Chapter 01

# 공간 음향(Spatial Audio) - 메타버스를 실현하는 오디오 기술

**정현주**_가우디오랩(주) 연구위원
**서정훈**_가우디오랩(주) 연구위원
**이태규**_가우디오랩(주) CTO
**오현오**_가우디오랩(주) CEO

현실과 가상 세계를 넘나드는 메타버스의 실감나는 사용자 경험을 제공하기 위해서 음향은 영상 못지 않게 중요한 요소이다. 인간의 청감 특성이 3차원 공간을 이해하는 원리 그대로를 재현하는 공간 음향(Spatial Audio)은 사용자와의 상호작용(Interaction)이 요구되는 가상/증강/확장현실 환경에서 보다 몰입감(Immersion) 있는 소리 경험을 제공하기 위해 요구되는 필수적인 기술이다. 본 고에서는 메타버스를 준비하며 현재 모바일 시장을 중심으로 연구 개발 및 활용되고 있는 공간 음향 기술에 대해 심도있게 다룬다.

## I. 서론

메타버스는 1992년 그 용어가 처음 등장한 이래[1], 일반적으로 “현실 세계와 같은 사회적, 경제적 활동이 가능한 3차원 가상공간” 정도의 의미로 사용되어 왔다[2]. 아바타(Avatar)라는 존재를 통해 실제 사용자의 페르소나가 반영된 인터넷 기반의 3차원 가상세계를 뜻하며, 사이버스페이스(Cyberspace)의 경우와 마찬가지로 점차 기술적, 문화적 용어로 확장되었다[3]. 미국의 미래 가속화 연구재단(Acceleration Studies Foundation: ASF)은 2006년 전 세계의 가상세계 연구를 집약하여 개최된 “‘메타버스 로드맵 회담(Metaverse Roadmap Summit)”을 통해 그 개념을 구체화, 확장시켜 현재에 이르고 있으며, 메타버스의 본질을 “가상적으로 확장된 물리적 현실과 물리적으로 영구화된 가상공간의 융합“으로 정의하고

* 본 내용은 정현주 연구위원(☎ 02-562-1968, sc@gaudiolab.com)에게 문의하시기 바랍니다.
** 본 내용은 필자의 주관적인 의견이며 IITP의 공식적인 입장이 아님을 밝힙니다.

있다[4].

메타버스 로드맵에 따르면 메타버스를 현실 세계의 대안 또는 반대로 보는 접근에서 벗어나 상호작용을 통해 두 세계의 교차와 결합으로 이해할 것을 제안하고 있다[4]. 현실/가상세계의 증강 기술과 시뮬레이션 기술의 발전, 외재적/내재적 기술의 발전이라는 양 축에 따라서 메타버스의 유형[5]을 분류하고 있으나, 각 유형은 명확히 분리되기보다는 그 경계점이 융합을 통해 허물어지는 경향을 보인다. 비록 현재까지는 기술적인 한계로 스크린 기반의 인터넷 가상세계 안에서 메타버스를 구현하고자 하는 플랫폼이 활성화되었으나, 동명의 소설을 원작으로 한 영화 '레디플레이어원'에서 보여준 바와 같이, 가상세계의 감각경험이 현실세계의 감각경험과 유사하거나 일치될 때 두 세계 간의 완벽한 융합이 가능하다는 점에서 시청각을 중심으로 한 확장현실 기반 기술은 메타버스의 기술적 근간을 형성하며 발전과 성장을 지속하고 있다. 글로벌 회계 · 컨설팅 그룹인 프라이스워터하우스쿠퍼스가 발간한 보고서[6]에 따르면 확장현실 시장은 2025년 537조 원에서 2030년 1,700조 원으로 급격하게 성장할 것으로 전망되고 있다.

따라서 지금까지 가상현실(Virtual Reality: VR), 증강현실(Augmented Reality: AR) 혹은 확장현실(Extended Reality: XR) 등으로 구분되던 개별 시장은 메타버스라는 더 담대한 그릇 안에 하나로 융합된다. 메타버스를 실현하는 확장현실의 핵심 기술은 화면(Video)과 음향(Audio) 기술로, 3차원 공간에서 사용자에게 더욱 사실적인 시각과 청각을 제공해 주는 것을 기술적인 지향점으로 삼고 있다. 즉, 인간의 오감 중 가장 큰 비중을 차지하고 있는 시각과 청각을 통해 가상세계의 몰입감(Immersion)을 제공하는 것을 목적으로 하며 주로 개인화된 HMD(Head-Mounted Display) 또는 헤드폰을 통해 재현된다. 이 중 헤드폰을 통해 재현되는 오디오 기술은 지금까지, 3D 오디오, 몰입형 오디오(Immersive Audio), 공간 음향(Spatial Audio) 등 여러 가지 용어로 표현되었으나, 최근 들어 애플에서 출시한 에어팟(프로)에 적용된 관련 기술이 공간 음향이란 용어로 소개되면서 급격하게 대중화의 길을 걷고 있다[7]. 가상현실, 증강현실 등의 기기에 적용되어 3차원 공간의 소리를 효과적으로 재현하기 위해 요구되는 일련의 기술들을 공간 음향 기술이라고 볼 수 있고, 그렇게 만들어진 사용자 경험 자체를 표현할 때는 몰입형 오디오라고 구분해서 부를 수 있지만 사실상 같은 의미로 해석된다[8].

헤드폰(이하 본 고에서 특별히 구분하지 않은 경우 헤드폰은 유 · 무선 연결을 막론하고

모든 종류의 헤드폰, 이어폰을 포함한다)은 알고 보면 가장 역사가 오래된 웨어러블(Wearable) 기기이며, 가장 오래된 원형의 확장현실 기기라고 말할 수 있다. 시끄러운 지하철에서 주변 소음을 차단하고 현실에는 없던 음악으로 나의 생각 공간을 이동시켜 주거나(VR), 지금 이 공간에 없는 누군가와 통화를 하면서 현실을 증강(AR)시킬 수 있는 특징들은, 광의의 메타버스이다. 하지만 현실과 가상의 융합을 위해서는 곧 사용자가 가상의 공간에 실재하는(Being There) 혹은 가상의 공간이 사용자의 실제 공간에 재현된 것과 같은(Being Here) 경험을 재현하는 것이 필수적이며, 이를 헤드폰 위에서 실현하는 것이 공간 음향이다. 따라서, 공간 음향은 메타버스를 실현하는데 필수적인 오디오 기술이라고 하겠다. 본 고에서는 이러한 공간 음향 기술에 대한 전반적인 기술 요소들을 소개한다. 이하 본 고에서 구분이 필요한 경우는 각각 가상현실, 증강현실이라고 부르고, 구분이 필요없는 경우는 확장현실로 표현한다. 또한, 이들 기술 기반을 통해 펼쳐지는 서비스는 메타버스라고 부르기로 한다.

## II. Place Illusion, Plausibility Illusion

확장현실에서 몰입감을 결정하는 요소는 매우 다양하다. 인지적 측면에서 몰입이란 현재의 물리적 환경과 확장현실로부터 제공되는 환경 사이의 경계를 모호하게 한다는 의미에서 환영(illusion)으로 간주하기도 한다. 실제로, 가상현실 콘텐츠를 경험하는 사람들은 가상 환경 안에 존재하지만, 실제 반응하는 방식은 실제 세계에서처럼 반응한다. 이러한 사용자들의 실제와 같은 반응들은 그만큼 가상 환경에 몰입했다는 의미이며, 동시에 사용자들은 가상 환경과 실제 환경 사이를 혼동하고 있다는 의미이기도 하다. 이러한 사용자들의 반응을 설명하기 위해 인지과학 영역에서는 “장소환영(Place Illusion: PI)”과 “그럴듯한 환영(Plausibility Illusion: Psi)”이라는 개념을 설정한다.

PI는 말 그대로 실제로 그 곳에 있는 듯한 “Being There(or Presence)”에 대한 인지의 정도이다[9]. 즉, 사용자가 스스로 가상의 장소에 있지 않다는 걸 알고 있음에도 불구하고, 얼마나 그 장소에 실제로 있는 것 같은 느낌을 갖는지에 대한 정도를 나타내는 개념이다[10]. 음향적으로는 가상환경에서 공간감을 느낄 수 있는 환경음, 바람소리, 실내(소)음, 해당 가상환경에서의 잡음 등의 조합에 의해서 특정 공간으로 몰입시키는 요소들을 포함한다. 경우에 따라서, 특정 청취자는 앞에서 언급한 요소들 중 특정 몇 가지 소리에 대해 더 예민한 감각을

가질 수 있고, 이러한 이유로 가상환경 상에서 해당 소리를 더 유심히 들어보거나 해당 소리에 대한 극한 상황까지의 시험을 진행할 수도 있다. 그 결과 극한 상황에서 가상환경 시스템이 해당 소리를 제대로 렌더링하지 못했을 때, 청취자의 PI가 붕괴되는 순간이 발생하고, 이는 몰입감을 저하시키는 요소로 작용한다[11].

Psi는 PI와 달리 청취자의 인지적 능력 자체와는 관련이 없고, 개개인이 해당 환경 내에서 '무엇'을 인지했는지에 대한 결과로 나타나는 환영이다[9],[10]. Psi는 몇 가지 조건이 충족되어야 발생하는데, 필수적인 요소는 아래 세 가지라고 알려져 있다[11].

- 사용자의 행동에 따라 반드시 가상환경에서 연관된 반응이 발생할 것
- 사용자의 실제 행동이 없더라도, 가상 환경은 사용자에게 직접적으로 반응할 것
- 해당 환경과 발생하는 사건(Event)들은 사용자의 경험과 실제환경에서의 상호작용에 기반했을 때 사용자의 지식과 기대감을 만족시킬 수 있는 것이어야 함

확장현실 오디오의 입장에서는 PI는 상호작용성(Interactivity) 또는 음상정위(Sound Localization)와 깊게 연관되어 있다. 따라서, 다음 장에서 설명하게 될 외재화는 Psi를 획득하기 위한 필수적인 기술이라고 할 수 있다.

## III. 오디오 외재화

외재화(Externalization)란 헤드폰을 이용한 바이노럴 렌더링(Binaural Rendering)[12]에서 해당 음향 장면 내의 소리가 머리 안에 맺히는 것이 아니라 헤드폰 밖 공간에서 들리는 것처럼 렌더링해 주는 기술로서, 바이노럴 렌더링 분야에서 여전히 활발히 연구가 이루어지고 있는 기술이다. 일반적으로, 무향실에서 측정한 일반화된 머리전달함수(Head-Related Transfer Function: HRTF)만을 이용하여 렌더링을 하는 경우, 해당 필터는 공간에 대한 정보를 갖고 있지 않고, 최종 사용자 개개인의 HRTF와 일치하지 않는 문제 때문에, 소위 음상이 머리 안쪽에만 맺히거나(In-Head Localization) 또는 좌/우 양 귀를 잇는 가상의 선 위에만 배치되는(Lateralization) 현상이 발생한다. 이는 가상/증강현실에서 사용자의 몰입감을 저하시키는 요소로 작용한다. 일례로, 가상현실에서 시각 정보는 객체가 사용자를 둘러싸고 있는 공간 상에 배치되어 있는 것을 보여주고 있는데, 렌더링되는 오디오 신호가 해당 공간에서 나는 것처럼 들리지 않고 음상이 모두 머릿속에 맺히게 될 경우 두 감각 사이

의 불일치가 발생하게 되고, 이는 사용자로 하여금 가상 환경에 몰입하지 못하게 하는 요인으로 작용한다.

쉬운 예로 심야시간 거실에서 블록버스터 영화를 헤드폰을 이용하여 감상하는 경우를 생각해 보자. 헤드폰을 통해 전달되는 소리는 청자의 머리 안에 배치된다. 당연히 스크린에서 보이는 화면과 헤드폰으로 전달되는 소리가 일치되지 않는 인지적 부조화가 발생한다. 이때 헤드폰으로 전달되는 오디오 신호가 영화가 재생되고 있는 공간의 특성을 반영해 마치 TV 또는 스크린 옆에 배치된 스피커에서 재생되는 것처럼 재현된다면, 훨씬 몰입감 있게 영화를 감상할 수 있을 것이다.

기존 오디오 업계에서 '음질'은 주로 원음에 대한 명료도(Clarity) 또는 충실도(Fidelity) 등에 초점을 맞추어 논의되어 왔다. 하지만 최근에는 음질의 의미를 기존의 좁은 의미에서 확장해 음향 공간을 충실히 제공하는지도 고려하고 있다. 이에 따라 외재화는 헤드폰을 이용한 바이노럴 오디오 렌더링 기술에서 음질을 결정하는 중요한 요소 중 하나로 고려되고 있다.

## IV. 공간 음향 라이브 저작 기술

Being There의 실현을 위해서는 우선 해당 현장 공간의 오디오를 효과적으로 취득하고, 저작하여 전송하는 것이 전제되어야 한다. 본 장에서는 라이브 콘서트 현장 등에서 공간 음향의 재현을 위한 음원 저작에 대해 살펴본다.

그룹 방탄소년단(BTS)이 게임 개발사 에픽 게임즈의 액션 빌딩 배틀로얄 게임 '포트나이트'의 파티로얄 모드 속에서 히트곡 '다이너마이트' 안무 버전 뮤직비디오를 처음 공개했다. 또한, 래퍼 트래비스 스캇 같은 유명 뮤지션도 '포트나이트'에서 가상 콘서트를 열어 2,770만 명 이상의 동시 접속자 수를 기록하기도 했다[13]. 게임이라는 가상현실 속에 BTS라는 실존하는 스타와 그들의 신곡이 등장하는 과정은 메타버스의 전형이라고 볼 수 있다. 음악이라는 콘텐츠의 속성상 가상현실 안에서 현실과 같은 콘서트 현장의 느낌을 실감나게, 그것도 쌍방향 소통이 가능하도록 실시간 라이브로 재현하는 기술은 점점 더 중요해지고 있다.

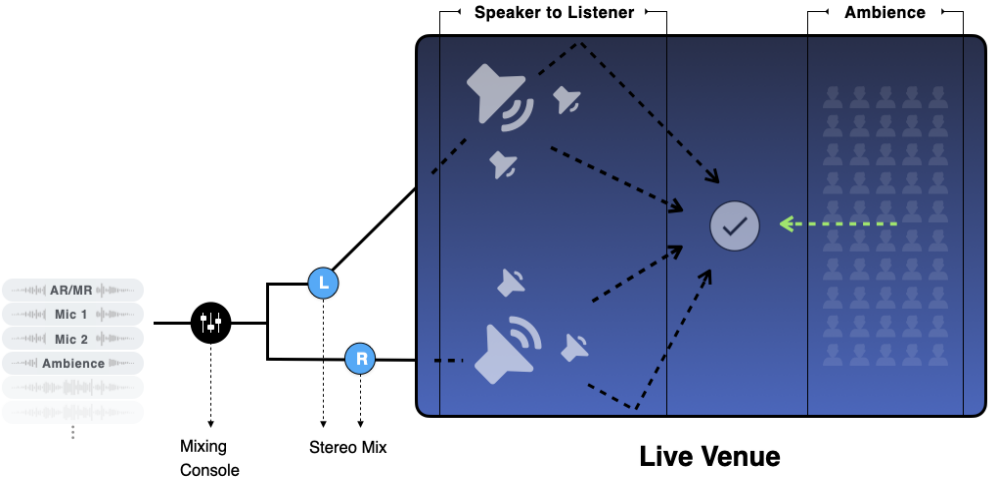
포트나이트의 게임 참여자와 같이 원격 사용자가 마치 공연장에 있는 것만 같은 "Being There"한 경험을 체험할 수 있도록 하는 다양한 기술들이 개발되고 있다. OTT(Over-The-Top, 주로 인터넷을 통해 방송 프로그램, 영화 등 미디어 콘텐츠를 스트리밍해주는 서비스)

와 같은 2D 동영상 서비스의 경우 다양한 카메라의 뷰를 제공하는 멀티뷰 기술, 모든 방향을 사용자가 자유롭게 볼 수 있는 360 영상 기술 등 기존의 2D 영상을 제공하면서 사용자가 몰입할 수 있는 수단을 추가적으로 제공하는 방식으로 발전하고 있다. 오디오 역시 이에 부합하도록 기존의 스테레오와 함께 멀티뷰, 360 영상에 걸맞은 몰입형 오디오를 제공하는 공간 음향 기술이 필요하다.

360 영상에서 공간 음향을 적용하는 가장 잘 알려진 방법은 앰비소닉스(Ambisonics) 캡처 방식이다[12]. 앰비소닉스는 음향 공간 상의 특정 위치에서 정의된 모든 방향의 오디오를 획득한 신호이며, 공간의 음장을 구면 조화 함수(Spherical Harmonics) 형태로 표현한 방식이다. 360 영상을 지원하는 Facebook, Youtube 등의 플랫폼에서는 입체음향 포맷으로 앰비소닉스를 채택하고 있기 때문에 많이 사용되고 있다.

그러나 공연장에서 앰비소닉스 방식을 사용하는 것에는 많은 한계가 존재한다. 현재 상용화된 대부분의 앰비소닉스 마이크는 높은 공연장의 음압을 감당하지 못한다. 모든 방향의 소리가 취득되어 반영되기 때문에 근접한 관객이나 스태프가 발생시키는 노이즈를 제어하거나 편집하기도 어려운 문제가 있다. 또한, 멀티뷰와 같은 다양한 위치에서의 입체 음향을 제공하기 위해서는 각 카메라 위치마다 마이크를 설치해야 한다.

공연장에서 실제 현장의 관객이 듣는 소리를 도식화하면 [그림 1]과 같다. 공연장에서는



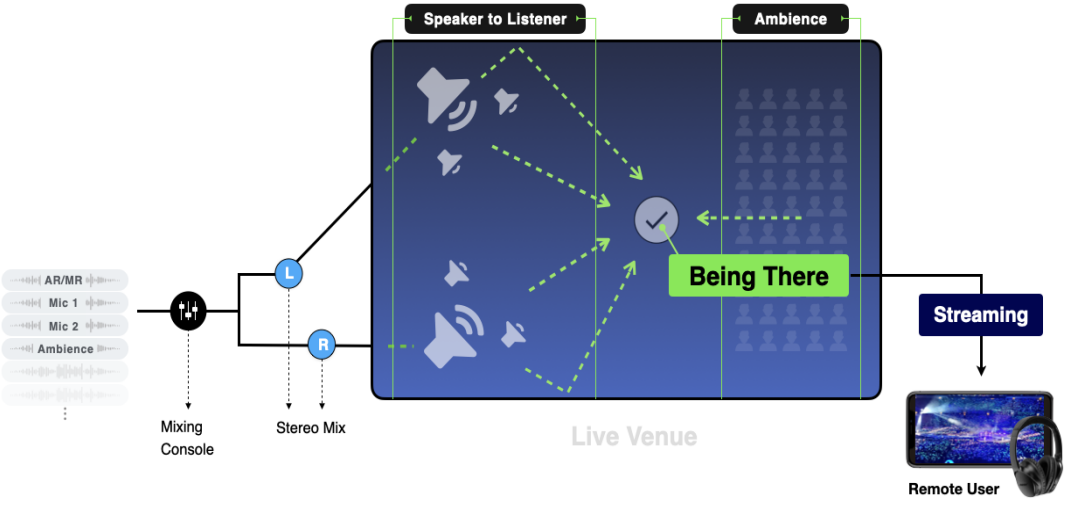

〈자료〉 Gaudio Lab, Inc., 2021.

[그림 1] 라이브 콘서트 현장의 관객 청취 음향 신호 모델링

보컬, 기타, 드럼, 관객의 환호성과 같은 공간음(Ambience), 현장음, 보컬을 제외한 반주인 MR(Music Recorded) 등을 레코딩하고, 라이브 사운드 엔지니어는 모든 소리를 관객이 조화롭게 잘 들을 수 있도록 스테레오로 믹스한다. 스테레오는 현장에 있는 스피커 어레이로 재생되고, 공간에 반사되어 현장의 관객에게 전달된다(Speaker to Listener 경로). 이 전달 과정은 공연장마다 다른 특징을 가진다. 예를 들어, 스타디움과 같은 대형 공연장의 경우 다른 관객들의 환호 소리와 같은 현장음과 함께 저음이 강하고 긴 잔향 시간을 갖는 반면, 아늑한 레코딩 스튜디오의 경우 짧고 작은 잔향이 발생하는 특징을 갖는다.

공연장에서는 각 악기별 소리와 현장음이 취득되기 때문에 객체 기반 오디오를 적용하기에 유리하다. 객체 기반 오디오는 개별 음원을 독립된 오디오 신호로 각각 전송하는 방식이다. 개별 음원이 공간 상의 해당 위치 정보(x, y, z 좌표 및 방향) 등을 포함하는 메타데이터와 함께 전송되고 재생 단에서는 사용자의 현재 위치, 방향에 따라 실시간으로 렌더링하여 재생하기 때문에 사용자의 6방향 자유도를 보장하는(6 Degrees-of Freedom: 6DoF) 확장현실 환경에서 가장 이상적인 포맷이라고 할 수 있다[12].

이와 같이 공연장 내에서 현장음을 직접 수음하는 과정의 문제를 해소하는 동시에 이미 확보 가능한 객체 신호를 이용하여 Being There를 재현하는 방안을 고려해볼 수 있다. 오디오 객체 신호의 조합으로 현장의 관객이 듣는 소리를 신호처리를 통해 재현(재창작)하는 기




〈자료〉 Gaudio Lab, Inc., 2021.

[그림 2] Being There Recreate System(BTRS)

술로 BTRS(Being There Recreate System)라는 기술이 있다[14]. 라이브 사운드 엔지니어가 믹스한 스테레오를 이용하여 Speaker to Listener를 처리하고, 음색에 민감한 공간음을 최적화해 처리함으로써 PC나 모바일로 소비하고 있는 사용자로 하여금 현장에 있는 듯한 느낌을 제공할 수 있다([그림 2] 참조).

최근 많이 시도되고 있는 언택트 콘서트의 경우 초록색 스크린으로 둘러싸인 크로마키 스튜디오에서 진행된다. 반면, 가상환경에 연출된 공간은 스타디움이나, 콘서트홀, 길거리 등 어느 공간이든 될 수 있다. 이 경우 공연이 이루어지는 실제 공간과 시청자가 감상 시 경험하게 될 가상공간의 음향 차이가 존재하게 된다. BTRS에서는 Speaker to Listener 경로를 연출된 가상공간으로 처리하는 기능을 제공하여, 제작 현장이 아닌 가상의 공간을 재창작할 수 있다.

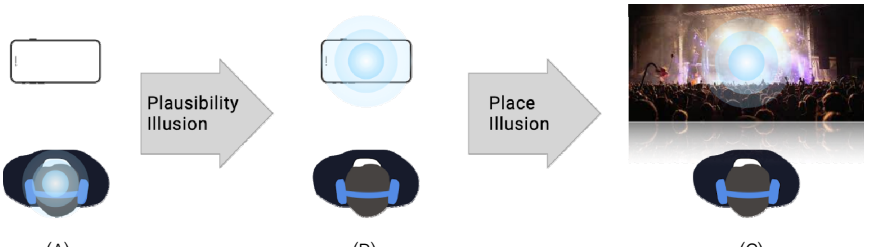
한편, TV나 PC, 스마트폰 등 기존의 2D 스크린을 통해 이와 같은 라이브 콘텐츠를 감상할 경우 시청자는 6DoF의 자유도를 만끽할 수 없다. 즉, Being There를 완전히 실현하는 것이 불가능하다. 그럼에도 불구하고 영상이 재생되는 위치, 즉 현장 공간의 특정 장면을 비추는 카메라의 위치에 대응하여 공간 음향을 재현하는 것은, 카메라 위치와 상관 없이 항상 고정된 음향 장면(Sound Scene)을 재현하는 것에 비교하여 한층 높은 몰입감을 제공할 수 있다. BTRS 기반으로 음원을 생성한 경우 이미 모든 위치에 대응하는 음향 장면을 재현할 수 있는 상태이므로, 2D 영상에서 화면 전환이 이뤄질 때마다 해당 화면에 적응적으로 공간 음향을 제공하는 것 역시 가능하다. 이를 VAA(View-Adaptive Audio)라고 부른다. 한편, 최근의 동영상 스트리밍 기술 가운데 연출자의 연출에 따른 장면전환에 의존하지 않고, 사용자가 직접 특정 카메라 위치를 선택하여 원하는 뷰를 시청할 수 있는 멀티뷰 기능이 점차 확대되고 있다. BTRS와 결합된 VAA 기술은 이와 같이 사용자가 임의로 선택한 멀티뷰에 대응하여 적절한 Psi를 제공하는데도 유용하게 사용할 수 있다.

## V. 공간 음향 기술의 모바일 응용

IV장에서 설명한 BTRS 및 VAA와 같은 공간 음향을 가상/증강현실뿐 아니라 스마트폰과 같은 모바일 환경에서도 효과적으로 재현하기 위한 기술이 요구된다. 본 장에서는 헤드폰 기반 모바일 환경에서의 공간 음향 재현에 대한 내용을 살펴본다.

애플이 그들의 Hearables[1] 기기인 에어팟에 공간 음향이라고 명명한 기능을 탑재하면서 [15], 일반 소비자 시장에서 공간 음향이 소위 캐즘을 훌쩍 뛰어넘어 전기 대중(Early-Majority) 사용자로 빠르게 나아가고 있다. 그동안 가상/증강현실을 위한 HMD 기기에 장착되곤 하던 머리 움직임을 인식할 수 있는 IMU(Inertial Measurement Unit) 센서가 헤드폰에 삽입됨으로써, 사용자의 머리 움직임에 실시간으로 반응하여 공간 음향을 제공하는 것이 가능해졌다. 이 기능을 탑재하면서 애플은 스마트폰(아이폰), 태블릿(아이패드), TV(애플TV) 등에서 재생되는 동영상에 대해서 영상에 일치하는 소리를 재현하여, 헤드폰만으로 극장에서와 같은 몰입감을 즐길 수 있다고 설명하고 있다.

Ⅲ장에서 설명한 것과 같이 일반적으로 헤드폰을 착용하면 머릿속에 음상이 맺히는 In-Head Localization이라는 문제에 봉착한다([그림 3] (A) 참조). 그렇기 때문에, 외재화를 통해 우선은 소리를 머리 밖으로 꺼내 줘야 한다. 이와 동시에 영상이 스마트폰에서 재생되고 있다면 소리 역시 스마트폰에서 재생되는 것처럼 느낄 수 있어야 한다. 스마트폰에 대한 위치 인식(이미 IMU가 내장되어 있다)과 사용자의 머리 및 회전에 대한 상대적인 인식이 필요하다. 이를 통해 사용자는 헤드폰을 쓰고 있다는 것을 잊고 스마트폰 자체에서 소리가 나는 듯한 착각의 경험을 하게 된다([그림 3] (B) 참조). 즉, Psi가 발생한다. 그리고 이는 영상 속 상황에 대한 몰입감을 극대화하여, 이제 현실 공간은 시청자의 머릿속에서 사라지고, 영상이 제공하는 가상세계 안으로 빠져들게 한다. Being There의 PI가 일어나는 것이다([그림 3] (C) 참조). 재생되고 있는 콘텐츠가 BTRS로 생성된 라이브 콘서트라고 하면, 사용




〈자료〉 Gaudio Lab, Inc., 2021

[그림 3] 공간 음향을 통한 Plausibility Illusion(Psi)과 Place Illusion(PI)

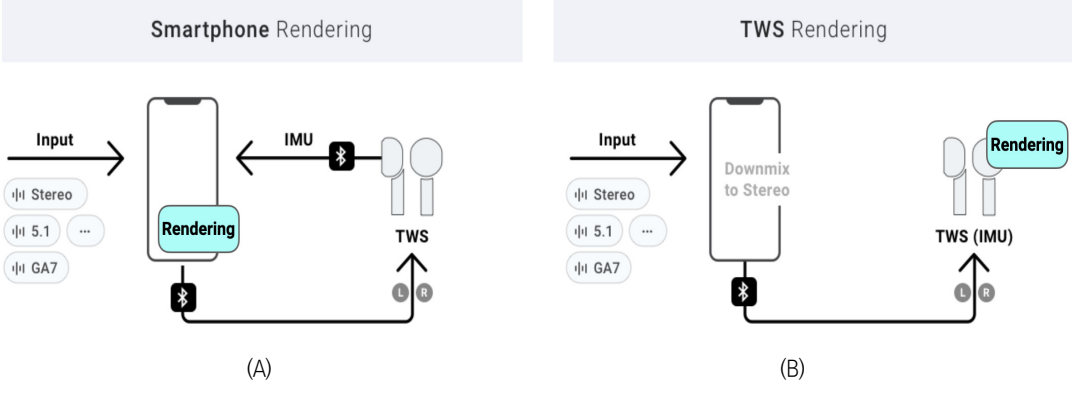
1) Hear와 웨어러블의 조합어로 귀에 착용하는 웨어러블이라는 의미. 한편 에어팟과 같이 양쪽 귀에 독립적으로 무선 연결이 되는 오디오 기기들을 TWS, True Wireless Stereo라고 부르기도 한다. 본 고에서는 필요한 경우를 제외하고 둘은 같은 의미로 사용한다.

자는 마치 공연장 안에 들어가서 함께 라이브를 즐기고 있는 것 같은 환영을 느끼게 된다.

이를 실현하는 데 있어서 필수 기술로 외재화가 필요하다고 설명하였는데, 이 외에 한 가지 더 기술적인 애로사항이 있다. Motion-to-Sound 지연(M2S Latency)이라고 부르는 문제인데, 사용자의 움직임이 실제 소리의 변화로 반영되어 사용자의 귀에 전달될 때까지의 경로가 유발하는 여러 시간 지연 요소의 합으로 표현되는 시간 지연 현상이다. 이 값을 일정 수준 이하로 낮추지 못하면, 더 이상 Psi가 실현되지 않는다.

무선 헤드폰이 연결되는 통로인 블루투스 통신이 제공하는 오디오 코덱의 호환성(5.1채널이나 MPEG-H, 애트모스와 같은 공간 음향 포맷을 전송하지 못한다)과 시중에서 흔히 볼 수 있는 Hearables의 연산 능력을 고려할 때, Being There 를 실현하기 위한 오디오 렌더링(바이노럴 렌더링이 필수 기술이다[12])을 스마트폰에서 수행하는 것이 바람직하다. 따라서, 머리 움직임을 수신하는 센서가 있는 Hearables로부터 스마트폰까지와, 센서 신호와 오디오 신호를 결합하여 바이노럴 렌더링을 스마트폰에서 수행한 후 스마트폰에서 다시 Hearables까지, 두 번의 무선 경로를 거치기 때문에 긴 M2S Latency의 문제를 안고 있다([그림 4] (A) 참조).

통상적으로는 M2S가 50msec 미만이어야 Psi가 일어난다고 알려져 있으나, 필자들의 실험에 따르면 일반 사용자의 사용 환경과 사용 시나리오를 고려할 때 200msec 미만으로만 유지하면 충분히 Psi가 발생하는 것으로 나타났다. 한 조사에 따르면, 실제 애플 에어팟에서의 M2S 지연이 약 204msec 정도라고 알려져 있다[16].




〈자료〉 Gaudio Lab, Inc., 2021

[그림 4] 공간 음향을 위한 Motion to Sound(M2S) Latency의 이해

한편, 공간 음향 처리를 Hearables 측에서 수행을 하게 되면, 앞서 언급한 M2S 문제를 현저하게 개선할 수 있다. IMU 값을 블루투스 무선 경로를 거치지 않고 바로 렌더링에 적용할 수 있기 때문이다([그림 4] (B)). 다만, 블루투스는 아직은 공간 음향 포맷을 수신할 수 없고 스테레오 신호만을 송수신할 수 있기 때문에 완전한 공간 음향을 실현하는 데는 한계가 있다. 음악, 방송물, 영화를 비롯하여 기존 시중에 유통되는 대부분의 콘텐츠가 스테레오인 점을 감안하면 Hearables에서 렌더링을 수행하는 형태의 구현 시나리오도 당분간은 유용할 것으로 전망한다. 대신, 이 경우 화면, 즉 스마트폰과 청취자 간의 상대적인 물리적 위치 관계를 정의할 수 있는 추가적인 정보 교환 방법이 고안되어야 한다.

오디오 렌더링에 있어 IMU를 활용하는 것은 공간 음향을 위한 Psi 완성을 위한 직접적 이유뿐만 아니라, 외재화를 향상시키는 효과도 있다. 사람은 애초에 소리의 공간 식별에 있어 머리의 미세한 회전을 활용하고 있기 때문이다. 동일한 입사각 위치에 있는 전, 후방의 대칭적인 음원에 대한 전, 후방 식별을 위해 사람은 무의식적으로 고개를 살짝 움직이면서 얻어지는 시차 신호(Parallax)를 시각 정보와 함께 사용하곤 한다. 이에 착안하여, M2S가 충분히 낮게 구현된 IMU 센서 기반의 바이노럴 렌더링을 수행하게 되면 외재화 효과를 혁신적으로 개선할 수 있다.

## VI. 증강현실 오디오

Hearables 기반의 공간 음향은 이를 소개할 증강현실 오디오(AR Audio)를 통해 보다 확장되고 완성된다.

이제는 잡음이 큰 환경에서 원치 않는 잡음을 차단할 수 있는 액티브 노이즈 캔슬링(ANC, Active Noise Cancelling) 기술이 적용된 헤드셋을 사용하는 것이 흔한 일이 되었다. ANC 기술을 대중적으로 가장 성공시킨 오디오 기기 전문 회사인 Bose는 Bose AR이라는 오디오 중심의 증강현실 플랫폼을 통해 새로운 비전을 제시하였다[17].

Bose AR은 선글라스 형태에 자이로 센서와 스피커가 결합된 형태로, 스마트폰과 블루투스로 연결되어 스마트폰에서 동작하는 애플리케이션에서 전달되는 각종 정보를 착용자가 오디오 형태로 수신할 수 있는 증강현실 오디오 플랫폼이다. 일반적으로 증강현실을 생각하면 떠오르는 기기의 형상은 홀로렌즈(Hololens), 매직립(MagicLeap)처럼 Glass 형태로,

착용자가 바라보는 현실에 가상의 객체를 시각적으로 형상화하여 보여주는 게 일반적이었다. 그러나 증강현실 오디오에서는 기기를 직접 쳐다보지 않아도 소리를 통해 원하는 정보를 편리하게 얻을 수 있어 활동성에 제약을 받지 않아도 된다는 장점이 있다. 최근 현대인들의 필수품이 되어가고 있는 TWS는 이와 같은 증강현실 오디오가 활용되기 좋은 플랫폼이며, 이미 애플의 시리, 아마존의 알렉사와 같은 AI 가상 비서는 이미 Hearables을 주요 활용처로 사용하고 있다. 애플의 에어팟은 증강현실 오디오 실현이라는 담대한 로드맵을 가지고 시작된 Wearables라고 생각할 수 있다.

증강현실 오디오 플랫폼에서 활용될 수 있는 애플리케이션은 다음과 같이 다양한 시나리오로 분류해볼 수 있다.

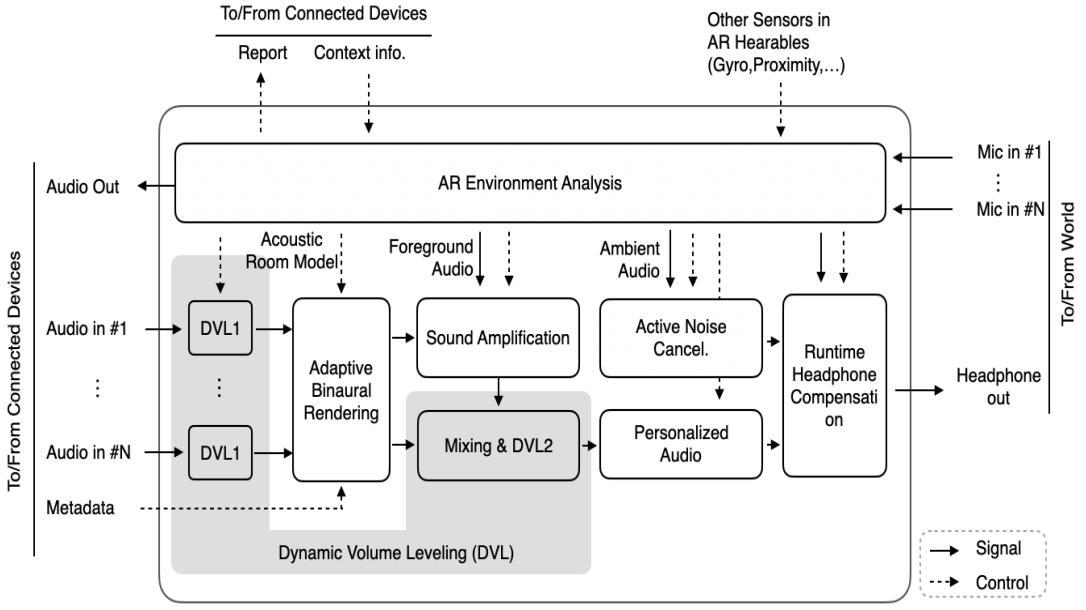
- Music: 공간 음향과 결합될 경우 재생되는 음악이 현재 공간의 특정 위치(스피커)나 가상의 객체(악기)에서 재생되는 것과 같은 환영(Being Here)을 제공. 또는 콘서트 장의 느낌을 재현(Being There)
- Navigation: 지도상에 위치가 표시되지 않더라도 원하는 곳을 찾아갈 수 있게 사용자의 위치를 기반(Location-based)으로 음성 안내를 제공. 소리의 방향이 실제 위치와 일치함으로써 사용자의 움직임에 맞춰 정확한 안내가 가능. 골퍼들을 위한 음성 기반 캐디 기능은 증강현실 오디오의 시장 사례
- Audio Guide: 박물관, 전시관 등에서 사용자의 위치를 기반으로 해당 작품에 대한 음성 가이드를 제공. 현장에서 별도의 기기를 대여할 필요 없이 개인이 착용하고 있는 증강현실 오디오 기기에서 바로 버추얼 도슨트 형태로 콘텐츠를 제공
- Blind Assistant: 시각장애인들이 일상적으로 생활할 수 있게 각종 시각 정보나 텍스트를 음성으로 변환하여 위치와 방향에 맞게 안내
- Personal Sound Amplification: 각종 보고서에 따르면, 현대인의 과도한 도시 인공잡음 노출, 장시간 헤드폰 사용 등으로 난청 인구가 급속히 늘어날 전망이며, 예방을 위한 국제보건기구(WHO)에서 권고안 제시[18]. 개인별 주파수 보정을 제공해 별도 보청기 없이 난청을 해소하고 예방
- Military: 군사 작전 수행 시 증강현실 기기를 착용한 병사는 현장 상황에 시선을 잃지 않고 각종 명령 및 후방 정보를 음향으로 수신하여 작전 수행이 가능
- Work: 산업 현장에서 증강현실 오디오를 통해 음성 통신 및 공간과 위치에 맞는 안내

기능을 제공. 작업자의 시야를 가리지 않고도 필요한 정보 전달이 가능

- AI Assistant: 애플 시리, 아마존 알렉사, 헤이 구글 등은 이미 장착되어 있으며, 위치/방향 정보 추가 및 주변 맥락 정보를 바탕으로 다양한 가치 창출이 가능

이상의 시나리오를 완성하기 위해서 갖춰야 할 증강현실 오디오 플랫폼의 구성 요소는 크게 하드웨어와 소프트웨어로 구분될 수 있다. 기본적으로 현실 세계의 정보를 센싱하거나 입력받기 위한 자이로 센서, GPS, 마이크 등의 입력 장치와, 정보를 처리하기 위한 연산장치와 메모리, 최종적으로 사용자에게 오디오 신호를 재생할 수 있는 트랜스듀서/스피커 등이 하드웨어에 포함된다. 소프트웨어에는 입력받은 센서 정보나 음성 입력을 분석하고, 출력 오디오 신호를 처리/생성하기 위한 알고리즘, 앞서 소개한 사용자 시나리오를 위한 애플리케이션 등이 포함될 것이다.

이와 같은 증강현실 오디오는 [그림 5]와 같은 기술 구성요소를 갖춰야 한다. 특히, 현실 세계의 음향 정보를 분석하여 새로 가공된 가상의 오디오 객체나 정보를 자연스럽게 증강 재현할 수 있는 처리 기술이 핵심이 될 것으로 예상되며, 이 부분에서 기술적인 과제가 많이




* 증강현실 오디오 실현을 위한 오디오 신호와 각종 센서 정보의 입/출력, 분석/처리부 등의 기술 구성 요소를 포함하는 SDK의 예
〈자료〉 Gaudio Lab, Inc., 2021.

[그림 5] 증강현실 오디오 기술 구성 요소

남아있다. 가령, 가상공간에서 객체를 렌더링하기 위해서는 현실 세계의 음향 환경은 분리/제외되어야 하기 때문에 ANC 기술로 차음된 헤드폰을 이용한 공간 음향이 적용된 오디오 재생만으로도 충분하다. 하지만 증강현실 오디오에서는 현실 세계의 소리가 사용자에게 그대로 전달되고, 이와 자연스럽게 어우러지는 증강 객체를 렌더링하여 재생하는 것이 몰입감을 제공하기 위해서 필수적이다. 때문에 현실 세계의 음향 환경이 분석되어 렌더링에 활용되어야 한다. 예를 들어, 강당과 같이 공간이 크고 울림이 많은 실제 공간에서 증강 객체는 긴 잔향음으로 렌더링되어야 하고, 사무실이나 침실같이 공간이 작고 잔향이 적은 곳에서도 그에 맞는 렌더링이 필요하다. 이와 같은 적응형(Adaptive) 바이노럴 렌더링(Binaural Rendering) 기술, 즉 Being Here 기술은 사용자가 현재 실재하는 공간의 BRIR(Binaural Room Impulse Response)을 지연시간 없이 실시간으로 분석/확인하는 것이 중요하며, 이를 위한 하드웨어 및 소프트웨어의 고도화가 요구된다.

## VII. 결론

본 고에서는 메타버스를 실현함에 있어 내가 그곳에 있는 듯한 경험(Being There)을 추구하는 가상현실과 실재하지 않지만 여기 있는 듯한 경험(Being Here)을 추구하는 증강현실을 구분하고, 각각이 오디오에서 갖는 의미와 이를 실현하기 위해 필요한 기술 요소들을 다뤘다. 혼합현실(Mixed Reality)이라는 개념 속에서는 이 둘은 혼재된다. 해석하는 방법에 따라 혼재한다는 말 자체가 모순일 수 있다. 그렇지만 하나의 기기에서 상황에 따라 두 가지를 함께 실현할 수 있다면 병립할 수 있는 개념이고, 그것이 메타버스로의 융합일 것이다. 예를 들어, 원격 회의를 하는데 그 장소가 내가 있는 현실 속 이곳 강의실이면 증강현실이고, 상대방의 장소인 그곳 혹은 원래는 없던 제3의 장소로 재현하느냐의 차이이다. 이를 자유자재로 바꾸는 것은 단순한 재미를 넘어 큰 가치를 만들 수 있다. 오디오는 그 단독으로도, 그리고 시각 등의 다른 감각과의 결합으로도 이를 실현하는데 필수적이며, 티핑 포인트에 가장 가까이 다가온 기술이기도 하다.

## 참고문헌


[1] Neal Stephenson, Snow Crash. Bantam Books, 1992.
[2] 서성은, 메타버스 개발동향 및 발전전망 연구, 한국컴퓨터게임학회논문지, No.12, 2008.
[3] 한혜원, 메타버스 내 가상세계의 유형 및 발전방향 연구, 디지털콘텐츠학회 논문지 제9권 제 2호, 2008.
[4] Smart, J.M., Cascio, J. and Paffendorf, J., Metaverse Roadmap Overview, 2007.
[5] 김상균, 메타버스 - 디지털 지구, 뜨는 것들의 세상, 플랜비디자인, 2020.
[6] Seeing is believing, How virtual reality and augmented reality are transforming business and the economy, PwC, 2019.
[7] 가우디오랩 블로그, AirPods Max와 함께 성큼 다가온 Spatial Audio 시대, 2021.
[8] 가우디오랩 블로그, 요즘 핫한 Spatial Audio, 2020.
[9] M. Slater, Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments, Philosophical Transactions of the Royal Society B, 2009, pp.3549-3557.
[10] R. Nordahl and N. C. Nilsson, The Sound of Being There: Presence and Interactive Audio in Immersive Virtual Reality, in The Oxford Handbook of Interactive Audio, Oxford University Press, 2018, pp.213-233.
[11] A. Rovira, D. Swapp, B. Spanlang and M. Slater, The Use of Virtual reality in the study of people's responses to Violent incidents, Frontiers in Behavioral Neuroscience, 2009, pp.1-10.
[12] 정현주, 오현오, VR/AR 오디오 기술 및 표준화 동향, 주간기술동향 1884호, 정보통신기획평가원, 2019.
[13] 동아일보 기사, "가상공간서 아바타 공연 관람"… IT-엔터 콘텐츠 융합 본격화, 2021.
[14] 이태규, 김대황 발표, 언택트 시대의 라이브 오디오기술 - "Being There" for Everyone, DEVIEW 2020, 2020.
[15] 공간 음향 소개, 애플 홈페이지, https://support.apple.com/ko-kr/HT211775, 2021.
[16] Kinicho(Volumetric Audio), Latency in Spatial Audio, https://medium.com/@kinicho/latency-in-spatial-audio-57236c15f243, 2020.
[17] SXSW 2018, Hear what you see. Experience Bose AR, https://schedule.sxsw.com/2018/events/OE2508, 2018.
[18] Hearing loss due to recreational exposure to loud sounds: a review, WHO Library Cataloguing-in-Publication Data, World Health Organization, 2015.