

MPEG 오디오 표준화 동향

*서정일, 강경옥

한국전자통신연구원 방통미디어연구부

e-mail : seoji@etri.re.kr, kokang@etri.re.kr

Status of MPEG Audio Standard

*Jeongil Seo, Kyeongok Kang

Broadcasting and Telecommunication Media Research Department Electronics and Telecommunications Research Institute (ETRI)

Abstract

This paper briefly introduces the current status of MPEG Audio Subgroup activities for standardizing a new audio coding technologies. Currently MPEG Audio Subgroup focused on spatial audio coding tools for compressing multiple audio objects and unified coding tools for presenting the consistence performance on speech and audio signal at the same time. Also a new MAF (MPEG Application Format) for interactive music was introduced at the 84th MPEG meeting.

I. 서론

1990년대 초반부터 시작된 MPEG(Moving Picture Experts Group)에서는 비디오 부호화 기술뿐만 아니라 오디오 부호화 기술을 꾸준히 개발해오고 있다. 표준화를 완료한 후 상용화로 이어져서 사용되고 있는 대표적인 기술이 mp3, AAC, BSAC, HE-AAC 등이 있으며[1]~[4], 국내 디지털 방송 및 온라인/오프라인 멀티미디어 서비스에 폭넓게 적용되고 있다.

2000년도 초반까지 진행된 지각 오디오 부호화(Perceptual Audio Coding) 기술에 기반한 고품질 오디오 부호화 기술이 AAC(Advanced Audio Coding)에 이르러 기술적인 한계에 도달하게 되었다. 그러나 이동방송 및 무선 통신 환경에서의 오디오 데이터의 전송 및 스트리밍 등에 대한 요구사항에 의해서 보다 낮은 비트율로 고품질의 오디오 신호를 압축하기 위한 기술들이 요구되어 오디오 신호의 특성을 파라미터화

하여 압축효율을 높이는 기술들이 표준화되고 있다. HE-AAC (High Efficiency AAC)는 고주파수 대역 신호를 파라미터화하여 표현하는 기술이며, HE-AAC V2 (High Efficiency AAC version 2)는 스테레오 채널 신호의 특성을 파라미터화하여 표현하는 기술이다 [4].

최근에는 3차원 공간상에서 음원을 인지하는 특성을 이용하는 공간 오디오 부호화(Spatial Audio Coding) 기술에 기반한 다채널 및 다객체 오디오 부호화 기술에 대한 표준화가 진행되고 있으며, 오디오 부호화 기술이 날로 진화함에 따라 음성 신호까지 우수한 음질로 부호화하기 위한 기술의 표준화가 논의되고 있다.

본 논문에서는 MPEG 오디오 서브그룹에서 최근에 표준화 되었거나 현재 표준화가 논의되고 있는 기술들에 대해서 소개하고자 한다. 추가로, 지난 84차 MPEG 회의에서 제안된 대화형 음악을 위한 MAF (MPEG Application Format)에 대해서도 간단히 살펴보고자 한다.

II. 본론

3차원 오디오 신호와 같이 다채널 및 다객체 환경에서는 전송에 필요한 대역폭이 채널의 개수나 오디오 객체의 개수에 비례하여 증가한다. 그러나 다채널 및 다객체 오디오 신호를 모노 또는 스테레오 채널로 다운믹스하여 압축하고 객체들과 채널들에 대한 정보는 별도의 부가정보로 표현하여 전송하면 압축효율을 극대화할 수 있다. BCC(Binaural Cue Coding)란 이름으로 Faller 등에 의해 제안된 공간 오디오 부호화 기술은 다채널 또는 다객체 오디오 신호를 최소한의 채

널로 다운믹스한 오디오 신호와 다채널 또는 다객체 오디오 신호의 음향 공간상의 단서(spatial cue)로 이루어진 부가정보로 분할하여 압축하는 파라미터 멀티채널 오디오 부호화 방식이다[5].

자유공간(free field) 상에서 단일 음원에서 발생된 음파는 청취자의 두 귀에 크기차와 도달 시간차를 가지며 도착하게 된다. 청취자는 이를 통하여 음원의 위치를 파악하므로 이를 양이효과(binaural effect)라고 부른다. 이때 두 귀 사이의 음향신호의 크기 차이를 ILD (interaural level difference)라 부르고, 두 귀에 도착하는 신호 도달 시간의 차이를 ITD(interaural time difference)라 부르며 이들을 양이단서(binaural cue)라고 통칭한다. 하나의 가상음원으로 발생하는 모든 음향 신호 성분들은 청취자의 귀에서 일정한 ILD와 ITD 값을 가지므로, 각 음향 신호 성분들은 자신만의 영역을 차지하며 이들의 크기는 전력(power)으로 표현할 수 있다. 만약 그림 1과 같이 세 개의 음원들이 시간-주파수 평면상에서 겹치지 않는 영역에서 존재한다고 가정하면, 이들이 생성하는 음향학적 이미지는 모든 음원들의 합신호를 이용하여 재현할 수 있다. 이러한 공간 오디오 부호화 기술을 이용하여 다채널 오디오 신호를 압축하는 기술이 MPEG Surround이며 다객체 오디오 신호를 압축하는 기술이 SAOC (Spatial Audio Object Coding)이다.

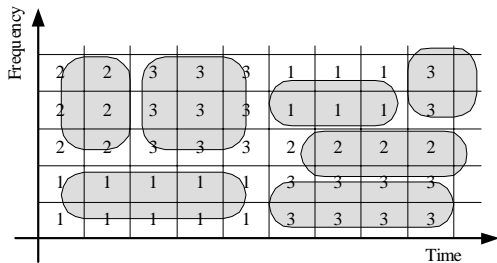


그림 1. 모노 합신호에 포함되어 있는 음원들과 이들의 영역들

Fig. 1. Each of the partitions associated with one of the sound sources contained in the mono sum signal.

2.1 MPEG Surround

MPEG Surround는 최소의 채널(모노 혹은 스테레오)로 합쳐진 다운믹스 오디오 신호와 멀티채널 오디오 신호에 대한 인간의 지각 특성을 나타내는 공간 파라미터(spatial parameter)들로 구성된 부가정보를 이용하여 멀티채널 오디오 신호를 부호화하는 기술이다[6].

그림 2는 MPEG Surround의 기본 개념도이다. 그림

에서 보는 바와 같이, MPEG Surround 부호화기는 N개의 멀티채널 오디오 신호를 입력받아 부가 정보로써 양이효과(binaural effect)에 바탕을 둔 두 귀 사이의 소리 크기 차이와 채널 사이의 상관도 등으로 표현되는 공간 파라미터를 추출한다. 추출된 공간 파라미터는 매우 작은 정보량(채널당 4kbps 이내)이므로 모노 또는 스테레오 오디오 서비스만 제공할 수 있는 대역폭에서도 고품질의 멀티채널 오디오 서비스를 가능케 한다. MPEG Surround 부호화기는 입력받은 멀티채널 입력신호로부터 다운믹스 신호를 생성하며, 이는 음향 엔지니어 등이 제작한 다운믹스(artistic/hand-mixed downmix) 신호로 대체할 수 있다. 생성된 다운믹스 신호는 기존의 고성능 오디오 압축 기술인 MPEG-4 AAC, MPEG-4 HE-AAC 및 MPEG-4 BSAC 등으로 부호화되어 공간 파라미터와 함께 전송된다.

MPEG Surround 복호화기는 전송된 오디오 비트 스트림과 공간 파라미터를 합성하여 N개의 멀티채널 오디오 신호를 복원한다. 멀티채널 오디오의 복호화가 불가능한 경우에는 기존의 오디오 복호화기를 이용하여 다운믹스 신호만을 복호화 할 수 있으므로 독립적인 서비스가 가능하다. 따라서 기존의 모노 또는 스테레오 오디오 부호화 시스템과 호환성을 제공할 수 있다.

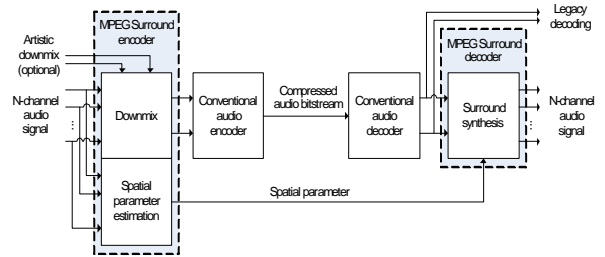


그림 2. MPEG Surround의 개념도

Fig. 2. The scheme of MPEG Surround

2.2 Spaital Audio Object Coding (SAOC)

SAOC는 다객체 오디오 신호를 모노 또는 스테레오 신호로 구성되는 합 신호와 공간파라미터들로 구성되는 부가 정보로 부호화 하는 기술이다. 기본적으로 MPEG Surround와 동일한 공간 오디오 부호화 기술을 이용한다.

그림 3은 SAOC 부호화기의 블록도이다. 멀티채널로 이루어지는 오디오 객체(MBO: Multichannel Background Object)는 MPEG Surround 부호화기를 이용하여 MPS 스테레오 다운믹스 신호와 MPS 비트 스트림으로 부호화한 후 모노 또는 스테레오 채널로 이루어지는 오디오 객체와 함께 객체 부호화(Object Encoder)를 진행한다. 객체 부호화 과정을 통해 SAOC

다운믹스 신호와 OLD(Object Level Difference)로 대표되는 공간파라미터들로 구성되는 객체 파라미터 생성한다. 최종적인 SAOC 비트스트림은 객체 파라미터와 MPS 비트스트림이 다중화된 형태로 구성되게 된다.

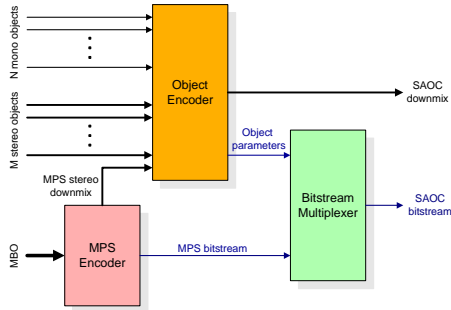


그림 3. SAOC 부호화기 블록도
Fig. 3. Block diagram of an SAOC encoder

SAOC 복호화기는 출력채널이 스테레오일 경우와 멀티채널일 경우를 구분하여 정의된다. 그림 4는 스테레오 출력일 경우에 대한 SAOC 복호화기의 구조를 나타내며, 그림 5는 멀티채널 출력일 경우에는 SAOC 부호화기가 SAOC 비트스트림을 MPS 비트스트림으로 변환하는 기능을 수행하기 때문에 “SAOC to MPS Transcoder”라 명명하였다. 외부에서 또는 SAOC 비트스트림내에서 제공되는 rendering matrix에 기반하여 SAOC 공간 파라미터를 MPS 공간 파라미터로 변환하거나 객체별 이득값을 결정하는 역할을 “SAOC parameter processing” 블록에서 수행하고, “Downmix preprocessing” 블록에서는 입력된 다운믹스 신호를 SAOC decoder일 경우에는 객체별 이득값이 반영된 출력 오디오 신호를 생성한다. 또한, MPS Binarual Decoder를 이용하여 출력신호를 생성할 경우 원하는 위치로 렌더링 시키기 위하여 외부로부터 HRTF 데이터를 받는 인터페이스를 제공한다.

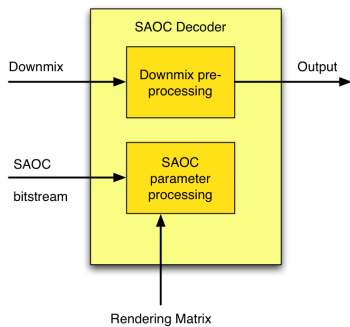


그림 4. 스테레오 출력을 위한 SAOC Decoder 블록도
Fig. 4. Block diagram of an SAOC decoder for stereo-to-stereo processing

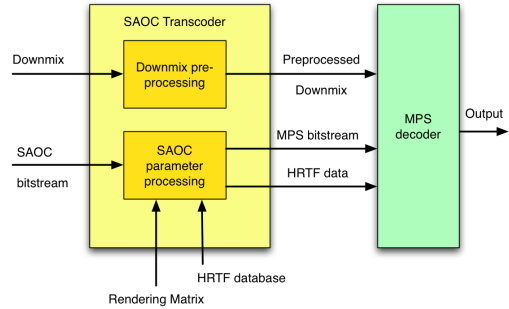


그림 5. SAOC to MPS Transcoder 블록도
Fig. 5. Block diagram of an SAOC to MPS Transcoder

SAOC는 대화형 음악 서비스뿐만 아니라 원거리 회의(tele-conference) 환경에서 다수의 화자(speaker)들의 신호를 효과적으로 압축하는데 이용될 수 있다. 그림 6은 다수의 화자들로 이루어진 SAOC 비트스트림들이 MCU(Multipoint Control Unit)라고 부르는 원거리 회의 서버에서 병합하여 합쳐진 비트스트림을 생성하는 과정을 보여준다.

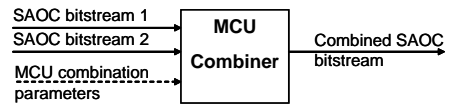


그림 6. MCU 병합기 블록도
Fig. 6. Block diagram of MCU combiner

2.3 Unified Speech and Audio Coding (USAC)

MPEG을 중심으로 진행된 오디오 코덱 기술과 ITU-T를 중심으로 진행된 음성 코덱 기술이 혁신적으로 발전함에 따라, 오디오 코덱은 음성코덱의 영역까지 음성 코덱은 오디오 코덱의 영역까지 응용영역을 확대해 나가고 있다. 또한, 핸드폰과 같은 이동 단말에서 음성통화와 방송 및 음악취청가 가능해짐에 따라 음성 코덱과 오디오 코덱이 모두 요구되는 환경이 나타나게 되었다. 이에 MPEG에서는 음성과 오디오 신호에서 모두 기존 음성 및 오디오 코덱보다 우수한 성능을 발휘하는 음성 및 오디오 통합 코덱(USAC: Unified Speech and Audio Codec)을 개발하고 있다. 이를 위하여 2007년 10월 상기와 같은 기술을 받기 위한 CfP(Call for Proposal)가 발표되었으며, 2008년 7월 RM0(Reference Model 0)를 선정할 예정이므로, 국내외 오디오 부호화 및 음성 부호화 전문 연구기관에서 RM0 선정을 위하여 기술개발을 추진 중이다.

MPEG에서 성능기준으로 삼고있는 Reference Codec은 HE-AACv2와 AMR-WB+이며 음성, 음악, 음성 및 음악 혼합신호로 구분되는 세 가지 category에서 모두 기존 코덱보다 우수한 성능을 나타내는 기술을 RM0로

선정할 예정이다.

감사의 글

본 연구는 정보통신부 및 정보통신연구진흥원의 IT 신성장동력핵심기술개발사업의 일환으로 수행하였음. [2007-S-004-01, 무안경 개인형 3D 방송기술]

참고문헌

- [1] ISO/IEC IS 11172-3 Coding of Moving Pictures and
- [2] ISO/IEC IS 13818-3 Generic Coding of Moving Pictures and Associated Audio, Part-3: Audio.
- [3] ISO/IEC IS 13818-7 Generic Coding of Moving Pictures and Associated Audio. Part-7: Advanced Audio Coding (AAC)
- [4] ISO/IEC IS 14496-3, Information Technology - Coding of audio-visual objects - Part 3: Audio, 2001.
- [5] C. Faller and F. Baumgarte, "Binaural Cue Coding: A Novel and Efficient Representation of Spatial Audio," in Proc. ICASSP 2002, Orlando, FL, May 2002.
- [6] ISO/IEC IS 23003-1, MPEG Audio Technology, Part 1: MPEG Surround, 2007.
- [7] <http://wg11.sc29.org>
- [8] <http://www.chiariglione.org/mpeg/>

2.4 Interactive Music AF (IM AF)

음악이 CD로 제작되기 전에 각 악기별로 구분되어 있는 상태로 제공하여 청취자가 자유롭게 악기별 크기를 제어함으로써 자신만의 음악으로 청취할 수 있는 서비스가 국내에서는 MUSIC2.0이란 이름으로 2007년 하반기에 서비스되고 있다. 이러한 대화형 음악 서비스를 표준화된 파일 포맷으로 정의하기 위한 기술이 2008년 4월 제84차 MPEG 회의때 제안되어 MAF under development로 포함되어 IM AF란 이름으로 표준화를 진행하기로 결정되었다.

또한, IM AF는 저작자가 동일한 악기 데이터를 가지고 다양한 형태의 오디오 장면을 제공할 수 있는 preset이란 기능을 제공할 수 있다. preset은 저작자에 의해서 미리 정의된 mixing matrix를 의미하며 일반 mixing, acoustic mixing, karaoke mixing등 다양하게 정의할 수 있다. 그림 7은 preset의 개념을 도식화하여 표현한 것이다.

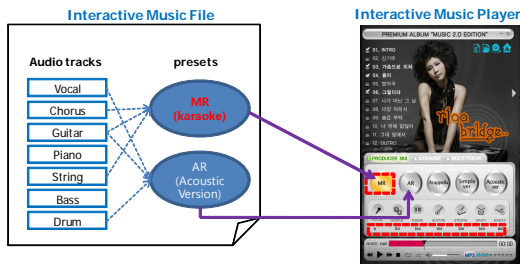


그림 7. IM AF에서의 preset 개념도

Figure 2. Conceptual model with presets of IM AF

IV. 결론

오디오 부호화 기술의 국제 표준화를 주도하고 있는 MPEG 오디오 그룹에서는 최근 들어 저대역폭 환경에서 고품질의 오디오 서비스를 제공하기 위한 파라메트릭 오디오 부호화 기술과 음성 및 오디오 신호에 모두 우수한 성능을 발휘하는 통합 부호화 기술에 역량을 집중하고 있다. 국내에서도 한국전자통신연구원, 삼성중기원, LG전자 등에서 예전과는 달리 주도적으로 참여하고 있다. 멀티미디어 어플리케이션에서 제품의 생산비용 중 로열티가 차지하는 비율이 날로 늘어가고 있으므로, 새로운 제품이나 서비스를 개발하고 상용화하기 이전에 핵심기술들에 대한 지적재산권을 확보하고 국제 표준화 활동을 통해 국제표준에 반영함으로써 로열티에 의한 부가적인 이득을 추구하는 것이 적절할 것으로 예측된다.