

6자유도 오디오 표준화 및 기술 동향
(기술보고서)

A Trend of 6 DoF Audio Standardization and
technology

(앞 표지)

기술보고서 초안 검토
위원회
기술보고서안 심의 위원회
미래미디어분과위원회
운영위원회

	성명	소속	직위	위원회 및 직위	기술보고서번호
기술보고서(과제) 제안	윤국진	ETRI	책임	미래미디어분과 위원장	FBMF-TR-008
	김홍국	광주과학기술원	교수	차세대오디오 WG 의장	FBMF-TR-008
	이용주	ETRI	책임	차세대오디오 WG 위원	FBMF-TR-008
	박호종	광운대학교	교수	차세대오디오 WG 위원	FBMF-TR-008
기술보고서 초안 작성자	정현주	(주)가우디오랩		차세대오디오 WG 위원	FBMF-TR-008

사무국 담당 김제우 KETI 센터장 운영위 간사 FBMF-TR-008

본 문서에 대한 저작권은 미래방송미디어표준포럼에 있으며, 미래방송미디어표준포럼과 사전 협의 없이 이 문서의 전체 또는 일부를 상업적 목적으로 복제 또는 배포해서는 안 됩니다.

본 표준 발간 이전에 접수된 지식재산권 협약서 정보는 본 표준의 '부록(지식재산권 협약서 정보)'에 명시하고 있으며, 이후 접수된 지식재산권 협약서는 미래방송미디어표준포럼 웹사이트에서 확인할 수 있습니다.

본 표준과 관련하여 접수된 협약서 외의 지식재산권이 존재할 수 있습니다.

발행인 : 미래방송미디어표준포럼 의장
 발행처 : 미래방송미디어표준포럼
 06130, 서울특별시 강남구 테헤란로 7길 22 신관 1108호
 Tel : 02-568-3556, Fax : 02-568-3557
 발행일 : 2021.11

서 문

1 기술보고서의 목적

이 기술보고서의 1차적인 목적은, 최근 전세계적으로 관심이 높아지고 있는 6자유도 오디오 관련 기술에 대한 기본 개념과 표준화 및 기술 동향을 요약/정리하고 주요 이슈 및 고려 사항을 제시함으로써, 관련 산업 종사자 및 표준화 담당자가 전반적인 사항을 쉽고 빠르게 파악할 수 있도록 하는 데 있다. 또한 이 기술보고서의 2차적인 목적은, 6자유도 오디오 관련 기술에 대한 국내 표준화가 필요한지에 대한 것을 분석하기 위한 기초 자료를 제공하는 데 있다.

본 보고서의 버전은 1.0이며 2021년까지의 6자유도 오디오 관련 기술의 개발 및 표준화 동향을 분석하는 것을 그 범위로 한다. 6자유도 오디오 관련 기술 개발과 표준화가 진행 중인 점을 고려하여, 2022년에 최신 동향을 반영한 버전 2.0을 발간할 예정이다.

2 주요 내용 요약

이 기술보고서의 5장에서는 6자유도 오디오의 개념에 대하여 간략하게 소개한다. 5.1절에서는 3자유도 오디오와 6자유도 오디오를 비교하여 설명하고, 5.2절에서는 6자유도 오디오에서 활용될 수 있는 대표적인 오디오 포맷에 대해 설명한다.

이 기술보고서의 6장에서는 6자유도 오디오 표준화 동향에 대해 기술한다. 6.1절에서 현재 표준화가 활발하게 진행 중인 MPEG-I audio 표준화 현황 및 내용에 대해 요약한다.

이 기술보고서의 7장에서는 6자유도 오디오 기술 동향에 대해 기술한다. 6자유도 오디오에서 활용될 수 있는 기술로 7.1절에서는 Acoustic Project에 대해 소개를 하고, 7.2절에서는 Resonance audio에 대해 소개한다.

이 기술보고서의 8장에서는 본 기술보고서의 결론 및 시사점에 대해서 기술한다.

3 인용 기술보고서와의 비교

3.1 인용 기술보고서와의 관련성

해당사항 없음

3.2 인용 표준과 본 기술보고서의 비교표

해당사항 없음

Preface

1 Purpose

The main purpose of this technical report (TR) is to summarize the base technologies and related requirements regarding immersive media like 360 video and audio, and to provide related issue and considerations for immersive media industry.

An additional purpose of this TR is to provide a base material for analyzing the technical availability for a domestic 360VR service standard.

2 Summary

Chapter 5 covers the 360VR file format (OMAF); section 5.1 and 5.2 analyzes OMAF requirements and standardization status. Chapter 6 and 7 analyze core technologies of system, audio and video components for a 360VR service. Chapter 8 covers various 360VR services and issues based on the IP and 5G network. Chapter 9 summarizes main requirements for a domestic 360VR service.

3 Relationship to Reference Standards

N/A

목 차

1	적용 범위	1
2	인용 표준	1
3	용어 정의	1
4	약어	2
5	6자유도 오디오 소개	3
	5.1 3자유도 오디오와 6자유도 오디오	3
	5.2 오디오 포맷의 종류와 활용	6
6	6자유도 오디오 표준화 동향	13
	6.1 6자유도 미디어 관련 MPEG 표준화 동향	13
	6.2 MPEG-I Audio 서비스 시나리오 및 고려사항	14
	6.3 MPEG-I Audio Architecture	14
	6.4 MPEG-I Audio Requirements	16
	6.5 MPEG-I Audio Encoder Input Format	18
	6.6 MPEG-I Audio Evaluation Platform	22
7	6자유도 오디오 기술 동향	24
	7.1 Project Triton and Acoustics	25
	7.2 Resonance Audio	29
8	결론 및 시사점	34
	8.1 국내 표준화 관점	34
	8.2 국내 서비스 및 산업 관점	34
부록	I -1 지식재산권 요약서 정보	36
	I -2 시험인증 관련 사항	37
	I -3 본 기술보고서의 연계(Family) 기술보고서	38
	I -4 참고 문헌	39
	I -5 영문기술보고서 해설서	40
	I -6 기술보고서의 이력	41

6자유도 오디오 표준화 및 기술 동향 (A Trend of 6 DoF Audio Technology and Standardization)

1 적용 범위

이 기술보고서는 6자유도 오디오 기술 및 표준화 동향을 종합하여 요약/정리하고 주요 이슈 및 고려 사항을 제시함으로써, 관련 산업 종사자 및 표준화 담당자가 전반적인 사항을 쉽고 빠르게 파악할 수 있도록 한다.

이 기술보고서의 8장의 결론 및 시사점은 미래방송미디어표준포럼 미래미디어분과에 참여하고 있는 내의 해당 분야 전문가 의견을 참고용으로 기재한 것으로서, 미래방송미디어표준포럼이나 미래미디어분과의 공식 의견이 아님을 밝혀 둔다.

2 인용 표준

해당 사항 없음

3 용어 정의

360VR 영상(360도 영상, VR360 영상)

하나의 영상 안에 360도 전방향의 화소가 모두 포함된 일종의 파노라마 영상이다. 360도 카메라(omnidirectional camera)를 사용하여 촬영하며, 재생 시 원하는 영역을 실시간으로 선택하면서 볼 수 있다. HMD(Head Mounted Display)를 착용하고 감상하면, 촬영 현장에 있는 듯한 몰입감을 느낄 수 있어 몰입형 영상이라고도 부른다. 이와 비슷하게, 180도 한 쪽 방향의 화소 모두를 포함하는 파노라마 영상을 180VR 영상이라고 함.

3 DoF

3차원 좌표계에서 원점을 중심으로 사용자 머리 회전(rotation)인 Yaw, Pitch, Roll이 허용된 움직임을 3 DoF 혹은 3자유도라 함.

3 DoF+

3차원 좌표계에서 원점을 중심으로 사용자 머리 회전이 허용된 3 DoF에 더해 제한적인 범위 내에서 사용자의 상하/좌우/앞뒤 움직임 허용된 것을 3 DoF+라 함.

6 DoF

3차원 좌표계에서 원점을 중심으로 사용자 머리 회전(rotation)인 Yaw, Pitch, Roll이 허용된 것을 3 DoF(3자유도)라 하며, 3자유도에 더하여 사용자의 상하/좌우/앞뒤 움직임까지 광범위한 범위에서 허용된 것을 6 DoF(6자유도)라 함.

HMD(Head mounted Display)

사용자의 머리에 장착한 디스플레이 장치를 통해 영상을 표시하는 장치로서, 자이로 센서를 함께 탑재하여 비행기 조종 시뮬레이션 등에 주로 사용되었으며, 360VR 영상의 시청 및 VR 게임을 위한 필수 장비임.

4 약어

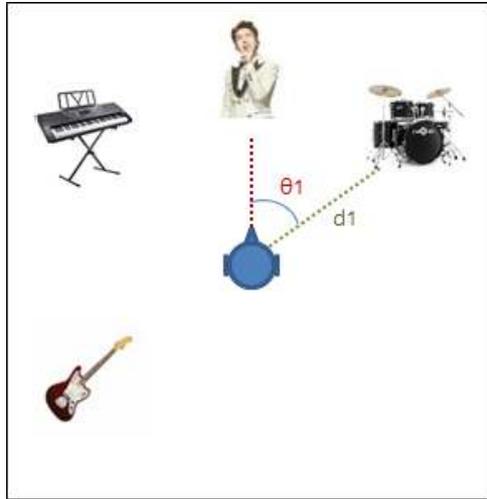
3GPP	3rd Generation Partnership Project
ADM	Audio Definition Model
AEP	Audio Evaluation platform
AR	Augmented Reality
CATT	Computer Aided Theater Technique
CfP	Call for Proposal
DDR	Diffuse to Reverberant Ratio
DoF	Degrees of Freedom
DVB	Digital Video Broadcasting
FDTD	Finite Difference Time Domain
FOA	First Order Ambisonics
HMD	Head Mounted Display
HOA	Higher Order Ambisonics
HRTF	Head Related Transfer Function
ILD	Interaural Level Difference
ITD	Interaural Time Difference
ITU	International Telecommunication Union
MPEG	Moving Picture Experts Group
MR	Mixed Reality
OMAF	Omnidirectional Media Format
OSC	Open Sound Control
RAVEN	Room Acoustics for Virtual Environments
RT60	Reverberation Time 60
SDK	Software Development Kit
TR	Technical Report
TTA	Telecommunications Technology Association
UHDTV	Ultra High Definition Television
VR	Virtual Reality
WD	Working Draft
WG	Working Group
Wwise	Wave Works Interactive Sound Engine

5 6자유도 오디오 소개

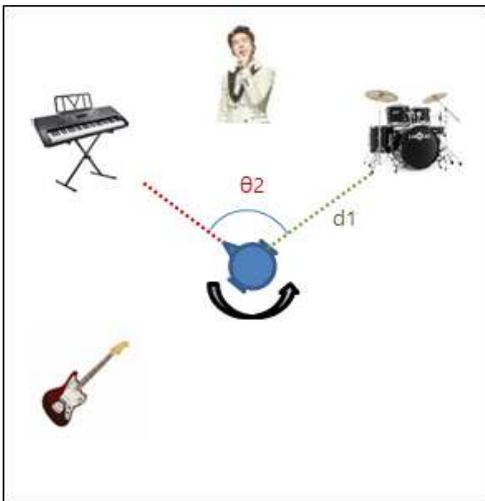
5.1 3자유도 오디오와 6자유도 오디오

최근 6자유도 오디오 기술에 대한 표준화와 기술 개발이 활발하게 진행되고 있다. 본 절에서는 6자유도 오디오가 어떠한 특징을 가지는지에 대해 3자유도 오디오와 비교하여 살펴본다.

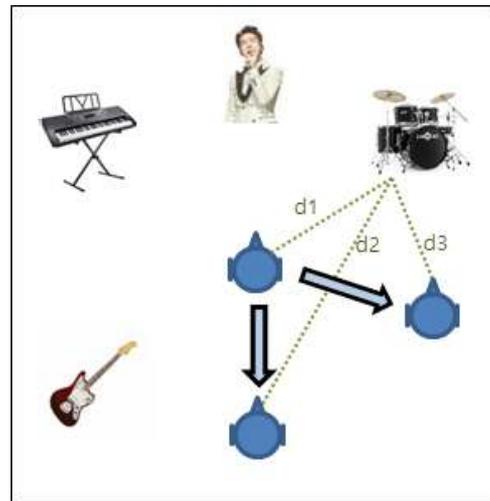
일반적으로 3자유도 오디오는 고정된 위치에서 방향을 변경할 수 있는 것을 나타내며, 6자유도 오디오는 3자유도 오디오와 같이 방향을 변경할 수도 있으며, 상하, 좌우로도 이동이 가능하여, 방향과 함께 거리가 변경될 수 있는 것을 나타낸다. 아래의 그림 5-1은 3자유도 오디오와 6자유도 오디오의 개념 및 차이를 그림을 통하여 나타낸 것이다.



(a) 가상 3차원 공간의 예



(b) 3자유도 오디오의 예



(c) 6자유도 오디오의 예

(그림 5-1) 3자유도 오디오와 6자유도 오디오의 예

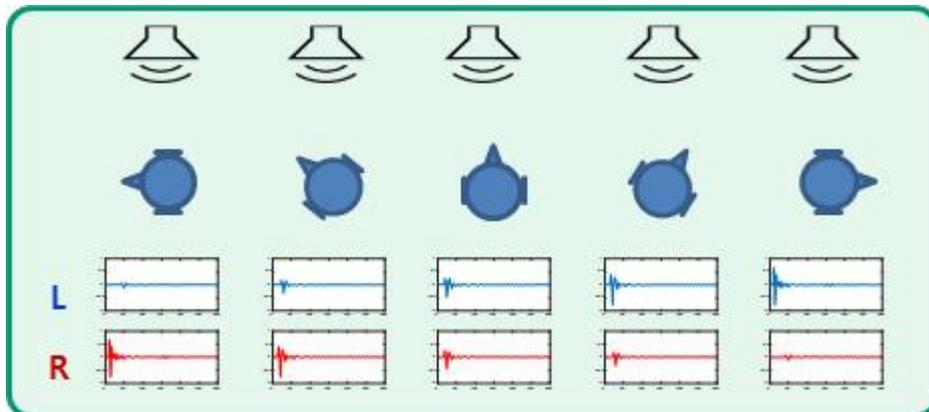
그림 5-1의 (a)는 다양한 음원이 존재하는 가상 3차원 공간의 예를 나타낸 것이다. 청취자를 중심으로 가수, 드럼, 신디사이저, 기타 등 4개의 음원이 존재한다. 청취자를 기준으로 정면에서 가수의 목소리가 들리고, 왼쪽에는 신디사이저 소리가 들리며, 오른쪽에서는 드럼 소리가 들리고, 왼쪽 뒤쪽에서 기타 소리가 들리게 된다.

그림 5-1의 (b)는 3자유도 오디오의 예를 나타내는데, 청취자는 고개를 좌우 상하로 변경할 수 있으며, 이에 따라 개별 음원이 청취자에게 들리는 각도가 변경된다. 그림 (b)와 같이 청취자가 고개를 왼쪽으로 돌린 경우, 왼쪽에서 들리던 신디사이저가 청취자의 정면에서 들리게 되며, 정면에서 들리던 가수의 목소리는 오른쪽에서 들리게 된다.

그림 5-1의 (c)는 6자유도 오디오의 예를 나타낸다. 청취자는 3자유도 오디오에서와 같이 고개를 좌우 상하로 변경하여 각도를 변경할 수도 있으며, 좌우 상하로 이동할 수도 있다. 청취자의 이동에 따라, 음원과 청취자 간의 방향과 함께 거리도 변경이 된다. 그림 5-1의 (c)에서 청취자가 뒤로 이동한 경우, 가수와 청취자 간의 거리가 멀어져서 가수의 목소리는 더 작게 들리게 될 것이며, 마찬가지로 신디사이저와 드럼의 소리도 작게 들리게 될 것이다. 기타와 청취자의 거리는 가까워져서, 기타의 소리는 좀 더 크게 들리게 될 것이다. 이와 같이 3자유도 오디오에서는 오디오 객체와 청취자 간의 방향만 변경이 되는 것이 일반적이며, 6자유도 오디오에서는 오디오 객체와 청취자 간의 방향과 함께 거리도 변경이 되는 차이가 있다.

6자유도 오디오 재현 환경에서 음원과 청취자 간의 방향이 변경되거나, 거리가 변경되면, 다양한 음향학적 변화가 발생하는데, 이들 중 주요한 몇 가지 특징에 대해서 아래에서 간단하게 살펴본다.

먼저, 음원과 청취자 간의 방향이 변경되면, 청취자의 양 귀에 입력되는 신호의 크기 비율이 변경되게 되는데, 이는 음원과 사람의 각도에 따른 음원과 사람 양 귀 간의 전달함수인 머리전달 함수를 통해 쉽게 이해할 수 있다. 아래의 그림 5-2는 음원과 사람의 머리 각도에 따른 머리전달 함수를 나타낸 것인데, 각도에 따라 양 귀에 입력되는 신호의 차이를 볼 수 있다.



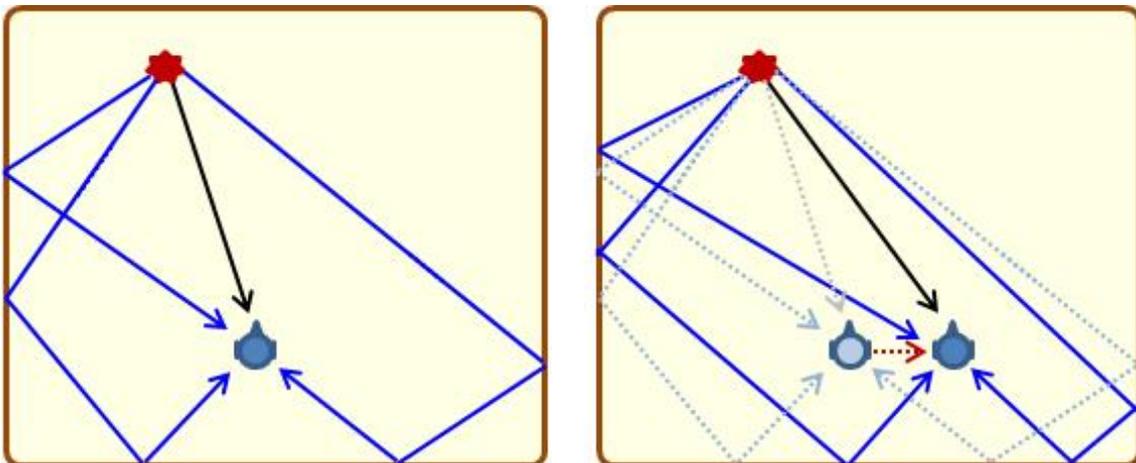
(그림 5-2) 음원과 청취자 간의 방향에 따른 양쪽 귀 입력 신호의 크기 차이

그림 5-2의 좌측 그림과 같이 청취자가 음원을 기준으로 왼쪽으로 고개를 돌리면, 오른쪽 귀와 음원과의 거리가 왼쪽 귀와 음원과의 거리보다 가까워지고, 청취자의 머리에 의해 왼쪽 귀로 입력되는 소리가 방해를 받아서, 오른쪽 귀로 입력되는 소리가 왼쪽 귀로 입력되는 소리에 비해 훨씬 크게 된다. 그림 5-2의 가운데 그림과 같이, 청취자가 고개를 돌리지 않고 음원을 정면으로 바라보는 경우, 음원과 오른쪽 귀 간의 전달 경로와 음원과 왼쪽 귀 간의 전달 경로는 대칭으로 서로 비슷하므로, 오른쪽 귀와 왼쪽 귀에 입력되는 신호의 크기는 유사하다. 그림 5-2의 오른쪽 그림과 같이, 청취자가 음원을 기준으로 오른쪽으로 고개를 돌리면, 앞서 보았던 것과 같은 효과로 인해 왼쪽 귀에 입력되는 신호의 크기가 오른쪽으로 입력되는 신호의 크기에 비해 훨씬 크게 된다. 이와 같이 음원과 청취자 간의 방향이 변경되면, 오른쪽 귀와 왼쪽 귀에 입력되는 신호의 크기 비율이 달라지게 된다.

또한, 음원과 청취자 간의 방향이 변경되면, 청취자의 양 귀에 입력되는 신호의 도착 시간도 변경이 된다. 이것 또한 머리전달 함수를 통해 쉽게 이해할 수 있다. 그림 5-2의 좌측 그림과 같이 청취자가 음원을 기준으로 왼쪽으로 고개를 돌리면, 오른쪽 귀와 음원과의 거리가 왼쪽 귀와 음원과의 거리보다 가까워지기 때문에, 오른쪽 귀에 소리가 먼저 도착하게 된다. 반대로, 그림 5-2의 오른쪽 그림과 같이, 청취자가 음원을 기준으로 오른쪽으로 고개를 돌리면, 왼쪽 귀와 음원과의 거리가 오른쪽 귀와 음원과의 거리보다 가까워지기 때문에, 소리가 오른쪽 귀 보다 왼쪽 귀에 먼저 도착하게 된다.

음원과 청취자 간의 방향이 변경되었을 때, 앞서 언급한 양 귀 간의 신호 크기 차이, 양 귀 간의 신호 도착 시간 차이 이외의 다른 변화도 있으나, 본 보고서에서는 지면의 문제로 인해 생략한다.

다음으로, 음원과 청취자 간의 거리가 달라지면, 거리의 변경에 따른 신호의 크기가 변경되고, 반사, 회절 등과 같은 음향의 전달 특성이 변경된다. 아래의 그림 5-3은 사용자의 이동에 따른 청취자에게 전달되는 소리의 경로 변화를 간단한 예로 나타낸 것이다.



(그림 5-3) 사용자의 이동에 따른 반사 경로 변경의 예

그림 5-3의 왼쪽 그림은 청취자가 임의의 위치에 서 있을 때의 소리의 전달 경로를 나타내며, 오른쪽 그림은 왼쪽 그림에서 청취자가 오른쪽으로 이동했을 때의 소리의 전달 경로를 나타낸다. 그림 5-3의 왼쪽 그림에서 검은색으로 표시된 경로는 음원의 소리가 청취자에게 직접 전달되는 경로인 직접음의 전달 경로를 나타내며, 파란색으로 표시된 경로는 음원이 벽면에 반사되어 청취자에게 전달되는 경로인 직접 반사음의 예를 나타낸다.

그림 5-3의 오른쪽 그림을 보면, 청취자가 붉은 점선을 따라 오른쪽으로 이동했을 때의 직접음과 직접 반사음을 나타낸 것이다. 검은색 선으로 나타낸 직접음의 전달 경로가 청취자가 이동하기 전보다 길어졌으며, 이에 따라 직접음의 크기가 적어지게 된다. 파란색으로 나타낸 직접 반사음의 전달 경로도 청취자의 위치가 이동함에 따라 전달되는 방향과 거리가 달라진 것을 볼 수 있는데, 이에 따라 직접 반사음의 좌우 크기 비율과 크기가 달라질 것이다. 이때 거리가 변경되는 것과 함께, 직접음 및 직접 반사음이 청취자에게 입사되는 각도도 변경이 되는 것을 볼 수 있다.

본 절에서 살펴본 것과 같이 3자유도 오디오에서는 오디오 객체와 청취자 간의 방향만 변경이 되는 것이 일반적이지만, 6자유도 오디오에서는 오디오 객체와 청취자 간의 방향과 함께 거리도 변경이 되는 차이가 있으며, 이에 따라 6자유도 오디오에서는 공간 음향을 재현하기 위하여 고려해야 하는 요소가 3자유도 오디오보다 많아진다고 볼 수 있다.

5.2. 오디오 포맷의 종류와 활용

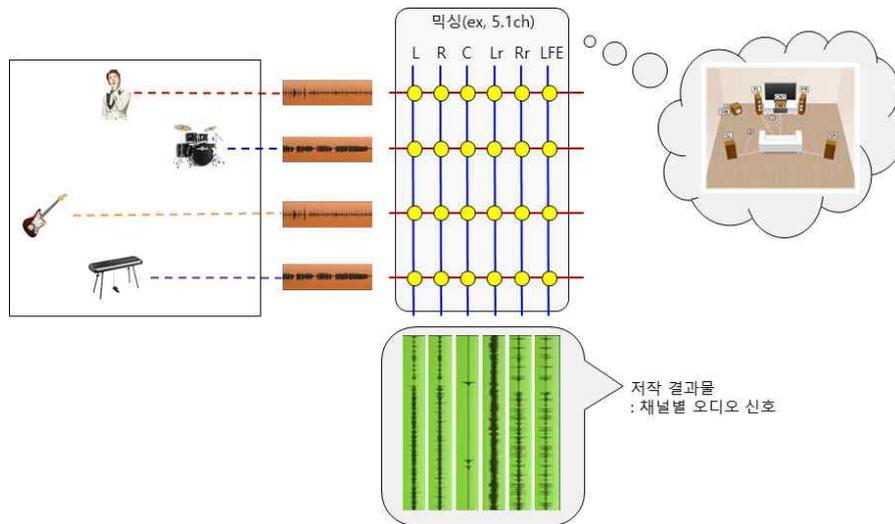
앞서 살펴본 것과 같이 6자유도 오디오에서는 청취자와 오디오 음원과의 방향과 거리에 대한 변화가 있을 수 있으며, 이에 따라 청취자에게 들리는 소리가 변경되게 된다. VR, 게임 등과 같은 콘텐츠 재생 환경에서는 이러한 변화에 따른 소리를 적절하게 생성하여 청취자에게 제공하는 것이 필요한데, 이를 위해서 사용할 수 있는 대표적인 오디오 포맷으로는 채널 기반 오디오, 객체 기반 오디오, 장면 기반 오디오가 있다.

객체 기반 오디오는 오디오 재현의 자유도가 아주 높은 오디오 포맷으로 6자유도 오디오에서 잘 활용될 수 있으며, 채널 기반 오디오와 장면 기반 오디오의 경우 오디오 재현 입장에서 자유도가 높지 않아서, 6자유도 오디오에서는 제한적으로 사용될 수 있는데, 조금 더 상세한 내용을 아래에서 기술한다.

5.2.1 채널 기반 오디오

채널 기반 오디오는 미리 정해진 오디오 재현 환경(스피커의 개수와 위치)에서 재생되도록 만들어진 오디오라고 생각할 수 있다. 최근까지 사용된 방송, 영화, 멀티미디어, 게임의 대부분 채널 기반 오디오를 사용하였는데, 스테레오, 5.1채널, 7.1채널, 10.2채널, 11.1채널 오디오 등이 대표적인 채널 기반 오디오의 예이다.

채널 기반 오디오에 대한 이해를 위해, 채널 기반 오디오를 생성하는 방법에 대하여 간단하게 그림으로 나타낸 것이 아래의 그림 5-4이다.



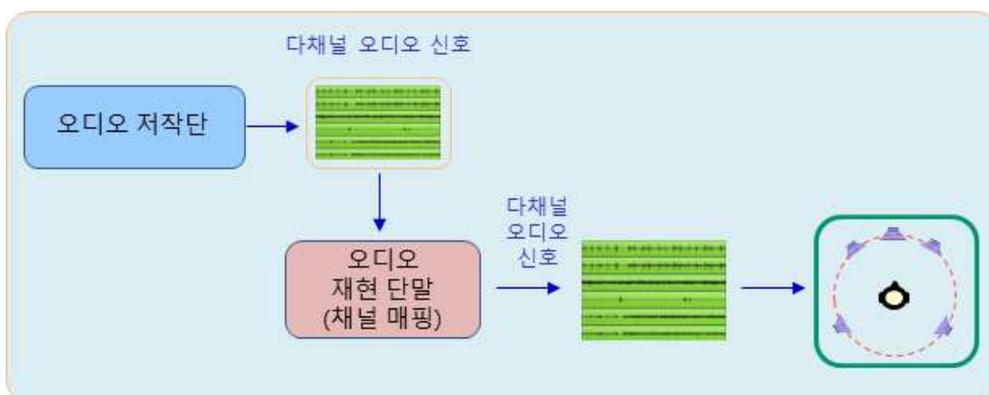
(그림 5-4) 채널 기반 오디오 저작 방법의 예

채널 기반 오디오를 생성하기 위해서는 앞서 기술한 것과 같이 오디오 재현 환경을 먼저 설정하는 것이 필요하다. 예를 들어 스테레오 환경에서 재현할 것인지, 5.1채널 환경에서 재현할 것인지 결정을 해야 하는데, 그림 5-4에서는 5.1채널 오디오를 재현 환경으로 설정한 것이다.

재현 환경이 결정되면, 주어진 오디오 음원을 활용하여 해당 재현 환경에 맞는 오디오를 생성하는 과정을 거치는데, 오디오 신호의 레벨 조정, 패닝(panning), 잔향 등 다양한 오디오 효과가 활용될 수 있다. 그림 5-4에서는 가수의 목소리, 드럼 소리, 기타 소리, 신디사이저 소리 등 4개의 음원으로 채널 오디오를 생성한 예를 나타낸다.

채널 기반 오디오 저작의 최종 결과물은 각 채널별 오디오 신호가 된다. 그림 5-4에서는 5.1채널 오디오 신호가 최종 결과물이 된다.

이러한 방식으로 저작된 채널 기반 오디오는 재생 단말에서는 단순히 채널 신호를 해당 스피커로 매핑하여 재생하는 방법으로 재현이 된다. 아래의 그림 5-5는 채널 기반 오디오가 재현되는 과정을 나타낸 예이다.



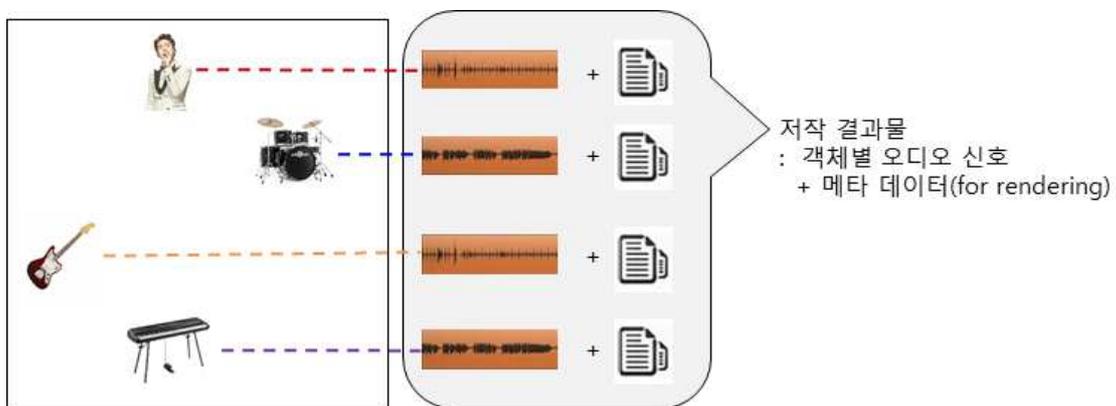
(그림 5-5) 채널 기반 오디오 신호의 재현

채널 기반 오디오의 장점은 단말이 매우 단순하다는 것으로, 앞서 기술한 것과 같이 오디오 신호를 단순하게 채널 매핑하여 재생하는 과정만 수행하기 때문이다. 만약 입력된 오디오 신호의 포맷과 오디오 재현 환경이 다른 경우(예를 들어 입력 신호는 5.1채널인데, 오디오 재현 환경이 스테레오 또는 7.1채널인 경우)에는 채널 다운믹스 또는 채널 업믹스 과정을 통해, 오디오 재현 환경에 맞는 채널 오디오 신호를 생성하는 과정을 수행한 후 재생을 하기도 한다. 채널 기반 오디오의 경우 사용자의 위치가 고정되었다고 가정하고 채널 신호를 저작하고, 사용자의 위치와 오디오 객체의 위치가 저작 단계에서 설정되고, 이에 따라 렌더링 된 결과를 채널별 오디오 신호로 저장하기 때문에, 단말에서 사용자와 오디오 객체의 위치 변경이 불가능하여 3자유도 오디오에서는 잘 활용이 될 수 있으나, 6자유도 오디오에서는 사용이 제한적이라 할 수 있다.

5.2.2 객체 기반 오디오

객체 기반 오디오는 객체 오디오 음원 신호와 객체 오디오 신호의 렌더링을 위한 정보를 포함하고 있는 메타 데이터로 구성된 오디오 포맷이다. 객체 오디오 신호의 렌더링을 위한 정보로는 오디오 객체의 3차원 공간에서의 위치 정보, 신호의 크기, 오디오 객체 신호의 방향성 등과 같은 정보가 될 수 있다.

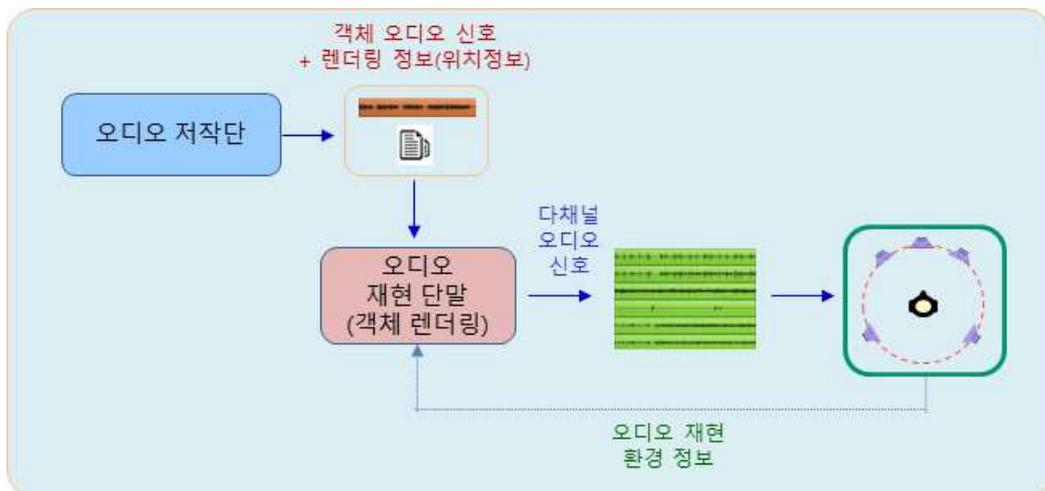
객체 기반 오디오의 경우 1990년대에 이슈가 되었는데, MPEG-4 system 규격에서 객체 기반 미디어 표준이 제정되면서, 객체 기반 오디오에 대한 개념도 표준화가 되었다. 이후 2010년대에 Dolby에서 객체 기반으로 영화 음향을 저작하고 재현하는 Dolby Atmos를 소개하면서 다시 이슈가 되었으며, 2013년 제정된 MPEG-H 3D audio 표준에도 객체 기반 오디오가 포함되었고, AC-4 표준에서도 객체 기반 오디오가 포함이 되었다. 참고로, MPEG-H 3D audio 표준에서는 객체 기반 오디오의 기본적인 렌더링 방식도 표준에 포함하였다. 객체 기반 오디오에 대한 이해를 위해, 객체 기반 오디오를 생성하는 방법에 대하여 간단하게 그림으로 나타낸 것이 아래의 그림 5-6이다.



(그림 5-6) 객체 기반 오디오 저작 방법의 예

객체 기반 오디오에서도 채널 기반 오디오와 마찬가지로 각 객체의 오디오 신호가 주어진다. 하지만, 객체 기반 오디오에서는 채널 기반 오디오와는 달리 오디오 재현 환경이 제공되지 않으며, 오디오 재현 환경을 고려하지 않고 저작을 한다. 객체 오디오 신호의 저작 단계에서는 객체 오디오 신호의 렌더링을 위한 정보가 설정되는데, 오디오 객체의 3차원 공간에서의 위치 정보, 신호의 크기, 오디오 객체 신호의 방향성, 객체가 존재하는 3차원 공간의 잔향 특성 정보 등과 같은 정보가 이러한 렌더링을 위한 정보가 될 수 있다. 객체 기반 오디오의 저작 결과물은 객체별 오디오 신호와 렌더링을 위한 정보가 포함된 메타 데이터가 된다. 메타 데이터를 저장하는 규격으로는 다양한 규격이 존재하고 있는데, ADM(Audio Definition Model), Dolby AC-4, MPEG-H 3D audio, DTD-X 등이 있다. 이러한 방식으로 저작된 객체 기반 오디오는 재생 단말에서 렌더링 되는데, 렌더링 이전에 실제 오디오가 재현되는 환경 정보를 받아서, 주어진 오디오 재현 환경에 가장 적합한 다채널 오디오 신호를 생성하게 된다. 아래의 그림 5-7은 객체 기반 오디오가 재현되는 과정을 나타낸 예이다.

객체 기반 오디오의 렌더링 단말은 객체 오디오 신호와 메타 데이터를 입력으로 받으며, 추가적으로 오디오 재현 환경에 대한 정보를 입력받는다. 입력된 정보를 기반으로 주어진 오디오 재현 환경에 가장 적합한 오디오 신호를 생성하는데, 최종적으로 생성된 신호는 주어진 오디오 재현 환경에서 각 채널로 매핑되는 다채널 오디오 신호가 된다. 객체 기반 오디오의 주요한 장점으로는 오디오 재현 환경에 독립적으로 저작할 수 있으며, 재현 시에는 주어진 오디오 재현 환경에 최적의 오디오를 렌더링 할 수 있다는 것을 들 수 있다. 또한, 단말에서 객체의 위치를 변경하거나, 소리의 크기를 변경하는 등과 같은 객체에 대한 제어가 가능하다는 점도 객체 기반 오디오의 중요한 장점 중 하나라고 할 수 있다. 하지만, 채널 기반 오디오 단말이 단순 채널 매핑만 하는 단순한 작업을 수행하는 것에 비해, 단말에서 객체 오디오 신호의 렌더링 정보에 따라 오디오 렌더링을 수행해야 하므로, 단말의 연산량 많아지는데, 이는 상대적인 단점이라 할 수 있다.

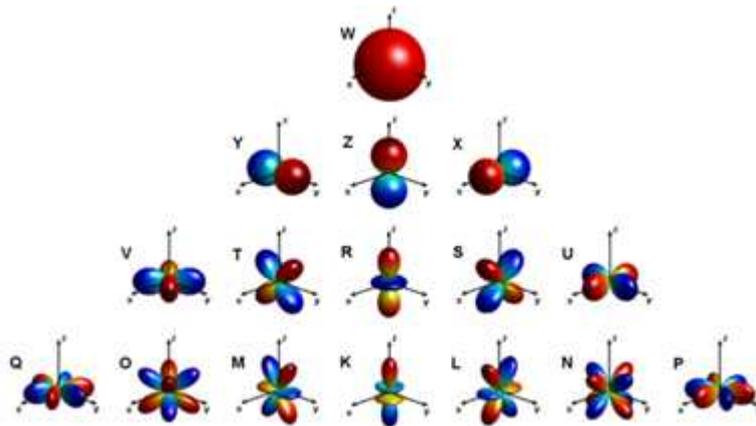


(그림 5-7) 객체 기반 오디오 신호의 재현

앞서 살펴보았던 것과 같이, 객체 기반 오디오의 경우, 오디오 객체의 위치는 저작 단계에서 미리 설정되어 있으나, 단말에서 변경 가능하며, 저작 단계에서 사용자의 위치는 정해져 있지 않으나, 단말에서 설정 및 변경 가능하며, 사용자의 위치/방향에 따라 단말에서 적절하게 오디오 렌더링 가능하다. 따라서, 사용자의 위치가 가변적이고, 오디오 객체의 위치도 변경될 수 있는 6자유도 오디오에서 잘 활용이 될 수 있다. 다만, 사용자가 지속적으로 움직일 수 있는 6자유도 콘텐츠의 특성을 감안하였을 때, 오디오 렌더링을 실시간으로 수행하기 위해 단말의 높은 연산 성능이 요구될 수도 있다.

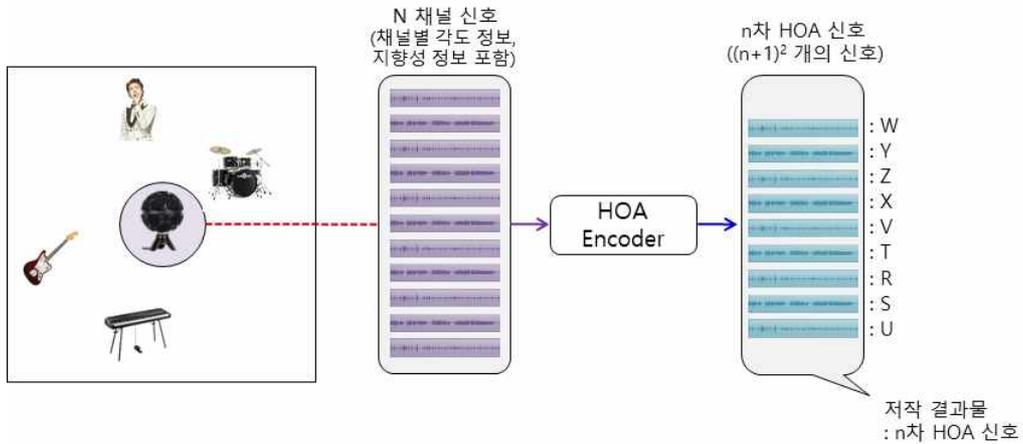
5.2.3 장면 기반 오디오

장면 기반 오디오는 한 지점을 기준으로 다양한 방향에서 오는 오디오 신호를 저장한 오디오 포맷으로, 장면 기반 오디오의 차수에 따라서 포함하는 방향의 개수가 다르게 설정된다. 아래의 그림 5-8은 장면 기반 오디오의 대표적인 예인 고차원 앰비소닉의 차수에 따른 직교 구면 고조파의 예를 나타낸 것으로, 장면 기반 오디오의 각 채널이 가지는 방향성을 나타낸다.



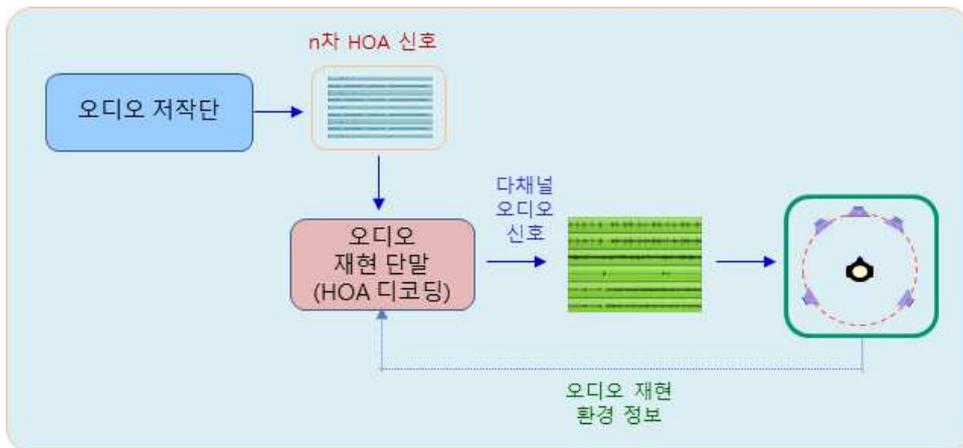
(그림 5-8) 고차원 앰비소닉의 차수에 따른 직교 구면 고조파의 예

장면 기반 오디오에 대한 이해를 위해, 장면 기반 오디오를 생성하는 방법에 대하여 간단하게 그림으로 나타낸 것이 그림 5-9이다. 장면 기반 오디오에서는 일반적으로 객체의 음원이 별도로 존재하는 것이 아니라, 동시에 다양한 방향에서 오는 오디오 신호를 녹음할 수 있는 다채널 마이크로폰을 활용하여 다채널 오디오를 녹음하는 것이 일반적이다. FOA(First Order Ambisonics) 마이크로폰 이나, HOA(Higher Order Ambisonics) 마이크로폰이 활용될 수 있는데, 이를 통하여 N 채널의 오디오 신호가 녹음된다. 녹음된 N 채널 오디오 신호는 HOA(또는 FOA) 인코더를 통하여 설정된 차수에 맞는 직교 구면 고조파 신호로 변환이 되며, 변환된 HOA 신호가 장면 기반 오디오의 저작 결과물이 된다.



(그림 5-9) 장면 기반 오디오 저작의 예

이러한 방식으로 저작된 장면 기반 오디오는 재생 단말에서 HOA 디코딩 되는데, 렌더링 이전에 오디오 재현 환경 정보를 받아서, 해당 오디오 재현 환경에 가장 적합한 다채널 오디오 신호를 생성하게 된다. 아래의 그림 5-10은 장면 기반 오디오가 재현되는 과정을 나타낸 예이다.



(그림 5-10) 장면 기반 오디오 신호의 재현

장면 기반 오디오의 렌더링 단말은 HOA 신호와 함께 오디오 재현 환경에 대한 정보를 입력받는다. HOA 재현 단말에서는 입력된 HOA 신호를 주어진 오디오 재현 환경에 맞게 디코딩하여 채널 오디오 신호를 생성하는데, 최종적으로 생성된 신호는 주어진 오디오 재현 환경에서 각 채널로 매핑되는 다채널 오디오 신호가 된다.

장면 기반 오디오의 경우 일반적으로 현장에서 녹음한 신호이기 때문에 녹음 현장 공간의 음향학적 정보가 잘 녹음이 되어, 높은 현장감을 제공할 수 있으며, 객체 기반 오디오와 유사하게 오디오 재현 환경에 맞는 디코딩이 가능하다는 장점이 있다. 장면 기반 오디오의 경우 녹음되는 위치를 기준으로 오디오가 재현되기 때문에, 청취자의 3차원 공간에서의 위치가 장면 기반 오디오의 녹음 위치와 일치하게 되므로, 사용자가 자유롭게 음

적일 수 있는 6자유도 오디오 환경에서의 오디오 재현에는 제한적일 수 있다. 최근에는 이러한 단점을 보완하기 위하여 공간 내에 여러 대의 HOA 마이크로폰을 사용하여 녹음하고, HOA 마이크로폰과 HOA 마이크로폰 사이의 임의의 위치에서의 오디오 신호는 가까운 위치에서 실제 마이크로폰을 사용하여 녹음된 신호들을 활용하여 보간하는 방법으로 생성하여, 6자유도 오디오를 지원하는 방법에 대한 연구가 이루어지고 있다.

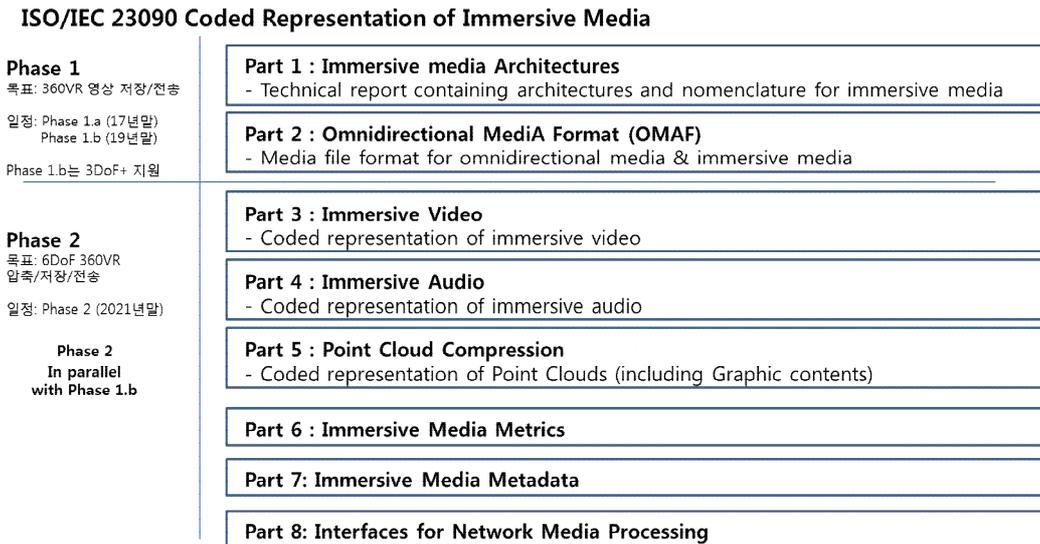
6 6자유도 오디오 표준화 동향

6자유도 오디오의 표준화와 관련하여 MPEG, DVB, 3GPP 등의 다양한 표준화 기관에서 관심을 가지고 표준화에 대한 조사를 수행하였는데, 표준화가 활발하게 이루어지고 있는 기관은 MPEG이 대표적이다. 본 장에서는 MPEG에서 이루어지고 있는 6자유도 오디오 표준화 현황에 대하여 살펴본다.

6.1 6자유도 미디어 관련 MPEG 표준화 동향

MPEG에서는 MPEG-I(ISO/IEC 23090 Coded Representation of Immersive Media) 표준화를 진행하고 있다. MPEG-I는 몰입형 미디어에 대한 MPEG 차세대 프로젝트를 말한다. 비디오, 오디오에서 시스템 요소에 이르기까지 전체 콘텐츠 에코시스템을 대상으로 하는 표준 기술로 구성된 프로젝트이며, 360 VR(Virtual Reality) 비디오와 MR(Mixture Reality) 및 6자유도(six degree of freedom, 6 DoF) 관련 콘텐츠를 포함한 몰입형 미디어 전체를 고려 대상으로 한다.

MPEG-I는 총 2단계의 Phase와 8개의 Part로 정의되어 있고, 각 Part 마다 서로 다른 표준으로 구성되어 있다. Phase 1에서는 360도 VR 영상 시청 시나리오를 목표로 진행하며 Phase 2는 6축 자유도를 가질 수 있는 6자유도 VR 영상도 포함한다.



(그림 6-1) MPEG-I 단계별 Part 구성

오디오의 경우 Part 4에 해당하며, 6축 자유도를 가질 수 있는 환경에서 활용될 수 있는 오디오 콘텐츠의 표현 및 재현 기술에 대한 표준화를 추진한다. 2021년말 표준화 완료로 목표로 하였으나, 여러 가지 환경적 문제로 인해 2022년에 표준화가 완료될 것으로 예상된다.

MPEG의 오디오 WG(Working Group)에서는 채널 오디오와 함께 장면 기반 오디오 및 객체 기반 오디오를 저장, 압축, 전송, 재현하는 표준인 MPEG-H 3D Audio 표준화를 2016년도까지 진행하였다. MPEG-H 3D Audio의 경우 장면 기반 오디오와 객체 기반 오디오를 지원하기 때문에, 3자유도 환경에서의 오디오 신호의 저장 및 재현을 대부분 지원한다고 볼 수 있다.

이러한 이유로, MPEG-I Immersive Audio에서는 6자유도 오디오를 제공하는 기술에 대한 표준화를 목적으로 진행이 되고 있다. MPEG-I Immersive Audio 표준에서는 채널, 객체 그리고 앰비소닉스(Ambisonics) 신호를 모두 활용하고 있는데, 이들 신호의 압축 및 복원을 위해서는 기존 기술인 MPEG-H 3D Audio 기술을 사용하기로 하였다. 따라서, MPEG-I Audio 표준화에서는 오디오 신호의 압축/복원과 관련된 부분은 제외하고, 6자유도 오디오 서비스를 위한 메타데이터(metadata) 및 렌더링(rendering) 기술에 대한 표준화를 진행하기로 하였다. 아래에서 MPEG-I Audio 표준화 내용 및 현황에 대해 상세하게 설명한다.

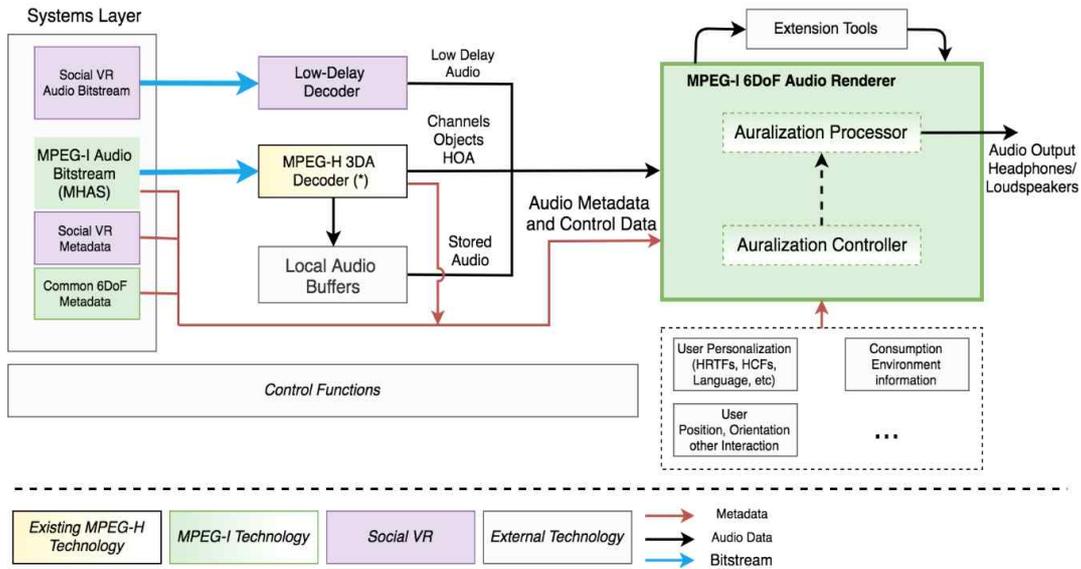
6.2 MPEG-I Audio 서비스 시나리오 및 고려사항

MPEG 118차 회의에서는 VR 3D audio의 서비스 시나리오와 오디오의 렌더링 측면에서 고려하여야 할 지연 시간 등에 대한 논의가 이루어졌는데, 아래와 같은 내용이 포함되어 있다.

- 시나리오
 - ✓ 360 Movie – 3 DoF
 - ✓ 360 Movie with head movement – 3 DoF+
 - ✓ Virtual Reality – 3 DoF(예: 콘서트 홀에서 좌석 선택)
 - ✓ Windowed 6 DoF
 - ✓ Virtual Reality – 6 DoF(예: 게임)
 - ✓ Augmented Reality
- 렌더링 지연 시간
 - ✓ 현재까지의 연구 결과와 Head tracking을 위한 시간을 5ms로 가정할 경우,
 - ✓ AR Application : 30 ~ 35ms 이내
 - ✓ VR Application : 50 ~ 55ms 이내
 - ✓ 구현을 위한 여유분을 고려하면, 알고리즘 지연은 더 낮아야 함

6.3 MPEG-I Audio Architecture

MPEG Audio WG에서는, 6자유도 오디오 서비스를 위한 시스템의 구조로 MPEG-I Audio 참조 아키텍처를 정의하고 있는데, 아래 그림 6-2가 이를 나타낸다.



(그림 6-2) MPEG-I Audio architecture

MPEG-I Audio architecture에서는 기존의 MPEG 기술과 MPEG-I Audio 표준화를 통하여 개발하는 기술을 구분하여 나타내고 있으며, social VR을 위한 기술 및 표준 외에 활용되는 독립적인 기술을 구분하여 표기하고 있다. 그림 6-2에서 볼 수 있듯이, MPEG-H 3DA Decoder를 그대로 활용하기로 하였는데, 이는 MPEG-I Audio 표준의 범위에 오디오 신호의 압축/복원에 대한 것을 포함하지 않는다는 것을 의미한다. MPEG-I Audio의 표준화 범위는 위의 그림에서 녹색으로 표기하고 있는데, MPEG-I 6 DoF Audio Renderer, MPEG-I Audio Bitstream, Common 6 DoF Metadata가 포함된다.

MPEG-I audio Bitstream과 Common 6 DoF Metadata에는 6 DoF 오디오 렌더링을 위한 메타 데이터가 포함이 된다. 오디오 신호에 대한 정보, 음향학적 3차원 공간에 대한 정보, interactive 정보 등이 포함이 될 것이며, 어떠한 정보가 메타 데이터로 정의되고 비트스트림에 포함되어야 하는지에 대한 부분이 표준화의 범위에 포함이 된다.

MPEG-I 6 DoF Audio Renderer는 복호화된 오디오 신호와 오디오 메타 데이터 및 제어 정보를 입력으로 받아서 6자유도 오디오 렌더링을 수행하는 부분인데, 바이노럴 렌더링을 위한 IR 생성, 도플러 효과 등과 같은 다양한 오디오 렌더링 기술이 포함될 것으로 예상된다. MPEG-I Audio architecture에서는 표준의 범위에 포함되어 있지 않으나, MPEG-I Audio의 렌더링을 위해 활용되는 기술을 소개하고 있는데, 머리전달 함수가 대표적인 예가 될 수 있다. 머리전달 함수는 MPEG-I Audio의 표준 범위에 포함되어 있지 않으나, MPEG-I 6 DoF Audio Renderer에서 활용하고 있으며, 성능 평가를 위하여 하나의 공통된 머리전달 함수를 사용하도록 하고 있다.

6.4 MPEG-I Audio Requirements

MPEG Audio WG에서는 MPEG-I Audio Architecture와 함께 요구사항을 제시하였는데, 6개의 카테고리에 대한 27개의 요구사항을 제시하였다. 제시된 6개의 카테고리 항목은 General, Rendering, Interface and Extensibility, Presentation mode, Social VR, Interoperability between 3 DoF and 6 DoF platforms 관련 요구사항으로 각 항목은 아래와 같다.

- General : 오디오의 음질 및 현상감과 관련된 요구사항

구분	순번	요구사항
일반적 요구사항 (General)	1	사용자가 음향 환경 및 사용자의 움직임과 일치하는 경험을 인식할 수 있도록 6자유도를 지원해야 한다. (예: 낮고 인지할 수 없는 모션 to 사운드 지연 시간).
	2	미디어 및 메타데이터의 효율적인 표현 및 압축을 지원해야 한다.
	3	음향 요소 및 음향 환경과 일치하는 오디오 장면의 렌더링을 지원하여 장면과 일치하는 사용자 경험을 제공해야 합니다.
	4	동적인 오디오 요소를 지원해야 한다. (예: 사용자 위치, 방향, 이동 방향 및 속도 또는 기타 오디오 장면 변경과 관련된 가청도).

- Rendering : 렌더러의 기능 및 동작과 관련된 요구사항

구분	순번	요구사항
렌더링 요구사항 (Rendering)	5	오디오 장면을 설명하는 메타데이터를 지원해야 한다.
	6	오디오 장면을 제어하고 제한하기 위한 메타데이터를 지원해야 한다.
	7	오디오 렌더링 매개변수(예: 소비 공간, 플레이어 기능 등)의 제어(예: 메타데이터 또는 인터페이스를 통해)를 지원해야 한다.
	8	시간(예: 0.5초마다) 및 공간(예: 하위 장면 내에서 또는 새로운 하위 장면으로 점프)에서 임의 액세스를 지원해야 한다.
	9	오디오 장면에서 두 개의 서로 다른 위치 사이에서 사용자가 점프하는 동안 오디오 렌더링에 전환 효과를 활성화하기 위한 메타데이터를 지원해야 한다. (예: 페이드 아웃, 페이드 인).
	10	메타 데이터는 오디오 줌을 활성화하는 기능을 지원해야 한다. (예: 돌출도 조정, 대화 향상, 피사계 심도 효과 시뮬레이션 등).
	11	메타 데이터는 오디오 객체가 부피를 가지는 것을 지원해야 한다.
	12	메타 데이터는 오디오 객체 및 채널의 방사 패턴 렌더링을 지원해야 한다.

	13	메타 데이터는 오디오 요소의 폐색을 지원해야 한다.
	14	오디오 장면에서 로컬 캡처된 오디오(예: 사용자 자신의 음성, 측음)를 지원해야 한다.
	15	오디오 개체(음원)의 정확한 3D 공간 위치 파악을 지원해야 한다. 위치의 차이는 인지할 수 있는 것과 관련이 있다.

- Interface and Extensibility : 렌더러의 입력 인터페이스 및 확장성과 관련된 요구사항

구분	순번	요구사항
인터페이스 요구사항 (Interface)	16	오디오 장면을 변경하기 위한 입력 인터페이스를 지원해야 한다.
	17	렌더링 기능의 확장이 가능해야 한다. (예: 외부 렌더링 도구에 대한 인터페이스, 확장 페이로드, 예약된 비트 필드 등).
	18	렌더러는 개인화된 HRTF를 지원해야 하며, 이러한 필터를 제공하기 위한 인터페이스를 포함해야 한다.
	19	렌더러는 헤드폰 이퀄라이징 filter를 지원해야 하며, 이러한 필터를 제공하기 위한 인터페이스를 포함해야 한다.

- Presentation mode : MPEG-I audio의 재현 방법과 관련된 요구사항

구분	순번	요구사항
확장성 요구사항 (Extensibility)	20	헤드폰 재생을 위한 6자유도 헤드 트래킹 바이노럴 렌더링을 지원해야 한다.
	21	스피커를 활용한 6자유도 헤드 트래킹 렌더링을 지원해야 한다. (예: 7.1 + 4H와 같은 몰입형 구성).
	22	사용자의 위치는 고정되고 가상 위치는 변경되는 사용 사례에 대해 스피커를 활용한 6자유도 렌더링을 지원해야 한다.
	23	6자유도 헤드 트래킹 바이노럴 헤드폰 재생과 스피커 재생의 조합으로 렌더링을 지원해야 한다.

- Social VR : Social VR 서비스 제공과 관련된 요구사항

구분	순번	요구사항
Social VR 요구사항 (Social VR)	24	가상 환경에서 다른 사용자의 음성 및 오디오 렌더링을 지원해야 한다. 음성과 오디오는 몰입형일 수 있다.

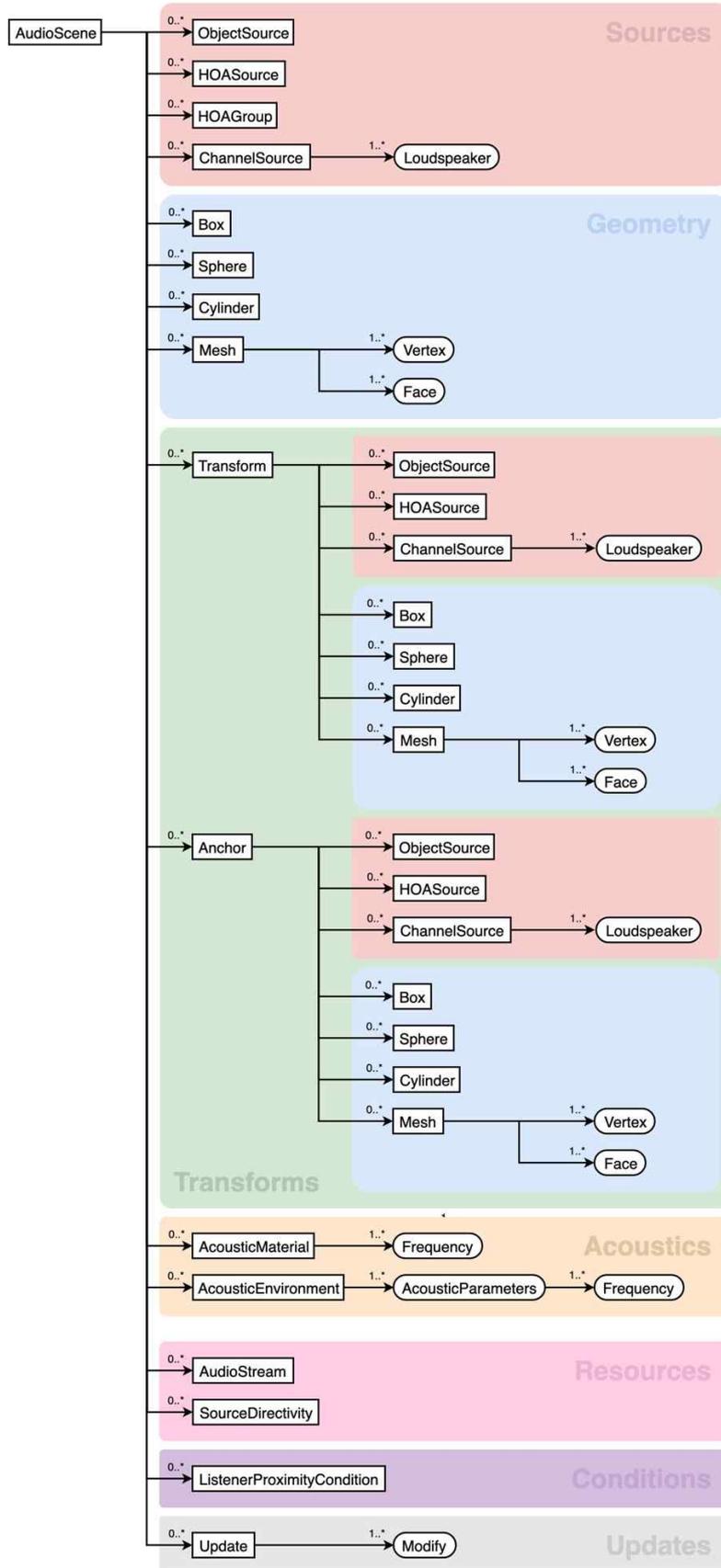
- Interoperability between 3 DoF and 6 DoF platform : 3자유도 플랫폼과의 호환성과 관련된 요구사항

구분	순번	요구사항
Social VR 요구사항 (Interoperability between 3 DoF and 6 DoF platform)	25	MPEG-H 3D Low complexity profile 오디오의 디코딩 및 재현을 지원해야 한다.
	26	MPEG-H 3D Audio Low Complexity Profile Decoder에서 MPEG-I Audio 콘텐츠의 디코딩을 지원해야 한다. (참고: MHAS 형식을 사용하면 이 요구 사항을 충족할 수 있다.)
	27	MPEG-I Audio 플랫폼은 3 DoF+, 3 DoF, 0 DoF와 같이 자유도가 감소된 환경에서 MPEG-I Audio 콘텐츠를 소비하는 것이 가능해야 한다.

6.5 MPEG-I Audio Encoder Input Format

MPEG Audio WG에서는 MPEG-I Audio 표준화를 위하여 제안 기술의 평가 목적으로 인코더 입력 포맷(EIF)을 규정하고 있다. EIF는 오디오 장면을 표현하는데 필요한 메타데이터로서, MPEG-I Audio 인코더의 입력 형식을 정의하고 있는 XML 파일이며, MPEG-I Audio 인코더는 이를 입력받아, 인코딩하여 하나의 비트스트림으로 생성한다. 이때, MPEG-I Audio 인코더는 EIF로부터 렌더링을 위한 정보를 추출하는 것이 가능하다.

EIF에서는 source, geometry, acoustics(geometry의 음향학적 특성 정보), resource, conditions, updates 정보를 구분하여 기술하고 있는데, 그 구조는 아래의 그림 6-3과 같다.



(그림 6-3) TIF의 구조

Source 정보에는 오디오 신호의 특성과 관련 정보에 대한 것이 포함되는데, 주요한 내용은 아래와 같다.

- Properties of audio signals – 오디오 신호의 특성에 대해 기술
 - ✓ object – 객체 오디오 신호
 - ✓ channels – 채널 오디오 신호
 - ✓ HOA – 앰비소닉 오디오 신호
- Metadata for signals – 오디오 신호와 관련된 정보에 대해 기술
 - ✓ 오디오 신호와 관련된 주요 정보

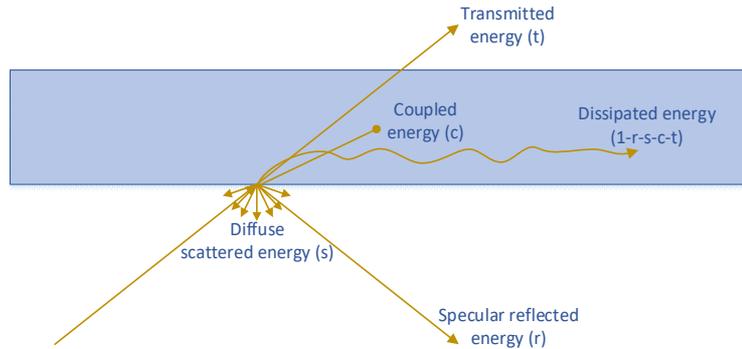
필드명	타입	설명
id	ID	아이디 정보
position	Position	3차원 공간 상의 위치 정보 (x, y, z)
orientation	Rotation	3차원 공간에서의 회전 정보 (Pitch, Yaw, roll)
active	Boolean	활성화를 위한 플래그
gainDb	Gain	소리의 크기 정보 (dB)
refDistance	Float > 0	기준 거리
signal	AudioStream ID	연관 오디오 신호의 아이디
extent	Geometry ID	오디오 객체의 크기 정보 표현
directivity	Directivity ID	오디오 신호의 방사 패턴
mode	Playback mode	재생 모드
play	Boolean	재생을 위한 플래그

Geometry 정보에는 가상 공간의 구성에 대한 정보가 포함된다. 가상 공간을 구성하고 있는 벽, 천장, 바닥, 물체 등에 대한 기하학적 정보가 기술되는데, 아래와 같은 요소들로 표현된다.

- Geometry information – 가상 공간의 구성 정보에 대해 기술
 - ✓ Box
 - ✓ sphere
 - ✓ cylinder
 - ✓ Mesh (Vertex, face)
 - ✓ 위의 구성 물체와 음향학적 특성 정보와의 연관성 정보

Acoustics에서는 물질의 음향학적 특성 정보와 공간의 음향학적 특징 정보를 기술하고 있는데, 이것은 상기 기하 공간의 구성 물체와 연결이 된다. 물질의 음향학적 특징 정보

는 4가지 요소들로 표현을 하고 있으며, 공간의 음향학적 특징 정보는 2가지 요소로 표현된다.



(그림 6-4) 물질의 음향학적 특성 정보의 요소

- Acoustic materials – 물질의 음향학적 특성 정보를 기술
 - ✓ r – specular reflection
 - ✓ s – diffuse scattering
 - ✓ t – transmission
 - ✓ c – coupling
 - ✓ $r + s + t + c \leq 1$

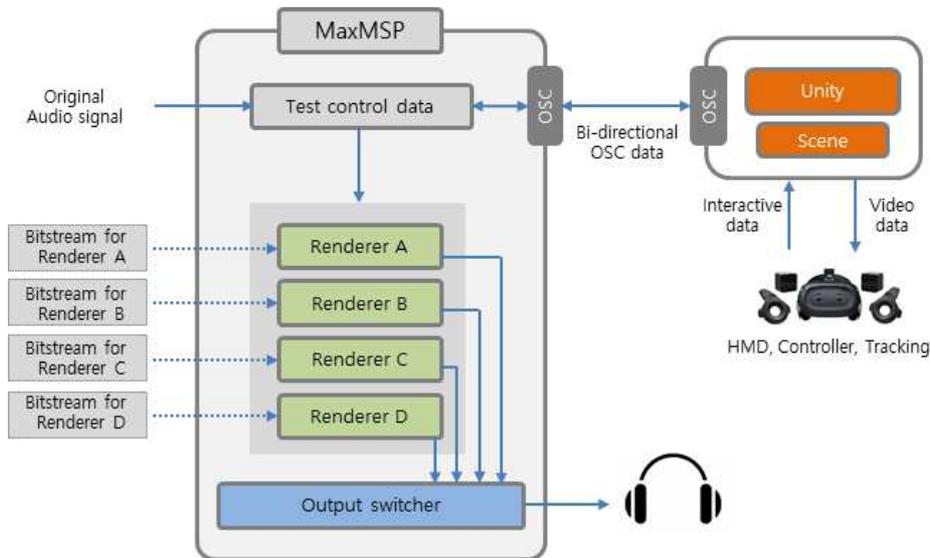
- Acoustic Environments – 공간의 음향학적 특성 정보를 기술
 - ✓ rt60 – reverberation time
 - ✓ ddr – diffuse to reverberant ratio

Resource 정보에는 렌더링에 사용되는 resource와 관련된 내용을 기술하는데, 실제 렌더러로 입력되는 오디오 스트림에 대한 정보와 렌더링을 위해 필요한 directivity sofa 정보가 대표적인 예가 될 수 있다. Conditions 정보에는 가상 공간에서의 장면 변화가 발생할 수 있는 조건 정보에 대해 기술한다. 예를 들어 ‘청취자가 특정 자동차와의 거리가 1.0m보다 적을 때’ 등이 이러한 조건 정보가 될 수 있다. Updates 정보에는 가상 공간에서의 발생할 수 있는 장면 변화에 대해 기술한다. 장면 변화는 다양한 형태로 발생할 수 있는데, 장면 변화가 발생하는 경우를 요약하면 아래와 같다.

- update type – 장면 변화가 발생하는 종류
 - ✓ 특정 시점에 업데이트 발생
 - ✓ 특정 조건에 업데이트 발생
 - ✓ 사용자의 interaction에 의한 이벤트성 업데이트 발생
 - ▷ eg) TV 전원 on/off
 - ✓ 사용자의 interaction에 의한 연속적인 업데이트 발생
 - ▷ eg) 움직이는 공

6.6 MPEG-I Audio Evaluation Platform

MPEG Audio WG에서는 MPEG-I Audio의 성능 평가를 위한 평가 시스템(MPEG-I AEP, MPEG-I Audio Evaluation platform)을 제공한다. 최초 제공된 평가 시스템을 기반으로 안정적이고 정확한 평가를 위한 지속적인 수정 및 보완이 진행되었다. 아래의 그림 6-5는 MPEG에서 제안한 평가 시스템을 간략화하여 나타낸 것이다.



(그림 6-5) MPEG-I Audio 평가 시스템의 구조

MPEG-I Audio 성능 평가 시스템은 청취자의 6자유도 움직임을 지원하기 위하여 청취자의 위치 정보와 같은 상호작용 정보를 제공해야 한다. 이를 위하여, 평가 시스템은 청취자에게 영상 정보를 제공하는 가상현실 재현 장비인 HTC Vive 장비, 비디오 렌더링 및 평가 제어를 위한 유니티(Unity) 프레임 워크, 그리고 오디오 렌더링을 위한 MaxMSP로 구성된다.

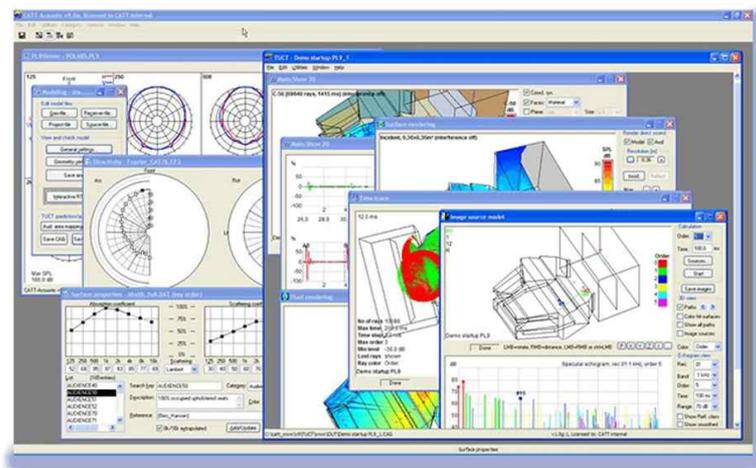
유니티 프레임워크는 6자유도 콘텐츠의 재생에 활용되는데, 평가자는 HMD를 통해 영상을 보면서 콘텐츠 내의 공간을 자유롭게 이동할 수 있으며, 컨트롤러를 활용하여 콘텐츠와 상호작용을 한다. 트래킹 장비를 통하여 측정된 사용자의 위치 정보 및 컨트롤러를 통한 사용자와 콘텐츠의 상호 작용 정보는 OSC(Open Sound Control) 통신을 활용하여 유니티 프레임워크에서 MaxMSP 응용 프로그램으로 전달된다.

MaxMSP 응용 프로그램은 MPEG-I Audio Renderer가 동작하는 환경을 제공하며, 사용자의 위치 정보와 인터랙션 정보 등을 유니티 프레임 워크로부터 전달받아, MPEG-I Audio Renderer로 전달한다. 또한, 여러 기관에서 구현한 MPEG-I Audio Renderer 들이 동시에 동작할 수 있는 환경을 제공하며, 여러 개의 MPEG-I Audio Renderer에서 생성된 6자유도 오디오들 중 하나를 평가자가 선택하여 청취할 수 있도록 하는 기능을 제공하며, 평가자의 평가 결과를 저장하는 기능을 포함하고 있다.

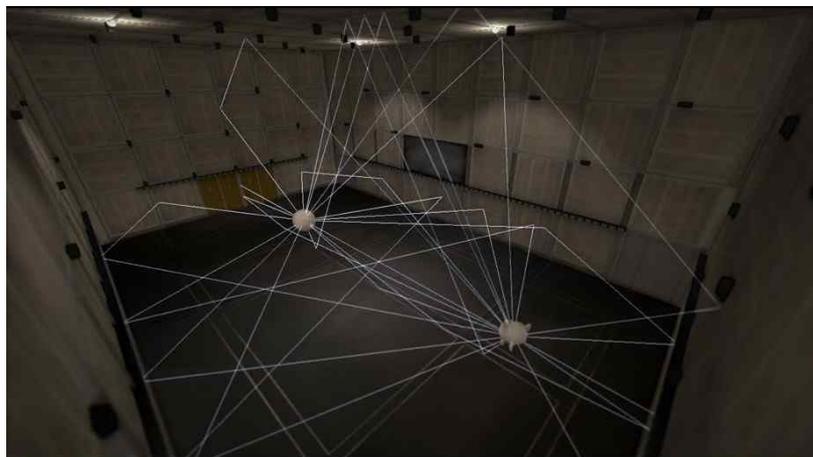
실제 6자유도 오디오 재현 알고리즘이 동작하는 부분은 MaxMSP 응용프로그램에서 동작하는 MAX External로 구현이 되는데, 그림 6-5에서 녹색으로 표기된 Renderer가 이에 해당한다. MAX External과 MAXMSP 응용 프로그램 간의 인터페이스를 미리 정의하였고, 기본적인 인터페이스를 지원하는 'skeleton' 코드를 제공하여, 각각의 기관이 독립적으로 MPEG-I Audio Renderer를 개발할 수 있는 환경을 제공하고 있다.

7 6자유도 오디오 기술 동향

앞서 살펴본 것과 같이 6자유도 콘텐츠 재현환경에서는 사용자가 이동하거나 회전함에 따라 오디오 객체와의 거리나 방향 등이 달라지며, 이에 따라 오디오 객체로부터 사용자에게 전달되는 직접음과 반사음의 경로가 달라진다. 이와 같은 요소들을 고려하여 6자유도 콘텐츠에서 오디오를 렌더링하고 재현하는 것은 복잡한 알고리즘과 높은 연산량을 요구하는 작업으로, CATT Acoustic, RAVEN(Room Acoustics for Virtual Environments), EVERTIMS 등과 같은 실시간 보다는 충실한 오디오 렌더링에 중점을 둔 오디오 신호처리 모듈로 개발이 되고 있었으며, 게임이나 가상 현실 등과 같은 실시간 연산이 요구되는 분야에서는 복잡한 기술이 적용되지는 못하고 있었다. 그러나, CPU의 성능이 좋아짐에 따라 비실시간으로 동작하는 오디오 렌더링 모듈을 실시간 렌더링 환경에서도 동작하도록 개발이 되고 있다.



(a) CATT Acoustic 실행 화면



(b) Blender에서 실행되는 Everims 화면
(그림 7-1) 6자유도 오디오 렌더링 SW의 예

최근에는 마이크로소프트, 구글 등과 같은 회사에서 6자유도를 가지는 게임과 콘텐츠 개발에 사용할 수 있는 오디오 재현 모듈에 대해 개발을 하고, 관련 소프트웨어를 플러그인 형태로 제공하고 있다.

CATT Acoustic, RAVEN, EVERTIMS, ODEON room acoustic SW 등의 경우, 오래전부터 개발이 되고 있는 것들로서, 다른 문서 등에서 많이 소개되었기 때문에, 본 보고서에서는 상대적으로 최근에 마이크로소프트와 구글에 의해 소개된 6자유도 오디오 기술인 Project Acoustics와 Resonance Audio에 대하여 소개한다.

7.1 Project Triton and Acoustics

7.1.1 Introduction of Project Triton and Acoustics

마이크로소프트에서는 최근 6자유도 오디오 기술의 개발을 위한 프로젝트인 Project Triton과 Project Acoustics에 대해 소개를 하였다. Project Triton의 경우, 6자유도에서의 공간 음향 렌더링을 위한 알고리즘에 대한 연구를 수행한 프로젝트로 볼 수 있으며, Project Acoustics의 경우, 개발된 알고리즘을 응용 프로그램에서 활용할 수 있도록 하는 기술에 대한 연구를 수행한 프로젝트라 볼 수 있다.

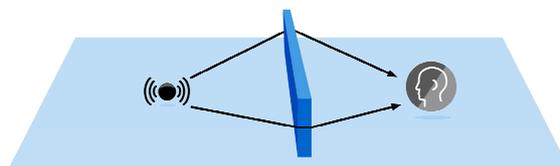
해당 연구와 관련된 논문으로 보이는 ‘Precomputed Wave Simulation for Real-Time Sound Propagation of Dynamic Sources in Complex Scenes’ 논문이 2010년도에 게재된 것으로 보아 관련 연구는 그 이전부터 수행된 것으로 보여진다.

마이크로소프트에서는 Project Triton을 ‘Project Triton is a physics-based audio design system that creates an acoustic rendering that sounds natural, while keeping CPU usage modest and providing designer control’라고 기술하고 있는데, 이는 CPU 사용량을 적정 수준으로 유지하면서, 물리학 기반으로 자연스러운 음향 렌더링을 생성하는 오디오 시스템을 목표로 한다는 것으로 해석할 수 있을 것이다.

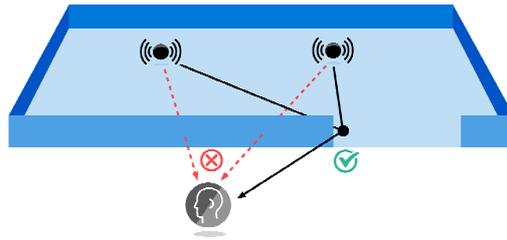
7.1.2 Sound Propagation Effects

마이크로소프트에서는 물리학 기반으로 자연스러운 음향 렌더링을 생성하기 위하여 obstruction, portaling, occlusion, reverberance, decay time 등과 같은 음향학적 효과를 고려하고 있다.

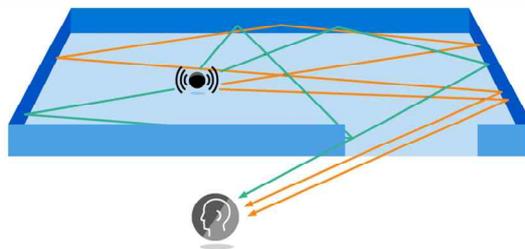
- Obstruction – Sound is weakened when it diffracts around obstructions.



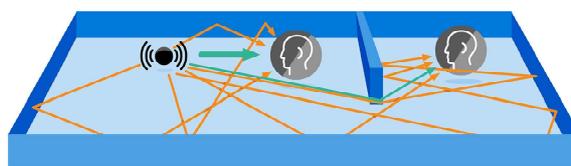
- Portaling (sound redirection) – Sound heard “through the wall” (red) from source direction is incorrect. Doors redirect sounds, which is more believable.



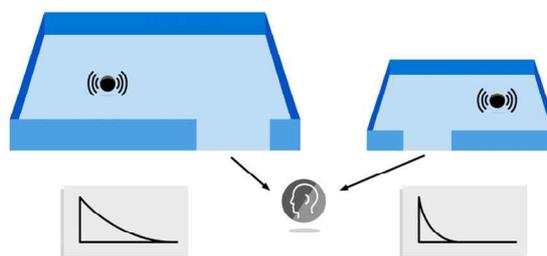
- Occlusion – Occlusion is total reduction in loudness from geometry, involving complex propagation and diffraction (diffracted paths in green).



- Reverberance – Left: Close to the source, direct path (green) is loud compared to reflections (orange), resulting in high clarity and low reverberance. Right: Behind the partition, direct path is weakened from diffraction, causing low clarity and high reverberance, conveying that the source is in a different room.



- Decay Time – Larger room reverberate longer.



7.1.3 Work Flow

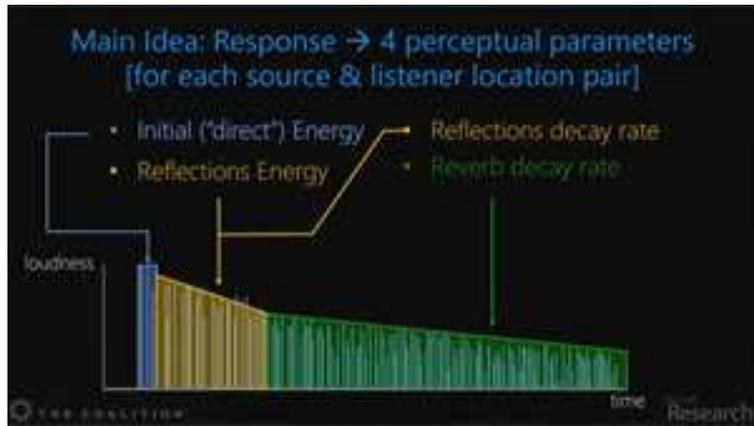
마이크로소프트의 6자유도 오디오 기술은 기본적으로 3차원 공간상에 샘플링 된 지점에서의 IR을 시뮬레이션하여 생성하고, 이를 파라미터화하는 과정을 수행하는데, 파라미터화된 값들은 오디오 렌더링 응용 프로그램에서 활용이 된다. Project Acoustics 에서는 이 과정을 Pre-baking, Bake, Runtime의 3단계로 나누어서 설명하고 있는데, 여기에서는 각 과정에 대해 간단하게 살펴본다.

- Pre-bake: Baking을 하기 위한 정보를 설정하는 단계로 볼 수 있다. 음향학적 3차원 공간 정보를 선택하고, 공간을 구성하는 물질의 음향학적 특성 정보를 설정한다. 음향학적 3차원 공간 정보에는 음원 객체의 위치 정보가 포함된다. 그리고, 해당 공간 내에서 임펄스 응답을 계산할 탐색 영역을 설정하고, 임펄스 응답을 계산할 위치를 샘플링 하여 설정하는 과정이 포함된다.



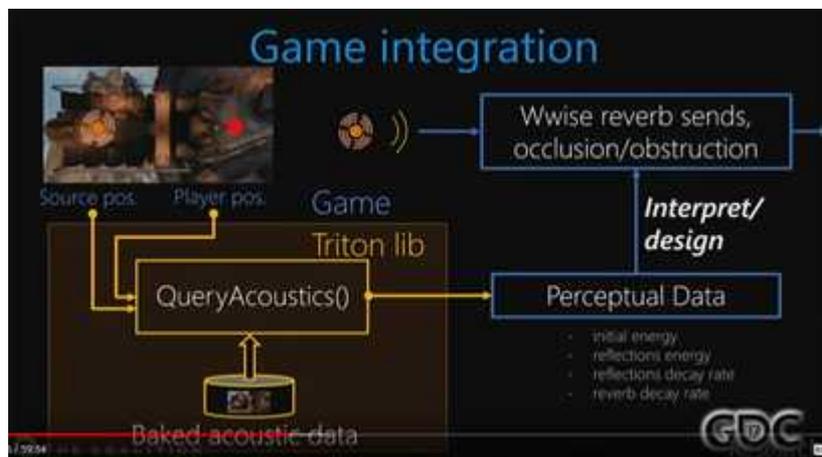
(그림 7-2) 3차원 공간 내에 임펄스 응답을 계산할 위치를 나타낸 예

- Bake: 실질적으로 공간의 음향학적 특성을 추출하는 과정이라 볼 수 있다. 앞선 pre-baking 과정에서 주어진 음향학적 3차원 공간 정보를 활용하여 샘플링 된 위치에서의 임펄스 응답을 시뮬레이션하여 계산한다. 이때, 임펄스 응답을 계산하는 기본적인 알고리즘은 FDTD(Finite Difference Time Domain)를 사용한다. 임펄스 응답은 앞서 샘플링하여 설정된 개수 만큼 생성이 되는데, 일반적으로 수많은 위치에 대한 임펄스 응답이 생성되므로, 많은 연산량이 소요되며, 임펄스 응답의 저장을 위해 많은 저장 용량이 요구된다. 다음으로, 생성된 임펄스 응답들을 분석하여 임펄스 응답을 파라미터화 하는 과정을 거치게 되며, 이후의 과정에서는 이 파라미터화 된 임펄스 응답이 사용되는데, 최초 생산된 임펄스 응답의 저장 용량 대비, 파라미터화 된 임펄스 응답의 저장 용량은 매우 적다.



(그림 7-3) 생성된 임펄스 응답을 활용하여 파라미터화 하는 개념을 나타낸 예

- Runtime: 파라미터화된 임펄스 응답을 응용에서 활용하는 단계이다. 파라미터화 된 임펄스 응답은 Project Acoustic에서 개발한 플러그인에서 로드되어 활용되는데, 해당 플러그인은 Unity와 같은 VR 콘텐츠 개발 SW에서 활용되며, 세분화된 제어가 가능하다.



(그림 7-4) VR 콘텐츠 개발 SW에서 파라미터화된 임펄스 응답을 활용하는 방법의 예

7.1.4 Support Platform

마이크로소프트에서는 해당 플러그인이 실행되는 platform에 대해 아래와 같이 소개하고 있다.

- Runtime Platform – Project Acoustics runtime plugin이 실행되는 플랫폼
 - ✓ windows
 - ✓ Xbox
 - ✓ Android
 - ✓ MacOS

- Editor platforms – Project Acoustics editor plugin이 실행되는 플랫폼
 - ✓ Windows
 - ✓ MacOS (Unity only)

또한 아래와 같은 SW를 제공하고 있다.

- Project Acoustics Unity plugin and samples
- Project Acoustics Unreal & Wwise plugins and samples – Windows

7.2 Resonance Audio

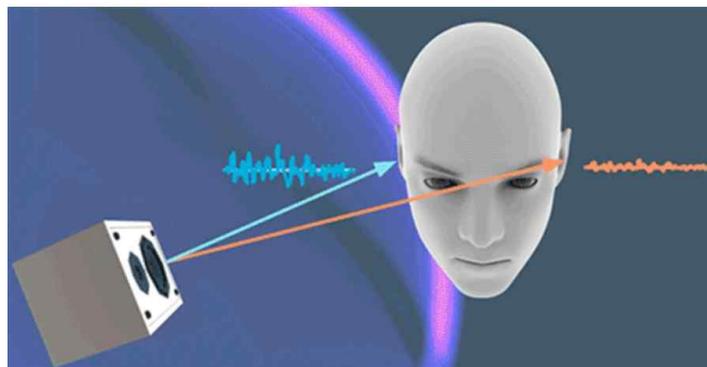
7.2.1 Introduction of Resonance audio

구글에서는 2017년도에 ‘Resonance Audio’라는 이름의 6자유도 오디오 재현을 위한 기술을 소개하였다. 구글은 Resonance Audio를 활용하여 VR, AR, 게임 환경에서 다이나믹한 공간 음향을 재현할 수 있다고 소개하고 있는데, 아래에서 Resonance Audio의 음향학적 기본 개념과 동작 방법에 대해 소개한다.

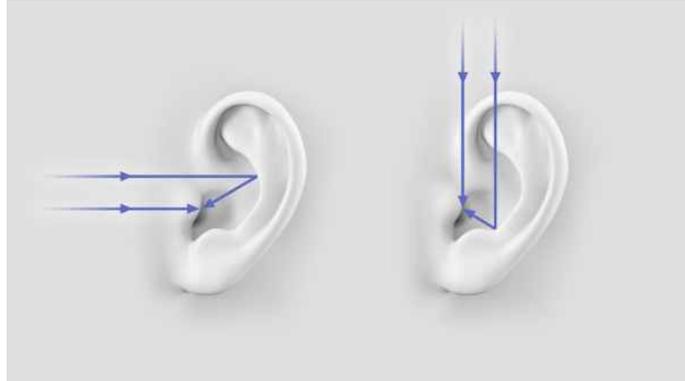
7.2.2 Fundamental Concepts

구글에서는 음향 신호가 사람의 귀와 주변 환경과 상호작용하는 것을 시뮬레이션 하는 방법으로 Resonance Audio가 동작한다고 소개하고 있으며, 상호 작용에 대해 아래와 같이 소개하고 있다.

- 음향 신호와 사람 귀와의 상호작용
 - ✓ Interaural time difference(ITD) – The time difference between sound wave arrival at our left and right ears helps us determine the horizontal location of low-frequency sounds
 - ✓ Interaural level difference(ILD) – These are differences in loudness and frequency distribution between our right and left ears, caused by the human head’s acoustic shadow.



- ✓ Spectral effects – Sounds coming from different directions bounce off of the inside of our outer ears in different ways



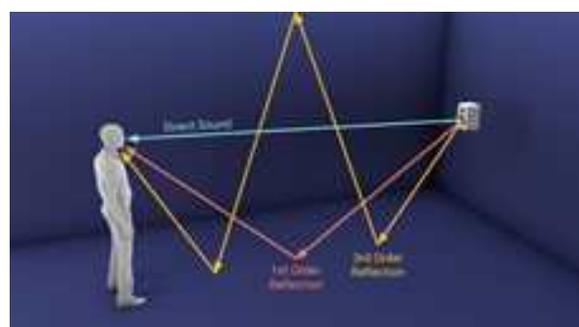
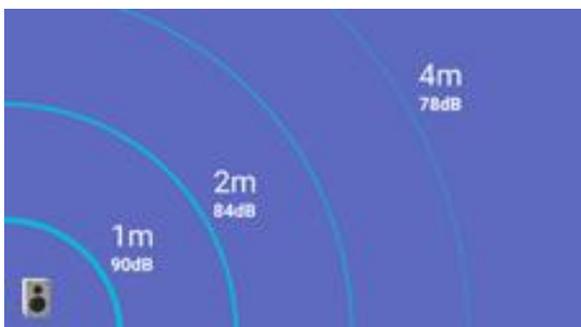
- ✓ using HRTF – To simulate real sound wave interactions with our ears, Resonance Audio uses head-related transfer functions (HRTFs).

- 음향 신호와 환경과의 상호작용

- ✓ Head movements and sound position – Moving our heads helps us to perceive relative changes in audio location.



- ✓ Early reflections and reverberations – Resonance Audio separates this mix of sound waves into the following three components (Direct sound, Early reflection, late reverberation)



- ✓ Occlusion – To add further realism, Resonance Audio can also simulate how real sound waves traveling between a source and listener are blocked by objects between them.



- ✓ directivity – A directivity pattern or shape represents the way in which sound emanates from a source in different directions



- ✓ Ambisonics – Resonance Audio uses a technology called Ambisonics to envelop the user in a sphere of sound.



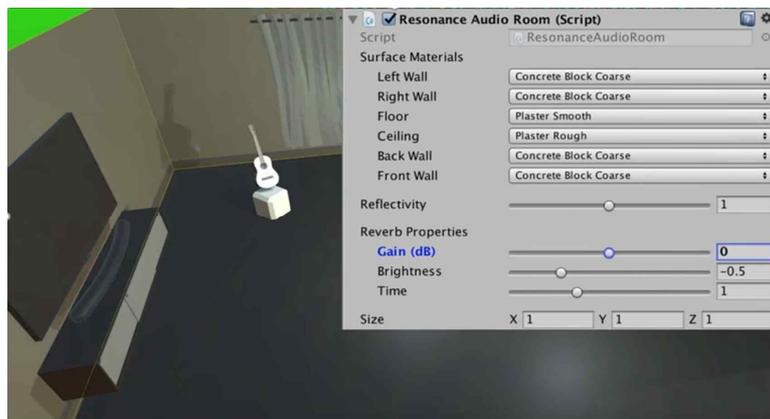
7.2.3 Work Flow

Resonance Audio는 기본적으로 실시간으로 IR을 생성하는 방법을 사용하며, 좀 더 정밀한 잔향을 생성하기를 원하는 경우에는 pre-processing 형태로 잔향을 미리 계산한 후에 활용하는 방법을 사용하고 있다.

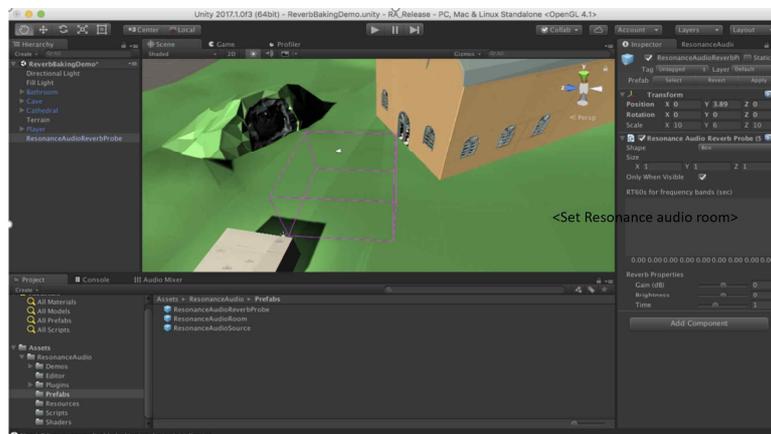
Resonance Audio를 실행하는 응용 플랫폼에 따라 이를 적용하는 방법은 차이가 있는데, 아래에서는 Unity에서 Resonance Audio를 활용하는 방법에 대해서 간단하게 기술한다. Unity에서는 Resonance Audio Component를 추가하고, 이들의 값을 설정하는 것으로 Resonance Audio를 활용할 수 있는데, Resonance Audio의 Components는 아래와 같은 것들이 있다.

- ResonanceAudioListener - 소리를 듣는 청취자에 해당
- ResonanceAudioSource - 음원 객체에 해당
- ResonanceAudioSoundfield
- ResonanceAudioRoom - ResonanceAudio가 적용되는 공간에 해당하며, 표면의 재질과 반사율 등과 같은 파라미터 설정
- ResonanceAudioReverbProbe - 공간의 더 미세한 모델링과 더 미묘한 리버브 효과를 위한 고급 옵션을 제공

아래의 그림 7-5와 7-6은 Resonance Audio에서 room component와 reverbProbe Component를 설정하는 그림의 예를 나타낸다.



(그림 7-5) Resonance Audio Room Component 설정 화면



(그림 7-6) Resonance Audio ReverbProbe Component 설정 화면

7.2.4 Support Platform

Resonance Audio는 아래와 같은 플랫폼에서 동작하며, 각 플랫폼에서 동작하는 방법에 대해서도 가이드를 제공하고 있다.

- Resonance Audio development environment or platform
 - ✓ Andorid Studio
 - ✓ FMOD
 - ✓ Resonance Audio SDK for IOS
 - ✓ Unity
 - ✓ Unreal
 - ✓ Audio Monitor VST
 - ✓ Resonance Audio SDK for WeB
 - ✓ Wwise

8 결론 및 시사점

8.1 국내 표준화 관점

6자유도 오디오 관련 기술은 오래전부터 개발이 진행되어 왔으며, 최근에도 활발하게 진행이 되고 있다고 볼 수 있는데, 이러한 기술들은 각 개발 주체별로 독립적인 방법과 규격을 활용하고 있어 6자유도 오디오를 위한 표준화의 필요성이 제기되고 있었다.

DVB, 3GPP 등과 같은 기관에서 Immersive Audio의 표준화에 대한 검토를 수행하였으며, MPEG에서는 관련 표준화가 진행이 되고 있는 상태이다. 현재 MPEG에서 6자유도 오디오 기술인 MPEG-I Audio 표준화가 활발하게 진행이 되고 있어, 향후에는 각 기관별로 다른 규격을 사용하는 문제를 해소할 수 있는 바탕이 마련될 것으로 기대가 된다. 다만, MPEG-I Audio 표준이 이미 각각의 방법과 규격으로 개발된 기존의 개별 시스템과 어떻게 호환성을 가지고 활용이 될 수 있을 지에 대해서는 시간을 두고 지켜볼 필요가 있을 것으로 생각이 든다.

한편, 국제 표준화 기관인 MPEG에서 6자유도 오디오에 대한 표준을 제정하고 있는 상태에서 국내에서 이와는 독립적인 6자유도 오디오와 관련한 별도의 표준을 제정하는 일을 수행하는 것은 시기적으로 적절하지 않은 것으로 생각이 된다.

MPEG-I Audio 표준의 제정이 완료되었을 때, 이를 분석하여, 필요하다면 MPEG-I Audio 표준을 기반으로 한국의 실정에 맞도록 표준을 수정하는 형태로 국내 표준을 제정하는 것이 효율적이라 생각된다. 현 시점에서 보면 현재 진행 중인 MPEG-I Audio 표준화에 국내 업체가 적극적으로 참여하여, 지적재산권을 확보하는데 노력하는 것이 좋을 것이라고 생각된다.

MPEG에서는 2021년 09월에 MPEG-I Audio의 기술 제안 및 평가를 위한 기관의 등록 절차가 있었고, Ericsson, Philips, Dolby, Nokia, Fraunhofer 등 12개의 업체가 등록을 하였다. 한국에서는 ETRI와 가우디오랩이 등록을 한 상태인데, 해당 업체의 적극적인 표준화 활동을 기대하며, 국내의 타 기관에서도 표준화에 많은 관심을 가지고 참여를 하였으면 한다.

8.2 국내 서비스 및 산업 관점

최근에 많은 VR/AR 콘텐츠 및 게임 등에서 6자유도 영상을 제공하고 있으나, 6자유도 오디오는 제공하지 못하는 경우가 많으며, 6자유도 오디오를 제공하더라도, 음향의 현장감은 아직 기대에 미치지 못한 상태이다. 이는 6자유도 오디오를 재현하는 기술이 여전히 개발 중인 것도 이유가 될 수 있으며, 이러한 기술의 개발이 6자유도 영상과 동기화되어 동작하는 환경에서 개발된 것이 아니라 6자유도 영상과는 별개로 오디오 신호의 처리에만 중점을 두고 개발이 이루어졌던 것도 이유가 될 수 있다고 생각된다.

최근에는 하드웨어 성능의 향상에 따라 EVERTIMS, Project Acoustics, Resonance Audio 등과 같은 6자유도 오디오 렌더링 프로그램이 영상과 동기화되어 실시간으로 재현되는 등 많은 발전이 이루어지고 있는 상태이다. 특히, Project Acoustics, Resonance Audio 등과 같은 SW는 Unity, Wwise 등과 같은 6자유도 콘텐츠 제작 환경에서 플러그인 형태로 동작을 하여, 6자유도 오디오를 더욱 쉽게 제작할 수 있는 환경을 제공하고 있다. 추후, 표준화가 진행 중인 MPEG-I Audio의 표준화가 완료되면, 이러한 발전은 더욱 진전될 것으로 생각된다.

국내에서도 6자유도를 제공하는 콘텐츠가 많이 제작되고 있는데, 6자유도 오디오의 실감 효과 제공 측면에서 보면, 아직은 부족한 면이 있다고 생각된다. 6자유도 영상과 함께 실감나는 6자유도 오디오를 제공하는 경우, 콘텐츠의 가치를 높일 수 있을 것으로 생각이 되므로, 콘텐츠 제작 업체에서는 앞서 소개하였던, Project Acoustics, Resonance Audio 등과 같은 6자유도 오디오를 제공하는 기술을 도입하여, 더욱 실감나는 콘텐츠를 제작하는 것이 필요할 것으로 생각된다.

한편, 국내에서도 6자유도 오디오 관련 기술 및 제품을 개발하는 업체가 있으나, 그 수는 많지 않은 상태이다. 이는 지금까지 6자유도 콘텐츠 관련 기술에서 영상 관련 부분이 많은 이슈가 되어, 영상 위주로 기술 개발이 진행되었던 것도 하나의 이유가 될 수 있다고 생각된다.

국외에서는 6자유도 관련 오디오 기술에 대한 관심 및 기술의 개발이 많이 이루어지고 있으며, 표준화도 진행되고 있으므로, 국내에서도 6자유도 관련 오디오 기술의 개발에 대한 관심을 가져야 할 것으로 생각이 된다. 이를 위해서는 산학연이 협력하여, 우수한 6자유도 오디오 재현 기술을 개발하고, 이를 제품화하는 노력이 필요할 것으로 생각되며, 이를 위한 정부의 투자도 필요할 것으로 생각된다.

부 록 1-1

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

지식재산권 요약서 정보

해당 사항 없음

부 록 1-2

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

시험인증 관련 사항

해당 사항 없음

부 록 1-3

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

본 기술보고서의 연계(family) 표준

해당 사항 없음

부 록 | -4

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

참고 문헌

[5 6자유도 오디오 소개]

- [1] ITU-R BS.2076-2, 'Audio Definition Model', Oct, 2019.
- [2] ETSI TS 103 190 V1.1.1, Digital Audio Compression (AC-4) Standard, April, 2014.
- [3] ISO/IEC 23008-3:2015, High efficiency coding and media deliver in heterogeneous environments), Part 3: 3D Audio, 2015.

[6 6자유도 오디오 표준화 동향]

- [1] ISO/IEC JCT1/SC29/WG11/N17267, "Requirements MPEG-I phase 1b", MPEG 120, Oct. 2017
- [2] ISO/IEC JCT1/SC29/WG11/N17285, "MPEG-I Visual activities on 6DoF and Light Field", MPEG 120, Oct. 2017
- [3] ISO/IEC JTC1/SC29/WG11/N17964, "WD of ISO/IEC 23090-2 2nd edition OMAF", MPEG 124, Oct. 2018
- [4] ISO/IEC JTC 1/SC29/WG 11 MPEG 2017/M16753, 'Thoughts on MPEG-I and 3D audio', April 2017
- [5] ISO/IEC JTC1/SC29/WG11/N19409, "Draft MPEG-I 6DoF Audio Encoder Input Format", July, 2020
- [6] ISO/IEC JTC1/SC29/WG11/N17177, "MPEG-I Audio Architecture and Evaluation for 6DoF," 2017. 10.
- [7] ISO/IEC JTC1/SC29/WG11, "Working Draft 1.0 of TR: Technical Report on Architectures for Immersive Media," N17060, July 2017.

[7 6자유도 오디오 기술 동향]

- [1] Dirk S., Michael V., "RAVEN: A Real-Time Framework for the Auralization of Interactive Virtual Environments", European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767.
- [2] D. Poirier-Quinot, B.F.G. Katz, M. Noisternig, "EVERTims: Open Source Framework for Real-Time Auralization in Architectural Acoustics and Virtual Reality," in Proc. Int. Conf. Digital Audio Effect(DAFx-17), Edinburgh, UK, Sept. 2017, pp. 323-328.
- [3] Nikunj Raghuvanshi, John Michael Snyder, "Parametric wave field coding for precomputed sound propagation", ACM Transactions on Graphics, July 2014 Article No. 38.

- [4] Gorzel, Marcin; Allen, Andrew; Kelly, Ian; Kammerl, Julius; Gungormusler, Alper; Yeh, Hengchin; Boland, Francis, “Efficient Encoding and Decoding of Binaural Sound with Resonance Audio”, AES Conference: 2019 AES International Conference on Immersive and Interactive Audio, March 2019.
- [5] <https://docs.microsoft.com/en-us/gaming/acoustics/what-is-acoustics>
- [6] <https://www.microsoft.com/en-us/research/project/project-triton/>
- [7] <https://resonance-audio.github.io/resonance-audio/>
- [8] <https://resonance-audio.github.io/resonance-audio/discover/concepts>

※ 상기 기재된 참고 문헌의 발간일이 기재된 경우, 해당 표준(문서)의 해당 버전에 대해서만 유효하며, 연도를 표시하지 않은 경우에는 해당 표준(권고)의 최신 버전을 따름

부 록 1-5

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

영문기술보고서 해설서

해당 사항 없음

부 록 1-6

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

기술보고서의 이력

판수	채택일	기술보고서번호	내용	담당 위원회
제1판	2020.11.29	제정 FBMF-TR-008	6자유도 오디오 표준화 및 기술 동향	미래미디어 분과위원회
오류정정				
오류정정				
제2판				