

6 자유도 오디오 재현 기술의 평가 시스템 및 방법

이용주, 이미숙, 임우택, 강경옥, 이태진
한국전자통신연구원 미디어부호화연구실
draball@etri.re.kr

Evaluation system and method of 6 degree of freedom audio reproduction technology

Yong Ju Lee, Mi Suk Lee, Wootak Lim, Kyeongok Kang, Taejin Lee
Media Coding Research Section
Electronics and Telecommunications Research Institute (ETRI)

요 약

6 자유도 오디오 재현 기술은 사용자가 6 자유도를 가지는 콘텐츠 또는 응용에서 음향학적 주변 환경과 사용자의 위치에 맞는 오디오를 재현하는 기술로서, 가상 현실, 게임 등에 적용이 될 수 있다. 6 자유도 오디오는 기존의 고정된 위치에서 재생되는 채널 기반 오디오와는 다른 특징이 있어 오디오 재생 장치와 헤드폰 또는 스피커만으로 수행하던 기존의 오디오 청취 평가 방법으로 비교하고 평가하는 것이 불가능하다. 최근 MPEG 오디오 서브그룹에서는 6 자유도 오디오 재현 기술에 대한 표준화를 진행하고 있으며, 이를 평가할 수 있는 시스템도 함께 제시하고 있으며, 2019 년 12 월에는 제안한 시스템을 활용한 파일럿 테스트가 진행이 되었다. 본 논문에서는 MPEG 에서 진행한 파일럿 테스트 시스템과 평가 방법에 대하여 살펴보고, 본 기관에서 평가한 내용 및 결과에 대해 기술하며, MPEG 에서 제시한 평가 시스템이 6 자유도 오디오 재현 기술의 평가에 적절한지에 대한 의견을 제시한다.

1. 서론

6 자유도 오디오 재현 기술은 사용자가 6 자유도를 가지는 콘텐츠 또는 응용에서 음향학적 주변 환경과 사용자의 위치에 맞는 오디오를 재현하는 기술이다. 6 자유도(6DoF)는 3 차원 직각 좌표계에서 X 축 중심의 좌우 회전(Roll: Tilting side to side on the X-axis), Y 축 중심의 앞뒤 회전(Pitch: Tilting forward and backward on the Y-axis), Z 축 중심의 위아래 회전(Yaw: Turning left and right on the Z-axis) 동작과 앞뒤(Surge: Moving forward and backward on the X-axis), 좌우(Sway: Moving left and right on the Y-axis), 위아래(Heave: Moving up and down on the Z-axis) 병진 동작(translational motion)을 포함하며 [1], 최근에 제작되는 많은 게임이나 가상 현실에서 6 자유도를 제공하고 있다.

6 자유도 콘텐츠에서 사용자가 이동하거나 회전하는 경우, 눈에 보이는 영상이 달라지는 것처럼, 사용자의 귀에 전달되는 오디오 신호도 달라지는 것이 일반적이다. 이는 사용자가 이동하거나 회전에 따라 오디오 객체와의 거리나 방향 등이 달라지며, 이에 따라 오디오 객체로부터 사용자에게 전달되는 직접음과 반사음의 경로가 달라지기 때문이다. 이와 같은 요소들을 고려하여 6 자유도 콘텐츠에서 오디오를 재현하는 것은 복잡한 알고리즘과 높은 연산량을 요구하는 작업으로, RAVEN, EVERTIMS 등과 같은 비 실시간 연산에 중점을 둔 오디오 신호처리 모듈로 개발이 되고 있었고, 게임이나 가상 현실 등과 같은 실시간 연산이 요구되는 분야에서는 복잡한 기술이 적용되지 못하고 있었다. 그러나 최근에는 마이크로소프트, 구글 등과 같은 회사에서 6 자유도를 가지는 게임과 콘텐츠 개발에 사용할 수 있는 오디오 재현 모듈에 대해 개발을 하고, 관련 소프트웨어를 플러그인 형태로 제공하고 있다[2-5].

6 자유도 오디오는 기존의 고정된 위치에서 재생되는 채널 기반 오디오와는 다른 특징이 있다. 먼저 사용자의 위치가 변경될 수 있으며 사용자의 위치에 따라 다른 소리가 재현되어야 한다는 점, 그리고 오디오만 단독으로 제공되는 것이 아니라 영상과 함께 주어지는 점, 마지막으로 오디오가 주어지는 영상 화면과 음향학 적인 동기화를 이루어야 한다는 점 등이 이러한 특성이라 할 수 있다. 이와 같은 특성으로 인해, 오디오 재생 장치와 헤드폰 또는 스피커만으로 수행하던 기존의 오디오 청취 평가 방법으로는 6자유도 오디오 재현 기술을 비교하고 평가하는 것이 불가능하다. 현재로서는 서로 다른 6 자유도 오디오 재현 기술을 비교하고 평가하는 시스템이나 방법에 대한 국제적인 가이드가 없는 상태이며, 향후 6 자유도 오디오 재현 기술의 발전을 위해서는 신뢰할 수 있는 평가 시스템과 방법이 제공되어야 할 것으로 생각된다.

최근 MPEG 의 오디오 서브그룹에서는 6 자유도 오디오 재현 기술에 대한 표준화를 진행하고 있으며, 이를 평가할 수 있는 시스템도 함께 제시하여 2019 년 12 월에는 제안한 시스템을 활용한 파일럿 테스트가 진행이 되었다. 본 논문에서는 MPEG 에서 진행한 파일럿 테스트 시스템과 평가 방법에 대하여 간략하게 설명하고, 본 기관에서 평가한 내용 및 결과에 대한 기술을 통해 MPEG 에서 제시한 평가 시스템이 6 자유도 오디오 재현 기술의 평가에 적절할지에 대한 의견을 제시한다.

2. 6DoF 오디오 재현 기술 평가 시스템

MPEG 에서는 MPEG-I 오디오 기술의 성능평가를 위하여 6 자유도 가상현실 장비 및 도구를 활용한 평가 시스템을 구현하여 공개하고 있다[6]. 그림 1 은 MPEG 에서 제안한 평가 시스템을 간략화 하여 나타낸 것이다.

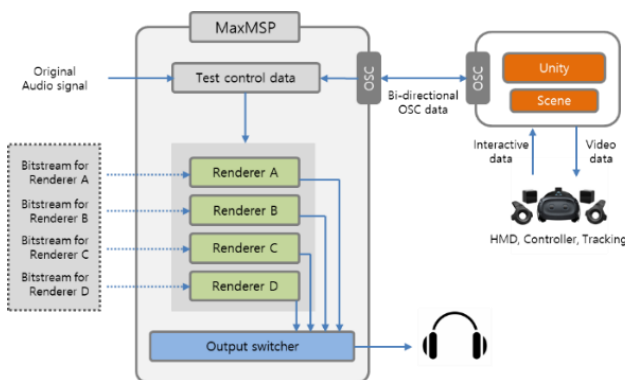


그림 1. MPEG-I audio 평가 시스템의 간략화 된 구조
Fig. 1. Simplified structure of MPEG-I audio evaluation system

평가 시스템은 청취자의 6 자유도 움직임을 지원하기 위하여 청취자의 위치 정보와 같은 상호작용 정보를 제공하는 가상현실 재현 장비인 HTC Vive 장비, 비디오 렌더링 및 평가 제어를 위한 유니티(Unity) 개발 도구, 그리고 오디오 렌더링을 위한 MaxMSP 로 구성된다[7].

유니티 프레임워크는 6 자유도 콘텐츠의 재생에 활용되는데, 평가자는 HMD 를 통해 영상을 보면서 콘텐츠 내의 공간을 자유롭게 이동할 수 있으며, 컨트롤러를 활용하여 콘텐츠와 상호 작용을 한다. 사용자의 이동 정보는 트래킹 장비를 통하여 측정되고 MaxMSP 프로그램으로 전달된다. MaxMSP 응용 프로그램은 사용자의 위치 정보와 인터랙션 정보 등을 Renderer 로 전달하며, 여러 기관에서 구현한 VST(Virtual Studio Technology)를 평가자가 선택하여 청취할 수 있도록 하는 기능과 평가 결과 등을 저장하는 기능이 포함된다. 실제 6 자유도 오디오 재현 알고리즘이 동작하는 부분은 MaxMSP 응용프로그램에서 동작하는 VST 로 구현이 되는데, 그림 1 에서 녹색으로 표기된 Renderer 가 이에 해당한다.

앞에서 살펴본 것과 같이, MPEG 에서 제시한 평가 시스템은 유니티, MaxMSP, VST 등으로 평가 시스템을 모듈화 하고, 각 모듈 간의 인터페이스를 정의하여 6DoF 영상 재생 모듈과 독립적으로 6DoF 오디오 재현 알고리즘이 동작할 수 있는 환경을 구축하였으며, 이로 인해 하나의 콘텐츠를 체험하는 동안 여러 개의 6DoF 오디오 재현 알고리즘을 서로 비교할 수 있는 환경을 제공한다.

3. 6DoF 오디오 재현 기술 평가 결과

MPEG 에서는 2019 년 12 월에 6 자유도 오디오 재현 기술의 평가 시스템과 평가 방법에 대한 검토를 위한 파일럿 테스트(Pilot test)가 수행되었다. 파일럿 테스트의 경우 그림 1 에 나타난 시스템 환경에서 테스트가 수행되어야 하나, 테스트 시스템의 구축이 완전하지 않은 상태라서 일부 변경 사항이 있는 상태로 테스트가 진행되었다. 원래의 테스트 시스템에서는 Renderer 를 위한 비트스트림이 Renderer 로 직접 제공되어야 하는데, 파일럿 테스트에서는 이 부분이 제공되지 않았다. 이로 인해 3 차원 공간에 대한 기하 정보가 Renderer 에 제공이 되지 못하였으며, 오디오 객체의 위치 정보는 MaxMSP 응용프로그램을 통해 Renderer 로 제공되었다.

평가 콘텐츠로는 ‘Singer in the Lab’, ‘Hospital’, ‘L-Shaped Room’, ‘Fountain Music’ 등 총 4 개의 6 자유도 360VR 콘텐츠가 사용되었으며, 아래의 표 1 에 기술한 4 개의 6 자유도 오디오 재현 기술에 대한 평가가 이루어졌다.

표 1. 파일럿 테스트에 사용된 6자유도 오디오 재현 기술
Table 1. 6DoF audio reproduction technology used in the pilot test

#	Systems Under Test
Cond1	GBRv2
Cond2	GBRv2AO with HRTFs offset by 20 degrees azimuth
Cond3	GBRv2LPF using 7 kHz LPF HRTFs
Cond4	Dual mono

4 개의 6 자유도 오디오 재현 기술은 모두 MPEG 에서 제공한 것으로 하나의 anchor 시스템과 3 개의 6 자유도 오디오 재현 기술로 구성된다. 3 개의 6 자유도 오디오 재현 시스템이라고하는 하나, 실제적으로는 하나의 6 자유도 오디오 재현 기술(GBRv2)과 이 기술에 약간의 변형(음상을 수평 방향으로 20 도 이동, 7kHz 저대역 필터링을 수행)을 가한 기술이다. 제공된 6 자유도 오디오 재현 기술의 상세 알고리즘에 대한 정보는 제공되지 않았다.

평가 방법으로는 MUSHRA VR 방법이 사용되었는데, 이는 일반적으로 오디오 평가에 사용하는 MUSHRA 방식을 VR 체험 환경에서 사용하도록 한 것으로, 360 VR 영상에 MUSHRA 평가 화면을 오버레이(overlay)하여 나타나도록 하고 평가자가 언제든지 여러 개의 6 자유도 오디오 재현 기술을 선택하여 청취하면서 점수를 책정할 수 있도록 하는 방식이다.

ETRI 에서는 4 명의 평가자가 평가를 수행하였으며, 각각의 콘텐츠 scene 에 대한 평가 결과는 그림 2 와 같다.

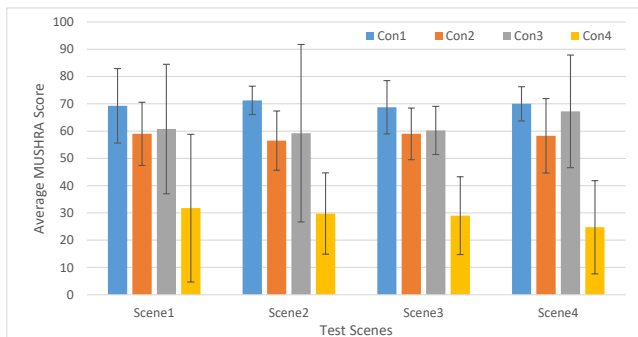


그림 2. MPEG-I audio 파일럿 테스트 평가 결과(95% 신뢰구간)
Fig. 2. The evaluation result of the MPEG-I audio pilot test (95% confidence interval)

평균 점수를 보면 4 개의 6 자유도 오디오 재현 기술이 4 개의 콘텐츠에서 항상 같은 순위를 가졌으며, 모든 Scene 에서 cond1, 2, 3 의 점수가 95% 신뢰구간에서 겹치는 것을 확인할 수 있다. 점수의 일관성 측면에서는 6 자유도 오디오 재현 기술에 대한 변별력이 있는 것처럼 보이기도 하나, 기술 간에 점수가 겹치는

것을 고려하면 변별력이 크다고 보기도 어렵다. 따라서 본 청취 평가 결과 만으로는 MPEG 에서 제시한 6 자유도 오디오 기술 평가 시스템 및 방법이 6 자유도 오디오 재현 기술을 평가하는 것에 사용되는 것이 적절한지에 대한 판단을 내리기에 어려움이 있다. MPEG 오디오 서브그룹의 다른 기관에서도 동일한 평가를 수행하고 있으므로, 그 결과를 포함하면 의미 있는 결과를 도출할 수도 있을 것이라 생각된다.

4. 결론

본 논문에서는 6 자유도 오디오 재현 기술의 평가를 위한 시스템과 방법으로 MPEG 에서 제시한 평가 시스템에 대해 살펴보고, 직접 평가를 수행하여 그 결과를 살펴보았다. MPEG 에서 제시한 평가 시스템은 유니티, MaxMSP, VST 등으로 평가 시스템을 모듈화 하고, 각 모듈 간의 인터페이스를 정의하여, 6 자유도 영상 재생 모듈과 독립적으로 6 자유도 오디오 재현 알고리즘이 동작할 수 있는 환경을 구축하였는데, 이는 하나의 콘텐츠를 체험하는 동안 여러 개의 6DoF 오디오 재현 알고리즘을 서로 비교할 수 있는 매우 효과적인 방법이라 생각이 된다. 평가에 사용된 MUSHRA VR 방법도 기존의 MUSHRA 평가에 익숙한 평가자들에게는 거부감이 없이 받아들여질 수 있을 것으로 생각이 되었다.

본 실험에서 한가지 아쉬운 점은 본 평가를 통해 수집한 청취 평가 결과 만으로는 MPEG 에서 제안한 6 자유도 오디오 재현 기술 평가 시스템 및 방법의 적절성에 대한 결론을 자체적으로 내릴 수는 없었던 것인데, 이는 MPEG 오디오 서브그룹의 타 기관에서 평가한 결과를 취합하여 해소할 수 있을 것으로 생각된다.

감사의 글

본 연구는 한국전자통신연구원 연구 운영 지원 사업의 일환으로 수행 되었음. [20ZH1200, 초실감 입체 공간 미디어, 콘텐츠 원천기술 연구]

참고문헌

[1] 한국정보통신기술협회, 정보통신용어사전.
[2] 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.

- [3] 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.
- [4] Nikunj Raghuvanshi, John Michael Snyder, "Parametric wave field coding for precomputed sound propagation", ACM Transactions on Graphics, July 2014 Article No.: 38
- [5] 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.
- [6] 이용주, 유재현, 장대영, 이미숙, 이태진, "체감형 미디어 서비스를 위한 공간음향 기술 동향", 전자통신동향분석 34 권 제 3 호 2019 년 6 월.
- [7] ISO/IEC JTC1/SC29/WG11/N17177, "MPEG-I Audio Architecture and Evaluation for 6DoF," 2017. 10.