AR 및 VR 을 위한 공간 오디오 렌더링

*김상욱†, **강경옥, **이태진
*중앙대학교, **한국전자통신연구원
*swkim4@cau.ac.kr, **kokang@etri.re.kr, **tjlee@etri.re.kr

Spatial Audio Rendering for AR and VR

*Sang-Wook Kim, **Kyeongok Kang, **Taejin Lee
*Chung-Ang University **Electronics and Telecommunications Research Institute

요 약

본 논문에서는 현실세계에서 사용되던 오디오 처리 기법을 가상현실과 증강현실로 확장하는 기술에 대해 제시한다. 메타버스 서비스 구축 등에 활용되는 가상현실 공간을 설계할 때에는 오디오 처리를 위해서 가상현실 공간내 사용자가 위치하는 장면에 따른 소리의 회절과 반사에 따른 잔향 효과를 고려해 줄 수 있어야 장면에 몰입된 사용자 경험이 가능하다. 증강현실 응용에서는 실제 정보와 증강된 효과를 제공하기 위해 가상과 실제 정보간의 위치 정합이 영상 또는 위치를 기반으로 하여 제공되어야 한다. 가상현실과 증강현실 지원을 위해 현실세계 오디오 재생 기술에 추가되어야 하는 기술들과 함께 진행중인 몰입형오디오 서비스를 제공하기 위한 국제표준 기술 개발의 현황을 살펴보고, 향후 추가로 기술이 개발되고 보완되어야 할 부분을 제시한다.

1. 서론

최근 들어 진행된 컴퓨팅 기술과 그래픽 가속기의 발전은 고속연산이 가능하게 하였고, 고품질의 그래픽 영상으로 제작된 합성 영상을 사용자가 즐길 수 있게 하였다. 현실세계에서의 생활을 가상현실과 증강현실을 지원하는 생활로 확장하고 있다. 현실세계에서의 오디오는 몰입감을 높여주기 위해, 재생에 사용되는 오디오 채널 수를 늘려준 다채널 오디오와 공간상에서 소리의 위치를 제어하는 객체 단위의 3 차원 오디오 제어 기술이 개발되었다. 가상현실과 증강현실에서의 오디오는 현실세계의 오디오 재생과는 다르게 고려 되어야 하는 점들이 있다[1],[2]. 예로, 현실세계의 오디오 재생에서는 극장에서 공연이나 영화를 감상할 때와 같이 장면과 동기되어 제공되는 소리를 동시에 듣게 되는 이외에는 눈으로 보이는 장면과 사운드의 동기를 필요로 하지 않는다. 이에 반하여, 가상현실의 경우에는 장면과 음원의 위치가 고정되어 설계되어 있는 장면에서, 사용자가 내비게이션을 하면서 체험을 하게 되고 이 경우 사용자에게 장면에서 보이는 공간에 속해 있다 라는 느낌을 제공해 주는 것이 필요하게 된다. 공간상의 느낌을 제공해 주기 위해, 건축 음향 기술의 적용이 필요하다. 증강현실의 경우에는 현실세계와 가상세계의 장면이 중첩되는 경우로 생각할 수 있다. 이러한 경우, 현실세계 또는 가상현실의 의미 있는 객체의 위치와 음원정보가 연결되어 제공되는 것이 필요하다. 사용자가 시야나 위치가 이동함에 따라 보이는 장면내에 있는 증강현실 정보를 내재한 객체들을 사용자에게 알려주는 방법도 제공되어야 한다[3],[4],[5]. 최근 메타버스를 새로운 기회영역으로 하여 관련된 연구 개발이 활발하게 진행되고 있다. 메타버스의 구성에는 가상현실과 증강현실 기술의 접목이 될 필요가 있어서 새로운 국제표준 기술도 개발 중에 있다[6],[7],[8]. 본 논문에서는 메타버스 구성에 필요한 가상현실과 증강현실을 위한 오디오 렌더링 기술에 대해 이야기한다.

본 논문의 구성은 다음과 같다. 2 절에서는 가상현실 및

증강현실을 위한 공간 오디오 렌더링 기술의 구조에 대해 다룬다. 3 절에서는 몰입형 오디오 제공을 위해 국제 기구에서 표준 개발 진행 중인 MPEG-I Immersive Audio 기술을 중심으로 몰입형 오디오 기술의 구성 요소 기술들을 이야기한다. 4 절에서는 표준기술 선정을 위한 테스트 결과를 보이고, 5 절에서 가상현실 및 증강현실을 위한 공간 오디오 렌더링 기술의 현 이슈들을 중심으로 결론을 이야기한다.

2. VR 및 AR 을 위한 공간 오디오 렌더링

가상현실내 공간에서 몰입형 공간 오디오를 렌더링하기 위해서는 음원과 청취자 간의 관계, 공간을 구성하는 각 객체의 특성 등을 알고 있어야 현장감 있는 재생이 가능하다. 가상현실 및 증강현실을 위한 공간 오디오 렌더링 기술의 기본 구조를 위한 인터페이스는 그림 1 에서 보인다. 오디오 입력과 함께 사용자의 위치와 움직임 정보, 관심 대상 객체의 위치와 움직임 정보, 장면 메타데이터를 입력으로 한다. 입력된 정보들을 가지고 렌더링하여, 가상의 공간에서 사용자 움직임을 고려한 사운드를 만들어 제공한다. 공간 오디오 장면 재생을 위하여 장면을 구성하는 공간의 물리적 음향 특성, 공간내 존재하는 객체와 장애물에 의한 회절 효과, 객체의 고속 이동시 도플러 효과 지원, 헤드폰 재생인지 멀티채널 스피커 재생인 경우에 따른 처리, 청취자 주변의 현장음과 원격지 사용자의 주변음에 대한 렌더링도 함께 고려되어 제공되어야 한다

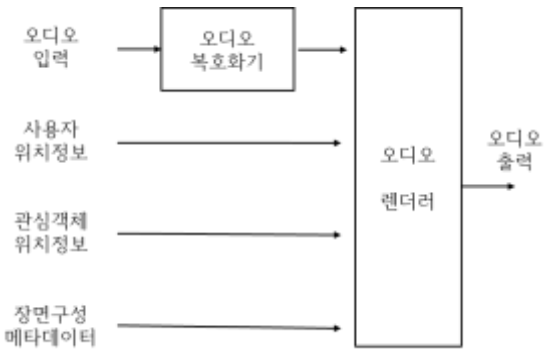

그림 1. VR 및 AR 을 위한 공간 오디오 렌더링 기본 구조

가상현실의 경우에서는 방의 크기와 모양, 재질에 따른 영향을 고려해 주어야 하고, 소리가 전달되는 경로에 책상이나 벽과 같은 장애물이 있을 경우, 소리의 반사와 회절 현상 등을 고려해 주어야 한다. 이러한 것은 장면을 구성하고 설계하여 사용할 때에, 미리 동작 가능한 시나리오를 구성하여 사용한다.

증강현실의 경우를 간단하게 구성하면 현실세계의 객체에 QR 코드를 부여하고, 관심있는 객체의 QR 코드를 이용해 반응하도록 하는 경우가 있을 수 있고, 복잡하게 구성한다면 현실세계의 장면을 인식하여 가상현실의 대상 장면을 구하고, 장면의 의미 있는 객체가 각각의 정보와 하이퍼링크되어 동작하는 경우가 있을 수 있다. 이 경우 현실세계를 분석하여, 가상현실의 장면을 구성하는 객체의 문맥정보 분석, 객체 동작 및 반응에 대한 물리 모델링(physical modelling) 등에 대한 기술이 필요하게 된다.

3. 몰입형 오디오 서비스를 위한 국제 표준기술

몰입형 오디오를 위한 표준 기술의 개발이 ISO/IEC JTC1/SC29/WG6 에서 진행 중이다. 2022 년 1 월 WG6 는 MPEG-I Immersive Audio CfP (Call for Proposals)결과를 가지고 RM0 (Reference Model 0) 기술을 결정하였다. 정해진 RM0 인코더의 대략적인 구조는 그림 2 와 같다.

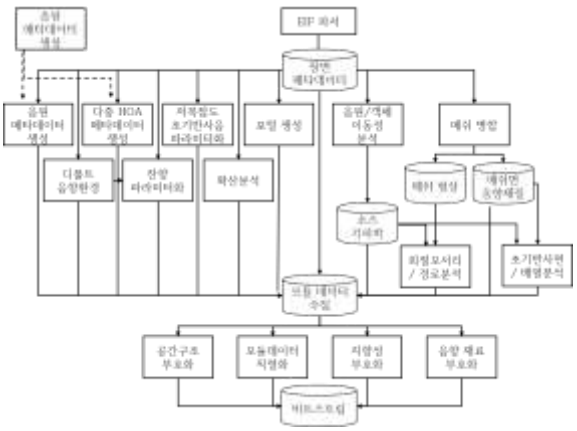

그림 2. MPEG-I Immersive Audio RM0 인코더 구조[7]

EIF (Encoder Input Format) 파서에서는 콘텐츠의 장면 정보를 구성하는 요소들을 추출한다. 공간의 기하학적인 구조, 음원의 위치, 방향성 등의 정보, 공간을 구성하는 객체들의 재료 특성, 공간의 음향 특성, 그리고 움직임 정보를 포함하는 업데이트 정보 장면 정보들이 구성 요소들로 사용된다. 음원 메타데이터 생성에서는 음원 신호의 특성 분석에 따른 거리감쇠 모델과 다중 음원을 가진 볼륨 음원의 확산 방사 특성 등을 가지고 렌더링 파라미터들을 만든다. 다중 HOA (High Order Ambisonics) 메타데이터 생성에서는 음원의 도착 방향 정보 DOA (Direction Of Arrival), 총에너지 크기, 총에너지와 직접파

에너지의 비율 DTR (Direct to Total energy Ratio)을 구한다. 잔향파라미터화에서는 RT60, 초기 지연, DDR (Diffuse to Direct Ratio)의 값에 기반하여 FDN (Feedback Delay Network)을 위한 감쇠 및 지연 파라미터, 그리고 음원 지향성에 따른 잔향의 방향성 필터 파라미터를 산출한다. 저복잡도 초기반사 파라미터화에서는 초기 반사음의 통계적 패턴 파라미터를 이용하여 초기반사음을 근사하여 추정해 준다. 포털 생성에서는 장면의 건축물 구조를 고려하여 방과 방 사이 혹은 외부와의 통로를 통하여 전달되는 음향에 대한 파라미터를 구한다. 음원이 있는 방과 일부 개방된 공간인 포털을 통하여 다른 공간의 청취자에게 전달되는 것을 모델링한다. 회절 모서리/경로 분석에서는 렌더러에서 회절효과를 구현하기 위한 회절경로를 구한다. 초기반사면 및 배열 분석에서는 렌더러의 초기반사음 연산을 가속화시키기 위한 파라미터들을 미리 추출한다. 이러한 각 모듈에서 생성된 파라미터들이 수집되어, 공간 구조, 음향 재료 등의 메타데이터 및 SOFA (Spatially Oriented Format for Acoustics) 파일의 지향성 정보가 양자화 처리되어 비트스트림을 구성한다.

렌더러에서는 그림 3 과 같이 비트스트림과 외부와 연결된 인터페이스로부터 디코딩 오디오, 클럭, 청취자 공간 정보 LSDF (Listener Space Description Format) 등을 입력 받아서 렌더링을 수행한 뒤 출력되는 오디오 신호를 구한다[7].

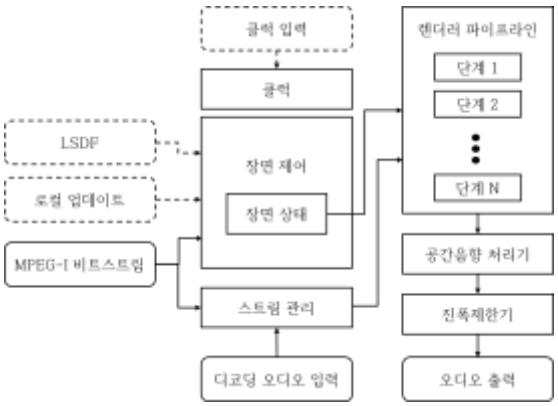

그림 3. RM0 렌더러 구조[7]

장면 제어에서는 로컬 혹은 외부의 모든 장면 정보의 변화를 처리해 준다. 비트스트림에 의해 전송된 장면 갱신 정보, 렌더러의 외부 인터페이스로부터 입력되는 청취자 위치 및 동적 갱신 정보, 그리고 AR 응용을 위한 LSDF 에서 정의된 음향 요소들을 가지고 현재의 장면 정보를 갱신한다. 스트림 관리에서는 장면 제어 정보의 음향 요소에 관련된 음향 신호를 입력하는 인터페이스를 제공한다. 렌더러는 스트림 관리로부터 제공된 음향 신호를 현재의 장면 정보를 이용하여 렌더링한다. 클럭은 장면의 현재 시간 정보를 장면 제어에 제공한다. 이 클럭 입력은 다른 외부 모듈들과의 동기신호가 될 수 있고, 렌더러 내부의 기준 시간이 될 수도 있다. 렌더러 파이프라인에서는 렌더링되어야 할 아이템을 가지고 선택된 재생 방법에 따라 신호를 제어하고, 공간음향 모듈과 진폭제한기를 거쳐서 오디오 출력 신호를 만들어 준다.

렌더러 파이프라인에서는 그림 4 와 같은 순서로 단계들을 진행한 뒤 공간음향 처리기에 입력 값을 전달해 준다. 필요에 따라서 일부 단계는 수행하지 않을 수도 있다.

그림 4. RM0 렌더러 파이프라인[7]

4. 성능 평가

표준 개발을 위해 진행된 주관적 성능 평가는 세가지 부분에서 진행되었다. 평가 1 에서는 객체, 채널, 3DoF HOA 기반 가상현실 사용 환경, 평가 2 에서는 객체, 채널 기반 증강현실 사용 환경, 평가 3 에서는 객체, 채널, 6DoF HOA 기반 VR 에 대해 진행되었다.

최종 평가는 7 개 기관에서 제출한 두 개씩의 렌더러를 대상으로 렌더링 결과에 대한 주관평가가 실시되었다. 주관평가에는 총 12 개 기관이 참여하였으며, 평가 1 에는 총 82 명, 평가 2 에는 총 39 명, 그 리고 평가 3 에는 55 명이 피험자로 참여하여 AB 비교를 통한 주관평가가 시행되었다. 평가 1 에서의 주관평가 시험 결과는 그림 5 와 같으며, 프라운호퍼, 에릭슨, 노키아 연합의 P13 이 최고 득점을 하여 RM0 base 기술로 선정되었다.

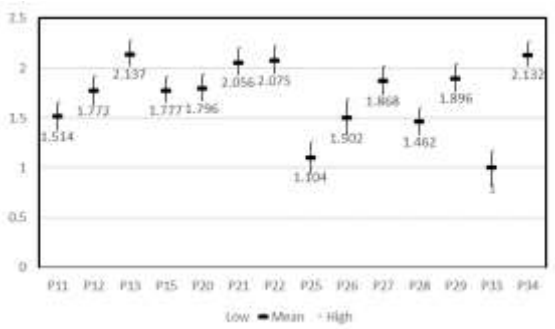

그림 5. 주관평가 시험 결과-평가 1 [9]

P13 과 동일한 렌더러 구조로 파라메터들을 달리 가져간 P21 와 P12 가 있다. P21 은 평가 2 에서 P12 는 평가 3 에서 최고 기술로 선정되어 P13 기반한 구조가 제일 좋은 성능을 보였다. 제안된 기술들 가운데 비트율 범주내에서의 우수한 기술로는 필립스, 돌비, 퀄컴 연합의 P27 기술이 선정되었고, P13 와 P27 을 기반으로 하여 RM0 를 구성하였다. 앞으로 진행되는 추가 기술제안에 따라 제안되는 기술들을 추가 검토하고, 채택여부를 결정하여 가상현실과 증강현실에서의 응용을 목표로 하는 국제 표준 오디오 기술을 제정할 계획이다.

5. 결론

지금까지 가상현실과 증강현실 응용에서 사용자에게 몰입형 오디오 재생을 제공하기 위한 기술의 이슈와 표준화 동향에 대해 다루었다. 가상현실과 증강현실에서는 모델링된 가상 공간구조에 대한 건축음향 기술이 제공되어야 하고, 공간상에서 사용자가 이동하거나 움직일 때 장면과 객체에 맞춤형 재생을 해주는 것이 필요하게 되어서, 몰입형 오디오 지원을 위한 새로운 국제 오디오 표준 기술이 개발 중에 있음을 살펴 보았다.

방송과 통신서비스 측면에서 볼 때에, 현재 개발 중인 국제 표준화 기술은 복잡도 측면에서 많은 개선이 필요하다. 새로운 표준은 일반인들이 사용할 수 있고 개인이 쉽게 방송할 수 있는 부호화 및 렌더러 솔루션 개발이 되어야 한다. 현재 개발중인 표준 기술은 복잡도가 높은데, 휴대용 기기를 지원하는 저전력 방식들의 개발이 되어야 사용 확대가 될 것이다. 이동하거나 움직임에 맞추어 사운드를 제공하기 위한 기술 개발도 필요하다. 센서 내장된 VR 헤드셋이 아닌 수단으로 사용자의 움직임 정보를 획득하는 방식에 대한 연구가 추가로 진행될 필요가 있다. 복잡도 문제가 해결이 되면, 메타버스 내에서도 실제 환경과 같은 경험을 하게 되는 것을 앞당기는 것이 가능하게 될 것이다.

Acknowledgement

본 연구는 한국전자통신연구원 연구운영비지원사업의 일환으로 수행되었음. [22ZH1200, 초실감 입체공간 미디어 · 콘텐츠 원천기술 연구]

참고문헌

[1] L. Savioja, J. Huopaniemi, T. Lokki, and R. Vaananen, "Creating Interactive Virtual Acoustic Environments", Jr. Audio Eng. Soc., v.47, no,9, pp.575-705, Sep. 1999.

[2] S. Liu and D. Manocha, "Sound Synthesis, Propagation, and Rendering: A Survey," eprint arXiv:2011.05538 (2020). https://doi.org/10.48550/arXiv.2011.05538

[3] N. Tsingos, E. Gallo, and G. Drettakis, "Perceptual Audio Rendering of Complex Virtual Environments," in Proceedings of SIGGRAPH, pp. 249-258, 2004.

[4] C. Schissler, C. Loftin, and D. Manocha, "Acoustic Classification and Optimization for Multi-Modal Rendering of Real-World Scenes," IEEE Tr. on Visualization and Computer Graphics, vol. 24, no. 3, pp.1246-1259, March 2017.

[5] H. Kim, L. Remaggi, A. Dourado, T. de Campos, P. Jackson, and A. Hilton, "Immersive Audio-Visual Scene Reproduction using Semantic Scene Reconstruction from 360 Cameras," Virtual Reality (2022) 26: 823-838, DOI: https://doi.org/10.1007/s10055-021-00594-3.

[6] S.R. Quackenbush and J. Herre, "MPEG standards for compressed representation of immersive audio," Proc. IEEE, vol. 109, no. 9, pp. 1578-1589, 2021.

[7] WG6 MPEG Audio Coding, WD0 v2 of ISO/IEC 23094-4, Immersive Audio, N0147, ISO/IEC JTC1/SC29/WG6, July 2022.

[8] 장대영, 강경옥, 이용주, 유재현, 이태진, MPEG-I Immersive Audio 표준화 및 기술 동향, ETRI Electronics and Telecommunications Trends, 2022. DOI: https://doi.org/10.22648/ETRI.2022.J.370306

[9] WG6 MPEG Audio Coding, Report on MPEG-I Immersive Audio Call for Proposals, N0119, ISO/IEC JTC1/SC29/WG6, Jan. 2022.