

# 체감형 미디어 서비스를 위한 공간음향 기술 동향

## Spatial Audio Technologies for Immersive Media Services

이용주 (Y.J. Lee, draball@etri.re.kr)

실감AV연구그룹 책임연구원

유재현 (J. Yoo, jh0079@etri.re.kr)

실감AV연구그룹 선임연구원

장대영 (D. Jang, dyjang@etri.re.kr)

실감AV연구그룹 책임연구원

이미숙 (M. Lee, lms@etri.re.kr)

실감AV연구그룹 책임연구원

이태진 (T. Lee, tjlee@etri.re.kr)

실감AV연구그룹 책임연구원

### ABSTRACT

Although virtual reality technology may not be deemed as having a satisfactory quality for all users, it tends to incite interest because of the expectation that the technology can allow one to experience something that they may never experience in real life. The most important aspect of this indirect experience is the provision of immersive 3D audio and video, which interacts naturally with every action of the user. The immersive audio faithfully reproduces an acoustic scene in a space corresponding to the position and movement of the listener, and this technology is also called spatial audio. In this paper, we briefly introduce the trend of spatial audio technology in view of acquisition, analysis, reproduction, and the concept of MPEG-I audio standard technology, which is being promoted for spatial audio services.

**KEYWORDS** Spatial Audio, Object audio, Higher Order Ambisonics

## 1. 서론

가상현실(VR: Virtual Reality) 기술은 가상체험, VR 콘텐츠, VR 게임, VR/AR(Augmented Reality) 교육 등의 서비스를 통하여 점차 우리 생활에 밀접하게 파고들고 있다. 2018년에 개봉된 영화 <Ready

Player One>은 미래의 가상현실을 배경으로 하고 있는데, 조만간 초고속 통신기술과 가상현실 기술을 통하여 많은 사람들이 가상세계에서 가상의 삶을 경험할 수 있을 것이라는 전망을 가능하게 한다.

가상세계에서 현실과 유사한 경험을 하기 위해서는 오감 정보를 획득/전송/재현하는 체감미디어

\* DOI: 10.22648/ETRI.2019.J.340302

\* 본 연구는 2019년도 ETRI 정부출연금사업(기관고유사업)의 일환으로 수행되었음[18ZR1100, 초실감 공간미디어 원천기술 개발].



본 저작물은 공공누리 제4유형

출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.

©2019 한국전자통신연구원

어 기술이 중요하다. 체감미디어에서는 오감 정보를 모두 다루고 있는데, 그중 시각과 청각 정보를 다루는 공간영상과 공간음향이 현실적으로 비중이 크고 중요하다고 할 수 있다.

본 고에서는 체감미디어 기술 중에서 현실과 유사한 음향을 재현하는 기술인 공간음향 기술의 동향에 대하여 살펴보고자 하며, 이러한 공간음향 기술이 향후 어떠한 진화를 이루어야 할지 전망해 보고자 한다.

## II. 체감미디어와 공간음향

### 1. 체감미디어의 개념

체감미디어는 현실의 오감 정보를 가상의 세계에 옮겨 놓은 미디어를 의미하며, 체감미디어를 통하여 우리는 현실과 유사한 간접경험을 할 수 있게 된다. 기본적으로 시각, 청각, 후각, 미각, 촉각의 정보를 모두 표현하는 것이 필요한데, 각각의 감각이 가지는 상대적인 중요도나 기술적인 한계로 인해 기술 개발 수준에는 차이가 있다.

오감 정보들 중 높은 중요도를 가지는 것으로 판단되는 시각과 청각의 경우 상대적으로 기술 개발이 많이 진행되었으며, 촉각 정보도 최근에는 여러 가지 장비를 통하여 표현할 수 있는 단계에 이르렀다. 후각 및 미각의 경우는 상대적으로 기술의 개발 단계가 높지는 않으나 기반연구가 꾸준히 진행되고 있다.

체감미디어에서 시각과 청각 정보를 잘 표현해 주기 위해서는 현실공간에서 가능한 한 많은 영상 및 음향 정보를 획득해야 하며, 이를 이용하여 가상공간에서 사용자의 움직임에 맞게 시각 및 청각 정보를 현실과 유사하게 모사해 주어야 한다. 좀 더 현실감 있는 시각 및 청각 정보를 재현하기 위한 연구는 이전부터 지속적으로 이루어져왔으며,

현재에도 많은 연구가 진행되고 있다.

본 장의 다음 절에서는 체감미디어 중 공간음향의 개념 및 사용되고 있는 기술에 대해 간단히 소개한다.

### 2. 체감미디어를 위한 공간음향

공간음향 기술이란 앞서 언급한 것과 같이 청취자가 현실과 유사한 음향을 경험할 수 있도록 청각적인 정보를 현실과 최대한 유사하게 재현하기 위한 기술이라 할 수 있다. 즉 인간이 현실세계에서 청각을 사용하여 소리의 위치를 파악하고 사무실, 실내체육관, 콘서트홀 등 특정한 공간의 특성을 인지하는 원리를 분석하고, 이를 기반으로 가상 세계에서도 청취자의 위치 및 방향 그리고 움직임에 따라 세밀하게 변화되는 현실적인 소리 및 공간을 느낄 수 있도록 공간의 소리를 획득/분석/재현하는 기술이다.

방송, 영화 등과 같은 기존의 미디어에서는 청취자의 위치가 고정되어 있다는 가정하에 음향이 제작되고 재현되는 것이 대부분이었다. 반면, 체감미디어에서는 사용자가 가상공간에서 움직이는 것을 고려하여야 하므로, 동적인 음원과 동적인 청취자의 상대적인 위치에 따라 시·공간의 파동 현상을 분석함으로써 동적으로 청취되는 음향을 생성하여야 한다.

이를 위해서는 음원의 음향신호의 특성과 함께 음원과 청취자가 속해 있는 공간의 구조 및 특성에 따른 반사음 및 잔향 특성, 그리고 청취자의 위치 및 방향에 따라 두 귀에서 청취되는 음의 변화, 즉 인간의 청각 특성을 고려한 음향신호처리가 필요하다. 이러한 음향의 처리기법을 객체기반 음향이라고 하며, 이를 위해서는 많은 음원과 음향의 전파 특성을 분석하는 복잡한 연산이 필요하게 되며,

일반적인 단말에서 이를 실시간으로 구현하는 것은 매우 어려운 일이 된다.

이에 대해 청취자를 중심으로 청취자에게 전달되는 모든 방향의 소리를 샘플링하여 획득하고 재생하는 것을 채널기반 음향이라고 하는데, 음향을 획득하는 마이크로폰 및 재생하는 스피커의 개수가 증가함에 따라 공간적인 음향의 재생 충실도는 비례하여 증가한다. 방송 및 영화의 오디오 채널 수가 점점 증가하는 것은 이러한 공간적인 음향의 재생 충실도를 증가시키므로써 몰입도를 증가시키기 위한 목적이라고 할 수 있다. 하지만, 충분히 만족할 만한 공간음향을 재생하기 위해서는 적어도 20개 이상의 스피커가 필요하며, 이는 음향의 획득, 제작, 전송, 재생에 있어 사업자들이나 사용자에게 적잖은 부담된다. 또한, 채널기반 음향의 경우 사용자의 위치가 고정된 환경을 기반으로 한 기술로서 사용자가 자유로이 움직일 수 있는 환경에서는 여러 가지 제약이 있을 수 있다.

채널기반 음향의 단점에 대해 장면기반 음향은 음향의 파동원리에 기반하여 공간음향을 수학적으로 해석 가능한 기하학적 방향성을 가지는 음향신호들로 표현하는 방식이다. 하나의 무지향성 마이크로폰만으로 구성되는 0차 앰비소닉부터  $x, y, z$  축 방향의 8자형 지향성 마이크로폰이 추가되는 1차 앰비소닉(FOA: First Order Ambisonics), 그리고 2차 이상의 차수를 가지는 고차 앰비소닉(HOA: Higher Order Ambisonics)이 있다. 앰비소닉은 차수  $n$ 이 증가함에 따라 공간적으로 균일하게 분포하는  $2 * n + 1$ 개의 마이크로폰이 추가되고, 이에 따라 각 차수에서는 모두  $(n+1)^2$ 개의 채널수가 사용된다[1,2]. 이러한 장면기반 음향은 수학적 해석이 가능한 특성에 의해 고차 앰비소닉을 사용하면 공간 필터링이 가능하여 객체기반 음향의 특징을 가

질 수 있으며, 각 채널의 조합에 의해 임의 채널을 구성할 수 있으므로 채널기반 음향의 특징을 함께 가질 수 있게 된다[3].

객체기반, 채널기반, 장면기반 음향 기술은 각각 장단점을 가지고 있으며, 상호 보완적인 요소를 포함하고 있으므로 이들을 적절히 조합함으로써 미래의 공간음향 기술을 확보할 수 있을 것으로 예상된다.

### III. 공간음향 획득 기술

음향신호를 획득하는 방법은 목적과 응용에 따라 매우 다양한데, 앞서 기술한 채널기반, 객체기반, 장면기반 음향과 같은 음향의 방식에 따라서도 획득하는 방법이 다를 수 있다. 본 장에서는 각 음향 획득 방법에 대해 간단히 소개한다.

#### 1. 채널기반 음향의 획득

채널기반 음향은 현재 사용되고 있는 다양한 채널포맷의 음향신호들을 획득하는 방법으로서 채널 개수 및 채널의 공간배치에 따라 여러 개의 마이크로폰을 공간상에 배치하여 청취하기에 적합한 멀티채널 음향신호를 획득하는 방법이다. 주로 사용되는 채널포맷으로는 모노, 2채널 스테레오, 5.1채널이 있으며, 영화 및 UHD TV(Ultra High Definition TV) 방송을 위하여 5.1채널보다 많은 채널 개수를 가지는 7.1, 11.1, 10.2, 14.2, 22.2, 30.2채널의 채널포맷이 등장하였다. 스테레오 및 5.1채널의 경우 2개 및 5개의 마이크로폰을 조합하여 단번에 획득하는 방법이 주로 음악 녹음에 사용되고, UHD TV 방송을 위하여 한국에서는 10.2채널, 일본에서는 22.2채널 마이크로폰을 개발한 예가 있지만, 대개는 다수의 근

접 마이크로폰 및 배경음 마이크로폰으로 녹음한 음향신호들을 Pro tools, Nuendo 등과 같은 DAW(Digital Audio Workstation)로 믹싱하여 다양한 채널포맷의 음향신호를 제작하는 방법이 사용되고 있다[4].

### 2. 객체기반 음향의 획득

객체기반 음향은 각각의 객체에서 나는 소리를 주변 환경의 영향을 받지 않은 독립적이고 순수한 상태로 획득하는 것이 좋다. 또한 객체에서 나는 소리는 방향에 따라 다르게 들릴 수 있으므로, 다수의 마이크로폰을 활용하여 소리의 지향성 패턴을 함께 녹음하는 것이 이상적이다. 현실적으로 오디오 객체의 방사 패턴을 녹음하는 것은 매우 번거로운 일이기 때문에 대부분의 객체기반 음향 획득은 근접 마이크로폰만으로 획득되고 있으며, 방사 패턴은 포함하고 있지 않다.

또한 객체기반 음향신호는 재생할 때 공간의 구조 및 흡음/반사 특성 등 공간의 특성과 음원의 위치 정보에 따라 반사음 및 잔향음 신호가 다르므로, 객체기반 음향을 재생환경에서 렌더링할 때 필요한 공간 정보를 메타데이터로 추가 전송할 필요가 있다. 이외에도 보다 세밀한 공간음향 렌더링을 위해서는 음향객체의 크기, 모양 및 흡음/반사 특성 정보도 함께 획득하여 전송할 필요가 있다.

### 3. 장면기반 음향의 획득

장면기반 음향을 획득하는 것은 각 차수의 앰비소닉 채널의 개수 및 지향성 특성을 가지는 마이크로폰을 하나의 점에 배치하여 녹음하는 것이 필요하지만, 복잡한 지향성을 가지는 마이크로폰을 구현하는 것이 불가능하고, 다수의 마이크로폰을 한

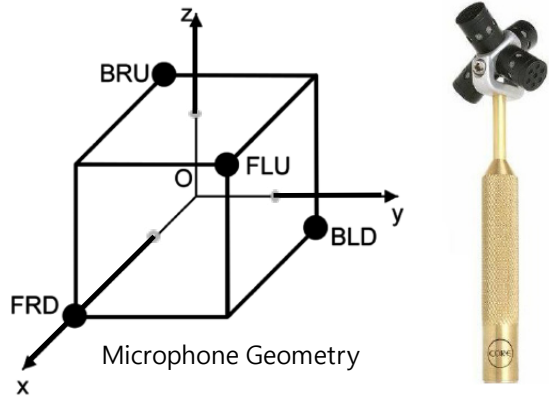


그림 1 1차 앰비소닉 마이크로폰 배치

출처 J.Y. Hong et al, "Spatial Audio for Soundscape Design: Recording and Reproduction," *Appl. Sci.*, vol. 7, 2017, Article no. 627, CC BY 4.0.

점에 배치시키는 것이 불가능하므로 충분한 개수의 단일지향성 마이크로폰을 구체 표면에 균일분포로 배치하여 A-포맷의 앰비소닉 신호를 획득한 다음 이들 신호들을 조합하여 이론적인 B-포맷의 앰비소닉 신호를 합성하는 방법을 사용하고 있다.

현재 주로 사용되고 있는 1차 앰비소닉 마이크로폰은 그림 1과 같이 네 개의 마이크로폰을 구체 표면에 균일한 분포로 배치하는 것으로 구현할 수 있고, 4개의 마이크로폰으로 획득한 A-포맷 신호 FLU, FRD, BRU, BLD로부터 B-포맷 신호 W, X, Y, Z는 식 (1)과 같이 구할 수 있다[5].

$$\begin{aligned}
 W &= FLU + FRD + BLD + BRU \\
 X &= FLU + FRD - BLD - BRU \\
 Y &= FLU - FRD + BLD - BRU \\
 Z &= FLU - FRD - BLD + BRU
 \end{aligned} \tag{1}$$

### IV. 공간음향 분석 기술

공간음향의 분석은 다수의 채널 신호를 압축하거나 파라미터화함으로써 렌더링 시 다양한 제어

기능을 실현하기 위해 필요한 절차라고 할 수 있다. 각 음향 방식에 따른 대표적인 분석 기술을 다음에 소개한다.

### 1. 채널기반 음향

채널기반 음향은 청취자를 중심으로 전방이나 수평면 혹은 반구의 표면에 배치된 다수의 스피커 재생을 기반으로 하고 있으며, 대개 인접한 스피커들을 이용하여 이득 혹은 위상 패닝 방법으로 팬텀 음원을 렌더링하게 된다. MPEG 오디오 기술에 있어 채널기반 음향의 분석 기술은 이러한 패닝 및 상관 파라미터, 즉 CLD(Channel Level Difference), ICC(Inter-Channel Coherence), CPC(Channel Prediction Coefficient)를 추출함으로써 다운믹싱된 신호로부터 원래의 채널 신호를 재합성할 수 있도록 하며, 이로써 전송되는 채널 수를 줄여 압축하는 방법을 제공한다. 또한 이러한 파라미터들을 조합함으로써 음원의 위치, 볼륨 등 다양한 제어가 가능하도록 한다[5].

한편 음원분리 기술을 활용하면 채널기반 음향을 객체기반 음향으로 변환할 수 있으며, 이를 통하여 객체기반 음향의 다양한 이점을 활용할 수 있는데, 아직은 음원분리 기술의 성능이 이러한 방식으로 활용하기에는 부족한 면이 있다.

### 2. 객체기반 음향

객체기반 음향에 있어서는 음향 자체의 분석은 필요하지 않으며, 렌더링하기 위한 파라미터를 추출하기 위한 분석이 필요하다.

우선 공간의 구조모델을 분석하는 것이 필요하며, 공간의 구조를 수학적으로 해석하기 위해서는 곡선으로 구성되어 있는 실제의 공간 정보를 정형

의 공간구조로 최적화 및 단순화하여야 한다. 즉 임의 공간의 음향의 전파특성을 분석하기 위해서는 공간을 이루는 각 면은 완전히 평탄하여야 하며, 각 모서리는 직선으로 이루어져야 한다. 또한 각 면은 특유의 흡음/반사/산란 특성을 가지게 되는데, 이들은 주로 재료를 만드는 업체에서 제공하는 데이터를 활용한다[6]. 음원의 지향성, 반사면의 흡음특성, 산란특성 등 객체기반 음향의 렌더링을 위해 필요한 파라미터들은 옥타브밴드 혹은 1/3 옥타브밴드로 나누어서 표현되며, 이에 의해 산출되는 반사음 및 잔향 특성은 주파수 종속적인 특성을 갖게 되며 청취자를 중심으로 도래방향을 가지는 반사음의 집합이 된다.

임의 공간에서 주어진 음원의 반사파를 추정하는 방법에는 파동에 의한 방법 및 음선에 의한 방법이 있다. 파동에 의한 방법은 파동방정식을 통하여 음원의 파형이 전파되는 현상을 예측하는 방법으로 매우 복잡하지만 상대적으로 정확한 해답을 얻을 수 있다. 음선에 의한 방법은 음원을 광원으로 각 벽면을 거울로 간주하여 샘플링된 음의 입자가 전파되는 경로를 추정하는 방법이며, 대표적으로 음선 추적(RT: Ray Tracing) 방식과 반사 음원(IS: Image Source) 방식이 있다[7].

### 3. 장면기반 음향

장면기반 음향은 앰비소닉 음향 기술의 수학적 기반에 의해 각 방향으로부터 도래하는 음원 성분에 대한 분석이 용이하다. 실제로 충분히 잘게 나누어진 주파수대역에 대해 음원의 도래방향 성분 및 지향성의 안정성 파라미터인 확산 성분을 추출함으로써 각 주파수대역의 신호성분을 분리된 음원 성분으로 간주하여 렌더링 및 제어하는 것이 가능하다.



장면기반 음향 분석 기술의 예로 핀란드 알토대학의 빌레 풀키 교수가 고안한 DirAC(Directional Audio Coding) 기술[8]과 호주 울런공대학 크리스찬 리츠 교수팀이 고안한 S<sup>3</sup>AC(Spatially Squeezed Surround Audio Coding) 기술[9]이 있다. 이들 기술은 이론적으로 채널기반 음향에도 활용할 수 있으며, 채널 다운믹스 및 업믹스에 의한 전송 채널 감축에 활용되며, 각 방향의 음원에 대해 독립적인 제어가 가능하므로 음향장면 회전, 압축, 음향 확대(zooming), 공간 필터링 등 다양한 기능으로 활용될 수 있다[3].

### V. 공간음향 재현 기술

공간음향 재현 및 제어 기술은 공간 정보를 가지는 오디오 콘텐츠의 재현에 있어 재현 환경의 변화 및 청취자의 위치에 대해 공간음향의 왜곡을 감소시키고, 원 제작자의 의도대로 모든 음원의 소리 및 실내효과에 의한 반사음 및 잔향음이 실제와 유사하게 청취되도록 제어하는 기술이다. 이러한 공간음향 재현 및 제어 기술은 기존의 멀티채널 오디오 장비에서도 필요했었지만, 6DoF(Degree of Freedom) 가상현실 환경에서는 필수적인 기능이라고 할 수 있다.

#### 1. 채널기반 음향

채널기반 음향은 그 역사적인 배경이 모노채널의 음향으로부터 공간감을 부여하기 위하여 점차 발전되어온 방식으로서, 그 태생적인 특징에 의해 이상적인 재생환경과 스위트스팟에 위치한 청취자에게는 이상적인 공간음향을 표현해 줄 수 있지만, 이상적인 환경에서 벗어나는 경우, 이를 보상해 줄 수 있는 기술은 매우 한정적이라고 할 수 있다. 주

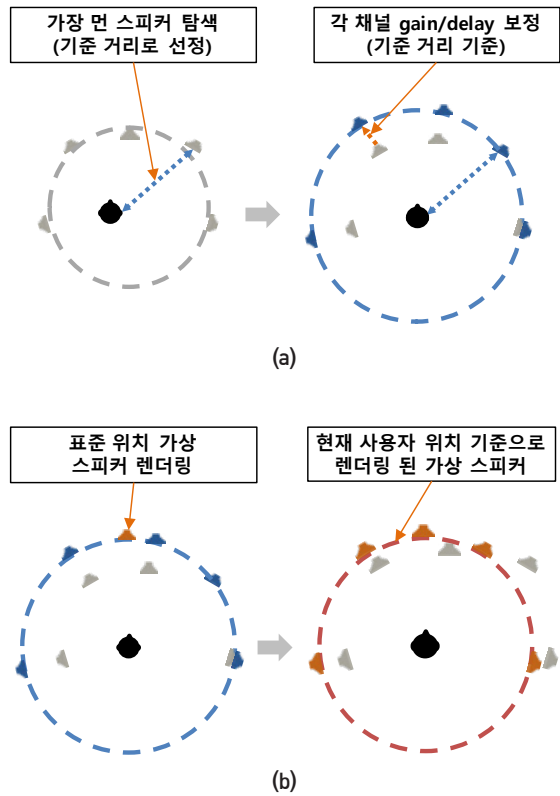


그림 2 멀티채널 스피커 보정 : (a) 스피커 거리에 따른 보정, (b) 스피커 위치에 따른 보정

로 재생환경의 변화, 즉 재생 스피커의 개수 및 배치의 변화에 대한 보상과 공간의 음향특성을 변화시키는 방법이 주로 사용되고 있다.

재생 스피커의 개수나 배치가 변화되는 경우, 원 제작자가 의도한 음향의 공간적인 이미지가 왜곡될 수밖에 없으므로, 그림 2와 같이 스피커의 배치 상황을 측정하여 각 스피커의 위치가 이상적인 위치에 있는 것처럼 각 채널 신호의 이미지를 조정하여야 하는데, 이를 위하여 각 채널 신호의 이득, 지연 및 주파수 특성을 조정하고, 실재하지 않는 스피커에 대해서는 인접 스피커의 패닝에 의해 가상 스피커를 생성하는 방법을 사용할 수 있다[10]. 이 외에도 공간의 잔향특성을 변화시키기 위하여 실내효과 필터를 사용할 수도 있다.

최근에는 스피커의 개수가 증가되고 스피커의 배치가 3차원 공간으로 확장되는 추세에 따라 기존의 두 채널 사이의 패닝에 의한 음향 렌더링 기술에서 3개 이상의 스피커 사이의 패닝을 사용하는 벡터기반 패닝(VBAP: Vector Based Amplitude Panning)이 활용되고 있으며[11], 주어진 유효청취 영역에 대하여 최적 음향재생을 보장하기 위하여 최소 오차를 가지는 응답을 찾아내는 LMS(Least Mean Square)를 이용한 렌더링 방법도 사용되고 있다.

## 2. 객체기반 음향

객체기반 음향은 모든 음원을 별도로 획득하고 음원의 위치 정보 및 지향성 특성을 메타데이터로 포함하기 때문에 공간음향의 재현 및 제어에 있어 재현환경 및 청취자의 움직임에 충실히 대응할 수 있는 가장 이상적인 방법이다. 하지만 이미 분석 단계에서 언급한 것과 같이 반사음 및 잔향음을 추정하기 위하여 음향에 최적화된 공간모델을 필요로 하며, 이는 실시간 응용서비스에 있어서는 커다란 장애물로 작용한다.

객체기반 음향은 모든 음원 및 공간모델을 최종 청취하려는 공간음향을 생성하기 위한 재료로 사용하므로, 공간모델이 주어지고 청취자의 위치가 정확히 주어지면 복잡한 음향의 전달함수 추정 및 렌더링에 의해 시공간적으로 가변하는 공간음향을 생성할 수 있다는 장점이 있다. 그러나 한편으로는 공간모델, 음원모델 및 청취자모델이 실제의 물리적인 모델을 근사화한 것으로 실제 응답을 정확히 생성할 수 없는 단점이 있다.

객체기반 음향의 실제 청취환경은 결국 헤드폰 또는 멀티채널 스피커 재생 환경이며, 이들 청취 시스템의 특성에 따라 성능이 좌우될 수 있으므로

청취 시스템의 특성 보정이 필요하다. 또한 청취모델의 경우, 인간의 두 귀의 물리적 특성의 차이로 인해 개인차가 있으므로 헤드폰 재생을 할 때는 개인의 청취모델을 고려하는 것이 필요하다. 이상적으로는 개인의 머리전달함수(HRTF: Head Related Transfer Function)를 직접 측정하여 안경처럼 개인의 청각 특성을 보정하는 것이 필요하지만, HRTF를 측정하는 것은 매우 힘든 일이므로 귀의 모양이나 머리 및 몸통의 특성값을 기반으로 모델링하는 방법이 주로 사용되고 있다[12].

## 3. 장면기반 음향

장면기반 음향은 객체기반 음향에 비해서는 다소 공간음향 성능이 떨어진다고 할 수 있지만, 수학적으로 해석가능하고 파라미터화할 수 있는 특징에 의해 여러 가지의 공간음향 재현 및 제어 방법을 활용할 수 있다. 즉, 각 방향의 음원들을 근사적으로 구분하여 파라미터화할 수 있으므로 청취자가 고정된 상황에서 머리의 회전에 의한 360 VR 체험에 잘 활용될 수 있으며, 상체의 움직임 정도까지는 보정할 수 있는 기술들이 제안되고 있다. 그러나 가까운 음원이 포함되는 경우 청취자의 움직임 보상에 대한 오차가 증가하게 된다.

다수의 HOA 마이크로폰을 사용하여 6DoF 가상 현실을 실현하는 방안으로, 주어진 공간의 샘플링된 지점에서 앰비소닉 마이크로폰으로 녹음하고, 그림 3과 같이 그 사이의 청취자 위치에 대하여 주변 장면기반 음향신호들을 인터플레이션에 의해 합성하는 기술들이 제안되고 있다[13].

## VI. MPEG-I 오디오 표준화

지금까지 살펴본 공간음향 기술들은 채널기반,

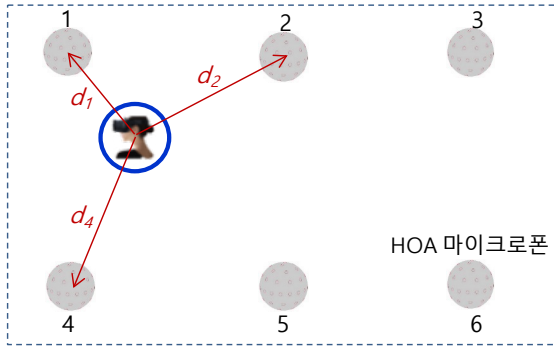


그림 3 다수의 HOA 마이크론 인터플레이션 개념

객체기반 및 장면기반 음향으로 구분하여 살펴보았지만, 각자 장점과 단점을 함께 가지고 있으며, 상호보완적인 특징도 있으므로 현실적으로는 적절한 공간음향 서비스를 제공하기 위해서는 채널기반, 객체기반, 장면기반 음향 기술을 함께 적용하는 것이 적절하다.

이후 절에서는 6DoF 가상현실 체험이 가능한 공간음향 서비스를 제공하기 위한 MPEG-I(Motion Picture Experts Group - Immersive) 오디오 표준화 동향에 대해 살펴보려고 한다.

## 1. MPEG-I 오디오 구조

MPEG-I 오디오는 기존의 MPEG 오디오 표준 기술과는 다르게 오디오 신호 압축 기술에 대해 다루지 않고 있으며, 최근 표준화가 완료된 MPEG-H 3D Audio 표준기술을 기반으로 6DoF 가상현실 서비스를 제공하기 위한 메타데이터 및 렌더링 기술을 표준화 대상으로 하고 있다. 6DoF 렌더링을 위해서는 음원에 대해 이동하는 청취자의 방향에 따라 음향의 특성이 변하는 음원의 지향성을 고려하여야 하며, 반사음 및 잔향음도 청취자의 위치에 따라 변화되는 것을 모사해 주어야 한다.

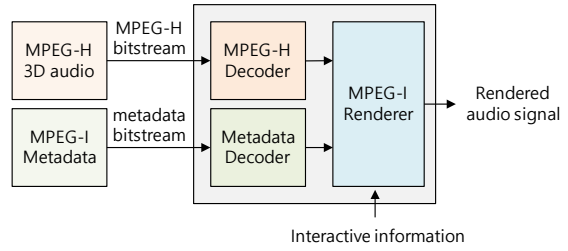


그림 4 MPEG-I 6DoF 오디오 구성

그림 4는 MPEG-I 오디오의 6DoF 오디오 서비스를 위한 구성을 간략하게 나타낸 것이다. MPEG-I 오디오에서는 음향신호의 압축 기술은 MPEG-H 3D 오디오 기술을 그대로 활용하고, MPEG-I 6DoF 렌더링을 위한 메타데이터가 추가된다. 재현 단말에는 MPEG-H 디코더와 함께 메타데이터의 해석을 위한 메타데이터 디코더 및 MPEG-I 렌더러가 포함된다. MPEG-I 렌더러에서는 오디오 신호와 메타데이터, 그리고 사용자 위치 정보와 같은 상호작용 정보를 입력으로 받아 최종 출력 오디오를 렌더링한다.

## 2. MPEG-I 오디오 평가 시스템

MPEG에서는 MPEG-I 오디오 기술의 성능평가를 위하여 6DoF 가상현실 장비 및 도구를 활용하여 평가 시스템을 구현하여 공개하고 있다.

평가 시스템은 청취자의 6DoF 움직임을 지원하기 위하여 청취자의 위치 정보와 같은 상호작용 정보를 제공하는 VR 장비인 HTC Vive 장비, 비디오 렌더링 및 평가 제어를 위한 Unity 개발 도구, 그리고, 오디오 렌더링을 위한 MAX7 도구로 구성된다 [14].

Unity와 MAX7 사이의 데이터 및 메시지 교환을 위하여 Unity를 위한 OSC(Open Sound Control) 플러그인이 사용되고 있으며, MPEG-I 렌더러는



MAX7 에서 동작하는 VST(Virtual Studio Technology)로 구현된다.

기술을 제안하는 기관은 제안 기술이 포함된 렌더러를 VST 형태로 구현해야 하는데, 각 사에서 구현한 여러 개의 VST가 평가 시스템에 등록이 되며, 평가 시스템에서는 여러 개의 VST들 중 하나를 선택하여 청취할 수 있도록 하여, 여러 개의 제안 기술에 대한 비교 및 평가가 가능하도록 하였다.

## VII. 결론

지금까지 공간음향 기술의 개념 및 동향 그리고 최근 공간음향 기술을 서비스하기 위한 MPEG-I 오디오 기술의 표준화 동향에 대하여 살펴보았다. 공간음향 기술의 역사는 매우 오래 되었지만, 그동안 주로 콘서트홀 등 음향공간 건축과 관련된 건축음향과 음향 콘텐츠 제작을 위한 음향 제작도구에서 중요하게 다루어진 부분이며, 음향 재생장비에서 다루기에는 매우 복잡한 기술로 여겨져 왔다. 그러나 가상현실 기술이 점차 발전하고 그 활용분야가 확대되면서 공간음향 기술이 재생 기술에 활용되는 상황으로까지 발전되었으며, 공간음향 재생을 위한 기술개발 및 표준화에 대한 요구가 점차 증가하고 있는 추세이다.

향후 가상현실 기술은 교육, 훈련, 의료, 정비, 홍보, 게임 등 다양한 분야에서 효과적으로 활용될 수 있을 것으로 예상되며, 이때 현실감 및 몰입감의 향상을 위하여 공간음향 기술의 중요성이 점차 확대될 것으로 예상된다. 또한 AR, MR(Mixed Reality) 기술의 확산을 기반으로 현실과 가상현실의 경계가 모호해지고, 서로 융합된 서비스가 점차 증가할 것으로 예상되므로 공간음향은 가상현실에서의 자연스러운 휴먼인터페이스를 위해 필수적인 요소가 될 것으로 전망된다.

### 용어해설

**공간음향** 음원과 청취자가 동적으로 움직이는 6DoF 가상현실에 있어서 청취자의 움직임에 대해 시공간적으로 현실과 유사한 체험을 가능하게 하는 3차원 음향처리 기술

**체감형 미디어** 혼합현실 입체 공간에서 시각, 청각, 촉각 등 오감의 감성 표현이 가능하며, 현실감과 몰입감 및 인터랙션을 제공할 수 있는 초실감 미디어

**MPEG-I** VR, AR, MR 등 차세대 몰입형 미디어에 적합한 비디오와 오디오 및 상호작용을 제공하기 위하여 입체 형상을 가지는 AV 객체의 획득, 부복호화/전송 및 렌더링 방법을 포함하는 차세대 몰입형 미디어 표준 기술

### 약어 정리

AR	Augmented Reality
CLD	Channel Level Difference
CPC	Channel Prediction Coefficient
DAW	Digital Audio Workstation
DirAC	Directional Audio Coding
DoF	Degree of Freedom
FOA	First Order Ambisonic
HOA	High Order Ambisonic
HRTF	Head Related Transfer Function
ICC	Inter-Channel Coherence
IS	Image Source
LMS	Least Mean Square
MPEG	Motion Picture Experts Group
MR	Mixed Reality
OSC	Open Sound Control
RT	Ray Tracing
S <sup>3</sup> AC	Spatially Squeezed Surround Audio Coding
UHDTV	Ultra High Definition TV
VBAP	Vector Based Amplitude Panning
VR	Virtual Reality
VST	Virtual Studio Technology

## 참고문헌

- [1] M. A. Gerzon, "Panpot Laws for Multispeaker Stereo," Convention of the Audio Eng. Soc., May. 1992, Paper no. 3309.
- [2] M. Poletti, "The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems," *J. Audio Eng. Soc.*, vol. 44, no. 11, 1996, pp. 948-963.
- [3] A. Politis, T. Pihlajamaki, V. pulkki, "Parametric Spatial Audio Effects," in *Proc. Int. Conf. Digital Audio Effect(DAFx-12)*, York, UK, Sept. 2012, pp. 1-8.
- [4] 장대영, 이용주, 유재현, 이태진, "고현장감 오디오 서비스를 위한 하이브리드 오디오 기술동향," *전자통신동향분석*, 제31권 제3호, 2016.6, pp. 81-90.
- [5] X. Zheng, "Soundfield Nabigation: Separation, Compression and Transmission," PhD thesis, University of Wollongong, 2013.
- [6] S. Siltanen, T. Lokki, L. Savioja, "Geometry Reduction in Room Acoustics Modeling," in *Proc. Institute Acoustics: Auditorium Acoustics. Int. Conf.*, Copenhagen, Denmark, May 2006, vol. 28, Pt.2.
- [7] 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.
- [8] V. Pulkki, C. Faller, "Directional Audio Coding: Filterbank and STFT-based Design," in *AES Convention*, Paris, France, May 2006.
- [9] B. Cheng, C. Ritz, I. Burnett, "A Spatial Squeezing Approach to Ambisonic Audio Compression," in *IEEE Int. Conf. Acoustics, Speech Signal Process.*, Las Vegas, NV, USA, 2008, pp. 369-372.
- [10] J.-H. Yoo, S. Suh, D. Jang, T. Lee, "A Study on Multichannel Sound for Off Sweet Spot Listening Position," in *Int. Meeting Inform. Display*, Busan, Rep. of Korea, Aug. 2017.
- [11] V. Pulkki, "Virtual Sound Source Positioning Using Vector Based Amplitude Panning," *J. Audio Eng. Soc.*, vol. 45, 1997, pp. 456-466.
- [12] M. Lins, R. Bomhardt, "Individualization of Head Related Transfer Function," *Lekar a tehnika Clinician Technol.*, vol. 44, no. 1, 2014, pp. 26-32.
- [13] J.G. Tylka, E.Y. Choueiri, "Soundfield Navigation using an Array of Higher-Order Ambisonics Microphones," in *AES Conf. Audio Virtual Augmented Reality*, Los Angeles, CA, USA, Sept. 2016.
- [14] ISO/IEC JTC1/SC29/WG11/N17177, "MPEG-I Audio Architecture and Evaluation for 6DoF," 2017. 10.