

실감 오디오 방송을 위한 초다채널 오디오 시스템 및 표준화 동향

Introduction and Standard Status of High Order Multichannel Audio System for Realistic Audio Broadcasting

서정일 (J.I. Seo) 실감음향연구팀 선임연구원
강경옥 (K.O. Kang) 실감음향연구팀 팀장

* 본고는 2011년도 정부(한국방송통신위원회) 재원으로 방송통신기술개발사업의 지원을 받아 수행된 연구임(과제번호: KCA-2011-11921-02001).

본고는 3DTV, UHDTV(Ultra High Definition Television)와 같은 실감방송 환경에서 실감 오디오 서비스를 제공하기 위한 초다채널 오디오 기술의 최근 연구 및 개발 동향을 소개한다. 스테레오와 5.1 채널로 대표되는 기존의 오디오 기술은 2차원 평면상에서만 음장을 형성할 수 있다는 표현의 한계를 가지고 있다. 3D 영화의 성공과 UHDTV로 대표되는 초고화질 비디오와 부합하기 위해서는 오디오도 3차원 공간상에서 표현되어야 하며 이를 위해서는 필연적으로 출력채널 수가 증가하여야 한다. 이러한 초다채널 오디오는 22.2 채널과 같은 대용량의 오디오 데이터를 압축하는 기술뿐만 아니라 다양한 오디오 출력 환경에 적응적으로 오디오 콘텐츠를 표현하는 기술에 대한 연구/개발이 필요하다.

2012
Electronics and
Telecommunications
Trends

스마트 미디어 시대의
방송통신 융합기술 특집

- I. 서론
- II. 초다채널 오디오 시스템
- III. 표준화 동향
- IV. 결론

I. 서론

스피커(speaker) 또는 헤드폰(headphone)을 이용하여 인간에게 음악, 영화, 뉴스 등을 제공하는 오디오 시스템은 라디오, 텔레비전, CD, MP3 플레이어 등을 통해 우리의 삶의 한 부분으로 자리잡고 있다. 글을 읽으면서 음악을 감상하는 등과 같이 오디오 시스템이 가지는 편리성은 홍수처럼 밀려들어 오는 첨단 멀티미디어 시스템의 틈바구니 속에서도 생명력을 가지고 존재할 것이다. 또한 비디오 신호와 동기되거나 함께 제공됨으로써 비디오 신호로 제공할 수 없는 정보를 제공하는 채널로 이용되고 있다. 최근 3차원 영화와 같이 실재와 동일하거나 더욱 과장된 현장감과 몰입감을 제공하기 위해 3차원 입체비디오와 부합하는 3차원 입체오디오를 제공하기 위한 기술개발이 진행 중이다.

멀티채널과 3차원 오디오는 서로 다른 속성을 가지고 있음에도 불구하고 혼용되어 사용되어 온 것이 사실이다. 멀티채널은 오디오 콘텐츠를 재생하는 오디오 시스템이 가지고 있는 출력채널의 수, 즉 스피커의 개수와 관련된 용어이며, 멀티채널이 3차원 오디오를 대변하지 않는다. 그리고 오디오 시스템의 출력신호(스피커 또는 헤드폰 신호)가 재현(reproduction)하는 음장이 3차원 공간을 표현하고 있을 때(음의 고저가 느껴지고 공간감이 느껴질 때) 3차원 오디오라고 부르게 된다. 그러나 3차원 오디오를 재현하기 위해서는 2개 이상의 스피커가 필요하게 되고, HRTF(Head Related Transfer Function)를 이용하는 바이노럴 3차원 오디오를 제외하고는 음의 고저감, 거리감, 공간감 등을 정확하게 재현하기 위해서는 많은 수의 스피커가 필요하게 된다. 이러한 이유로 멀티채널과 3차원 오디오가 관계를 가지며 혼용되어 사용되게 된 것이다.

본고에서는 3차원 오디오에 대한 논의는 제외하고 멀티채널 오디오 시스템이 현재의 5.1 채널에서 어떠한 형태로 진화하고 있는가에 대해서 설명한다. 본고의 구

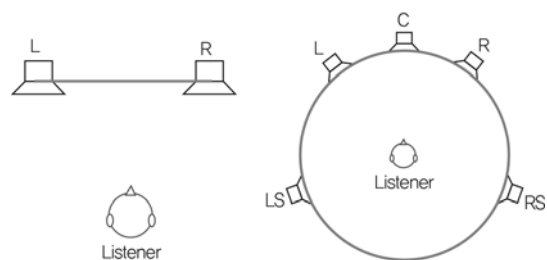
성은 다음과 같다. II장에서는 초다채널 오디오 시스템을 5.1 채널부터 시작하여 NHK 22.2 채널 시스템까지 소개하고, III장에서는 초다채널 오디오와 관련된 표준화 동향을 소개하고, IV장에서 결론을 맺는다.

II. 초다채널 오디오 시스템

1. 5.1 채널

5.1 채널은 5채널 서라운드 멀티채널 사운드 시스템의 일반적인 명칭이며, 현재 홈시어터와 극장용 사운드 시스템으로서 보편적으로 사용되고 있는 시스템이다 [1]. 5개의 전 대역(full bandwidth) 채널과 1개의 저주파 채널(0.1채널)을 사용한다. 또한 5.1 채널은 디지털 방송과 음악의 표준 서라운드 사운드 오디오 구성 요소이다. 모든 5.1 채널 시스템은 front left, center, front right, surround left, surround right, sub woofer 채널로 구분되는 동일한 스피커 채널과 구성을 사용한다.

(그림 1)과 같이 5.1 채널 사운드는 기존의 2채널 스테레오에서 제공하지 못했던 청취자의 측면과 후면으로부터의 사운드 재생을 가능하게 함으로써 전방에만 머물렀던 가상음상의 정위 영역이 360도 전 방향으로 확장됨으로써, 보다 입체감이 풍부해지고 실감 있는 음장을 전달할 수 있는 오디오 재생 기술이 마련되었다는 점에서 큰 의미를 가진다. 과거 스테레오 사운드 이후의 3채널 서라운드, 4채널 서라운드 재생 방식이 극장에 도입되어 가정에서의 오디오 청취 환경과 차별화되는 특



(그림 1) 스테레오와 5.1 채널의 가상 음상정위 영역의 차이

성임을 강조하였는데, 5.1 채널도 마찬가지로 극장에서 처음 사용되고 이후 가정으로 도입되는 형태로 확산되었다.

2. Auro 3D

Auro 3D는 벨기에의 Auro Technologies사에서 제안한 극장 및 홈시어터용 멀티채널 시스템이다[2]. (그림 2)와 같이 5.1 채널 시스템을 기본으로 하고 center 스피커를 제외한 나머지 4채널에 상향 스피커(height speaker)를 추가한 9.1 채널이 기본 포맷이며, 천정 스피커(ceil speaker)까지 추가한 10.1 채널이 제안되어 있다.



(a) Auro 3D 9.1 채널

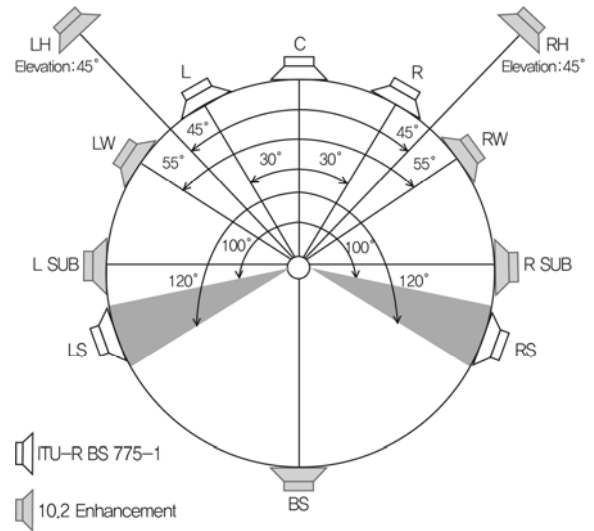


(b) Auro 3D 10.1 채널

(그림 2) Auro 3D의 홈시어터용 9.1 채널 및 10.1 채널 스피커 포맷

3. T. Holman 10.2 채널

THX의 T. Homan은 2001년도에 C. Kyriakakis와 함께 10.2 채널을 제안하였다[3]. 당시 10.2 채널에 대



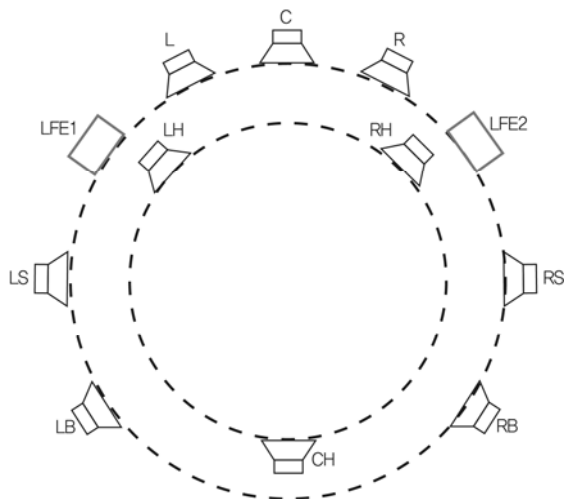
(그림 3) THX 10.2 채널 스피커 배치도

한 슬로건은 “5.1 채널의 두 배만큼 우수하다”였으며, 사운드 엔지니어들에게는 제작 과정에서의 보다 큰 자유도(flexibility)를, 청취자에게는 좀 더 몰입할 수 있는 오디오 공간을 제공하는 것을 목적으로 하고 있었다. 이러한 점에서 NHK 22.2 채널, Dolby Digital EX(6.1 채널), SDDS(7.1 채널), DTS-ES(6.1 채널) 등과 비교되고 있으며, 근래에는 14채널 혹은 그 이상을 전송할 수 있는 Dolby Digital Plus 및 Dolby TrueHD와도 비교되고 있다.

THX 10.2 채널은 (그림 3)과 같이 5 채널(center, left, right, left surround, right surround)에 수평면상의 해상도를 향상시키기 위한 3채널(left wide, right wide, back surround), 상향 2채널(left height, right height), 좌우측 서브우퍼를 추가한 구조이다.

4. ETRI/Samsung 10.2 채널

3DTV와 UHD TV(Ultra High Definition Television)와 같은 실감방송에서 기존 HDTV의 5.1 채널 시스템보다 사실적이고 고품질의 사운드를 제공하기 위하여 한국전자통신연구원과 삼성전자가 제안한 10.2 채널 시스템은 (그림 4)와 같은 스피커 배치를 가진다[4].

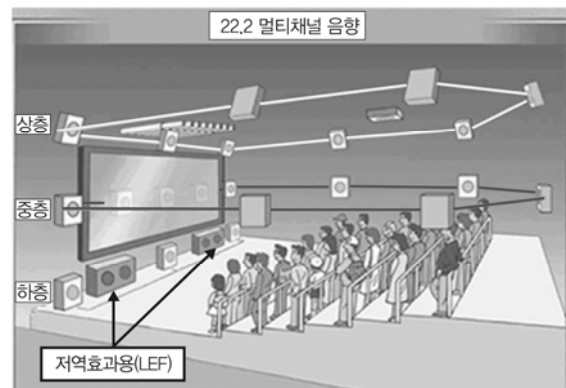


(그림 4) ETRI/Samsung 10.2 채널 스피커 배치도

ETRI/Samsung 10.2 채널 시스템은 전방에 수평면 3개 채널과 수직면 2개 채널로 총 5채널이 배치되며, 측 후면에 총 4개 채널, 천장에 1개 채널이 배치되고, 바닥면에 저주파 담당 2개 채널이 구성되어 총 12개 채널이 사용되며 저주파 담당 채널을 0.1로 간주하여 10.2 채널이 된다. 5.1 채널 및 7.1 채널 시스템과 호환성을 제공하면서 청취자를 중심으로 3차원 음장을 제공할 수 있다는 장점을 지니며, 일부 스피커들에 대해서는 배치 영역을 제공하므로 오디오 엔지니어와 청취자의 재생 환경이나 취향에 따라 자유로이 스피커의 위치를 변경할 수 있다는 장점을 지닌다.

4. NHK 22.2 채널

5.1 채널 오디오 시스템은 영화에서뿐만 아니라, DVD-Video, DVD-Audio, SACD, 디지털 방송에 이르기까지 다양한 분야에서 널리 사용되고 있다. 그러나 비록 스테레오 시스템에 비하여 향상된 공간감을 제공하기는 하지만 보다 넓은 청취공간을 형성하는 데 있어서 5.1 채널 오디오 시스템에는 다양한 제약 사항이 뒤따른다. 특히, 유효 청취 영역(sweet spot)이 좁게 형성되고, 수직 음상을 제공할 수 없기 때문에 극장과 같이 넓은 청취공



(그림 5) NHK 22.2 채널 시스템의 구성

간에는 적절하지 않다.

NHK 22.2 채널 오디오 시스템의 구조는 (그림 5)와 같다. 그림과 같이 22.2 채널 오디오 시스템은 세 개의 스피커 층으로 이루어져 있다. 위층은 9개의 채널, 가운데 층은 10개의 채널, 아래층은 3개의 채널과 2개의 LFE 채널로 이루어져 있다[5].

III. 표준화 동향

1. 초다채널 오디오 국내외 표준화 동향

NHK 기술연구소는 SMPTE를 통해 2007년 UHDTV를 위한 비디오 신호규격을 제정하였고[6], 2008년에 22.2 채널 오디오 재생시스템을 UHDTV를 위한 오디오 신호규격으로 표준화함으로써 가장 앞선 행보를 보이고 있다[7]. 그러나 NHK의 22.2 채널 오디오는 기본적으로 극장과 같은 환경을 고려하고 있기 때문에, 가정 환경에 적용하기 위해서는 스피커 설치공간의 문제가 발생할 수 있다. 이에 따라 국내에서는 NHK의 22.2 채널 음상정위와 음질 재현 성능을 최적화하여, 가정 환경을 고려한 10.2 채널 오디오를 국내 UHDTV 오디오 신호 규격으로 제정하였다[8].

ITU에서는 멀티채널 오디오 재생 기술에 대한 음질과 기능적 요구사항을 정의해 두고 있고, 5.1 채널부터 시

작하여 NHK 22.2 채널, WFS(Wave Field Synthesis) 등 다양한 멀티채널 오디오 재생 기술에 대한 보고서를 받고 있다. 현재는 이러한 다양한 멀티채널 오디오 재생 기술 중 하나를 선택하기보다는, 이러한 멀티채널 오디오 포맷의 구조를 표현하기 위한 메타데이터 구조 표준화를 진행하고 있으며, 2013년 중 표준화 완료를 목표로 하고 있다[9].

2. MPEG

MPEG와 ITU-T가 공동으로 설립한 JCTVC에서 HEVC(High Efficiency Video Coding) 표준화가 시작된 2010년 하반기에 HEVC가 목표로 하는 대화면 고화질 비디오와 부합하는 오디오 서비스에 대한 기술적 검토가 시작되었으며, NHK의 22.2 채널 오디오 시스템에 대한 소개와 함께 3D Audio란 이름의 AhG(Adhoc Group)가 Audio 서브그룹에 설립되었다. 당시 NHK에서는 자사의 22.2 채널 오디오 시스템의 국제표준 반영을 위하여 SMPTE, ITU-R, ATSC 등과 같은 다양한 국제표준화 단체에 새로운 오디오 표준에 대한 필요성을 제시하던 시점이었으며, 디지털 시네마 분야에서는 DTS, Dolby 등에서 기존 5.1/7.1 채널 서라운드 오디오 시스템을 획기적으로 개선할 수 있는 3차원 오디오 포맷에 대한 연구 및 기술개발이 활발히 진행되고 있었다. 또한, 국내에서는 차세대방송표준포럼을 중심으로 UHD-TV를 위한 비디오 및 오디오 신호규격을 제정하기 위한 연구가 진행되고 있었다.

상기와 같은 국내외 기관들의 차세대 오디오 서비스에 대한 연구/개발/표준화 활동과 보조를 맞추어 3D Audio 표준화에 대한 활발한 논의가 벌어졌으며 2013년 상반기로 예정된 CFP 발표를 기점으로 본격적인 표준화가 시작될 것이다[10].

가. Application Scenario

3D Audio가 목표로 하는 응용 분야들은 다음과 같다.

- Home theater: HEVC가 목표로 하는 대화면 고 해상도 디스플레이에 부합할 수 있도록 함께 제공되는 오디오는 고음질을 가지면서도 안정적이고 넓은 음장 영역을 표현할 수 있어야 한다. 특히 비디오 이미지와 오디오 이미지가 일치될 수 있도록 정확한 영상재현 성능 및 음상의 고저감을 표현할 수 있어야 한다.
- Personal 3DTV: 22.2 채널과 같은 다채널로 대표되는 홈시어터뿐만 아니라 동일한 오디오 콘텐츠를 태블릿 PC와 같은 소형 개인 단말에서 헤드폰이나 스피커 어레이를 이용하여 원음장과 유사한 음질과 음장감을 제공할 수 있어야 한다.
- TV for smart phone: 헤드폰을 이용하여 3차원 오디오 콘텐츠를 원음장과 유사한 음질과 음장감으로 재생할 수 있어야 한다.
- Multichannel audio program: 오디오 CD와 같이 오디오 단독으로 다채널 재생 환경을 이용하여 실감 있고 몰입감 있는 오디오 서비스를 제공하여야 한다.
- Telepresence: 원격회의 시스템, 가상현실 시스템 등과 같이 네트워크를 이용하여 원격에서 가상공간을 재현할 때 오디오도 3차원 공간상에서 재현되어야 한다.

나. 표준화 요구사항(Requirements)

상기 응용 분야에 3D Audio 표준이 사용되기 위해서 만족해야 할 요구사항들이 논의되고 있다. 3D Audio에서는 요구사항들을 주요(primary) 요구사항, 부가(secondary) 요구사항, 바람직한 특징(desirable feature) 세 가지로 구분하여 정의하고 있다. 본고에서 제시되는 요구사항은 2012년 7월까지 논의된 결과이며, 변경되거나 삭제될 수 있으며 새로운 요구사항이 추가될 가능성도 있다.

먼저 표준화에서 중요하게 다루어질 주요 요구사항(primary requirements)은 다음과 같다.

- High quality: 22.2 채널과 같은 초다채널 환경에서 높은 음질(perceptual transparent quality)을 제공하여야 한다.
- Localization and envelopment: 거리감을 포함하여 렌더링되는 음상의 정확한 표현과 풍부한 음장 포위감(envelopment)을 제공하여야 한다.
- Rendering on setups with fewer loudspeakers: 입력 오디오 신호의 채널 수보다 적은 수의 채널로 재생/렌더링될 때에도 최적의 음질과 음장감을 제공하여야 한다.
- Flexible loudspeaker placement: 입력 오디오 신호의 채널배치와 다른 위치에 스피커가 위치할 때에도 최적의 음질과 음장감을 제공하여야 한다.
- Audio/visual alignment & consistency: 비디오와 오디오 객체의 재현되는 공간상의 위치가 동일하여야 한다.
- Latency: 생방송과 같은 환경을 지원하기 위해 인코더/디코더에 의한 지연이 적어야 한다.
- Audio program inputs to the submitted encoding systems: 인코더로 입력되는 오디오 신호의 형태는 채널신호, 오디오 객체신호, High Order Ambisonics(HOA)와 같은 오디오 장면(audio scene) 신호가 될 수 있다.
- Transcoding for low bandwidth devices: 저대역(낮은 음질과 공간해상도) 단말을 위하여 오디오 비트스트림으로부터 저대역 오디오 신호를 적절한 복잡도로 추출할 수 있어야 한다.

RMO(Reference Model 0) 선정 단계에서는 고려하지 않겠지만 다음으로 중요하게 다루질 요구사항(secondary requirements)은 다음과 같다.

- Efficiency for decoding on different setups: 입력 오디오 신호와 상이한 오디오 출력 환경에서도 효율적으로 변환하여 적절한 음질을 제공하여야 한다.
- Computational complexity: 여러 가지 응용 환경

에서 적절한 복잡도를 가져야 한다.

- Issues of backward compatibility: 기존 5.1 채널 재생 환경에서도 변환하여 재생할 수 있어야 한다.
- Interactivity: 오디오 객체를 제어하는 것과 같이 대화형 오디오 서비스에서의 interactivity를 제공할 수 있어야 한다.

마지막으로 3D Audio 코덱의 장점으로 부각될 수 있는 특징(desirable features)들은 아래와 같다.

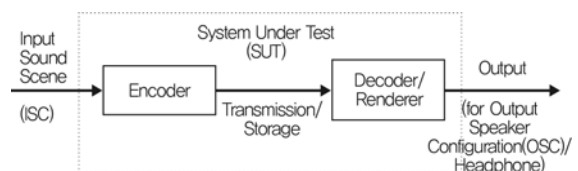
- Compressed representation that allows extraction of lower-bitrate lower quality representations: 비트스트림 레벨에서 낮은 비트율과 낮은 음질의 신호를 손쉽게 추출할 수 있어야 한다.
- Low latency and low computational complexity: 양방향 통신 환경이나 telepresence와 같은 응용 환경을 위하여 적은 지연과 낮은 복잡도를 가져야 한다.

다. 표준화 대상 기술 분야

2013년 상반기에 공지될 CFP에 대하여 여러 기관에서 제출될 기술 제안서의 성능을 평가하기 위한 절차와 기준을 논의 중에 있으며, 현재까지 합의된 개괄적인 시스템 구성도는 (그림 6)과 같다.

(그림 6)에 사용된 용어들에 대한 정의는 다음과 같다.

- SUT(System Under Test): 3D Audio 콘텐츠를 특정 비트율에서 인코딩/디코딩/렌더링하는 시스템을 의미하며, CFP에 대응하여 제안된 개별적인 시스템들을 지칭한다.



(그림 6) MPEG 3D Audio 인코더/디코더/렌더러 시스템 구성도

- ISC(Input Sound Scene): SUT로 입력되는 오디오 신호를 의미하며 채널신호, 오디오 객체신호, 오디오 장면신호가 될 수 있다. 또한 세 가지 신호들이 조합된 형태로도 입력이 가능하다.
- OSC(Output Speaker Configuration): SUT가 출력하는 오디오 신호를 의미하며 ISC와 상이한 형태도 가능하다.

상기와 같은 형태로 오디오 코덱이 구성된다면 전술한 표준화 요구사항을 만족시키기 위해서는 다음과 같은 기능블록들이 필요하다고 예측할 수 있다.

- 개별채널 신호 또는 객체신호를 고효율/고음질로 압축하기 위한 모노 또는 스테레오 코어코덱: 전통적으로 MPEG에서는 MPEG 표준기술을 활용하는 것을 권고하기 때문에 MPEG-4 HE-AAC (Advanced Audio Coding) v2나 MPEG-D USAC (Unified Speech and Audio Coding)가 활용될 가능성이 높다.
- 스케일러블 채널구조: 방송 환경과 같이 제한된 대역폭에서 다양한 포맷(예: 22.2, 10.2, 5.1)을 지원하여야 할 경우 스케일러블 채널형태로 비트스트림을 구성할 수 있으며, 이로부터 발생할 수 있는 음질저하를 상쇄시킬 수 있는 기술이 필요하다.
- 채널신호와 객체신호를 모두 제어할 수 있는 hybrid 코덱 구조: 객체신호만으로 제공되는 오디오 서비스는 일반적이지 않지만 객체신호가 채널신호와 공존하는 형태로의 서비스는 DTS와 Dolby에서 제안하고 있다. 따라서, 채널신호와 객체신호를 모두 제어할 수 있는 hybrid 코덱 구조가 3D Audio 표준에 채택될 가능성이 높다.
- 다양한 출력채널(OSC) 환경과 스피커의 배치 환경에 적응적인 렌더러 기술: 렌더러가 표준의 범주에 포함되는 것에 대해서 공감대가 형성되고 있으며, 코덱 자체보다는 렌더러가 3D Audio 표준에서 중요한 부분을 차지할 것으로 예측된다.

또한, 디코더와 렌더러가 분리된 형태가 아닌 결합된 형태가 될 것으로 예측되며, MPEG Surround나 SAOC(Spatial Audio Object Coding)에서 표준화된 다양한 채널변환 툴들이 3D Audio 표준에서도 채택될 가능성이 높다.

IV. 결론

5.1 채널 이상의 채널 수를 가지는 오디오 시스템을 초다채널 오디오로 정의하여 관련한 시스템과 표준화 동향을 소개하였다. 비디오에 비하여 오디오는 기술 개발속도에 비하여 상용화 속도가 이에 미치지 못하는 실정이다. 그러나 이러한 주변상황을 성급히 판단하여 기술개발 및 표준화를 머뭇거리다가는 과거와 같이 핵심 기술에 대한 특허권을 선점 당해 막대한 로열티를 지불하는 결과를 초래할 수 있다. 현재 표준화가 진행 중인 MPEG-H 3D Audio와 Dolby를 중심으로 극장 사운드로 소개되고 있는 초다채널 및 객체기반 오디오 서비스에 대한 핵심 기술을 확보할 필요가 있다.

용어해설

UHDTV(Ultra-High Definition Television, 초고해상도 텔레비전) 일본 NHK에서 SHV(Super Hi-Vision)란 이름으로 제안되었으며 HDTV 해상도의 4배(4K) 또는 16배(8K)의 해상도를 가지는 고해상도 비디오 서비스와 22.2채널 오디오를 이용한 실감 오디오 서비스를 제공함.

MPEG-H NHK의 SHV와 같은 초고해상도(4K/8K) 대화면 디스플레이(100인치 이상)와 초다채널 오디오 채널(22.2채널 등)을 이용하는 실감 몰입형(immersive) 멀티미디어 서비스를 위한 MPEG의 새로운 표준화 프로젝트.

약어 정리

AAC	Advanced Audio Coding
AhG	Adhoc Group
HEVC	High Efficiency Video Coding
HOA	High Order Ambisonics
HRTF	Head Related Transfer Function
ISC	Input Sound Scene

OSC	Output Speaker Configuration
RM0	Reference Model 0
SAOC	Spatial Audio Object Coding
SHV	Super Hi-Vision
SUT	System Under Test
UHDTV	Ultra High Definition Television
USAC	Unified Speech and Audio Coding
WFS	Wave Field Synthesis

참고문헌

- [1] ITU-R BS.775-1, "Multichannel Stereophonic Sound System with and without Accompanying Picture," 1994.
- [2] <http://www.auro-technologies.com/>
- [3] T. Holman, "The History and Future of DSPs in Consumer Audio Equipment — Part II: Emerging Areas and the Future," *ICCE*, 2008.
- [4] Y.W. Lee et al., "Virtual Height Speaker Rendering for Samsung 10.2-Channel Vertical Surround System," *131st AES Convention*, Oct. 2011.
- [5] K. Hamasaki et al., "5.1 and 22.2 Multichannel Sound Productions Using an Integrated Surround Sound Panning System," *NAB BEC Proc.*, Apr. 2005.
- [6] SMPTE 2036-1, "Ultra High Definition Television-image Parameter Values for Program Production," 2009.
- [7] SMPTE 2036-2, "Ultra High Definition Television-audio Characteristics and Audio Channel Mapping for Program Production," 2008.
- [8] TTAK.KO-07.0098, "초고선명 디지털 TV 오디오 신호," 2011.
- [9] 서정일 외, "멀티채널 사운드의 역사와 기술개발 동향," *한국방송공학회지*, vol. 16, no. 4, 2011, pp. 8-19.
- [10] ISO/IEC JTC1/SC29/WG11 (MPEG), "Draft Call for Proposal for 3D Audio," N12997, July 2012.