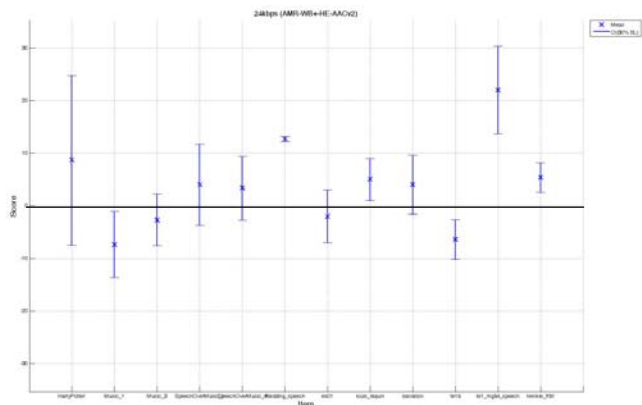


그림 7. 24kbps mono 에서 AMR-WB+와 HE-AAC v2의 test sequence 별 주관 평가 점수 차이

ITEM		20kbps (mono)	24kbps (mono)
speech	es01	AMR=AAC	AMR<AAC
	louis_raquin_15	AMR>AAC	AMR>AAC
	wedding_speech	AMR>AAC	AMR>AAC
	te1_mg54_speech	AMR>AAC	AMR>AAC
mixed	twinkl_ff51	AMR=AAC	AMR=AAC
	SpeechOverMusic_1	AMR>AAC	AMR>AAC
	SpeechOverMusic_4	AMR>AAC	AMR=AAC
	HarryPotter	AMR>AAC	AMR>AAC
music	Salvation	AMR> AAC	AMR>AAC
	te15	AMR< AAC	AMR<AAC
	Music1	AMR<AAC	AMR< AAC
	Music3	AMR<AAC	AMR<AAC

표 1. 주요 비트 전송률에서의 코덱 별 성능 비교

2. 코덱별 성능 분석

실험 결과를 통해 코덱 별 성능을 분석해보면, 일반적으로 음성 특성이 많이 포함된 신호일수록 AMR-WB+의

성능이 우수하며, 고주파 대역의 신호가 포함되어 있거나 transient가 자주 발생하는 신호에 대해서는 HE-AAC v2가 우수함을 알 수 있다. 특히 도시된 20, 24kbps보다 낮은 전송률에서는 AAC core 모듈의 블록 변환된 주파수 계수에 대한 심리음향모델을 통한 비트 할당 방식보다, AMR-WB+ 기반의 ACELP 또는 TCX의 모델링 및 잔차 신호 양자화 방식이 우수하다. 그러나 HE-AAC v2의 성능 향상에 큰 공헌을 하였던 PS와 SBR 모듈에 비해 AMR-WB+의 고주파 대역 코딩 방식과 채널 신호 파라미터 부호화 방식은 그 성능에서 약세를 보이고 있으며, 향후 USAC 표준 기술은 이러한 VT의 단점을 모두 보완할 수 있는 방향으로 진행될 것으로 예측된다.

II. 결론

본 논문에서는 현재 MPEG audio group 에서 진행중인 USAC 표준화 기술과 관련하여, 목표 기술인 HE-AAC v2와 관련 기술에 대한 배경을 설명하였다. 또한 USAC 표준화 기술 성능 평가에 사용될 test sequence 별 최신 코덱 성능을 평가하여 코덱 별로 요구되는 개선 사항을 정리하였다.

감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 IT 원천 기술개발사업의 일환으로 수행하였음. [2008-F-011-01, 차세대 DTV 핵심기술 개발]

참고문헌

- ISO/IEC IS 11172-3 Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbit/s. Part-3: Audio
- 이태진 외 3 인 “MPEG-2 AAC 를 이용한 인터넷 오디오 방송”, SK Telecommunications Review, 제 12 권, 5 호, 2002.
- ISO/IEC IS 13818-3 Generic Coding of Moving Pictures and Associated Audio: Audio
- ISO/IEC IS 13818-7 Generic Coding of Moving Pictures and Associated Audio. Part-7: Advanced Audio Coding (AAC)
- 미키 스케이치 저, 고성제, 김종옥 역, “다양한 영상-음성을 자유자재로 부호화하는 MPEG-4의 세계”, 영풍문고, 1999, pp. 115~124.
- ISO/IEC IS 14496-3, Information Technology-Coding of audio-visual objects – Part 3: Audio, 2001.
- ISO/IEC IS 14496-3, Information Technology-Coding of audio-visual objects – Part 3: Audio/Amd. 1, 2005.
- ISO/IEC IS 14496-3, Information Technology-Coding of audio-visual objects – Part 3: Audio/Amd. 2, 2005.
- 이태진 외 2인, “파라미터 기반의 오디오 부호화 기술 및 표준”, SK Telecommunications Review, 제 17 권, 5 호, 2007.
- N. Iwakami, et al., “High-quality audio-coding at less than 64kbit/s by using transform-domain weighted interleave vector quantization (TWINVQ),” Proc. IEEE ICASSP-95, May, 1995, pp. 3095~3098.
- S-W. Kim, et al., “Fine grain scalability in MPEG-4 Audio,” Proc. 111th AES Convention, preprint #5491, Nov., 2001.
- P. Ekstrand, “Bandwidth Extension of Audio Signals by Spectral Band Replication,” in IEEE Benelux workshop, Nov., 2002.
- Martin Wolters, et al., “A closer look into MPEG-4 High Efficiency AAC,” Proc. 115th AES Convention, Oct., 2003.
- 김인철, “지상파 및 위성 DMB에서의 오디오 부호화 기법”, 방송공학회지, 제 11권, 3호, 2006, pp. 51~63.
- J. Breebaart, et al., “Parametric Coding of Stereo Audio,” EURASIP Journal of Applied Signals Proc., 2005.
- Jimmy Lapierre, Roch Lefebvre, “On Improving Parametric Stereo Audio Coding,” 120th AES Convention, May, 2006.
- ITU-R BS.1534, “Method for the subjective assessment of intermediate quality levels of coding systems”.
- Stefan Meltzer, Gerald Moser, “MPEG-4 HE-AAC V2 – audio coding for today’s digital media world,” EBU technical review, Jan., 2006.