



가상현실의 완성, 가상현실 오디오



서정훈

가우디오랩(주) 최고책임과학자
p@gaudiolab.com

서울대학교 전기컴퓨터공학부 박사
독일 RWTH-Aachen Institut für Technische Akustik 박사후연구원
서울대학교 초실감음향기술센터 연구원
가우디오랩(주) 최고책임과학자
관심분야: Binaural Hearing, Psychoacoustics, VR Audio



오현오

가우디오랩(주) 대표이사
o@gaudiolab.com

LG전자 Digital TV연구소 오디오 팀장
연세대학교 전자공학과 연구교수
(주) 굿데이투인벤트 대표이사
가우디오랩(주) 대표이사
관심분야: 3D Audio, Psychoacoustics, VR Audio

성장하는 가상현실, 그리고 오디오

구글, 페이스북, 삼성전자 등 국내외 유수의 기업들이 가상현실 (Virtual Reality, VR)과 관련된 사업 영역을 넓혀가고 있다. 언제쯤 상용화가 될까 싶은 생각이 들기도 했지만, 이미 You Tube에서는 360 영상을 감상할 수 있고 오쿨러스 (Oculus Rift)나 바이브 (Vive)와 기기를 이용해 가상현실 게임을 즐길 만큼 VR은 우리들 가까이 와 있다.

가상현실이 남다른 대중의 관심을 모으는 이유는 HMD 기반의 가상현실 콘텐츠가 혁신적인 사용자 경험을 부여하기 때문이다. 이 사용자 경험은 크게 두 종류로 나뉘는데 현장감과 몰입감이다. 기존 2D 및 3D 영상 출력 방식은 고정된 디스플레이가 한정하는 고정 범위의 시각을 통하여 제3자의 입장에서 미디어를 접하게 된다. 그러나 HMD의 경우 헤드 트래킹 기술과 눈 위에 직접 투사하는 양안 (Stereoscopic) 3D가 사용자의 움직임에 따른 전 방향의 시선에 대응하는 영상을 제공하기 때문에 사용자는 마치 그 장소에 있는 것 같은(Being There) 1인칭 시점의 경험이 가능하다. 이와 같이 사용자가 존재하는 현실과 가상 공간의 경계를 허무는 것을 이머전 (Immersion)이라고 부르고, 이를 가능하게 하는 기술을

이머시브 테크놀로지 (Immersive Technology)라고 부른다.

결국 가상현실은 대중에게 낮은 비용으로 거리, 시간, 혹은 장애를 초월하여 현장감과 몰입감이 충분한 경험을 제공할 수 있는 것이다. 그러나 최근 가상현실 기술의 발전을 보면 대부분 영상에 대한 기술 개발에 집중되고 있다. 소리가 우리 경험의 절반이라는 말을 되새긴다면, 가상현실이 진정한 이머시브를 제공하기 위해서는 가상현실에 적합한 VR 오디오 기술이 필수적으로 요구된다. 특히 가상현실을 위한 대표적 기기인 HMD를 착용하는 형태의 환경에서는 헤드폰/이어폰만을 이용한 VR 오디오가 필수적으로 요구된다.

HMD 및 헤드폰을 이용한 VR 오디오 시나리오에서는 사용자가 고개를 돌리거나 걷는 등의 6축(6 DOF; Degree of Freedom)의 움직임에 따른 반응이 이뤄져야 한다. 그림 1은 이와 같은 반응형 렌더링이 제대로 구현되지 않은 채 기존의 방법으로 입체음향만을 제공할 때의 문제를 예시한다. 따라서, VR 오디오에서는 몰입감 있는 소리를 제공하기 위한 3D 입체음향 기술과 함께 이에 대한 interaction을 실시간으로 재현할 수 있는 기술이 함께 요구된다.

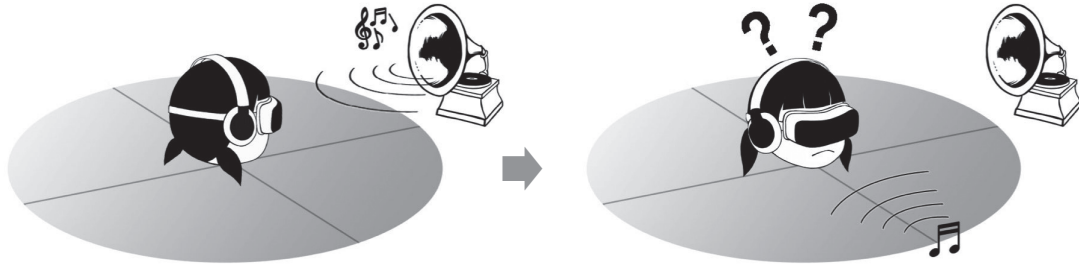


그림 1. 반응형 렌더링이 되지 않아 몰입감이 저하된 예

Gaming VR vs. Cinematic VR

현재의 VR 시장은 크게 둘로 나뉘볼 수 있다. 하나는 게임이 가상현실로 들어온 가상현실 게임이고, 다른 하나는 실사 영상을 기반으로 한 VR (이하 Cinematic VR)이다. 게임 플랫폼에서는 기존과 마찬가지로 게임상에서 진행되는 모든 구성요소와 배경 등의 환경이 합성된 가상 세계이므로, 오디오 관점에서도 모든 음원은 개별 음원객체 (Sound object)에 해당한다. 기존의 평면적인 게임 공간이 360도 공간으로 확장됨으로써 3차원 공간을 정확히 묘사하는 이머시브 오디오 기술은 VR 게임에서도 매우 중요해졌다. 캐릭터의 음성, 캐릭터의 움직임에 의해 발생하는 소리, 배경음 등 사운드에 관련된 모든 요소들은 게임 상의 시나리오에 따라 제작자의 의도대로 합성하고 배치되어야 한다.

Cinematic VR 은 사용자가 직접 촬영한 360 영상부터 기존의 영화, 스포츠 중계, 라이브 공연 등과 같은 미디어가 대체되는 영역을 의미한다. 이때, 기존의 이와 같은 미디어와는 다르게 VR화된 실사 영상 콘텐츠인 Cinematic VR에서는 고개 움직임은 물론 사용자의 이동 등의 행동에 따라 다른 영상과 소리가 제공되고 시나리오가 달라짐으로 인해 완전히 새로운 형태의 엔터테인먼트 미디어로 해석되고 있다. 할리우드에서는 스티븐 스필버그 등 유명 감독들이 이미 VR 을 기반으로 한 새로운 형태의 영화를 제작하기 시작했고, 이 과정에서 VR

오디오의 중요성은 다시 한번 주목되고 있다. 예를 들어, VR 영화와 같은 경우, 제작자의 의도와 다르게 사용자의 시선과 관심이 현재 장면에서 원하지 않는 방향을 향하고 있을 수 있는데, 이 때, 사용자를 다시 제작자가 의도한 방향으로 유도하는 행동 유도에 있어서 3차원 공간에서 정확한 위치에서 재생되는 소리는 가장 효과적인 연출 기법이 된다. 또한 가상현실 상에 펼쳐지는 환경에 적합한 공간감을 제공하기 위해서는 현장의 동시 녹음한 개별 음원, 후시 녹음, Foley 등에 추가되어, 현장감을 담고 있는 앰비언스 사운드의 역할이 더욱 중요해졌다. 그리고, 이와 같은 음원들을 효과적으로 VR에 맞게 믹싱하고 마스터링하는 후반 작업에 사용할 수 있는 새로운 VR 오디오 저작물이 시급하게 요구되고 있는 상황이다.

Binaural Hearing

사람은 양쪽 귀에 들어온 신호에 담겨있는 단서(cue)들을 추출하여 소리의 방향을 인지한다. 양쪽 귀에 들어온 신호를 바이노럴 (Binaural) 신호라고 하는데, 바이노럴은 “양이(兩耳)” 즉, “귀가 두개인”, 혹은 “두 개의 귀를 가진”이라는 의미이다. 바이노럴 신호가 가진 단서는 크게 두 가지로 나누어 설명할 수 있다.

첫 째는 양쪽 귀에 들어오는 신호의 차이를 인지하는 바이노럴 큐(Binaural cue)이다. 바이노럴 큐는 주



로 두 신호의 각 주파수별로 나타나는 레벨 차이와 시간 차이 등에 기인한 것으로서 수평면 상에서 소리의 방향(Azimuth)을 인식하는데 이용된다. 바이노럴 큐는 따라서 양이 레벨차 (Interaural level difference, ILD)와 양이 시간차 (Interaural time difference, ITD)라고 정의된 값을 가지게 된다. 만일 오른쪽에 중소리 음원이 있었다고 하면 오른쪽 귀에는 바로 소리가 전달 되는 반면 왼쪽 귀에는 양 쪽 귀 사이의 거리에 따른 시간 지연을 두고 전달된다. 또한 왼쪽 귀에는 머리, 몸통, 귀바퀴 등의 회전과 반사를 통해 신호가 전파되기 때문에 상대적으로 더 큰 감쇄(Head shadowing)가 발생한다. 따라서 사람은 주로 이러한 시간차(ITD)와 레벨차(ILD)를 이용하여 수평면 상의 방향을 인지한다.

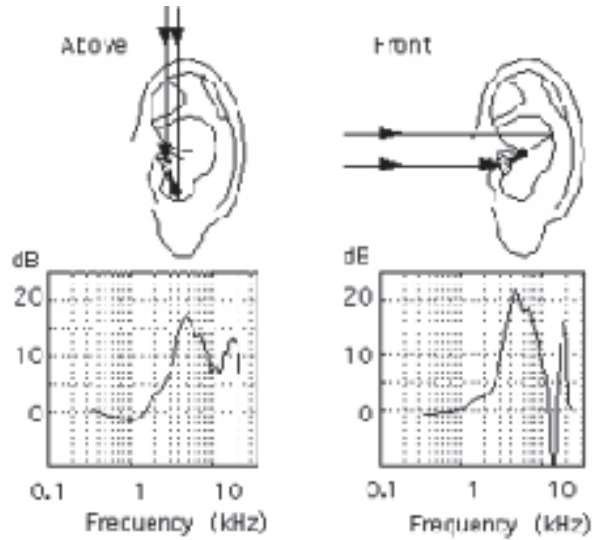


그림 3. 모노럴 큐 (출처: <http://interface.cipic.ucdavis.edu/sound/tutorial/psych.html>)

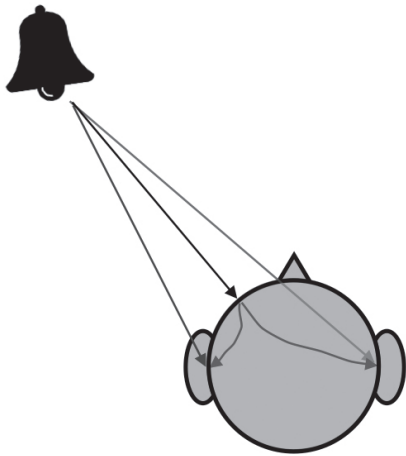


그림 2. 두 개의 귀를 가진, 바이노럴 (출처: 가우디오랩)

둘째는 양 쪽 귀에 공통으로 존재하는 모노럴 큐(monaural cue)로 신호 자체의 특성을 이용하여 지각하는 단서이다. 모노럴 큐의 경우 특히 소리의 고도(elevation)를 인지하는데 사용되는데 음원 위치의 고도각에 따라 입력된 신호의 주파수 특성이 달라지기 때문이다. 한편, 우리 귀가 수평면과 평행하게 위치하기 때문에 앞서 설명한 바이노럴 큐에는 높이에 대한 단서가 없다. 따라서 모노럴 큐

에 의존하여 고도를 인지해야한다. 예를 들면 귀높이에서 발생한 소리와 그보다 높은 위치에서 발생한 소리는 귓바퀴(외이)에서의 반사에 의한 공명 주파수(resonance frequency)가 달라지게 되고 그에 따른 스펙트럼 상의 peak나 notch가 달라지는 특징이 나타나는데, 사람의 뇌는 이를 이용하여 고도를 인지하는 것이다 (그림 2).

Binaural Rendering vs. Binaural Recording

바이노럴 렌더링은 앞서 설명한 binaural hearing의 원리에 따라 바이노럴 신호를 합성하는 과정이다. 개별 음원의 위치로부터 양쪽 귀까지 도달할 때까지의 경로에 대한 전달함수를 머리전달함수 (Head-Related Transfer Function, HRTF) 라고 하는데, 이를 이용해 바이노럴 렌더링을 수행할 수 있다. 머리전달함수는 인공적으로 잔향이 발생하지 않도록 설계된 무향실 (anechoic room)에서 측정하는데, 사람은 무향 공간에 익숙하지 않기 때문에, 바이노럴 렌더링 구현 시 무향실에서 측정된 머리전달함수만을 이용하여 렌더링된 결과는 매우 어색하게 들린다. 이를 보상하기 위해 합성된 렌

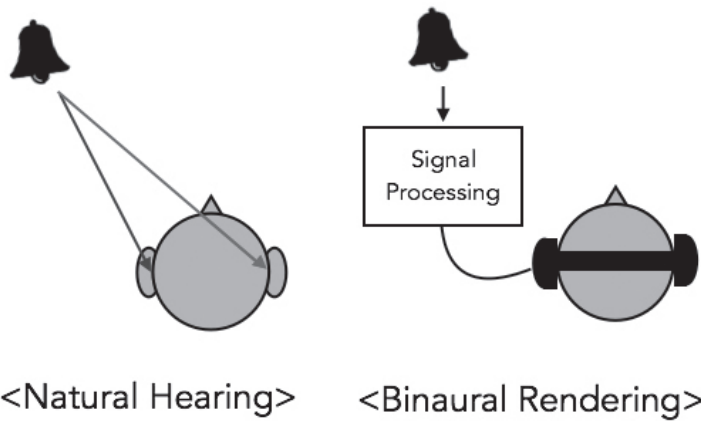


그림 4. Binaural Hearing 과 Binaural Rendering (출처: 가우디오랩)



그림 5. 앰비소닉을 위한 구형 마이크로폰 어레이 (출처: <http://mhacoustics.com>)

터링 결과에 가상 공간에 적합한 인공 잔향을 더하거나, 무향실이 아닌 청음 공간에서 측정된 양이공간전달함수 (Binaural Room Transfer Function, BRTF) 를 이용하기도 한다. 단, 양이공간전달함수는 특정 공간에서 측정된 잔향이 포함된 전달함수이기 때문에, 재현하고자 하는 공간이 측정 공간과 다를 경우 사용하기 어렵다.

한편 바이노럴 기술이 풀어야 할 또 다른 숙제는 머리 전달함수가 사람마다 각각 다른 개인별 고유의 데이터라는 것이다. 사람마다 양 귀 사이의 거리, 머리의 크기, 귓바퀴의 모양 등이 다르기 때문에 다른 사람의 머리전달함수를 사용하여 합성한 신호를 들었을 경우 음원의 의도한 위치에 정위되지 않을 수도 있다. 측정된 머리전달함수에 이러한 개인별 차이를 반영하는 과정을 머리전달함수의 개인화 (Personalization) 라고 한다.

바이노럴 신호를 생성하는 또 다른 방법은 인체 모형의 귀 위치에 마이크를 부착하거나 실제 사람 귀 입구에

마이크를 삽입하여 직접 현장음을 녹음하는 바이노럴 레코딩이다. 바이노럴 레코딩은 실제 공간에서의 소리 전달 과정을 거의 모두 담고 있어서 현장의 공간과 머리 전달함수가 모두 반영된 음원을 바로 획득하여 사용할 수 있는 장점이 있다. 하지만, 사용자의 신체 구조적 특징을 반영하는 개인화 작업이나 사용자의 움직임을 반영하는데 제한이 있는 단점 또한 존재한다. 표 1은 바이노럴 렌더링과 바이노럴 레코딩의 장단점을 비교한 것이다.

바이노럴 레코딩 방법은 이와 같은 환경에 대응하기 적합하지 않다. 또한 라이브 레코딩이나 스포츠 중계처럼 실제 존재하는 공간이 아닌, 공상과학 영화 속 현실과 같이 실제 존재하지 않는 공간에 대응하는 소리를 재현하는 것 역시 바이노럴 레코딩으로는 구현하기 어렵다. 이와 같은 이유로 VR 오디오에서는 바이노럴 렌더링을 통한 이머시브 오디오 재현이 적합하다.

바이노럴 렌더링과 바이노럴 레코딩의 중간 지점에 위

표 1. 바이노럴 렌더링과 바이노럴 레코딩 비교

Binaural Rendering	Vs.	Binaural Recording
0	Computing Efficiency	00000
000	Immersive	0000
00000	Interaction	00
00000	Personalization	0



치한 방식이 마이크로폰 어레이를 이용한 레코딩 방식이다. 주로 구형 마이크로폰을 이용해 앰비소닉 신호로 변환하여 최종 바이노럴 렌더링을 하는 방법인데, 현장의 잔향, 공간의 느낌 등을 정확히 담을 수 있고 동시에 사용자의 움직임 역시 반영할 수 있는 반응형 렌더링도 가능하다. 다만, 공간 해상도의 부족, 렌더링을 위해 필요한 높은 연산량의 문제는 여전히 단점으로 남아있다.

맺음말

Sound is half the experience. 스타워즈의 영화 감독 George Lucas의 말이다. 그의 말처럼 우리 경험의 절반은, 때로는 절반 이상이, 소리에 의해 결정된다. 특히, 그 경험의 플랫폼이 가상현실로 넘어온 경우라면 그

중요성은 더욱 중요해진다.

“가상현실”이라는 단어가 이제는 더이상 공상과학영화나 책에서만 보는 말이 아니라 시나브로 현실과 공존을 시작하였지만, 여전히 해결해야 할 과제들은 많이 남아있다. 흐려져야 할 현실과 가상의 경계는 여전히 존재하고 있고, 많은 엔지니어와 미디어 제작자들은 끊임없이 경계를 무너뜨려 지금의 경험이 현실 세계로부터의 것인지, 가상 세계로부터의 것인지 구분하지 못할만큼 높은 몰입감을 제공하기 위해 고민하고 조금씩 앞으로 나아가고 있다. 언제쯤 상용화가 될까를 생각하는 동안 이미 현실로 들어와버린 VR이라면 저 경계가 없어질 순간도 그리 멀지는 않다고 생각하는 것이 그저 장밋빛 미래에 대한 상상에 그치는 건 아닐 것이다.🌀