차세대 실감방송서비스를 위한 MPEG-H 3D Audio 표준화 동향

서 정 일, 장 대 영, 이 태 진 / ETRI 오디오연구실

1. 서 론

3DTV와 UHDTV(Ultra High Definition Television)로 대변되는 차세대 실감방송 서비스는 대화면 고해상도(4K 또는 8K) 비디오와 다채널 및 객체기반 3차원 오디오를 이용하여 기존의 HDTV에서 제공하는 방송서비스와는 차별된 현장감과 몰입감을 제공하는 것이 목표이다. 본 글에서는 차세대 실감방송 서비스를 제공하는데 필요한 여러 가지 오디오 기술 등 가운데 MPEG에서 최근 표준화가 활발히 진행 중인 MPEG-H 3D Audio 표준에 대해서 논하고자 한다.

UHDTV로 대표되는 차세대 실감방송에서 오디오의 역할은 대화면 고해상도 실감 비디오와 부합하면서 디스플레이가 표현할 수 없는 공간에서도 오디오를 통해 3차원 공간감을 표현하는데 있다. 따라서 기존의 스테레오 및 5.1채널 서라운드 오디오가 표현할 수 있는 2차원 공간감을 3차원으로 확장하여, 시청자를 중심으로 3차원 모든 방향에서 둘러싸는 공간감을 제공하여야 하며, 대화면 디스플레이에서 표현되는 비디오 객체와 동일한 공간상의 위치에서 음상이 정확하게 표현되어야 한다. 상기와 같은 목적을 위하여 일본 NHK에서는 그림 1과 같은 22.2채널 오디오 시스템을 개발하여 일본의 UHDTV방송 표준인 SHV(Super HiVision)에 적용하였다. NHK 22.2 채널 오디오 시스템은 상층, 중층 및 하층의 3개의 계층으로 22개의 스피커들을 3차원 공간상에 배치한다. 특히 대화면 디스플레이가 위치하는 전방에는 스피커들을 촘촘히 배치하여(총 11채널) 대화면 UHD 영상과 공간상의 동일한 위치에 음원들이 재현되도록 하였다. 또한 2개의 서브우퍼 채널을 이용하여 좌우가 분리된 풍부한 저음역대를 재생할 수 있다.

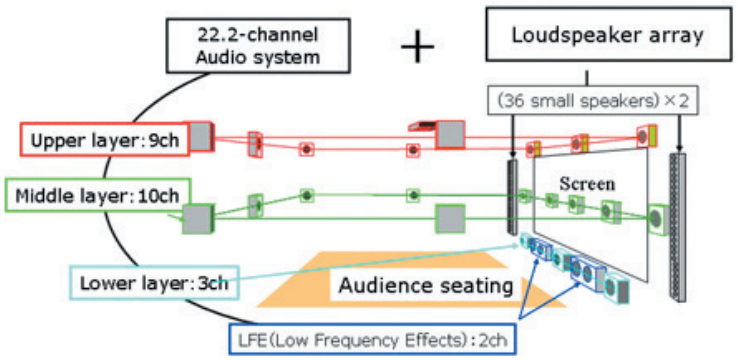

그림 1 NHK 22.2채널 오디오 시스템

Dolby 에서는 기존의 5.1채널 및 7.1채널 영화음악 시스템이 가지는 표현력의 한계를 극복하기 위하여, 천정

에 스피커를 추가로 배치하여 9.1채널까지 오디오 채널을 확장하고 관객에게 정확한 음상감을 표현하는 것이 필요한 음원들은 음원객체(sound object)로 분리하여 전송하고 표현하는 Dolby ATMOSTM 기술을 개발하였다(그림 2 참조). 또한 DTS에서는 기존의 채널신호 기반의 영화음악을 제작환경을 객체기반으로 변환하기 위하여 객체기반 영화음악 콘텐츠 저작포맷인 MDA(Multi Dimensional Audio)를 제안하고 영화제작사, 영화음악 제작사, 솔루션 제공업체 등을 통한 보급을 추진하고 있다.

그림 2 Dolby ATMOS의 Beds(9.1채널)와 객체의 개념도(출처: www.dolby.com)

한편 국내에서는 ETRI와 삼성전자가 협력하여 국내 UHDTV 서비스를 위한 10.2채널 멀티채널 오디오 포맷을 제안하여 TTA, ITU-R 등에 표준화를 완료하였으며, 소닉티어에서는 영화음악 시스템의 현장감을 극대화하기 위한 14.2채널 및 30.2채널 멀티채널 포맷을 개발하여 '명량', '국제시장' 등과 같은 한국영화에 적용하고 전용관을 통한 서비스를 진행하고 있다.

그러나 22.2채널과 오디오 객체를 함께 방송하고 수신기를 통해 출력할 수 있다고 하더라도 일반가정에서 22개의 스피커를 구비하여 표준위치에 설치하여 청취하는 것은 기대하기 어렵다. 따라서 전송된 다채널 오디오 신호를 시청자가 구비하고 있는 스피커 배치환경에 최적으로 변환하여 재생하는 기술이 동반되어야 하며, 최근 스마트폰을 통한 멀티미디어 콘텐츠의 소비가 급증하고 있으므로 스마트폰과 연결되는 헤드폰을 통한 다채널 오디오 신호를 3차원 공간상에 표현할 수 있는 기술도 함께 필요하다.

전술한 영화음악 관련 기술개발 동향과 UHDTV와 같은 실감방송 서비스를 위하여 새롭게 제시되고 있는 오디오 서비스에 대한 요구사항은 아래와 같이 3가지 정도로 요약할 수 있다.

- 고차 다채널 오디오(high-order multichannel audio): NHK 22.2채널, ETRI/Samsung 10.2채널, Dolby 9.1채널, Auro 3D 9.1채널과 같이 5.1채널 이상의 스피커들이 시청자를 중심으로 3차원 공간상에 배치되는 고차 다채널 오디오 포맷을 지원해야 한다.
- 객체기반 오디오(object-based audio): 전통적인 오디오 채널신호와 분리된 오디오 객체 신호들을 전송하고 단말에서 원하는 3차원 공간상에 표현할 수 있어야 한다. 또한 시청자의 취향이나 선택에 따라 오디오 객체신호를 제어할 수 있어야 한다.
- 자유로운 다운믹싱 및 재현(flexible down-mixing and rendering): 수신된 고차 다채널 및 객체 신호들을 시청자가 구비한 스피커 배치환경에 최적으로 다운믹싱하거나 채널 포맷을 변환하여 재생할 수 있어야 한다. 또한 헤드폰을 통해서도 저작자가 의도한 3차원 음장감을 최적으로 표현할 수 있어야 한다.

이러한 요구사항들을 만족시키기 위하여 MPEG에서는 MPEG-H 3D Audio란 이름으로 2012년부터 표준화 활동을 시작하였으며 2015년 상반기에 Phase I에 대한 국제표준을 발간할 예정이다.

2. MPEG 오디오 표준 개발 현황

MPEG-H 3D Audio는 MPEG 오디오 서브그룹에서 표준화되었던 오디오 코덱툴들을 취사 선택하여 구성하였으므로 MPEG-H 3D Audio 표준을 이해하기 위해서는 MPEG 오디오 표준들이 어떻게 개발되어 왔는지에 대한 내용을 파악할 필요가 있다. 따라서 본 절에서는 MPEG-H 3D Audio 표준개발 이전까지 MPEG 오디오 서브그룹에서 진행한 표준화 현황을 간략하게 살펴보고자 한다.

비디오 및 오디오 신호에 대한 압축 부호화 표준화를 담당하는 MPEG에서는 1997년 AAC(Advanced Audio Coding) 오디오 부호화 표준을 제정한 이후 AAC를 핵심 부호화 도구(core)로 사용하면서 부호화 성능을 개선하기 위한 툴들에 대한 표준화를 진행하였다. 스테레오 오디오 신호를 위한 PS(Parametric Stereo), 고주파수 대역 신호를 효과적으로 부호화하기 위한 SBR(Spectral Band Replication), 멀티채널 오디오 신호를 스테레오 재생 시스

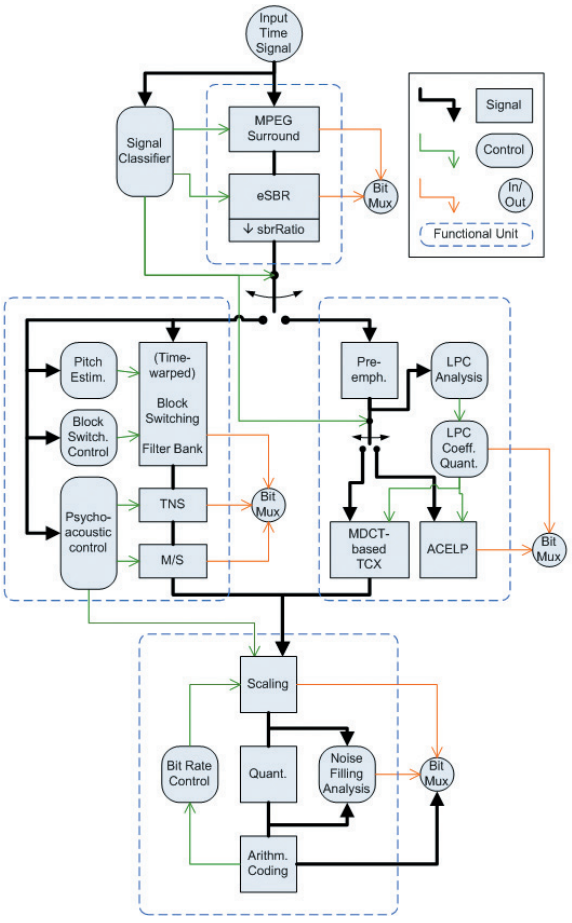

그림 3 USAC 인코더 블록도(출처: ISO/IEC 23003-3)

템과 호환성을 유지하면서 간략하게 압축하기 위한 MPS(MPEG Surround) 등이 순차적으로 표준화되었으며, 이들은 모두 AAC와 역호환성을 제공하는 것이 특징이었다.

한편 2008년부터는 높은 비트율에서 AAC 보다 우수한 음질을 나타내고, 낮은 비트율에서는 오디오 신호뿐만 아니라 음성신호에 대해서도 우수한 음질을 나타내는 USAC(Unified Speech and Audio Coding) 기술을 표준화하였다. USAC은 MPEG에서 최근까지 개발된 다양한 오디오 부호화 툴들(AAC, SBR, MPS 등)과 음성신호에 대하여 강점을 가지는 CELP(Code Excitation Linear Prediction)계열의 AMR-WB+(Extended Adaptive Multi-Rate-Wideband) 기술을 효과적으로 접목하여 다양한 비트율에서 세계최고 수준의 부호화 성능을 제공하였다. 표준제정 초기에는

그림 4 AAC 패밀리 프로파일 개념도(출처: www.voiceage.com)

USAC이 이동방송, 오디오북, 음악 플레이어 등과 같은 스테레오 오디오 신호를 이용하는 서비스로 응용분야를 한정하였기 때문에 스테레오 오디오 신호에 대한 압축방법만 표준에 명시되어 있지만, 인코더와 디코더를 스테레오 쌍(stereo channel pair)으로 확장하면 다채널 오디오 신호에 대해서도 적용이 가능하다.

그림 3은 USAC 인코더의 블록도를 나타낸다. 그림 3에서 확인할 수 있듯이 USAC은 두 개의 스위치를 이용하여 음악 신호를 위한 AAC 코어, 음성신호를 위한 ACELP 코어, 음성과 음악이 혼재된 신호를 위한 TCX 코어가 선택되어 동작하며, 스테레오 신호를 위해서는 MPS툴을 고주파수 대역 신호를 위해서는 eSBR(enhanced SBR)툴이 제공된다. 또한 각 툴들에 의해서 부호화된 결과 파라미터들은 산술부호화 방식을 이용하여 재압축된다. USAC에 ACELP, TCX 등과 같은 새로운 툴들이 적용되어 있지만 기본적으로는 AAC 계열의 툴들이 재활용되고, 기존 AAC 계열 툴들을 활용하면 쉽게 구현될 수 있으므로 xHE-AAC(extended High Efficiency AAC) 프로파일이란 이름으로 AAC 패밀리 프로파일에 추가되었다. 그림 4는 AAC, HE-AAC, HE-AAC v2, xHE-AAC 프로파일의 포함관계(하위 호환성 관계)를 나타낸 블록도이다.

3. MPEG-H 3D Audio 오디오 표준 개발 과정

2010년경부터 UHDTV를 위한 22.2채널 오디오 시스템을 개발한 NHK는 자사의 오디오 포맷을 MPEG 표준에 채택시키기 위하여 다채널 오디오 신호에 대한 새로운 부호화 및 재생방식에 대한 표준화를 MPEG에서 수행해야 한다고 주장하였으며, 비슷한 시기에 진행된 고품질 비디오 부호화 표준인 HEVC(High Efficiency Video Coding)에 버

금가는 오디오 품질을 제공할 수 있는 실감 오디오 기술에 대한 표준화 필요성이 MPEG 내외부에서 부각되었다. 이러한 배경으로 MPEG에서는 HEVC와 동일한 표준 프로젝트(MPEG-H)내에서 3D Audio란 이름으로 다채널 고음질 오디오 부호화 및 재생방법에 대한 표준화를 진행하기로 결정하고 2013년 1월에 기술을 제안 받기 위한 CfP(Call for Proposal)을 발표하였다.

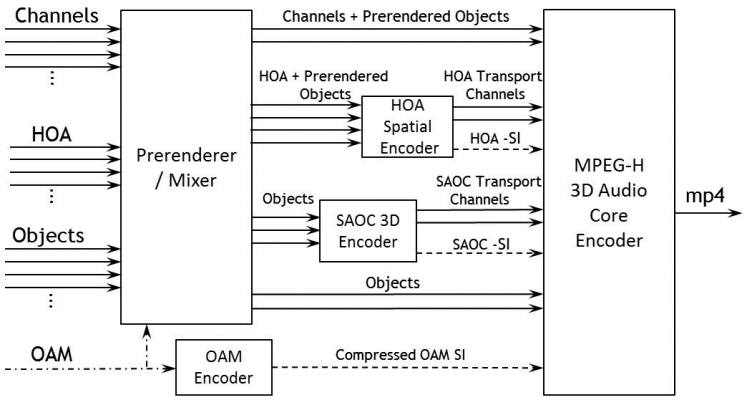

그림 5 MPEG-H 3D Audio 인코더 블록도(출처: MPEG-H 3D Audio DIS)

MPEG-H 3D Audio는 전통적인 채널기반의 오디오 신호뿐만 아니라 오디오 객체 신호와 3차원 음장을 직접 녹음하고 표현하는 고차 앰비소닉스(HOA: High Order Ambisonics)신호도 처리하는 것을 목표로 하였다. 또한 부호화 기술뿐만 아니라 단말의 재생환경에 맞게 디코더 출력 신호를 변환하는 다운믹스(downmix) 기능 및 자유 렌더링(flexible rendering) 기능도 표준화의 범주에 포함시켰다. 여기서 다운믹스는 입력되는 다채널 오디오 신호의 채널 수를 줄이는 과정(예: 22.2 채널 신호를 5.1채널로)을 지칭하며, 자유렌더링은 단말에 구비되어 있는 스피커 배치환경이 표준환경과 상이하거나 개수가 다를 경우에 최적으로 입력 다채널 오디오 신호를 변환하는 과정을 의미한다. 이때 채널 수뿐만 아니라 각 채널들에 대한 스피커 위치도 가변될 수 있다. 단 HOA 신호는 HOA 파라미터인 공간 하모닉스(spatial harmonics)를 해석하고 합성하는 별도의 처리과정이 존재하기 때문에 독립적으로 표준화를 진행하기로 결정하였다.

이러한 배경으로 2013년 5월말 ETRI, Fraunhofer-IIS, Fraunhofer-IDMT, Sony는 채널과 객체신호(CO) 분야에 부호화 및 재생기술을 제안하였고, Orange Labs, Technicolor, Qualcomm은 HOA 분야에 자사의 기술을 제안하였다. 2개월 동안 청취평가를 수행한 후 2013년 7월에 열린 제 105차 MPEG회의에서 CO 분야에서는 Fraunhofer-IIS 기술이, HOA 분야에서는 Technicolor에서 제안한 기술이 참조모델(Reference Model)로 선정되었다. 그러나 표준화 진행 과정 중에 HOA 신호를 CO 분야에 선정된 USAC을 이용할 경우 성능이 향상됨을 확인하여 CO 분야와 HOA 분야를 통합하여 단일 표준으로 진행하는 것으로 합의하였다.

2015년을 전후하여 일본의 UHDTV 방송규격, 미국의 차세대 방송규격인 ATSC 3.0, 유럽의 UHDTV 방송규격에 대한 표준화가 진행될 예정이었으므로 MPEG-H 3D Audio는 서둘러 표준화를 진행하였으며, 2014년 7월 DIS(Draft International Standard)가 승인되었고 2015년 상반기에 IS로 발간될 예정이다. 또한 128kbsp 이하의 낮은 비트율에서 다채널 오디오 신호를 서비스하기 위한 기술들은 MPEG-H 3D Audio Phase 2란 이름으로 표준화를 진행하고 있으며, 2016년경 표준화가 완료될 예정이다.

4. MPEG-H 3D Audio 표준 개요

먼저 3D Audio 인코더와 디코더의 동작을 간략하게 설명하면, 그림 5와 같이 3D Audio 인코더는 채널신호, 객체신호, HOA 신호, 객체신호의 렌더링 정보를 포함한 메타데이터(OAM: Object Audio Metadata)를 입력 받아 부호화한다. 오디오 객체신호는 인코더단에서 미리 렌더링하여 채널신호에 더하거나, 개별 채널 신호로 부호화 하거나, SAOC(Spatial Audio Object Coding) 3D 부호화기를 통해 압축 부호화한다. 그리고 HOA 신호는 HOA 공간 부호화기를 통하여 HOA 계수들을 채널신호로 변환하는 과정을 수행한다. 이렇게 처리된 채널신호, 객체신호, HOA 신호들을 USAC의 성능을 조금 향상 시킨 USAC 3D 부호화기를 이용하여 압축하여 부호화함으로 3D Audio 비트스트림이 출력된다.

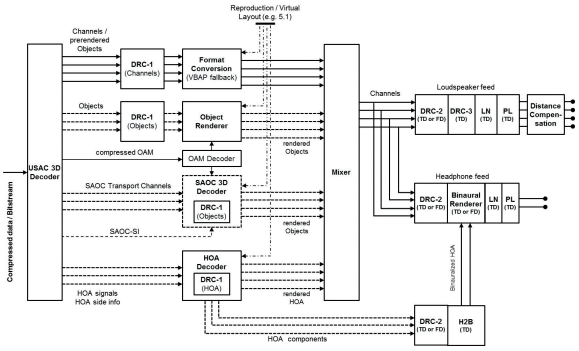

그림 6 MPEG-H 3D Audio 디코더 블록도(출처: MPEG-H 3D Audio DIS)

그림 6과 같이 3D Audio 디코더는 USAC 3D 디코더를 이용하여 복원된 채널신호, 객체신호, SAOC 정보, HOA 신호들을 출력 스피커 포맷에 맞도록 포맷 변환기, 객체 렌더러, SAOC 3D 디코더, HOA 디코더에서 변환한 후 믹싱하여 스피커 출력신호를 출력한다. 만약 헤드폰을 통한 출력이 요구될 경우에는 바이노럴 렌더러가 추가로 동작하여 스테레오 헤드폰 신호를 출력한다. 추가로 스피커나 헤드폰을 통해 출력되는 오디오 신호의 크기를 제어하기 위한 DRC(Dynamic Range Control)툴들이 디코더 단계마다 동작할 수 있다.

3D Audio 표준을 구성하는 오디오 코덱들에 대해서 설명하면 다음과 같다. USAC 3D는 USAC의 대부분의 기능을 활용하면서 다채널 오디오 신호를 효과적으로 부호화하기 위한 QCE(Quad Channel Element) 모드를 추가한 것이 특징이다. QCE는 4채널 신호를 모노신호로 축약하여 부호화하는 기술로써 USAC 스테레오 부호화 도구인 MPS212를 병렬로 활용하여 압축 효율을 극대화하였다.

전술한바와 같이 오디오 객체신호는 인코더단에서 채널신호에 미리 렌더링되어 더해지거나 모노 채널신호로 독립적으로 부호화된다. 또한 객체신호의 개수가 많거나 전송 비트율이 낮아 추가적인 압축이 필요할 경우에는 SAOC(Spatial Audio Object Coding)를 간략화 한 SAOC 3D를 이용하여 압축하여 부호화한다. 복원된 오디오 객체신호를 스피커를 이용하여 3차원 공간상에 렌더링하는데 필요한 각 객체신호별 재생위치는 메타데이터로 전달된다.

HOA 신호에 대해서는 입력된 HOA 계수들을 오디오 채널 신호로 변환하는 HOA 공간 부호화기와 이를 다시 HOA 계수로 복원하고 스피커 출력신호로 변환하는 HOA 디코더기를 통하여 표현된다. 이러한 과정을 통하여 채널신호, 객체신호, HOA 신호가 인코딩 및 디코딩되며 22.2 채널의 경우 1.2Mbps에서 원음과 구별할 수 없는 음질을 제공할 수 있다.

기존의 MPEG 오디오 표준들이 오디오 신호를 압축하고 복원하는 코덱 기술에 한정되어 있던 것과는 달리 3D Audio는 단말의 다양한 오디오 재생환경에 최적으로 오디오 신호를 재생하는데 필요한 렌더링 기술도 함께 표준화하였다. 포맷변환기는 채널신호의 스피커 포맷과 단말의 스피커 포맷을 고려하여 USAC 3D 디코더의 출력신호 중 채널신호를 스피커 출력신호를 변환하는 과정을 수행한다. 단순하게는 정해진 매트릭스를 이용하여 다운믹스(예: 22.2채널에서 5.1채널로)하는 과정일 수도 있으며, 표준 스피커 포맷과 다른 위치에 다른 개수의 스피커가 배치되는 환경(예: 8개의 스피커가 사면체방의 각 모서리에 위치하는 경우)에서 입력되는 다채널 오디오 신호의 음장감을 유지하면서 변환하는 과정일 수도 있다. 이러한 동작 과정은 능동 다운믹싱, VBAP(Vector Based Amplitude Panning) 렌더링, 허상스피커 렌더링 기법 등을 이용하여 구현된다.

다채널 스피커 출력신호를 헤드폰을 통해 3차원으로 표현하는 바이노럴 렌더러는 바이노럴 렌더러로 입력되는 신호의 형태에 따라 주파수영역 바이노럴 렌더러(FD-Bin)와 시간영역 바이노럴 렌더러(TD-Bin)로 구분하여 표준화되었다. 그림 7과 같이 FD-Bin은 QMF(Quadrature Mirror Filter)대역 신호로 입력되는 다채널 스피커 출력신호에 BRIR(Binaural Room Impulse Response) 필터를 QMF 대역에서 적용하여 바이노럴 출력신호를 생성한다. 이때 직접음과 초기반사음에 대해서는 대역별로 다른 길이를 가지는 BRIR(Binaural Room Impulse Response) 필터를 적용하며, 잔향음에 대해서는 인공잔향기를 이용하여 표현한다. 또한 청감상 둔감한 특성을 가지는 고주파수 대역에 대해서는 하나의 이득과 지연만으로 표현되는 TDL(Tap Delay Line)로 간략하게 표현함으로써 바이노럴 렌더링에 필요한 연산량을 획기적으로 감소시켰다. TD-Bin은 상대적으로 긴 길이를 가지는 FIR필터를 FFT 변환을 통해 주파수 영역에서 고속으로 연산하는 전통적인 방법을 이용한다.

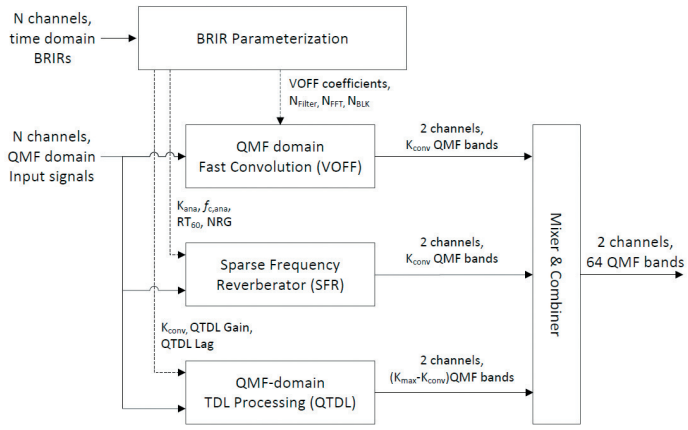

그림 7 주파수영역 바이노럴 렌더러 블록도(출처: MPEG-H 3D Audio DIS)

5. MPEG-H 3D Audio 표준의 향후 전망

HEVC, MMT(MPEG Media Transport)와 함께 MPEG-H 프로젝트를 통해 개발된 3D Audio는 차세대 UHDTV 방송 표준에서 실감 오디오 서비스를 위해 필요한 핵심기술이다. 최근 ATSC에서는 ATSC 3.0이란 이름으로 미국 UHDTV 방송표준을 논의하고 있으며, 오디오에 대해서도 2014년말 CfP가 발표되었고, 2015년 1월 Dolby, DTS와 더불어 3D Audio가 후보 기술로 제안되었다. 3D Audio는 우수한 코덱 성능, 다양한 렌더링 기능, 상대적으로 저렴한 로열티를 무기로 Dolby 및 DTS와 경쟁하고 있으며, 로열티에 민감한 제조업체를 중심으로 많은 지원을 받고 있다.

유럽의 DVB(Digital Video Broadcasting)에서는 2014년 7월 UHDTV를 위한 요소기술들에 대한 백서를 발표하고 UHDTV 표준을 제정하는데 필요한 요구사항을 정의하고 있는 상황이다. 오디오에 대한 요구사항 중 다채널과 오디오 객체를 지원하는 것이 필수사항으로 포함되어 있고, 미국에 비해서 Dolby의 영향력이 상대적으로 덜하기 때문에 표준화가 순조롭게 진행된다면 3D Audio가 표준에 채택될 가능성이 매우 높다고 할 수 있다.

일본의 경우에는 자국의 SHV 방송 표준 제정을 거의 완료해가고 있는 단계이며, 오디오 코덱은 기존 HDTV 방송에 사용하였던 AAC를 사용하는 것으로 확정하였지만 DRC, 포맷 변환기 등과 같은 3D Audio 표준에 포함된 툴들을 적용하는 것에 대해서 논의되고 있다.

6. Conclusion

MPEG-H 3D Audio는 AAC부터 시작하여 MPEG에서 표준화 되었던 모든 오디오 코덱 툴들과 단말의 스피커환경에 최적으로 재생하기 위한 렌더링 기술들이 총망라되어 있는 표준 기술이다. 방송표준으로의 진입을 통해 전체가 활용될 가능성도 높으며, 바이노럴 렌더러와 같이 독립적으로 활용될 가능성도 배제할 수 없다. 따라서 정부를 중심으로 오디오 서비스와 관련된 산학연 관련 기관에서 3D Audio 표준에 대한 연구와 제품개발에 많은 투자가 필요한 시점이다. 또한 오디오 객체신호를 이용한 다양한 서비스 발굴을 통하여 점점 규모가 감소하고 있는 국내 오디오 산업을 활성화하는 계기를 마련해야 할 것으로 사료된다.

감사의 글

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발 사업의 일환으로 수행하였음 [과제번호: 14-000-02-001, 초고품질 콘텐츠 지원 UHD 실감방송/디지털시네마/사이니지 융합서비스 기술 개발].

참고문헌

[1] ISO/IEC 23003-3, Information technology - MPEG audio technologies, Part 3: Unified speech and audio coding, 2012.

[2] N13411, Call for Proposal for 3D Audio, Geneva, Swiss, January 2013.

[3] N14064, Submission and Evaluation Procedures for 3D Audio Phase 2 Submissions, Geneva, Swiss, November 2013.

[4] N14747, Text of ISO/IEC 23008-3/DIS, 3D Audio, 2014.