

EXPERT
KNOW-HOW137th AES Convention 참관기
MPEG-H 3D를 중심으로

+ 정인 MBC 중계부 차장

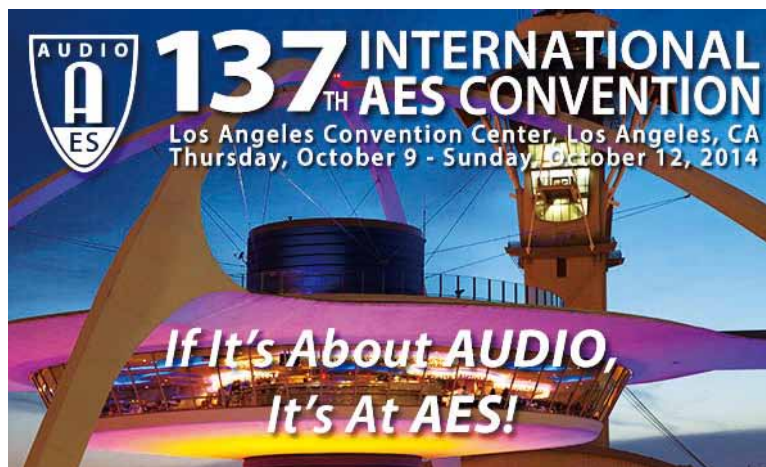


그림 1. AES Convention Banner



그림 2. 필자(좌)와 SBS 김지완(우)

(사)방송기술교육원이 주최하는 해외교육 사업의 일환으로 MBC의 필자와 SBS의 김지완 씨가 함께 10월 9일부터 12일까지 Los Angeles Convention Center에서 개최되는 137회 AES(Audio Engineering Society) Convention에 참석하게 되었다. 이런 행사가 있다는 것은 대략 인지하고 있었으나 예상보다 훨씬 큰 규모이고 전 세계 음향 엔지니어들의 관심과 열정에 방송음향 종사자의 한 사람으로서 대단히 영광스러운 자리였다. [그림 3]과 같이 행사의 종류도 매우 다양해서 모든 교육과정에 참가할 순 없었고, 관심도가 높은 주제를 우선순위로 4일간 쉼 없이 달려 나갔다.

커리큘럼은 크게 다음과 같이 구성되어있었다.

- Special Tracks
 - Broadcast & Streaming Media
 - Game Audio
 - Live Sound Seminar
 - Networked Audio
 - Product Design
 - Recording & Production
 - Sound for Picture

- Paper Sessions
- Workshops
- Tutorials
- Historical
- Project Sound Expo
- Live Sound Expo
- Technical Tours
- Exhibition

■ 2014.10.9(THU) - Day 1

- #1 Tutorial 2 '3D AUDIO'
- #2 Technical Tour 3 'Staples Center'
- #3 Broadcast/Streaming Media Session 2 'Loudness for Streaming and Radio'
- #4 Live Sound Seminar 2 'The Creative Technologist : Career Paths of the Audio Professional'
- #5 Technical Tour 4 '20th Century Fox'

■ 10.10(FRI) - Day 2

- #6 Broadcast/Streaming Media Session 4 'Audio Issues for 4K and 8K Television'
- #7 Technical Tour 6 'Dolby Theater'
- #8 Technical Tour 7 'Walt Disney Concert Hall'
- #9 Technical Tour 8 'Capitol Studio'

■ 10.11(SAT) - Day 3

- #10 Technical Tour 9 'Henson Recording Studio'
- #11 Live Sound Expo 9 'The Small Venue Monitor Mix'
- #12 Project Studio Expo 13 'Listen Up, And Learn!-Track 2'
- #13 Broadcasting/Streaming Media Session 13 'SMPTE : Audio Issues for Live Television-Overcoming the Challenges of Live Television Broadcast in Today's Wild, Wild World'
- #14 Product Design Session 14 'Specifying and Selecting Loudspeaker Drivers'
- #15 Recording and Production Session 5 'An Afternoon with Geoff Emerick'

■ 10.12(SUN) - Day 4

- #16 Recording and Production Session 6 'Inside the Mix by Dave Reitzas'
- #17 Technical Tour 11 'The Dub Stage'
- #18 Sound for Picture 6 'Superstars of Production Sound Recording'
- #19 Sound for Picture 7 'Recording Production Sound - A Master Class'

붉은색으로 표시된 부분이 방송 음향에 대해 다룬 이론 교과이다.

[그림 4]와 같이 음향에 관한 이론교육·토론·명사강의·질문 등으로 구성된 세션과 참가자들이 직접 헤드폰을 쓰거나 PA를 이용한 믹싱 기법강의·Masterpiece Review 등으로 구성된 세션, LA 인근에 있는 음향 명소들을 직접 찾아가 설명을 듣는 Technical Tour, 약 300여 개 업체가 참가한 전시회 등으로 크게 구성되어 있다. 여타 과정도 좋았지만 특히 기억에 남는 건 Technical Tour로 일반인들이 접근할 수 없는 구역들에 직접 가볼 수 있었고, 미디어상에서나 접하던 스타 엔지니어들에게서 친히 장비·시설 등에 관하여 설명을 들을 수 있다는 점이 매우 인상 깊었다. Technical Tour는 'Staples Center, 20th Century Fox, Dolby Theater, Walt Disney Concert Hall, Capitol Studio, Henson Recording Studio, The Dub Stage' 등에 방문하였고, 이곳들은 하나같이 깊은 역사와 전통으로 자긍심이 뛰어났으며, 빠르게 변화 발전하는 음향 환경에서도 자신들만의 노하우와 역사 깊은 장비들을 가지고 뛰어난 퀄리티의 사운드를 만들어 나간다는 점이 인상 깊었다. 실제로 반세기 넘게 탑클래스로 인정받아 오고 있는 녹음 스튜디오들에서는 아날로그 AMU와 아날로그 멀티 트랙 레코더, 에이징이 잘 된 중저가대의 모니터 스피커, 50년이 넘는 마이크와 아웃보드 장비들을 아직까지 애용하고 자랑으로 삼는다는 점이 신형 장비 의존증이 심한 우리의 상황과 비교가 되었다.



그림 4. 좌상부터 시계방향으로 Tutorial, PSE, Technical Tour, Exhibition



그림 5. Recording Studios (좌상부터 시계방향: 20th Fox, Capitol, Henson, Multi-Track Recorders)

지금부터는 AES에서 진행된 강의내용을 바탕으로 현재 방송계의 화두가 되고 있는 UHD(4K, 8K) 방송의 Audio Format에 관하여 다루어 보고자 한다. NHK의 SHV(Super Hi-Vision) 방송을 필두로 하여 차세대 방송 표준으로 4K, 더 나아가 8K를 거론한다. 영상으로는 장족의 발전과 머지않은 상용화, 대량의 수요층이 예상되지만 음향적인 측면에서는 22.2CH이라는 현실성이 떨어지는 모호한 포맷으로 음향발전에 의문을 품고 있는 것이 사실이다. 필자는 14년 동안 방송음향에 매진하고 있는 종사자로서 과연 현실적인 음향 발전 방향이 무엇인가에 대해 생각해 보기로 한다.

현재 우리나라의 DTV 송출 방식은 ATSC 방식으로 유럽의 DVB-T 방식과 비교가 된다. UHDTV의 송출 방식은 아직 확정이 되지 않았고, 일본의 SHV 방송과 인천 아시안게임의 4K 방송은 유럽의 DVB-T2 방식을 이용해 송출이 되었다. ATSC 3.0의 UHDTV 오디오 표준으로 는 MPEG의 MPEG-H 3D, Dolby의 AC-4, DTS의 DTS-UHD가 제안되고 있지만, 아직 구체적인 표준안은 나오지 않은 상태다. 그중 현재 UHDTV의 시스템과 비디오 압축 코덱으로 사용되고 있는 MPEG-H에 대해 알아보자.

■ MPEG-H standard

- **System** : MPEG-H Part 1 : **MMT** (MPEG media transport)
 - A media streaming format similar to the Real-time Transport Protocol that is adaptable to different networks
- **Video** : MPEG-H Part 2 : **HEVC** (High Efficiency Video Coding)
 - A video compression standard that doubles the data compression ratio compared to H.264/MPEG-4 AVC and can support resolutions up to 8192×4320v
- **Audio** : MPEG-H Part 3 : **3D Audio - CFP** (Call for Proposal - 2015.2월 결정)
 - An audio compression standard for 3D audio that can support many loudspeakers

MMT와 HEVC(H.265)로 불리는 시스템과 비디오의 코덱은 거의 표준화가 완료된 상태이고, 다만 오디오에서 MPEG-H 3D로 일컫는 코덱은 내년 2월의 MPEG 표준화 회의에서 결정될 예정이다. 오디오의 표준화가 늦어진 이유는 비디오의 경우와 다르게 오디오 코덱은 이미 압축 기술 자체는 수렴 단계에 이르렀고, MPEG 계열의 코덱들(AAC, MPEG Surround, SAOC)과 그밖에 시장에서 널리 사용되는 상용 오디오 코덱들(AC3, DTS)이 차세대 멀티미디어 시장에서 사용하기에 큰 불편함이 없기 때문이기도 할 것이다.

구분		HDTV		4K-UHDTV		8K-UHDTV	
화면당 화소수		1920×1080		3840×2160		7680×4320	
수평시야각		30°		55°		100°	
시청거리		3.0H		1.5H		0.75H	
		2m	2.5m	2m	2.5m	2m	2.5m
TV크기	대각선	55'	67'	110'	134'	220'	268'
	가로	121cm	148cm	242cm	296cm	484cm	592cm
	세로	68cm	83cm	136cm	166cm	272cm	332cm

표 1. HD, 4K, 8K 비교표

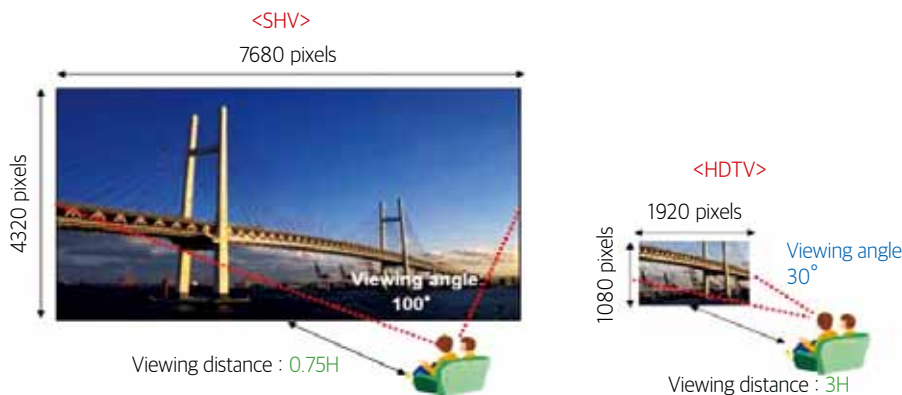


그림 6. SHV(8K)와 HD 화면 비교

[표 1]에서 보는 바와 같이 UHD는 HD 대비 최대 16배의 해상도를 제공하고 있다. 100°에 이르는 시야각과 기존 HD가 2m 거리의 픽셀 인지화면이 55라면 8K는 같은 거리에서 무려 220°에 이른다. 즉, 거실의 한 벽면 전체가 하나의 디스플레이가 되고, 화면상의 객체가 좌우뿐 아니라 높낮이에서도 차별적인 소리를 제공할 수 있어야 한다. 기존의 서라운드 방식이 너비(W)와 깊이(D)의 2D 오디오라면 UHD 오디오는 여기에 높이(H)의 개념을 더해 진정한 3D 오디오를 제공한다는 것이다.

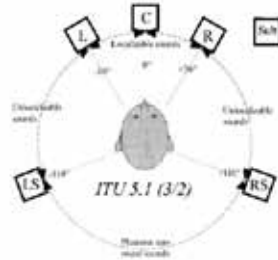


그림 7. Legacy 5.1 Speaker placement

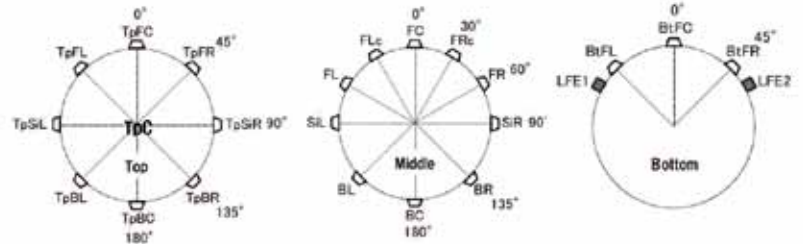
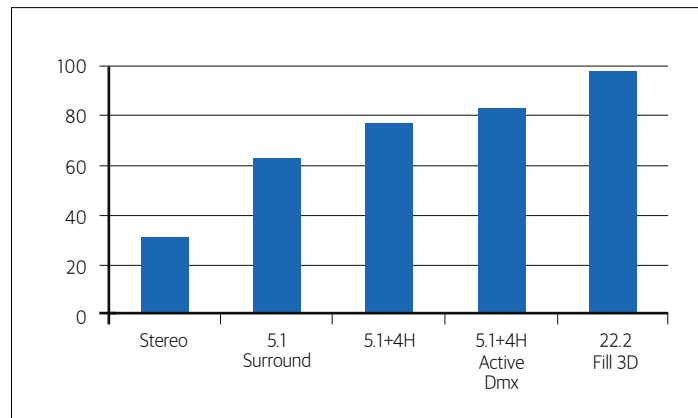


그림 8. 22.2 Speaker Placement

[그림 8]에서와 같이 22.2는 9개 스피커의 Top Layer와 10개의 Middle Layer, 3개의 Bottom Layer와 2개의 LFE로 구성되어 있다.



그렇다면 일반 가정환경에서는 현실적으로 제대로 된 스피커의 구성이 힘든 22.2 오디오 포맷을 왜 제안하고 있는 것일까? [그림 9]와 같이 몰입도와 가상현실의 오디오 측면에서 22.2를 표현해 낼 수 있는 최대치로 잡고 2.0 5.1 7.1 9.1 11.1 등과 같은 포맷들로 변환 가능하도록 한다는 것이 표준화의 핵심 화두이기도 하다.

그림 9. Overall Sound Quality Impression

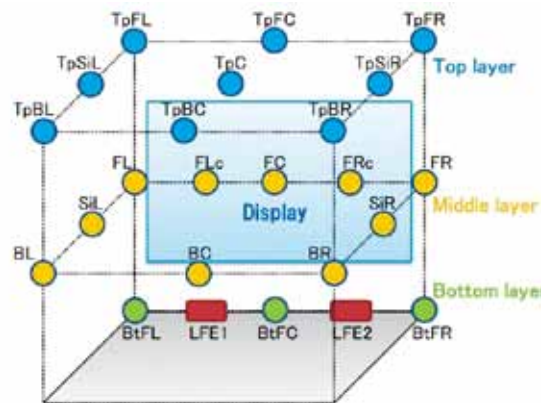


그림 10. 22.2 Audio Layout

Syntactic element	Channel mapping
SCE	FC
CPE	FLc, FRc
CPE	FL, FR
CPE	SIL, SIR
CPE	BL, BR
SCE	BC
LFE	LFE1
LFE	LFE2
SCE	TpFC
CPE	TpFL, TpFR
CPE	TpSL, TpSR
SCE	TpC
CPE	TpBL, TpBR
SCE	TpBC
SCE	BtFC
CPE	BtFL, BtFR

그림 11. NHK 22.2 to 16ch SDI

방송 환경은 월드컵과 올림픽 등 국제 스포츠 행사를 통해 진화한다고 해도 과언이 아니다. 2014 인천아시안게임에서는 지상파 3사가 NTT/NEC HEVC Encoder를 이용해 DVB-T2 방식으로 4K 방송을 실시간으로 송출하였다. 오디오는 MPEG2-AAC Stereo와 5.1을 사용하였다. NHK는 2016년 브라질 리오 올림픽에서 SHV(8K) 방송을 22.2로 제작해 방송할 계획이다. 아직 MPEG-H 3D가 표준화되지 않은 관계로 MPEG4-AAC를 이용해 [그림 11]과 같이 Pair 채널은 Co-Relation을 이용해 하나로 묶어서 기존의 SDI 오디오 포맷과 같이 16채널로 구성하였다. 과연 우리나라는 2018년 평창올림픽을 대비해 어떤 기술을 선보일지 기대가 된다.



그림 12. NHK 22.2 Evaluation Room



그림 13. Dub Stage 22.2 Mixing Hall

MPEG-H 3D Audio가 목표로 하는 응용분야들은 아래와 같다.

- Home Theater : HEVC가 목표로 하는 대화면 고해상도 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 음상 재현 성능 및 음상의 고저감을 표현할 수 있어야 한다.
- Personal 3DTV : 22.2채널과 같은 다채널로 대표되는 홈시어터뿐만 아니라 동일한 오디오 콘텐츠를 태블릿 PC와 같은 소형 개인 단말에서 헤드폰이나 스피커 어레이를 이용하여 원음장과 유사한 음질과 임장감을 제공할 수 있어야 한다.
- TV for Smart Phone : 헤드폰을 이용하여 3차원 오디오 콘텐츠를 원음장과 유사한 음질과 임장감으로 재생할 수 있어야 한다.
- Multichannel Audio Program : 오디오 CD와 같이 오디오 단독으로 다채널 재생환경을 이용하여 실감 나고 몰입감 있는 오디오 서비스를 제공하여야 한다.

주요 요구사항(Primary Requirements)은 다음과 같다.

- High quality : 22.2채널과 같은 고다채널 환경에서 높은 음질(perceptual transparent quality)을 제공하여야 한다.
- Localization and envelopment : 거리감을 포함하여 렌더링 되는 음상의 정확한 표현과 풍부한 음장포위감(envelopment)을 제공하여야 한다.
- Rendering on setups with fewer loudspeakers : 입력 오디오 신호의 채널 수보다 적은 수의 채널로 재생/렌더링 될 때에도 최대한 가능한 음질과 임장감을 제공하여야 한다.
- Flexible loudspeaker placement : 입력 오디오 신호의 채널배치와 다른 위치에 스피커가 위치할 때에도 최대한의 음질과 임장감을 제공하여야 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(Channel-Based), 객체신호(Object-Based), High Order Ambisonics (HOA)와 같은 오디오 장면(Audio Scene) 신호가 될 수 있다.
- Rendering for Headphone Listening : 헤드폰 환경에서 청취가 가능하여야 하며, 이는 출력 다채널 신호를 바이노럴(binaural) 포맷으로 변환 가능한 처리 기술을 필요로 한다. 예를 들어, 개인화된 HRTF(Head Related Transfer Function) 함수를 지원하는 방법도 표준화 대상 기술이 된다.

다음으로 중요하게 다루어질 요구사항(Secondary Requirements)은 다음과 같다.

- Computational complexity : 여러 가지 응용환경에서 적절한 복잡도를 가져야 한다.
- Interactivity : 오디오 객체를 제어하는 것과 같이 대화형 오디오 서비스에서의 interactivity를 제공할 수 있어야 한다.

위에서와 같이 MPEG-H는 단순히 압축을 얼마나 더 잘하느냐가 아니라, 실제 재생 환경에서 쓸모가 있는 오디오 코덱을 만들고자 하는 것이다. 이와 같은 고민 속에 자연스럽게 등장한 것이 객체 오디오 신호(Object-Based)이다. 여기서 객체란 사운드 장면을 구성하는 각 음원을 의미한다.



그림 14. Band Mixing Sound Image

[그림 14]와 같은 밴드 믹싱 사운드에서 기타, 드럼, 베이스 등이 각각 하나의 객체라고 볼 수 있다. 기존의 오디오 신호를 다루는 패러다임은 이와 같은 객체들이 한데 어우러진 사운드 장면을 재생 단의 스피커에 1:1로 매칭되는 채널(Channel-Based) 단위로 묶어서 처리하는 것이었다. 이와 같은 채널 대신, 객체 단위로 신호를 저장 전송하고, 사용자 단에서 이를 다시 재구성하여 사용자의 스피커 구성 환경에 독립적으로 제작자가 원하는 사운드 장면을 제공할 수 있다. 즉, 사용자가 22개 스피커를 가지든, 헤드폰으로 청취하든 그 환경에 맞게 렌더링할 수 있다.

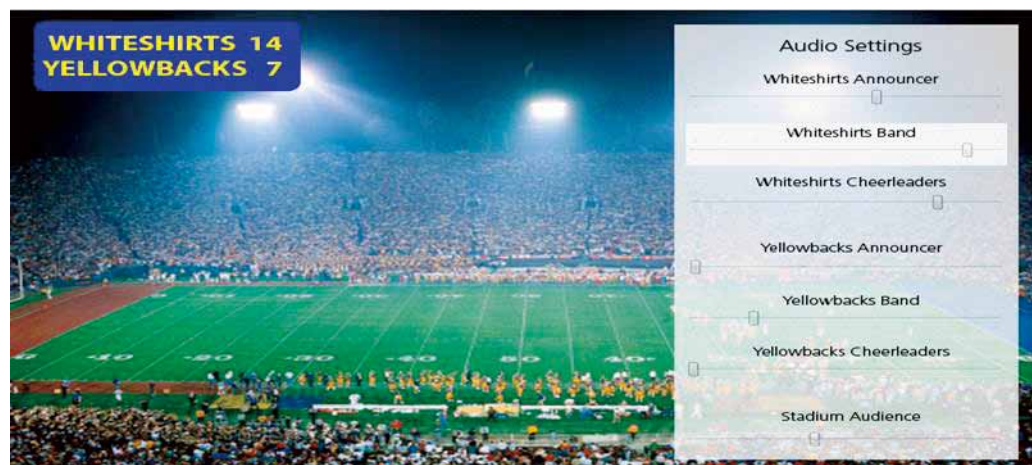


그림 15. Object-Based Interactive Audio

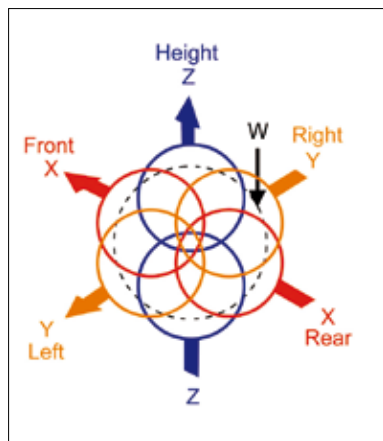


그림 16. Ambisonics Format

[그림 15]와 같이 스포츠중계에서 마치 비디오게임의 오디오처럼 사용자가 자신의 취향에 맞게 듣고 싶은 사운드를 조절한다는 개념이다. 3D 오디오 표준화의 논의 가운데, 채널과 객체 이외에 새로운 입력 신호로 거론된 것이 HOA(Higher Order Ambisonics) 신호이다. Ambisonics란 이미 70년대에 SoundField사에서 개발된 녹음 기법으로 [그림 16]과 같이 동축상에 한 대의 무지향성 마이크와 두 대의 쌍지향성 마이크를 조합하여 세팅하고 여기에 앰비소닉 엔코더를 삽입해 수음하여 음원에 방향과 위치 정보를 포함시키는 방식이다.

HOA(Higher Order Ambisonics)는 Ambisonics를 다계층으로 구성하는 방법으로 동그런 캡슐 안에 Ambisonics 마이크를 여러 개 구성하여 입체감을 좀 더 세밀하게 표현해 낼 수 있다. 높이의 개념이 도입된 3D 오디오에서 매우 유용한 패러다임으로 대규모 관현악단과 같은 연주에서 지휘자의 포지션에 HOA mic를 이용해 녹음을 한다면 악기 각각의 위

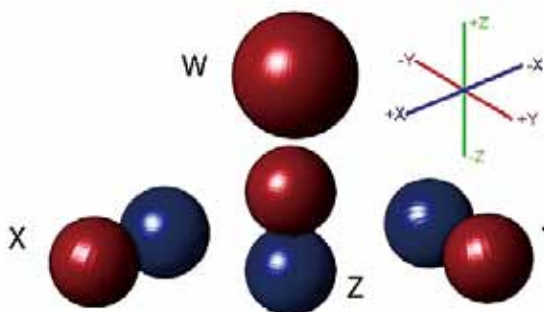


그림 17. Ambisonics B-Format

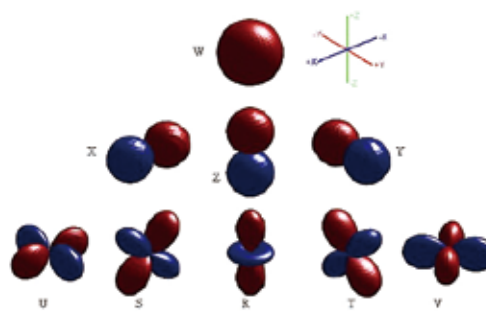


그림 18. HOA Format



그림 19. SoundField Microphones (B-Format type과 HOA type)

치를 비교적 정확히 표현할 수 있고 후반 작업 시에도 입체적인 정보를 제공할 수 있다.

최근의 MPEG 표준화 회의에서는 독일의 Fraunhofer IIS가 Channel과 Object 개념을 묶은 CO의 선점기업이 되었고, HOA는 프랑스의 Technicolor와 Orange가 동맹을 하여 특허권을 획득하였다. 여기에 다시 CO와 HOA를 통합하려는 움직임도 비치고 있다. [그림 20]은 디코더의 블록 다이어그램으로 보이는 바와 같이 크게 Channel, Object, HOA 개념으로 나뉘고, 특히 Object-Based는 Metadata의 역할

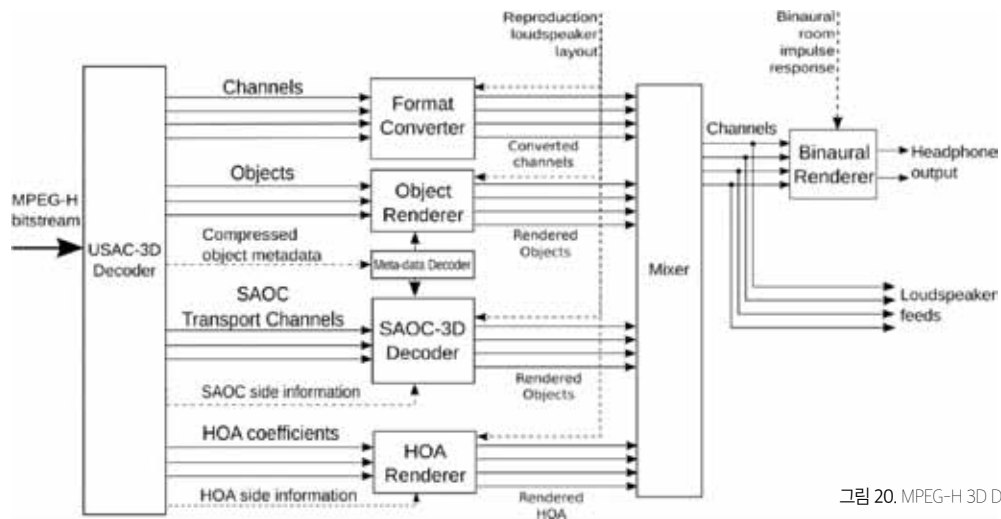


그림 20. MPEG-H 3D Decoder Block Diagram

Bitrates in kbs for;	Good	Recommended	Transparent
22.2 Channels	256	512	1200
5.1 Channels	96	160	256
2.0 Channels	32	56	160

표 2. Bitrates 비교표

이 무엇보다 중요하게 되었다.

지금까지 대략적인 MPEG-H 3D의 개념을 살펴보았다. 필자는 5.1 채널이 등장한 시기부터 드라마, 다큐멘터리 등의 5.1 콘텐츠 제작에 직접 참여하였으며, 스포츠와 공연의 실시간 5.1 중계에도 참여한 바 있다. 5.1 콘텐츠의 제작은 단순한 비교로 스테레오 제작에 비해 3배 이상의 시간을 필요로 한다(사실 훨씬 더 많은 시간이 소요되기도 한다). 그러면 22.2는 12배의 시간이 필요할까? 불행히도 그럴 듯싶다. 게다가 작은 영화관 크기의 전문 믹싱 공간도 필요하다. HD 방송과 맞물려 오디오계의 화두가 되었던 지상파 5.1은 콘텐츠 부족과 결정적으로 수동적인 수용자들에게 능동적인 스피커 시스템과 음향장비의 조작을 요구하여 입체음향에 목말라하는 오디오 마니아를 제외하고는 근래 잊혀져버린 존재가 되었다. 분명 22.2는 UHD 방송에 있어서 매력적인 존재인 건 분명하나 현실적인 지상파방송 제작에서는 회의적이다. 개인 디바이스 재생이 더 중요해질 미래의 방송환경에 맞춰 Headphone과 같은 Format Rendering의 개념은 큰 장점이 될 수 있고 실제로 ETRI(한국전자통신연구원)는 연세대학교, 윌러스표준연구소와 공동 개발한 차세대 방송환경에서 다채널(22.2채널) 오디오 콘텐츠를 스마트폰의 헤드폰으로 현장감 있게 재생할 수 있는 '멀티밴드 헤드폰 재생 기술'을 개발하여 MPEG-H 3D Audio 국제 표준으로 채택되기도 했다. 객체 기반 전송도 여러 분야로 응용해 볼 수 있는 매력적인 개념이다.

교육과정 중 어느 외국 엔지니어가 말한 다음의 내용을 마지막으로 마무리하고자 한다.

‘전 세계 어디나 엔지니어의 고충은 똑같다. 산업과 맞물려 방송환경은 주기적인 진화를 하여야 하고 엔지니어들은 주어진 과제를 해결해 나가기 위해 오늘도 끊임없이 노력한다.’ 

참고자료 및 이미지 출처

- [1] Seungkwon Beack et al., "Overview of MPEG-H 3D Audio Standard Activities" - ETRI 2013
- [2] Jürgen Herre et al., "MPEG-H Audio - The New Standard for Universal Spatial/3D Audio Coding" - IAL 2014
- [3] Takehiro Sugimoto et al., "Bit Rate of 22.2 Multichannel Sound Signal Meeting Broadcast Quality" - NHK 2014
- [4] Simone Füg et al., "Design, Coding and Processing of Metadata for Object-Based Interactive Audio" - Fraunhofer 2014
- [5] <http://www.gooddaylab.com> [6] <http://www.whydsp.org>