일반논문 (Regular Paper)
방송공학회논문지 제28권 제2호, 2023년 3월 (JBE Vol.28, No.2, March 2023)
https://doi.org/10.5909/JBE.2023.28.2.213
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

MPEG-I AEP 기반 실시간 6 자유도 공간음향 렌더링 시스템

강 경 옥[a)‡], 유 재 현[a)], 장 대 영[a)], 이 용 주[a)], 이 태 진[a)]

A Real Time 6 DoF Spatial Audio Rendering System based on MPEG-I AEP

Kyeongok Kang[a)‡], Jae-hyoun Yoo[a)], Daeyoung Jang[a)], Yong Ju Lee[a)], and Taejin Lee[a)]

요 약

본 논문에서는 가상환경에 위치한 청취자의 움직임에 대응하여 실시간으로 6DoF 공간음향을 제공하는 공간음향 렌더링 시스템에 대해 소개한다. 본 시스템은 MPEG-I Immersive Audio CfP 대응을 위하여 MPEG-I AEP를 개발환경으로 사용하여 구현되었으며 인코더와, 디코더를 포함하는 렌더러로 구성된다. 인코더는 인코더 입력 포맷(EIF) 파일에 포함된 가상공간 장면의 공간적 오디오 파라미터와, SOFA 파일로 제공되는 음원의 지향성 정보 등의 메타데이터를 오프라인으로 부호화하여 비트스트림으로 전달하는 역할을 하며, 렌더러는 전달된 비트스트림을 수신하여 청취자의 위치에 따라 실시간으로 6DoF 공간음향 렌더링을 수행한다. 개발된 렌더링 시스템에 적용한 주요 공간음향 처리 기술로는 음원 효과 및 장애물 효과 처리 기술이 있으며, 그 외 시스템 동작에 필요한 기술로는 도플러 효과 및 음장효과 처리 기술 등이 있다. 개발된 시스템에 대한 성능평가 결과로서 자체 주관평가 결과를 소개한다.

Abstract

In this paper, we introduce a spatial sound rendering system that provides 6DoF spatial sound in real time in response to the movement of a listener located in a virtual environment. This system was implemented using MPEG-I AEP as a development environment for the CfP response of MPEG-I Immersive Audio and consists of an encoder and a renderer including a decoder. The encoder serves to offline encode metadata such as the spatial audio parameters of the virtual space scene included in EIF and the directivity information of the sound source provided in the SOFA file and deliver them to the bitstream. The renderer receives the transmitted bitstream and performs 6DoF spatial sound rendering in real time according to the position of the listener. The main spatial sound processing technologies applied to the rendering system include sound source effect and obstacle effect, and other ones for the system processing include Doppler effect, sound field effect and etc. The results of self-subjective evaluation of the developed system are introduced.

Keyword : MPEG-I, Immersive Audio, VR, Encoder, Renderer

a) 한국전자통신연구원 초실감메타버스연구소 미디어연구본부(Media Research Division, Hyper-Reality Metaverse Research Laboratory, ETRI)
‡ Corresponding Author : 강경옥(Kyeongok Kang)
E-mail: kokang@etri.re.kr
Tel: +82-42-860-5521
ORCID: https://orcid.org/0000-0002-7673-652X

※This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [22ZH1200, The research of the basic media·contents technologies]
· Manuscript December 23, 2022; Revised February 14, 2023; Accepted February 14, 2023.

Copyright ⓒ 2023 Korean Institute of Broadcast and Media Engineers. All rights reserved.
"This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered."

Ⅰ. 서 론

통신 및 미디어 기술의 발전에 따라 실감을 제공하는 미디어 서비스는 더 이상 먼 미래에 체험할 수 있는 서비스가 아니며, 방송, 통신, 모바일 기기를 이용하여 가상현실(Virtual Reality; VR) 및 증강현실(Augmented Reality; AR) 서비스를 통해 몰입감(또는 실감)을 체험하는 것도 가능하게 되었다. 더 나아가서는 이러한 VR/AR 서비스에 이어 메타버스(metaverse) 서비스에 대한 관심이 고조됨에 따라 실내음향(room acoustics) 기술을 가상환경에 적용하기 위한 기술개발 및 표준화가 활발히 진행되고 있다.

방송, 통신 환경이나 모바일 서비스를 통해 제공되는 3D 오디오 또는 3차원 음향이라 불리는 기존의 공간음향 서비스에서는 청취자 위치 기반으로 최적의 공간음향을 제공하거나 대화형(interactive) 공간음향을 제공하지만, 청취자가 위치한 청취공간에서 자유롭게 이동하는 경우에 청취자의 위치 변경이나 머리 움직임에 따른 공간음향의 변화를 제대로 반영해 주지 못하는 것이 주지의 사실이다.

청취자가 위치한 공간에서 음원 소스와 공간에 의한 실내음향 효과(공간음향)를 제대로 전달하거나 표현하기 위해서는, 음원 소스 모델링, 전달 매질(실내음향) 모델링 및 청취자 모델링이 중요하다고 할 수 있다[1]. 그림 1에서 볼 수 있듯이 음원 소스 모델링 요소로는 소스의 지향성(directivity) 특성을 들 수 있고, 전달 매질 모델링의 주요 요소는 청취자가 위치한 공간의 충격응답(room impulse response) 특성을 들 수 있다. 청취자 모델링의 경우에는 인간이 공간감을 인지할 때 사용하는 특성에 해당하는 양이레벨차(Interaural Level Difference; ILD), 양이시간차(Interaural Time Difference; ITD) 및 머리전달함수(Head Related Transfer Function; HRTF) 등을 고려할 수 있다.

청취자가 위치한 공간의 실내음향(room acoustics), 즉 공간의 충격응답 특성이라고 할 수 있는 공간음향 특성을 제대로 표현하기 위한 연구 또한 오랫동안 진행되어 오고 있다. 실내음향 연구는 크게 물리 기반의 파동 방정식을 이용한 접근법인 파동(wave) 기반 모델링 방법과 기하음향(geometry acoustics; GA) 기반 모델링 방법으로 분류할 수 있다[2][3]. 이중 GA 기반 모델링 방법은 빛(light)을 광선(ray)로 표현하는 것과 마찬가지로 소리(sound)도 음선(ray)으로 표현할 수 있다는 것이다. 즉 GA 기반 모델링에서는 소리의 파동 성질을 무시하기 때문에 이와 관련된 모든 현상은 사라지게 되어 회절과 같은 현상을 표현할 수 없다. 이와 같은 약점에도 불구하고 소리의 고주파수 특성을 잘 표현하고 무엇보다도 파동 기반 방식에 비해 대상 주파수와 무관하게 상당히 적은 계산량을 필요로 한다는 장점이 있어 기하음향 방식은 현실적으로 널리 적용되고 있다. 기하음향의 대표적 방식은 가상음원(image-source) 방식과 레이 트레이싱(ray-tracing) 방식을 들 수 있으나 구체적인 내용은 본 논문의 주제와 벗어나 생략하기로 한다.

이와 같은 GA 기반 모델링 방법이 적용된 솔류션으로는 Unity를 위한 SDK 형태로 배포된 Google의 공명 오디오

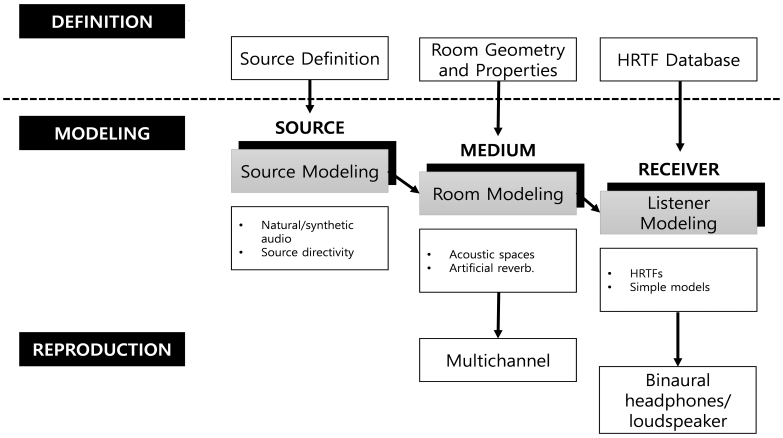


그림 1. 가상 음향환경 구현 절차[1]
Fig. 1. Process of implementing virtual acoustic environments[1]

(Resonance Audio)[4] 기술을 들 수 있고, 충격응답 생성 소프트웨어로는 EVERTims[5]와 RAVEN[6] 등이 대표적이라 할 수 있다.

VR, AR 또는 더 나아가 메타버스 서비스 등과 같은 가상환경에서는 일반적으로 청취자가 HMD (Head Mounted Display)를 머리에 착용한 상태에서 가상환경을 체험하게 된다. 이때 청취자는 머리를 자유롭게 움직일 수 있어야 하며, 가상공간에서 제한없이 이동(병진운동)할 수 있어야 청취자가 몰입감 있게 가상환경을 체험할 수 있다. 여기서, 머리 움직임(pitch, yaw, roll)을 3 자유도(Degree of Freedom; DoF, 3DoF)라고 하며, 머리 움직임에 공간상의 이동(병진운동)까지 포함하는 것을 6 자유도(6DoF)라고 한다.

한편, 국제표준화 기구인 ISO/IEC JTC1 산하의 SC29/WG6 (MPEG Audio Coding)에서는 '21년 4월 제3차 WG6 회의에서 MPEG-I Immersive Audio (ISO/IEC 23090-4) 표준화를 위한 기술제안 요구서(Call for Proposals; CfP)[7]를 발간하고, '21년 11월 각 표준화 참여기관으로부터 기술제안을 받아 기술선정을 위한 주관평가를 수행하였다. 이를 토대로 '22년 1월 제6차 WG6 회의에서 참조모델(Reference Model 0, RM0)을 선정, '22년 10월 제9차 WG6 회의에서 작업문서(Working Draft; WD)를 확정하였으며 12월에 이에 대응하는 RM 소프트웨어를 공개하였으며, 현재 핵심실험(Core Experiment; CE)이 진행 중에 있다. MPEG-I Immersive Audio 표준화는 청취자의 자유로운 이동과 인터랙션이 가능한 6DoF 공간음향 서비스를 위한 몰입형 실감음향 렌더링 기술 및 렌더링에 필요한 메타데이터 부호화에 대한 표준화를 목표로 하고 있다.

본 논문에서는 가상환경에 위치한 청취자의 움직임에 대응하여 실시간으로 6DoF 공간음향을 제공하는 공간음향 렌더링 시스템에 대해 소개한다. 이 시스템은 MPEG-I Immersive Audio CfP에 대응하는 과정에서 개발된 시스템으로서 MPEG-I AEP (Audio Evaluation Platform)[8]에서 동작한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서 본 시스템의 동작 플랫폼인 AEP의 주요 특성과 AEP 환경에서 개발된 실시간 6DoF 공간음향 렌더링 시스템의 구조를 기술한다. 3장에서 본 연구의 주요 음장 처리 기술의 특성에 대하여, 4장에서 개발된 시스템에 대한 성능평가 결과로서 자체 주관평가 결과를 각각 기술한다. 끝으로 5장에서 결론과 향후 과제에 대하여 기술한다.

II. 6DoF 공간음향 렌더링 시스템의 구조

공간음향은 임의공간의 기하학적인 구조에 의해 청취자

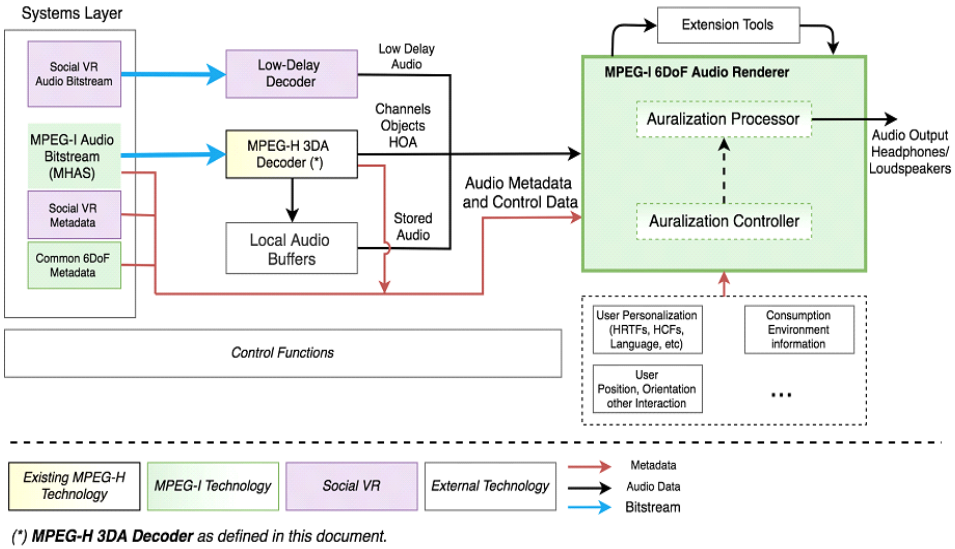


그림 2. MPEG-I Immersive Audio 기본 구조[9]
Fig. 2. MPEG-I Immersive Audio Reference Architecture[9]

의 위치에 대응하여 변화하는 다양한 음향효과를 실제와 유사하게 제공하는 음향처리 기술로서, 기존의 멀티채널 음향으로는 실현이 쉽지 않으며, 객체기반 음향을 통하여 각 객체 음원의 위치, 형태, 움직임에 따라 변화하는 음향효과를 생성하여 실시간으로 제공함으로써 시각적인 정보와 일치하는 음향 이벤트를 느끼도록 하여 실감을 제공하는 기술이다.

전술한 바와 같이 SC29/WG6에서는 몰입형 미디어 서비스의 확산에 대응하여 몰입형 미디어 표준기술의 개발을 위하여 MPEG-I Immersive Audio 표준화를 진행하고 있다. MPEG-I Immersive Audio 구조 및 요구사항 문서[9]에서 명시한 MPEG-I Immersive Audio 기본 구조는 그림 2와 같고, 녹색으로 색칠된 영역이 표준화 대상이다.

SC29/WG6에서는 MPEG-I Immersive Audio 표준화에 있어 각 기관에서 제안하는 기술(인코더와 렌더러로 구성)의 주관평가를 위하여 오디오 평가 플랫폼인 MPEG-I AEP를 개발하였으며, 그 구조는 그림 3과 같다.

그림에서 황색으로 채워진 영역인 인코더(Encoder)와 렌더러(Renderer)가 각 기관에서 제안하는 기술 영역에 해당한다. 인코더는 인코더 입력 포맷(Encoder Input Format; EIF)[10]으로 표현되는 6DoF 공간음향 렌더링을 위한 메타데이터 등을 오프라인으로 부호화하는 역할을 하며, 렌더러는 전달된 비트스트림을 수신하여 실시간으로 6DoF 공간음향 렌더링을 수행한다. 좀 더 구체적으로 설명하면, AEP는 가상현실 장비인 HTC Vive의 HMD를 기반으로 시각 정보 렌더링을 위한 Unity 도구와, 6DoF 몰입형 실감음향 렌더링을 위한 Max/MSP 개발환경을 통합하여 구성되어 있으며, 실감음향 렌더러는 AEP 환경에서 비트스트림과 함께 실시간으로 사용자의 위치, 업데이트 이벤트를 입력받아 인터랙션이 포함된 실감음향을 생성하여 재생하여야 한다. 실감음향 렌더링을 Max/MSP를 사용하는 대신 Unity 만을 사용하는 것으로 구현할 수 있으나, AEP의 개발 목적이 다수 개의 인코더 및 렌더러 조합에 대해 주관평가를 통한 평가이기 때문에 실감음향 렌더링을 위해 별도의 Max/MSP 개발환경을 사용하게 된 것이다. 따라서, AEP에서는 Unity와 Max 사이의 실시간 6DoF 공간음향 렌더링을 위한 최소한의 메시지 교환을 위해 OSC (Open Sound Control)를 사용하며 비디오와 오디오 신호와의 동기를 보장하고 있다. 사용자의 실시간 위치 정보 및 가상현실 공간의 변동사항을 업데이트 이벤트로 실감음향 렌더러에 전달하기 위한 Max patch를 개발하여 제공하며, 주관평가를 위한 GUI (Graphic User Interface)를 제공한다.

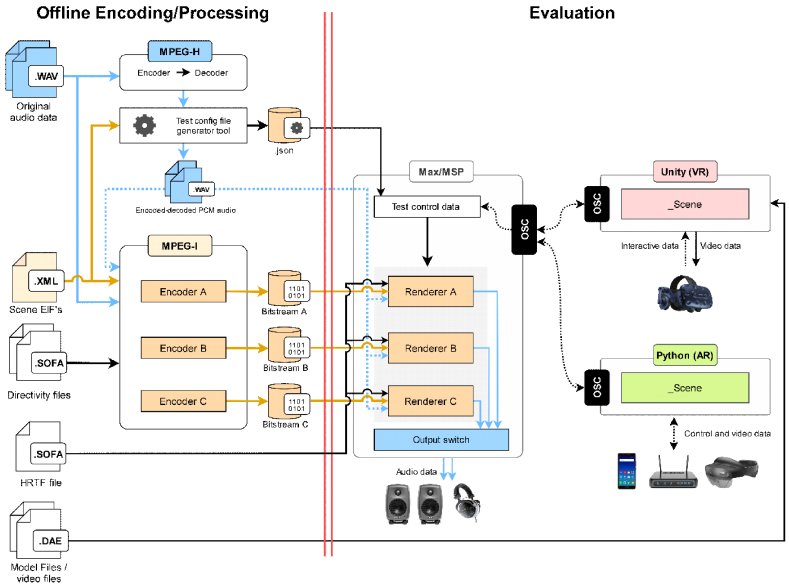


그림 3. MPEG-I AEP 개요 및 처리과정[8]
Fig. 3. Overview of MPEG-I Audio Evaluation Platform and Process[8]

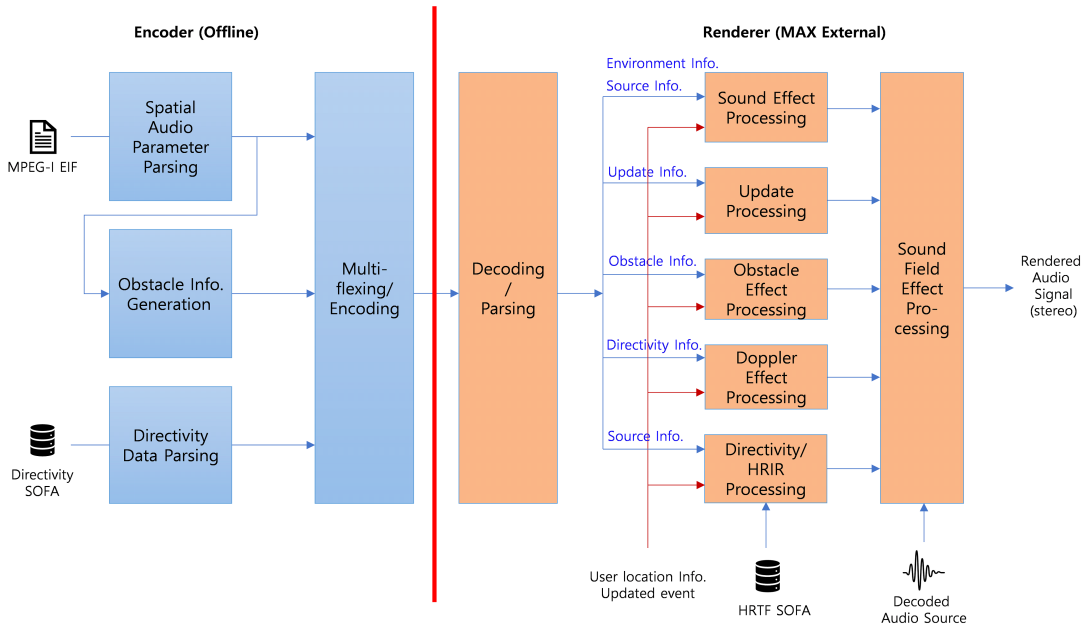


그림 4. 6DoF 공간음향 시스템 기능 구조
Fig. 4. Functional block-diagram of a 6 DoF spatial audio system

그림 4는 본 연구에서 제안하는 6DoF 공간음향 렌더링 시스템의 기능 구조를 나타내는데, VR 서비스만 고려하고 있고 MPEG-I AEP 기반으로 구현되었기 때문에, 그림 3에서와 마찬가지로 오프라인 인코더(좌)와 렌더러(우)의 두 부분으로 구성된다.

인코더는 EIF 파일에 포함된 각 테스트 장면(scene)의 공간적 오디오 파라미터와, SOFA (Spatially Oriented Format for Acoustics) 파일로 제공되는 음원의 지향성 정보를 파싱하여 비트스트림을 생성한다. 특히 인코더는 EIF에 포함된 기하 정보인 면(face)들 중에서 렌더링 시 장애물이 될 수 있는 면을 추출하여 비트스트림에 포함시키고 있는데, 이는 렌더러가 현재 음원과 청취자 사이에 장애물이 있는지를 판단할 때, 추출된 면에 대해서만 장애물 여부를 판단할 수 있도록 하여, 렌더러의 연산량을 감소시킬 수 있다.

렌더러는 MPEG-I AEP에 따라 Max 패치와 연동되는 Max External로 구현되어야 하며, Max 패치에서 전송된 디

표 1. 렌더러 기능 블록의 주요 역할
Table 1. The main role of each function block in renderer

Name	Main role
sound effect processing	Operations such as RT60, HOA, and extent sound source processing are performed
update processing	L1, L2, and L3 update operations are performed.
obstacle effect processing	Coordinates of the Mesh or Transform object are updated for the time period of the corresponding frame in case of a moving obstacle. Then, it searches for obstacles among the transmitted obstacle candidates for the straight path between the sound source position and user position of the corresponding frame, and applies the transmission and diffraction loss of the face detected as an obstacle.
Doppler effect processing	The Doppler effect is implemented by pitch shifting method using waveform on the time axis.
directivity processing	It calculates the angle between the sound source object and the listener, finds the most similar directivity information to the calculated angle from the given directivity information and processes it.
HRIR processing	The relative coordinates between the listener and the sound source are calculated, and the HRIR data of the appropriate location is read from the SOFA file and processed.
sound field effect processing	Audio signal processing such as distance attenuation and reverberation operations are performed.

코딩된 음향신호를 입력받아 청취자의 현재 위치와 가상현실 공간의 다양한 형태 및 변화에 따른 음향 효과를 실시간으로 생성한다. 렌더러는 3차원 음향 재생을 위한 음장효과 및 바이노럴 처리를 기반으로 자유롭게 움직이는 청취자에게 장애물에 의한 음향 효과, 지향성과 확장(extent) 음원의 음향 효과, 도플러 효과 등을 제공한다. 렌더러의 개발 환경은 다음과 같다.

- OS: Windows 10
- Unity: V.2019.2.19f1
- Max: Max 8 (64bit)

그림 4의 각 기능 블록의 기능은 표 1에 요약 정리하였으며, 지향성 및 HRIR 처리블록을 제외한 나머지 블록에 대한 자세한 특성에 대해서는 아래 3장에서 기술한다.

Ⅲ. 6DoF 공간음향 처리 기술 특성

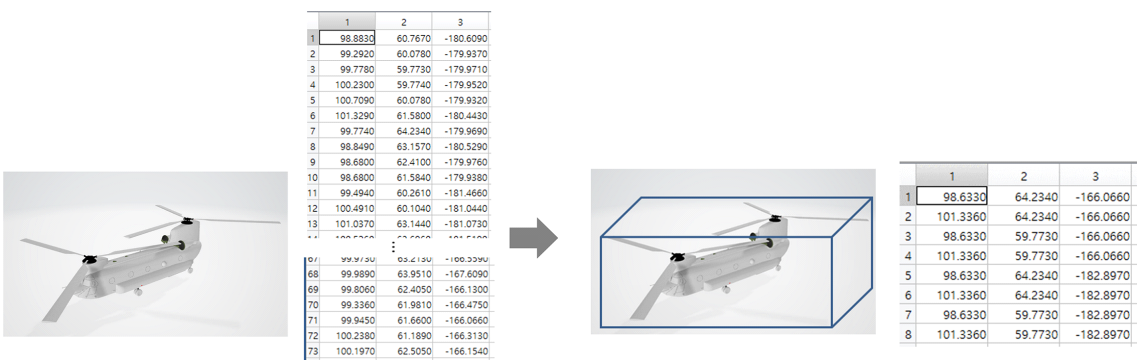
본 장에서는 앞장에서 기술한 6DoF 공간음향 렌더링 시스템에 적용한 공간음향 처리 기술에 대해 기술한다. 이 기술들은 MPEG-I Immersive Audio CfP에 대응하기 위하여 구현되었으며, 이중 주요 공간음향 처리 기술로는 음원 효과 및 장애물 효과 처리 기술 등이 있고 그 외 시스템 동작에 필요한 기술로는 도플러 효과 및 음장효과 처리 기술 등이 있다. 이들을 그림 4의 구조 순서대로 그 기술적 특성에 대해 간략하게 기술한다.

1. 음원 효과 기술

6DoF 공간음향 서비스에서 실제와 같은 음향을 재생하기 위해서 다양한 음향 신호 처리 방식이 사용되는데, 일반적인 음원을 재생하는 방식인 점음원 처리를 비롯해서, 확장 음원으로 구분되는 선 음원, 면 음원, 부피 음원 등 다양한 음원의 형태가 존재한다. 가령, 높은 곳에서 떨어지는 폭포 소리나, 긴 해변가의 파도소리, 징이 울리는 소리 등은 하나의 점으로 표현하기보다는 확장 음원의 형태로 표현될 때 더욱 더 청취자에게 사실적인 음향 효과를 제공해 줄 수 있게 된다.

확장 음원 효과 처리를 위한 첫 번째 단계는 메쉬(mesh) 영역의 단순화로, 복잡하게 주어지는 메쉬 정보를 8개의 좌표를 가지는 직육면체로 단순화한다. 즉, 콘텐츠 제작 시에 작게는 몇십 개에서 많게는 몇천 개에 이르는 메쉬 좌표 정보를 그대로 사용하지 않고 각 공간 좌표들의 최대, 최소값을 이용하여 직육면체로 단순화하면 구조 정보를 해석하기 위한 연산량에서 많은 이점을 갖게 된다. 가령 그림 5와 같이 총 73개의 메쉬 좌표로 이루어진 확장 음원 객체가 주어진다고 가정하면, 전체 좌표 정보를 스캔하여 최대값과 최소값을 추출하고 그로부터 8개의 직육면체 꼭지점 좌표를 생성한다.

콘텐츠 제작 시 확장 음원은 모노 혹은 멀티채널로 제공될 수 있는데, 확장 음원 처리를 하기 위해 사용자의 위치에 따라 단순화한 확장 음원 영역 내부에서 음원이 사용자를 따라서 이동하는 방식을 공통적으로 적용한다. 모노 확장 음원의 경우 그림 6에서처럼 오른쪽에서 왼쪽으로 이동하



	1	2	3
1	98.8830	60.7670	-180.6090
2	99.2920	60.0780	-179.9370
3	99.7780	59.7730	-179.9710
4	100.2300	59.7740	-179.9520
5	100.7090	60.0780	-179.9320
6	101.3290	61.5800	-180.4430
7	99.7740	64.2340	-179.9690
8	98.8490	63.1570	-180.5290
9	98.6800	62.4100	-179.9760
10	98.6800	61.5840	-179.9380
11	99.4940	60.2610	-181.4660
12	100.4910	60.1040	-181.0440
13	101.0370	63.1440	-181.0730
⋮			
68	99.9890	63.9510	-167.6090
69	99.8060	62.4050	-166.1300
70	99.3360	61.9810	-166.4750
71	99.9450	61.6600	-166.0660
72	100.2380	61.1890	-166.3130
73	100.1970	62.5050	-166.1540

	1	2	3
1	98.6330	64.2340	-166.0660
2	101.3360	64.2340	-166.0660
3	98.6330	59.7730	-166.0660
4	101.3360	59.7730	-166.0660
5	98.6330	64.2340	-182.8970
6	101.3360	64.2340	-182.8970
7	98.6330	59.7730	-182.8970
8	101.3360	59.7730	-182.8970

그림 5. 메쉬 영역 단순화의 예(왼쪽: 주어진 메쉬 데이터, 오른쪽: 직육면체로 단순화한 확장 음원 영역)
Fig. 5. Example of mesh area simplification (left: given mesh data, right: simplified extent source area as a cuboid)

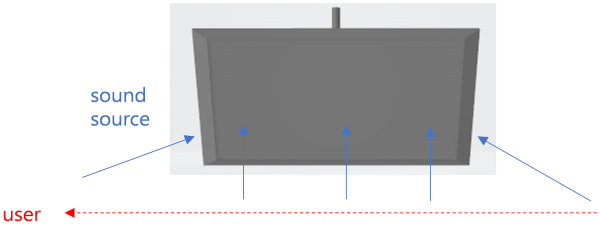


그림 6. 모노 확장 음원에 대한 음원이 사용자를 따라 움직이는 처리 방식
Fig. 6. A processing method in which the sound source for a mono extent sound source moves along the user

는 사용자의 움직임에 따라서 음원의 위치 또한 확장 음원 영역의 오른쪽 가장자리에서 내부를 거쳐 왼쪽 가장자리로 이동시켜 준다. 즉, 사용자가 확장 음원 영역 외부에 있을 때 및 확장 음원 영역 내부와 평행 이동하는 영역에 있을 때에는 사용자와 가장 가까운 확장 음원 영역 가장자리로 확장 음원을 위치시킨다. 그러나 사용자가 확장 음원 영역 내부에 있을 때에는 사용자의 위치와 확장 음원의 위치를 동일시하는 방식을 적용하여 사용자가 어느 위치로 이동하더라도 확장 음원은 해당 영역 내에서 실시간으로 이동하여 청취자로 하여금 넓은 범위의 음원으로 인식할 수 있도록 음원 효과를 처리한다.

멀티채널 확장 음원은 확장 영역 내에 음원의 개수가 두 개 이상 존재하는 경우로, 그림 7과 같이 평면을 기준으로 상, 중간, 하의 3개 층으로 구분하여 Top Let(TL), Top(T), Top Right(TR), Left(L), Center(C), Right(R), Bottom Left(BL), Bottom(B), Bottom Right(BR) 9개의 음원이 배치될 수 있으며, 음원 객체(OBJECT)를 기준으로 처리하느냐 또는 사용자(USER)를 기준으로 처리하느냐에 따라서 두 가지 경우로 나누어 처리될 수 있다.

멀티채널 OBJECT 확장 음원을 처리하기 위해서는, 직육면체로 단순화된 확장 영역을 9x3개의 sub-box 영역을 만들고(cube), 재생 채널에 해당하는 sub-box 내에서 사용자의 위치에 따라 사용자를 따라가며 각각의 음원을 이동시킨다. 그리고 멀티채널 USER 확장 음원을 처리하기 위해서는, User(사용하는 HMD)의 시야(Field of View; FOV)에 들어오는 확장 영역 안으로 위에서 설명한 9개의 멀티채널 음상을 정위한다. 즉, 사용자의 위치에 따라 단순화된 확장 영역 내에서, HMD의 FOV에 기반하여 사용자를 따라가며 음원 각각을 이동시키며 사용자가 확장 영역 내부로 들어가게 되면, 모노 확장 음원 영역으로 작용하여 해당 영역 내에서 모노 확장 음원 효과와 같이 처리한다.

그리고 6DoF 공간음향 서비스에서 사용되는 오디오 신호의 종류에는 채널 음원 신호, 객체 음원 신호 및 HOA 음원 신호가 있다. 채널 음원은 전통적인 모노와 스테레오에서부터 5.1채널, 10.2채널, 22.2채널 등 스피커의 수와 재생 채널의 수가 일치하는 포맷이며, 객체 음원은 가수 목소리, 피아노 소리 등과 같이 특정 사운드에 대한 오디오 신호가 개별적으로 존재하는 오디오 신호이다. 한편, High Order Ambisonic (HOA) 음원은 구형 고조파(spherical harmonics)로부터 얻어지는 'W, X, Y, Z, ...' 등과 같은 많은 채널로 구성되는 오디오 신호를 말한다. VR 응용 환경에서는 이 3가지의 오디오 신호가 모두 활용되는 경우가 있는데, 특히 HOA 음원은 전체적인 콘텐츠 장면(scene)에 대한 음장을 가장 손쉽게 제작하고 재생할 수 있다는 점에서 장면기반(scene-based) 오디오 렌더링 방식으로 많이 사

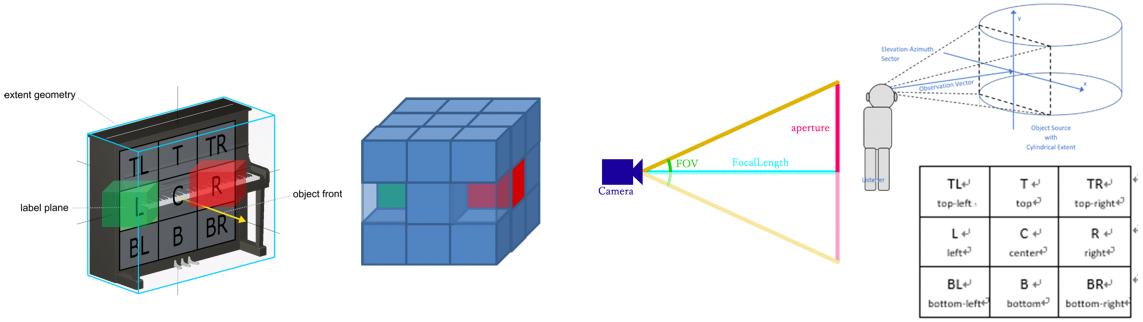


그림 7. 멀티채널 OBJECT 확장 음원에 대한 처리(좌), USER 확장 음원에 대한 처리(우)
Fig. 7. Processing for multi-channel OBJECT extent sound source (left), USER extent sound source (right)

용된다. 그러나 HOA 음원은 정밀한 신호를 얻기 위해서 차수(order)를 높여야 하며, 차수가 증가할수록 구형 고조파 신호가 증가하게 되어 오디오 신호 처리가 복잡해지는 단점이 존재한다. 따라서 연산량도 적으면서 VR과 같은 실시간 재생 환경에서는 ETSI TS 126 260 V15.0.0 (2018-10)[11] 에서 정의한 ESD (Equivalent Spatial Domain) 표현(representation) 정보를 활용하여, HOA 음원을 반지름이 1인 단위 구면 상에 균일하게 분포된 가상 스피커로 매핑시켜 줄 수 있다. 즉, 그림 8과 같이 ESD 표현 정보를 이용하면 HOA 음원을 별도의 가공 없이 공간 상에 매핑해 줄 수 있으며, 바이노럴 재생 방식에서는 각 음상 위치에 대해 HRTF 렌더링만 해주면 HOA 음원에 대해 효과적으로 음원을 처리할 수 있다.

특히, HOA 음원은 많은 채널 수로 인해 공간 상의 배경음으로서 재생되는 경우가 많으며 이 음향은 청취자의 위치 및 지향방향과 무관하게 항상 일정하게 재생되어야 하는데, VR 시스템에서 실시간으로 변화하는 청취자 위치 및 지향방향 정보에 무관하게 ESD 기반의 HOA 채널별 음상 정위 정보를, 항상 청취자 정면을 0도로 하여 거리 1미터의 상대적인 음상 정위 정보로서 적용하면 청취자에게 언제나 일정한 HOA 음원의 음장 재생 효과를 제공할 수 있다.

또한, 수신된 EIF의 정보에 따라 VR 청취 공간에 대한 잔향 정보인 RT60 값에 따라서 사용자에게 적합한 잔향 음원 효과를 처리한다. 이때, 청취 공간에 대한 메쉬 정보가 복잡하고 여러 개가 제공된다면, 전술한 바와 같이 메쉬 영역의 단순화를 통해 효과적으로 RT60에 기반한 잔향 효과를 청취할 수 있어 사용자는 어느 공간에서든 사실적인 음원 청취 효과를 얻을 수 있다.

2. 업데이트 처리 기술

MPEG-I Immersive Audio EIF 문서에서는 L1, L2, L3와 같이 세 종류의 업데이트를 정의하고 있는데 그 개념은 다음과 같다.

표 2. 업데이트의 구분
Table 2. Classification of updates

Updates	Definition
L1	A timed update (L1) is executed by the renderer once at a fixed, predefined time.
L2	A conditional update (L2) is initiated by the renderer when a certain condition is met.
	A triggered update (L2) is manually triggered from an external entity (using an OSC message).
L3	A dynamic update (L3) is an update from an external entity that includes the values of the attributes to be updated.

이를 위하여, L1 업데이트에서는 EIF 파일의 특정 시간별 위치, 방향 정보로부터 음원의 움직임 정보를 획득 및 분석하고 이로부터 프레임별 실시간 좌표를 계산하여 음상

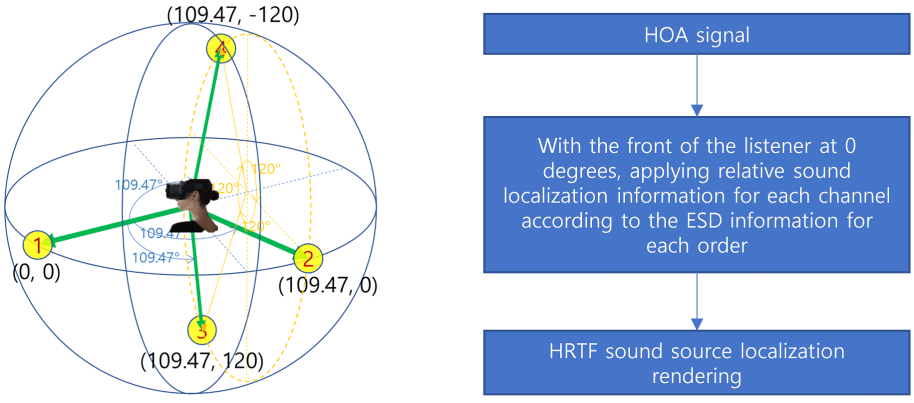


그림 8. Ambisonic 음원의 ESD 정보 기반 음상 배치 예(왼쪽: 단위 구 상의 ESD 음원 배치 정보, 오른쪽: 신호처리 과정)
Fig. 8. Example of sound source localization based on ESD information of Ambisonic sound source (left: ESD sound source information in unit sphere, right: signal processing step)

을 정위하며, L2 업데이트에서는 EIF 정보를 기반으로 해당 오디오 신호의 온/오프 재생을 적용한다. 그리고, L3 업데이트에서는 수신된 위치, 방향 및 이득(dB) 정보에 따라 현재 프레임의 각 객체에 적용한다.

3. 장애물 효과 처리 기술

3.1 장애물 탐색

공간음향에 있어 장애물 효과는 음원과 공간구조와 청취자 위치에 대하여 종속적이며, 이동하는 청취자 환경에서는 실시간으로 장애물 효과처리를 수행하여야 한다. 장애물 여부를 판단하기 위해서는 공간구조를 구성하는 개별 메쉬에 대하여 임의 음원 및 청취자 위치 사이의 직선 경로가 메쉬를 관통하는지 여부를 판별하는 기하학 알고리즘을 적용할 수 있다. 아래 그림 9와 같이 공간의 구조물을 구성하는 하나의 메쉬에 대하여 음원 P와 청취자 Q를 잇는 직선경로에 대해 메쉬의 내부에 관통하는 점 R이 있으면 장애물이라고 판단할 수 있다. 이때 메쉬는 앞과 뒤가 구분될 수 있는 평면이며, 벡터값의 방향에 따라 노멀 성분의 방향을 앞면으로 정의하고 앞면으로 입사하는 빔은 메쉬에 의해 영향을 받는 것으로 간주할 수 있다. 이러한 장애물 판단 알고리즘은 기하학의 스칼라삼중곱(Scalar Triple Product)에 의해 메쉬의 한 모서리와 P-Q 직선에 의해 만들어지는 평형육면체의 부피를 계산함으로써 판단될 수 있으며, 부피가 음수인 경우 장애물 메쉬에 관통하지 않으므로 장애물이 아닌 것으로 판단할 수 있다. 그러나 P-Q의 연장선이

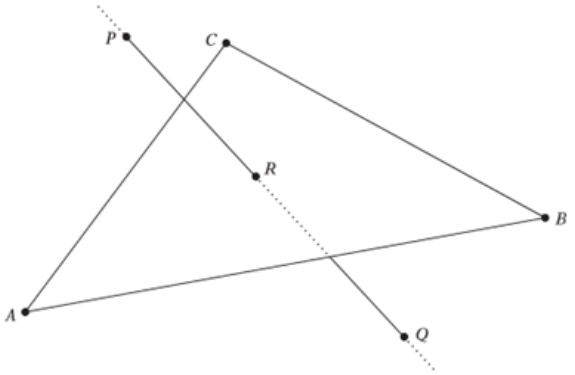


그림 9. 장애물 검출을 위한 기하학적 개념
Fig. 9. Geometrical concept of occlusion detection

장애물 메쉬에 관통할 때에도 부피가 양수로 계산되는 경우가 있으므로 P-R 직선의 길이 및 R-Q 직선의 길이와 P-Q 직선의 길이를 비교함으로써 장애물 여부를 확정할 수 있다.

장애물 여부가 결정되면 장애물에 의해 발생하는 투과손실 및 회절손실을 아래 그림 10의 개념으로 고려할 수 있는데, 투과손실의 경우에는 콘텐츠 저작단계에서 주어지는 음향재료의 특성값에 포함되어 있으므로 적용하면 되지만, 회절의 경우에는 장애물의 분석에 의한 회절경로를 추가적으로 연산하여야 한다.

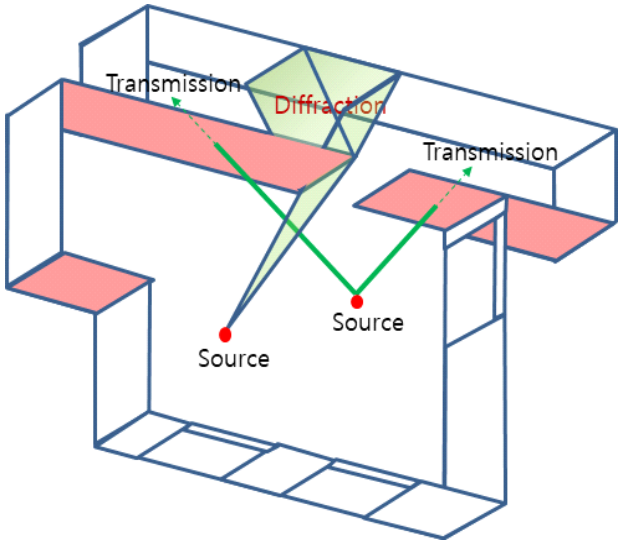


그림 10. 공간에서의 장애물 처리 개념
Fig. 10. Concept of occlusion processing for given space

장애물 효과는 공간을 구성하는 모든 메쉬에 대하여 장애물 여부를 판단하고 프레임 단위로 투과손실 및 회절경로, 회절손실에 대한 연산을 수행하여야 하는데, 가상현실을 구성하는 메쉬의 개수는 대부분 매우 많기 때문에 이를 실시간으로 판별하고 장애물 효과처리를 수행하는 데는 어려움이 있다.

본 연구에서는 장애물이 될 수 있는 장애물 후보 메쉬를 인코더에서 먼저 선별하고 장애물 효과처리에 필요한 정보를 미리 생성하여 비트스트림으로 전송한 후 렌더러에서 이를 이용하여 장애물 여부 판단 및 장애물 효과처리를 고속으로 수행할 수 있도록 하였다. 인코더에서 수행하는 장애물 후보 선별작업은 콘텐츠에 포함되어 있는 모든 객체

신호와 청취자의 이동 범위 정보를 기반으로 장애물 판단을 수행하여 장애물이 될 수 있는 모든 메쉬를 장애물 후보로 등록하고 장애물 후보의 정보를 추출하여 비트스트림에 추가하게 된다. 이때 비트스트림에 포함되는 장애물 후보 메쉬의 부가정보의 종류는 다음과 같다.

- 장애물 후보 메쉬의 Vertex 좌표 정보
- 장애물 후보 메쉬의 투과율 및 주파수 대역 정보
- 장애물 후보 메쉬의 업데이트 여부 및 메쉬 Index 정보

장애물 후보 메쉬 선별의 인코딩 과정 수행에 이점은 다음과 같다.

- 장애물 판단이 필요한 메쉬의 수 감소: 약 1/10 감소
- 메쉬의 방향정보에 의한 회전 연산을 인코더에서 미리 수행함으로써 렌더러 연산 감소

3.2 장애물 효과 처리

렌더러의 장애물 효과처리는 비트스트림을 디코딩 및 파싱하여 얻은 장애물 후보 메쉬 정보를 이용하여 현재의 개별 음원 및 청취자 위치값에 의해 생성되는 직선경로에 의해 장애물 후보 메쉬가 관통되는지 여부를 판단하는 것으로 시작한다. 장애물 후보 메쉬 중 음원 청취자 직선경로에 의해 관통된 메쉬들은 장애물이 되며, 각 장애물의 투과율 및 주파수 대역 정보에 의해 음원 신호를 필터링함으로써 장애물에 의한 투과효과가 생성된다. 음원 청취자 직선경로에 여러개의 장애물 메쉬가 존재하는 경우에는 투과율을 중복 적용하여야 하며, 장애물 메쉬가 업데이트에 의해 이동하는 물체인 경우에는 장애물 후보 메쉬의 Vertex 좌표 정보는 메쉬 혹은 그룹객체인 트랜스폼 객체의 좌표 정보에 대해 상대적인 좌표를 사용하여 업데이트된 메쉬의 위치에 대하여 장애물 효과처리를 수행한다. 장애물 효과처리의 개념 및 구조는 아래 그림 11과 같다.

장애물 효과를 구현함에 있어 투과율만 적용하는 경우에는 장애물의 경계에서 음향신호가 갑자기 단절되는 현상이 발생할 수 있으며, 이는 매우 부자연스럽게 들리게 된다. 음원과 청취자 사이에 장애물이 존재하는 경우, 장애물에 의한 투과 현상과 함께 고려하여야 하는 것이 장애물에 의한 회절 현상이다. 회절 현상에 의한 손실은 음원과 청취자 사이의 직선경로와 회절경로의 거리 비율을 활용하여 계산하는 것이 일반적인데, 본 연구에서는 복잡도의 문제로 이러한 회절경로를 산출하지 않고, 장애물의 개수, 즉, 회절점의 개수에 비례하여 회절손실을 증가시키는 방법을 사용하여 회절 현상을 구현하였다.

4. 도플러 효과 처리

도플러 효과는 움직이는 음원과 청취자간의 거리 변화에 따라 소리의 음색이 변화되어 들리는 현상으로, 움직이는 속도가 빠르면 소리의 변화는 더욱 심해진다. 속도에 따른 소리의 주파수의 변화는 아래와 같은 수식으로 간단하게 나타낼 수 있다. 이때, c는 소리의 속도이고, v_0 및 v_s는 각각 청취자 및 음원의 속도이다.

$$f' = \frac{c \pm v_0}{c \mp v_s} f \tag{1}$$

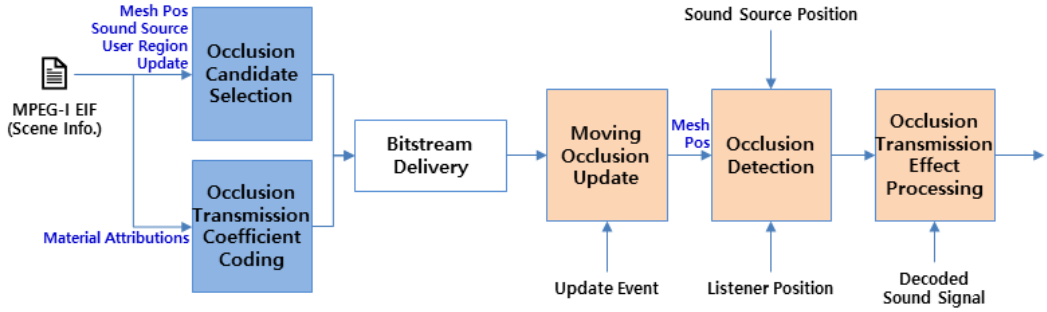


그림 11. 음향학적 장애물 효과 처리 구조
Fig. 11. Block-diagram of acoustic occlusion effect process

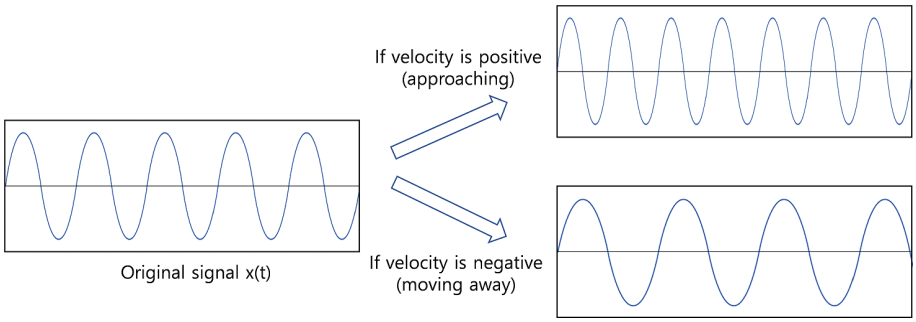


그림 12. 시간축 상의 파형을 활용한 피치 이동 방법의 예
Fig. 12. Example of pitch shifting method using waveform on the time axis

이러한 속도에 의한 주파수의 변화는 피치 이동(pitch shifting) 방법을 이용하여 시뮬레이션할 수 있는데, 이 방법은 주로 오디오 신호의 시간 축상 파형을 활용하여 이루어진다. 본 연구에서도 기존에 활용되는 시간 축상의 파형을 활용하여 도플러 효과를 구현하였다.

5. 음장 효과 처리 기술

소리가 음원에서 사용자에게 전달되는 경로는 다양하다. 음원에서 사용자에게 직접 전달되는 직접음과 벽, 바닥, 천장 등에 의해 반사되어 전달되는 정반사음 등이 대표적이며, 확산음과 잔향도 이러한 주요 전달 경로에 해당한다. 아래에서 각 전달 경로에 따른 처리에 대해서 간략하게 설명한다.

5.1 직접음의 처리

직접음의 경우 음원에서 청취자에게 직접 전달되는 소리이며, 직접음의 처리를 위해서는 음원과 청취자 간의 거리 정보 및 각도 정보가 활용된다. 음원과 청취자 간의 거리 정보는 음원과 청취자의 3차원 공간상의 정보가 활용되며, 아래의 식으로 계산이 된다.

$$dist_{cur} = \sqrt{(x_s - x_l)^2 + (y_s - y_l)^2 + (z_s - z_l)^2} \quad (2)$$

여기서, $dist_{cur}$는 청취자와 음원과의 거리, x_s, y_s, z_s는 음원의 3차원 공간 좌표, x_l, y_l, z_l은 청취자의 3차원 공간 좌표를 나타낸다.

거리 정보는 음원의 소리 크기를 계산하는 데 사용된다. MPEG-I Immersive Audio에서는 오디오 객체의 소리 크기가 0dB가 되는 기준 거리(reference distance)를 정보로 제공하는데, 이를 활용하여 특정 거리에서의 음원의 소리 크기는 아래와 같이 계산된다.

$$gain[\mathrm{dB}] = 20 \times \log\left(\frac{dist_{ref}}{dist_{cur}}\right) \quad (3)$$

여기서, $dist_{ref}$는 기준 거리를 나타낸다.

음원과 청취자 간의 각도 정보는 수평 각도 정보와 수직 각도 정보로 나누어서 계산되는데, 이 또한, 음원과 청취자의 3차원 공간 좌표를 활용하여 계산한다. 음원과 청취자 간의 각도 정보는 머리전달함수를 가져오는 데 활용되며, 해당 각도의 머리전달함수가 원 신호와 컨볼루션 된다.

5.2 초기 정반사음 및 확산음 처리

초기 정반사음(specular reflected sound)의 경우 반사면에 따라 다르게 설정이 된다. 공간의 모양이 다르고, 반사면의 위치가 다르면 다른 정반사음이 생성되는데, 이는 주어지는 3차원 공간의 기하정보를 활용하여 image source

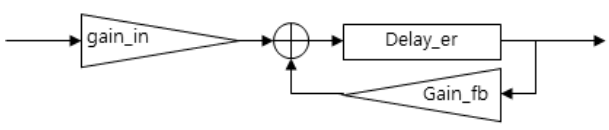


그림 13. 정반사음 처리를 위한 콤 필터
Fig. 13. Comb filter for specular reflections

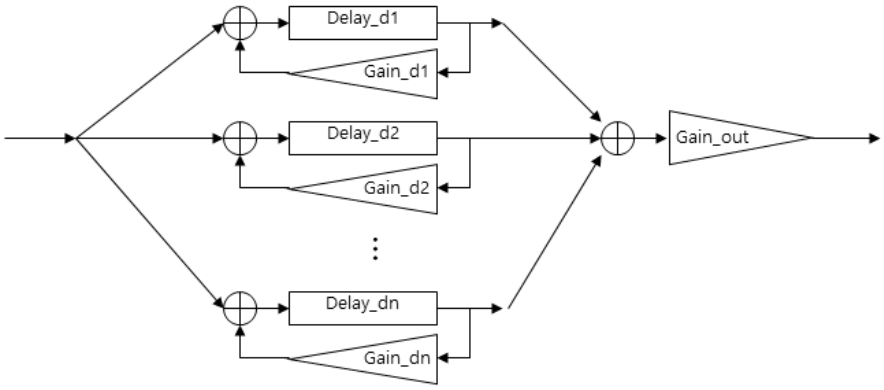


그림 14. 확산음 처리를 위한 콤 필터
Fig. 14. Comb filters for diffuse reflections

method 등을 활용하여 계산할 수 있다. 본 연구에서는 실시간 연산을 고려하여 아래의 그림 13과 같이 간단한 하나의 콤 필터(comb filter)로 이를 구현하였다.

확산음(diffuse reflected sound) 또한 직접 반사음과 마찬가지로 3차원 공간의 기하정보에 의해 달라질 수 있는데, 본 연구에서는 실시간 연산을 위하여 아래 그림 14와 같이 콤 필터를 여러 개 사용하는 방법으로 이를 구현하였다.

5.3 잔향 처리

잔향음은 반사와 확산으로 인해 발생하는 음향 전달 경로이나, 반사된 회수가 많고 전달 경로가 상대적으로 길어 직접음 대비 소리의 크기도 크지 않다. 잔향 또한 3차원 공간의 기하 정보에 의해 달라질 수 있는데, 직접 반사음 대비 연산량이 아주 크게 증가하기 때문에, 본 연구에서는 올패스필터(all pass filter)를 활용하여 그림 15와 같이 간략하게 구현하였다.

Ⅳ. 성능평가

1. 주관평가 환경 및 절차

본장에서는 앞에서 기술한 6DoF 공간음향 렌더링 시스템의 성능 확인을 위해 수행한 주관평가 결과를 소개한다. 본 주관평가는 MPEG-I Immersive Audio CfP[7] 대응의 일환으로 수행되었으며, HMD를 착용한 상태에서 장시간에 걸쳐 평가가 이어지기 때문에, 주관평가는 SC29/WG6에서 제시한 테스트 지침 문서[12][13]에 따라 수행되었다. CfP 문

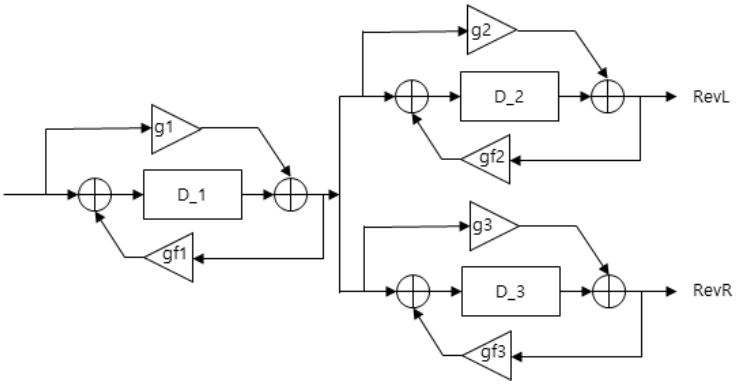


그림 15. 잔향 처리를 위한 올패스 필터
Fig. 15. All pass filters for reverberation

서에서 정의한 Test 1(Virtual Reality)에 대한 A-B 테스트 방법론을 사용하여 13개 테스트 장면(14개 테스트 속성)에 대하여 8명의 평가자가 평가를 수행하였다. 평가를 위한 테스트 환경 셋업은 표 3, 평가 장면 구성은 표 4와 같으며, 이는 MPEG-I Immersive Audio CfP에서 CfP 대응 기술들의 주관적인 비교 평가를 위하여 선정한 조건에 해당한다. 좀 더 부연 설명하면, 다수의 기관에서 제안된 기술을 AEP를 통하여 주관평가 성능을 비교 평가하기 위하여 표 3에서 정의한 평가 환경을 공동으로 사용하고, 표 4에서 정의한 평가 장면(콘텐츠)을 대상으로 각 속성의 표현 정도를 평가하기 위한 것이다. 표 4에서 볼 수 있듯이 평가 장면 중 'Park'에 대해서는 2개의 속성을 평가하여 전체 평가 속성은 14개이다. 평가 대상 기술(렌더러, condition)은 7개 기관에서 복잡도 차이에 따른 2개의 렌더러를 제안하여 총 14개로 구성되어 있으며, 본 연구의 결과물에 해당하는 렌더러는 P25와 P33이다.

20대에서 50대에 걸친 8명의 평가자가 평가에 참여하였으며, 그중 3명은 VR 환경에서의 평가 경험이 없는 평가자

표 3. MPEG-I Immersive Audio CfP에 따른 주관평가 환경 셋업
Table 3. Set-up for subjective test environment defined by the MPEG-I Immersive Audio CfP

Components	Detailed Information
Hardware	• VR system: HTC Vive • Headphone: DT-990 Pro • External Sound Card: Focusrite Scarlet 2i2 (3rd Generation) with ASIO ad_asio Focusrite USB 2.0 driver • VR capable PC (mandatory system): ▪ Graphics Card: NVIDA GeForce RTX 3090 ▪ Processor: Intel(R) Core (TM) i9-10900K, 10 Core(s), 20 Logical Processor(s) ▪ Memory: 64GB RAM ▪ Storage: SSD • Analog Headphone Amplifier: Sony PHA-2
Software	• OS: Windows 10 • Unity: V.2019.2.19f1 • Max: Max 8 (64bit) • AEP: master branch of the "ImmersiveAudio/AudioEvaluationPlatform" project in the MPEG Gitlab • Renderer: 14 renderers from 7 proponents
Test environment	• Test Space: 3x3 m2 • Noise level: 47.3 dB(A) measured by NTi Audio XL2 with calibration using B&K 4231

표 4. MPEG-I Immersive Audio CfP에 따른 평가 장면 및 그 속성
Table 4. Test scenes and their attributes by the MPEG-I Immersive Audio CfP

No	Item (Test scene)	Attribute
1	Park	Size/shape of a sound source
2	SingerInTheLab	Directivity
3	Park	Occlusion (by obstacles; partially translucent & hard)
4	SimpleMaze	Diffraction (around corners ···)
5	Hospital	Coupling/sound propagation between adjacent rooms
6	DowntownBus	Early reflections (from walls, rocks, ···)
7	VirtualBasketball	Late reverberation
8	Cathedral	Transition between different environments
9	Beach	Distance attenuation and far field effects (behavior in a distance)
10	ParkingLot	Doppler
11	Battle	Interaction (M) + L1 + 6DoF localization (N)
12	Canyon	Interaction (M) + L3 + 6DoF localization (N) (grab radio and go to "tent" area)
13	Recreation	Teleport transitions between very different acoustic environments
14	OutsideHOA	3DoF HOA

이고 나머지 5명은 부호화 또는 공간음향에 대한 전문가이다. 평가절차 문서와 가이드라인 문서에 따라 다음과 같은 사항은 유지하면서 사전평가(훈련)와 본평가를 수행하였다.

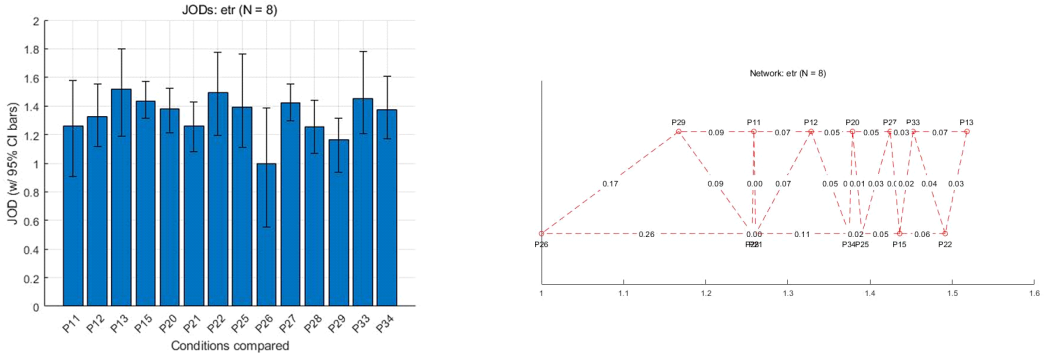
- 장면-타스크별(장면에 따른 주요 속성평가 타스크) 사용자 안내
- 훈련단계: 음향 효과, 시험 환경 및 시험 내용에 대한 숙지
- 테스트 계획 및 평가자에 대한 교육

2. 주관평가 결과

평가 결과 분석은 SC29/WG에서 제공한 Thurstone Case V analysis[14] 툴을 사용하였으며, 이에 따른 막대 그래프와 네트워크 그래프는 아래 그림 16과 같다.

그림에서는 전체 분석 결과를 JOD (Just Objectionable Differences)로 설명하고 있다. 테스트 및 평가절차 문서[15]에 따르면 1 JOD의 차이는 75%의 사람들이 더 낮은 조건보다 더 높은 조건을 선택했음을 의미한다. 막대 그래프의 오차 막대는 95% 신뢰구간(CI)를 나타내고 있다. 인접 비교('conditions'사이의 비교) 사이의 거리만 보여주는 네트워크 그래프에서 수평 스케일은 JOD, 수직 스케일은 임의의 값을 나타내는데, 파란색 선은 유의미한 차이를 나타내고(그림 15의 경우에는 존재하지 않으나) 빨간색 선은 유의미하지 않은 차이를 나타낸다고 할 수 있다.

이번 평가에 참여한 평가자의 평가 소요시간 등의 정보는 다음과 같다. 종합적으로 볼 때, 1명의 평가자가 1개의 평가장면(속성)에 대하여 14개의 렌더러(condition)를 모두 비교하여 A-B 평가를 하기 위해서는 91개의 trials(91개의



(a) (b)

그림 16. 전체 테스트 장면에 대한 A-B 주관평가 결과의 Thurstone 모델 분석: 막대 그래프 (a) 및 네트워크 그래프 (b)
Fig. 16. Thurstone model analysis of A-B test results for all scenes: bar plot (a) and network graph (b)

표 5. 평가자 정보 및 평가 소요시간
Table 5. Subject information and test time

Subject No.	Category	Age	Gender	Time (incl. breaks)		Time (net)	
				Training	Test	Training	Test
1	C. Expert	42	M	1:52	2:46	1:32	1:48
2	B. Expert	55	M	2:27	1:38	2:06	1:14
3	B. Expert	42	M	3:19	2:03	1:41	1:31
4	Beginner	24	F	1:33	1:49	1:17	1:23
5	Beginner	22	F	2:15	1:41	1:48	1:14
6	Beginner	26	M	3:48	1:54	2:09	1:30
7	B. Expert	49	M	2:24	2:14	2:27	1:16
8	C. Expert	35	M	2:07	2:09	1:34	1:40

A-B 쌍)을 평가하여야 한다. 그러나 이는 평가소요 시간, 평가자의 피로, 전체 평가시간 등의 문제로 현실적으로 불가능하다. 따라서, 이번 SC29/WG6에서 설계한 주관평가의 경우에는 한 평가자가 전체 장면을 대상으로 44개의 trials 만을 대상으로 한 결과 8명의 평가만으로 결과를 분석하는 것은 의미가 있다고 보기 어려운 문제가 있어 여기에서는 구체적 분석은 생략한다. 본 연구 결과의 렌더러 중 P33은 3번째 수준에, P25는 6번째 수준에 있는 것을 확인할 수 있으며, 두 렌더러의 기술적 차이는 복잡도에 있다.

V. 고 찰

지금까지 MPEG-I AEP 기반 6DoF 공간음향 렌더링 시스템의 구조, 기술적 특성 등을 살펴보고, 시스템의 성능평가의 일환으로 주관평가 결과를 소개하였다. 오디오 부호화 등과 같이 음질평가가 중요한 요소인 주관평가의 경우에는 전통적으로 MUSHRA와 같은 평가방법을 사용하여, 고품질의 원음, 저품질의 앵커신호에 비교하여 평가 대상 신호를 원음 대비 수준으로 평가한다. 그러나, VR 환경에서 HMD를 착용한 상태에서 비주얼 장면 또는 신호에 동기되거나 그렇지 않다고 하더라도 음향 신호를 평가하는 것은 간단한 일이 아니라고 할 수 있다. 그 중 첫 번째 이유는 MUSHRA 방법과 달리 원음과 같은 기준신호가 없다는 것이고, 주관평가에 참여한 평가자의 수에서의 제한 및 모든 평가장면에 대한 충분히 균등한 평가의 어려움 등으로 인해 전체 비교 대상 기술(렌더러)간에 충분한 비교 평가가 이루어지지 못했다는 제한점 또한 존재한다. 또한 HMD 착용의 불편함과 VR 영상 기반의 인터랙션으로 평가가 이루어짐으로 인한 두통 등 그동안 진행해온 평가 방식과의 차이로 인해 평가자들이 적응하기 어려웠고, 따라서, 제시한 주관평가 결과만으로 본 연구에서 제안한 시스템의 성능을 판단하는 것에는 다소 제약이 따른다.

또한, 본 연구의 시스템은 렌더링시 사용자의 움직임에 따른 6DoF 기반의 상호작용을 사용자에게 실시간으로 제공하여야 하기 때문에, 고품질의 공간음향 제공에 필수적인 주어진 공간에서의 상세한 수준의 음장 모델링을 제공하는 것에 현실적인 제약이 존재할 수 밖에 없다.

이와 같은 본 연구가 가지는 몇 가지의 제약에도 불구하고, 현재 차세대 서비스로 부각되고 있고 여러 표준화 단체에서 표준화의 필요성을 제기하고 있는 메타버스 서비스를 고려할 때, 가상현실, 증강현실 또는 현실의 물리공간과 가상공간이 함께 존재하는 혼합현실 등에서 6DoF 공간음향을 몰입감 있게 제공하기 위한 출발점을 제시한 것에 의미가 있다고 평가할 수 있다.

SC29/WG6에서 추진하고 있는 MPEG-I Immersive Audio 표준화에서는 2022년 10월 회의에서 작업표준(WD)을 제정하였고 2023년 10월 회의에서 위원회 표준(Committee Draft; CD) 및 2024년 초에 국제표준(IS) 제정을 목표로 하고 있다.

끝으로, 대부분의 공간음향 기술이 정방형 형태의 공간을 주로 다루고 있으나, 현실에서 느끼는 공간음향을 보다 실감있게 제공하기 위해서는 이러한 형태의 공간음향 뿐만 아니라 임의 형태의 공간에서도 실감있는 공간음향을 제공하기 위한 연구가 진행되어야 할 것이며, 공간상에서 AV신호 사이에 공간적인 동기 뿐만 아니라 시간적인 동기를 보장할 수 있도록 신호압축에 사용할 코덱의 지연시간 또한 충분히 고려 및 연구가 필요한 분야라고 할 수 있다.

참 고 문 헌 (References)

[1] Lauri Savioja et al., “Creating Interactive Virtual Acoustic Environments,” *Journal of Audio Engineering Society*, Vol.47, No.9, pp.675-705, September 1999.
http://www.aes.org/e-lib/browse.cfm?elib=12095

[2] Lauri Savioja et al., “Overview of geometrical room acoustic modeling techniques,” *Journal of Acoustic Society of America*, Vol.138, No.2, pp.708-730, August 2015.
doi: https://doi.org/10.1121/1.4926438

[3] Room Acoustics Modeling with Interactive Visualizations (by Lauri Savioja), https://interactiveacoustics.info/ (accessed Feb. 9, 2023)

[4] Resonance Audio, https://resonance-audio.github.io/resonance-audio/ (accessed Feb. 9, 2023)

[5] EVERTims, https://evertims.github.io/ (accessed Feb. 9, 2023)

[6] RAVEN, https://www.akustik.rwth-aachen.de/go/id/dwoc/lidx/1/file/183613 (accessed Feb. 9, 2023)
RAVEN: A real-time framework for the auralization of interactive virtual environments

[7] ISO/IEC JTC1/SC29/WG6 N0056, “MPEG-I Immersive Audio Call for proposals,“ Virtual, April 2021.

[8] ISO/IEC JTC1/SC29/WG6 N100, “MPEG-I Immersive Audio Documentation for the Audio Evaluation Platform, Version 2” Virtual, October 2021.
[9] ISO/IEC JTC1/SC29/WG6 N0028, “MPEG-I Immersive Audio Architecture and Requirements,” Virtual, January 2021.
[10] ISO/IEC JTC1/SC29/WG6 N0054, “MPEG-I Immersive Audio Encoder Input Format,” Virtual, April 2021.
[11] ETSI TS 126 260 V15.0.0 (2018-10), “5G; Objective test methodologies for the evaluation of immersive audio systems (3GPP TS 26.260 version 15.0.0 Release 15), 2018
[12] N0101 (N20921), "MPEG-I Immersive Audio Subjective Test Logistics document", 5th WG6 meeting (136th MPEG meeting), October 2021, Virtual.
[13] ISO/IEC JTC1/SC29/WG6 N0105, "MPEG-I Immersive Audio additional instructions for Test Supervisors and Test Subjects," Virtual, October 2021, Virtual.
[14] ISO/IEC JTC1/SC29/WG6 m55106, "Thoughts on evaluation procedure for MPEG-I CfP," Virtual, October 2020..
[15] ISO/IEC JTC1/SC29/WG6 N0084, "MPEG-I Immersive Audio Test and Evaluation Procedures," Virtual, July 2021.

저 자 소 개

강 경 옥

- 1985년 2월 : 부산대학교 물리학과 (이학사)
- 1988년 2월 : 부산대학교 물리학과 (이학석사)
- 2004년 2월 : 한국항공대학교 항공전자공학과 (공학박사)
- 2006년 4월 ~ 12월 : 영국 University of Southampton ISVR 방문연구원
- 2014년 8월 ~ 2015년 7월 : KAIST 문화기술대학원 방문연구원
- 1991년 2월 ~ 현재 : ETRI 미디어연구본부 미디어부호화연구실 책임연구원
- ORCID : https://orcid.org/0000-0002-7673-652X
- 주관심분야 : 실감음향, 음향신호처리, MPEG 오디오 표준화

유 재 현

- 2003년 2월 : 홍익대학교 전자전기공학부 학사
- 2005년 2월 : 서울대학교 전기컴퓨터공학부 석사
- 2022년 8월 : KAIST 전기및전자공학부 박사
- 2005년 2월 ~ 현재 : ETRI 미디어연구본부 미디어부호화연구실 책임연구원
- 2011년 11월 ~ 2012년 11월 : 일본 NHK 방송기술연구소 방문연구원
- ORCID : https://orcid.org/0000-0002-9326-9353
- 주관심분야 : 3D 오디오 및 음장 재생 기술, 오디오 신호 처리

장 대 영

- 1991년 2월 : 부경대학교 전자공학과 (공학사)
- 2000년 2월 : 배재대학교 컴퓨터공학과 (공학석사)
- 2008년 2월 : 배재대학교 컴퓨터공학과 (공학박사)
- 2004년 10월 ~ 2005년 9월 : 일본 동경전기대학/DiMagic Co. Ltd. 방문연구원
- 2019년 4월 ~ 2020년 3월 : 호주 울런공대학교 방문연구원
- 1991년 1월 ~ 현재 : ETRI 미디어연구본부 미디어부호화연구실 책임연구원
- ORCID : https://orcid.org/0000-0003-2572-4374
- 주관심분야 : 실감음향, 객체기반 오디오, 디지털 방송, 대화형 미디어

저 자 소 개

이 용 주

- 1999년 2월 : 경북대학교 전자공학과 (공학사)
- 2001년 2월 : 경북대학교 전자공학과 (공학석사)
- 2001년 2월 ~ 현재 : ETRI 미디어연구본부 미디어부호화연구실 책임연구원
- ORCID : https://orcid.org/0000-0003-0753-1616
- 주관심분야 : 오디오 신호처리, 바이노럴 오디오 렌더링

이 태 진

- 2014년 : 충남대학교 전자전파정보통신공학과 공학박사
- 2002년 ~ 2003년 : 일본 Tokyo Denki University, 방문연구원
- 2000년 ~ 현재 : ETRI 미디어연구본부 본부장, 책임연구원
- ORCID : https://orcid.org/0000-0002-6067-511X
- 주관심분야 : 오디오 부호화, 실감음향, 오디오 신호처리