

## 제 8 장. 메타버스 디스플레이 기술

학습목표
<ul style="list-style-type: none"><li>▪ 학습내용: 해당 차시에서 학습할 학습주제(목차)를 제시해 주세요.</li><li>▪ 학습목표: 해당 차시 학습을 통해 <u>학습자들이 달성해야 할 목표</u>를 학습내용과 연계하여 작성해 주세요.</li></ul>

### ▶ 학습내용

1. 디스플레이 기술은 어떻게 메타버스를 창조했을까?
2. 메타버스를 구현하는 실시간 컴퓨터 그래픽스 기술
3. 실감 멀티모달 출력 기술
4. 초실감이 가능한 상호작용 기술

### ▶ 학습목표

1. 메타버스 디스플레이 기술을 이해하고 설명할 수 있다.
2. 실시간 컴퓨터 그래픽스 기술을 이해하고 설명할 수 있다.
3. 멀티모달 출력기술을 이해하고 설명할 수 있다.
4. 초실감 상호작용을 이해하고 설명할 수 있다.

## 1. 디스플레이 기술은 어떻게 메타버스를 창조했을까

지금의 메타버스가 있게 해준 가장 핵심적인 기술을 꼽으라고 한다면 헤드 마운티드 디스플레이 (HMD, Head Mounted Display) 기술을 들 수 있다. 머리에 장착할 수 있는 디스플레이 장치로 양쪽 눈에 서로 다른 영상을 제시해 마치 실제 공간 속에 존재하는 것 같은 깊이감을 제공하는 기기다. 얼핏 보면 단순히 헬멧에 디스플레이를 달아서 양쪽 눈에 서로 다른 영상을 송출하기만 하면 되는 단순한 기기처럼 보이지만, 사실은 굉장히 복잡한 광학 기술이 적용되어 있다. HMD의 기술발전 속도에 진정한 메타버스 플랫폼이 등장하는 시기가 결정된다고 해도 과언이 아니다.

HMD 기술에 메타버스의 미래가 달려 있는 것이다. 이 때문에 구글, 오쿨러스, HTC Vive, 소니, 밸브, 삼성과 같은 각종 기업들이 차세대 VR·AR HMD를 만들기 위해 많은 노력을 기울이고 있다. 디스플레이 업체들도 HMD에 사용될 디스플레이 기술의 확보와 특허, 그리고 국제 표준을 선점하기 위한 노력을 기울이고 있다.

HMD가 나오기 전에도 VR 기기를 만들려는 다양한 시도가 이루어졌다. 1962년 출시된 입체영상 기기 ‘센소라마(Sensorama)’는 양쪽 눈에 서로 다른 영상을 비추어 시청자가 입체 영상을 느끼게 하는 장치였다. 입체 영상 뿐만 아니라 음향, 진동, 냄새까지 모방해낸 시스템으로 최초의 3D영화관이기도 하다.

1966년에는 당시 워싱턴 대학교 교수였던 토마스 퍼니스가 제작에 참여한 공군조종사를 위한 훈련장치 ‘플라이트 시뮬레이터’가 나왔다. 미국 VPL 리서치는 1987년에 사용자의 손의 움직임을 VR에 적용할 수 있는 ‘데이터글로브’를, 1989년에는 사이버페이스(Cyber face)라는 이름의 VR헤드셋을 출시했다. 1990년대의 VR이 게임 콘텐츠에 접목되면서 1991년 미국 게임 제작사 세가(SEGA)가 사용자의 머리 움직임을 게임 조작에 반영하는 ‘세가 VR’이라는 최초의 게임용 VR헤드셋 장치를 선보였다.

HMD는 초창기에 군사 장비 분야, 특히 조종사를 위한 장비에 주로 쓰였다. 조종사의 헬멧에 디스플레이를 부착하고, 각종 정보를 보여주는 형태다. 일반 사용자가 HMD를 접하게 된 것은 게임 분야로 닌텐도가 1995년 제작한 ‘버추얼 보이’가 대표적이다. 눈을 갖다 대는 디스플레이 장치와 디스플레이 장치를 지지하는 받침대, 게임 조작을 위한 패드 등으로 구성돼 있다. 휴대용으로 제작됐지만, 무게가 2Kg에 달해 책상 위에 놓고 사용하거나 누워서 사용했다. 하지만 적색 LED만을 사용해 다양한 색상 표현이 어려웠고, 3D 초점이 잘 맞지 않아 두통을 유발해 출시 1년 만에 사라졌다.

메타버스에서 중요하게 여기는 목표 중 하나는 몰입감과 현존감을 극대화하는 것이다. 두 용어에 대해 간단히 설명하자면, 메타버스에서의 몰입감이란 메타버스 세계에 너무나 심취한 나머지 자신이 가상공간에서 무엇인가 하고 있다고 생각하지 못하고 현실처럼 생각하고 행동하게 되는 정도를 말한다. 또한 메타버스에서의 현존감은 메타버스 내에서 어떤 행위를 할 때, 마치 그 가상공간에 내가 실제로 존재하는 것 같다고 느끼는 정도를 말한다.

몰입감과 현존감을 극대화하는 가장 쉬운 방법은 메타버스를 이용하는 사용자에게 실제 같은 경험을 제공하는 것이다. 사람은 시각 경험에 크게 의존한다. 따라서 3D로 구현된 디지털 세상을 우리의 현실처럼 이질감 없이 볼 수 있게 하는 것이 메타버스의 몰입감과 현존감을 극대화하기 위해 달성해야 하는 최우선 목표가 되는 것은 당연하다. 그렇다면 고품질의 디스플레이에 고해상도의 3D디지털 이미지를 송출하기만 하면 되는 것일까?

안타깝게도 일반 디스플레이를 통해 3D 디지털 이미지를 내보낼 경우, 평면에 투사된 3D

디지털 이미지를 보는 느낌이 들어 현실과 구분되는 이질감을 느낄 수 밖에 없다. 가까이 있는 물체와 멀리 있는 물체를 바라볼 때는 서로 다른 깊이감을 느껴야 하는데, 평면 디스플레이에는 이러한 깊이감이 존재하지 않아서 발생하는 문제라고 할 수 있다.

이러한 문제를 해결하기 위해 1968년, 컴퓨터 그래픽 기술의 선구자인 미국의 컴퓨터 과학자 이반 서덜랜드(Ivan Edward Sutherland)가 스테레오스코피(Stereoscopy)와 HMD를 이용한 몰입형 가상현실 기술을 선보였다. 가상현실(VR)기기의 시초이기도 한 이반 서덜랜드의 HMD는 전투기 조종사가 사용하는 헬멧처럼 눈앞에 설치된 렌즈에 다양한 정보를 표시할 수 있었다. 비록 꽤 무겁고 부피가 큰 시범적인 형태의 기기이긴 했지만, 실제 디스플레이 깊이보다 디지털 이미지가 더 멀리 있는 것처럼 느끼도록 유도했다. 또한 왜곡된 이미지를 보정한 양안시차가 있는 이미지를 디스플레이에 렌더링하여 양안시차에 의한 3D 입체감을 제시했다. 이때부터 몰입형 가상현실에 대한 구체적인 가능성이 논의되기 시작했다.

그 이전까지만 해도 컴퓨터 그래픽은 단순 도형에 불과했고, 모니터가 아닌 다른 기기로 디지털 화면을 본다는 것은 상상조차 하기 어려웠다. 이반 서덜랜드가 만든 HMD기기는 모니터 화면을 넘어서 사용자가 컴퓨터 세상에 직접 들어가 여러 프로그램을 제어한다는 개념을 만들어냈다. 그가 만든 기술 덕분에 오늘날 가상현실 게임이 탄생했고, 사용자가 가상현실에서 상호작용할 수 있는 센서 기술이 등장했기에 그의 발명은 디스플레이 기술에 새로운 지평을 열었다. 이반 서덜랜드는 1960년대에 가상현실에 대해 “최종적인 디스플레이는 물론 공간이 될 것이며, 그 공간에 있는 모든 물체를 컴퓨터가 통제할 수 있을 것이다” 라고 말했다.

가상현실에 대한 사람들의 관심이 커진 것은 1980~1990년대에 등장한 만화, 영화, 애니메이션과 같은 서브컬처의 힘이 컸다. 1995년 상영된 극장 애니메이션 <공각기동대>는 이후 등장하는 SF 영화는 물론이고, 정신이 가상세계나 다른 인조인간의 몸속에 완전히 들어가는 풀 다이브(Full Dive, 완전몰입형 가상현실) 개념이 생기는 데 결정적인 역할을 했다. HMD를 이용한 몰입형 가상현실 기술이 가능하다는 사실이 알려지면서 수많은 연구자와 기술자들이 상용화를 위한 기술개발에 돌입했고, 2000년대 초반에는 곧 메타버스 세상이 도래할 것이라는 얘기까지 나왔다. 그러나 안타깝게도 HMD가 상용화하려면 디스플레이 기술이 더 발전해 산재한 문제들이 해결되어야 하는데, 많은 연구를 거쳤음에도 해결되지 않았고 그로 인해 메타버스에 대한 사람들의 관심도 사라지게 되었다. 최근에 메타버스라고 많은 플랫폼이 나오고는 있지만, 이전보다는 나아졌어도 아직 완벽하게 문제 해결이 되지 않아 진정한 메타버스를 상용화하는데 어려움을 겪고 있다. 그렇다면 과연 어떤 문제들이 남아있어서 메타버스 세상의 실현이 아직은 불가능한 것인가?

먼저, HMD에서 제공하는 양안시차에 의한 3D 입체감은 디스플레이를 렌즈의 초점거리보다 가까운 거리에 둘 때 얻을 수 있는 가상의 이미지, 즉 즉 정립허상을 이용한 것이다. 상의 방향이 물체의 방향과 같은 경우를 정립이라고 하고, 실제로 빛이 모이지 않고 생긴 상을 ‘허상’이라 하므로 ‘정립허상’이란 ‘똑바로 선 허상’을 의미한다. 정립허상은 실제 디스플레이 이미지를 확대된 형태로 보는 것과 같다. 이때 확대된 디지털 이미지의 해상도가 낮으면 픽셀과 픽셀 사이에 존재하는 검은 영역인 블랙 매트릭스가 확대되어 이미지가 마치 모기장을 통해 보는 것과 같이 변하는 스크린도어 효과가 발생하게 된다. 이는 마치 오래된 TV를 시청하는 것과 같은 답답함을 주어 사용자의 몰입감과 현존감을 떨어뜨린다. 따라서 고해상도 디스플레이를 통해 양질의 디지털 이미지를 사용자에게 제공하는 것은

HMD기기의 필수 조건이다.

실제로 메타 (페이스북) 산하 기업 오쿨러스사가 2013년 공개한 초기 HMD모델인 오쿨러스 DK1은 한쪽 눈에  $640 \times 800$ 의 해상도를 제공했는데, 2021년 모델인 오쿨러스 퀘스트 2는 한쪽 눈에  $1832 \times 1920$ 의 해상도를 제공해 이러한 문제를 많이 완화했다. 해상도가 4K이상이면서도 MPRT(Moving Picture Response Time) 응답속도가 빠른 HMD를 출시하여 이러한 문제를 완전히 해결할 것이라고 자사 제품의 로드맵을 소개하는 회사들도 있다. 하지만 해상도가 증가할수록 고해상도 디지털 이미지를 생성하기 위해 더 좋은 성능의 연산장치가 필요해진다. 이에 따라 더 많은 전력을 소모하게 되며, 하드웨어의 소비자 가격이 상승하게 된다. 따라서 일반 소비자도 구매할 수 있는 적정 가격의 HMD가 나오기까지는 좀 더 많은 연구와 자본의 투자가 필요할 것으로 보인다.

HMD를 상용화 하는데 따르는 또 다른 걸림돌은 바로 HMD가 크고 무겁다는 것이다. 초기의 HMD들은 휴대하기가 힘들었고, 개인 사용자가 집에서 오랜 시간 사용하기에는 무게로 인한 피로 유발이 심각했다. 그러던 중 2013년 개인이 장시간 사용할 수 있을 만큼 경량화된 몰입형 HMD인 오쿨러스 리프트 DK1이 등장했고, 이와 동시에 다시 한번 전 세계가 메타버스에 주목하기 시작했다. 이어서 2015년에 출시된 오쿨러스 리프트 DK2는 비구면 렌즈 한 개와 디스플레이를 통해 사용자에게 입체감을 전달했다. 하지만 이들 역시 상용화 되기에는 여전히 그 한계가 뚜렷했다. 우선 초점거리를 50mm 이하로 줄일 수 없는 제약이 있어서 400g이나 되는 디스플레이가 눈앞으로 60mm 가량 튀어나오도록 HMD가 설계되었다. 이 때문에 디스플레이 무게를 지탱하기 위한 벨트를 머리 전체에 단단히 고정된 채로 HMD를 사용해야 했고, 장시간 사용 시 두통과 피로가 심각하다는 사용자 평가가 잇따랐다. 결국 오쿨러스 리프트 DK2는 장시간 사용이 힘들다는 세간의 평을 피할 수 없었다. 이러한 문제를 해결하려면 렌즈의 크기와 무게, 렌즈와 디스플레이 사이의 간격, 디스플레이 패널의 크기와 무게를 모두 줄여야 했다. 이를 위해서는 고수준의 광학 기술에 대한 연구와 적용이 필수적이었다.

최근 HMD에 적용된 기술들을 보면 렌즈의 크기를 줄이기 위한 연구들은 어느 정도 성공한 것으로 보인다. 초창기 HMD모델에 사용된 렌즈는 볼록 렌즈였다. 렌즈의 크기를 줄이면서 시야각을 확보하기 위해서는 렌즈 자체의 굴절률 또는 렌즈의 두께를 키워야 했다. 하지만 전자의 경우 재료의 특성 자체를 바꿔야 하므로 그 증가 폭이 제한적일 수 밖에 없었고, 후자의 경우 구면수차각이 발생하여 디지털 이미지의 선명도를 떨어뜨리기 때문에 렌즈의 두께를 무분별하게 증가시킬 수 없었다.

그래서 다음 세대의 HMD들은 기존의 볼록 렌즈 대신 프레넬 렌즈(Frenel Lens)를 사용하기 시작했다. 프레넬 렌즈는 기존 광학 렌즈의 곡면을 연속적인 동심원 형태의 동심홈(Concentric Groove)으로 대체한 광학 부품으로, 각 홈이 개별 굴절면처럼 작용해 평행광선을 일반적인 초점 길이로 굴곡을 줄 수 있다. 이러한 특성으로 인해 프레넬 렌즈는 기존 광학렌즈보다 훨씬 얇은데도 일반 렌즈와 비슷하게 빛을 모을 수 있어 HMD에 들어가는 렌즈의 크기를 줄일 수 있게 되었다. 하지만 프레넬 렌즈는 이미지의 선명도가 손상되기 때문에 해상도가  $4K \times 4K$  이상인 디스플레이 패널을 사용하더라도 화질이 좋지 않다. 또한 볼록 렌즈를 기반으로 한 광학계를 사용하기 때문에 여전히 렌즈와 디스플레이 사이의 거리는 어느 정도 보장되어야 한다. 최근의 HMD에서는 이러한 프레넬 렌즈의 한계를 극복하기 위해 반투과 및 반사편광 기술을 이용한 팬케이크 렌즈를 사용하고 있다.

오쿨러스 HMD에 사용된 팬케이크 렌즈는 부분 반사체, 1/4 파장판 으로 이루어진 제1 렌

즈부와 반사형 편광판을 포함하는 제2 렌즈부로 이루어져 있다. 디스플레이로부터 생성된 빛은 제1렌즈부와 제2렌즈부 사이에 편광과 반사가 이루어지면서 경로가 접히게 되고, 이를 통해 렌즈와 디스플레이 사이의 거리를 획기적으로 줄일 수 있다. 하지만 팬케이크 렌즈에도 단점이 있다. 반투과 및 반사 편광 기술의 특성상 렌즈와 광효율이 낮아 단위 면적에 들어오는 빛의 양인 휘도(Luminance)가 프레넬 렌즈를 사용했을 때에 비해 상대적으로 부족하다. 그러므로 휘도를 보상하기 위한 기술들이 추가로 접목되어야 한다. 하지만 디스플레이 패널의 해상도를 늘릴수록 더 좋은 화질의 선명한 이미지를 제공할 수 있다는 점과 HMD의 소형화에 크게 기여한다는 점이 너무나 큰 장점이 있어서 차세대 HMD들은 모두 팬케이크 렌즈를 기반으로 할 것으로 전망된다.

이외에도 HMD 상용화를 방해하는 다양한 문제가 존재한다. 눈 바로 앞에 디스플레이를 두고 사용하기 때문에 눈이 빨리 피로해지는 문제, 사용자가 바라보는 가상 물체의 위치와 디스플레이의 실제 위치가 달라서 발생하는 수렴-초점 불일치 문제, 디스플레이 크기의 한계로 인한 시야각의 제한 문제 등이 대표적이다. 그러나 최근의 기기 발전 속도와 다양한 연구 결과를 종합해 볼 때, 이러한 문제들이 해결되는 데에는 오랜 시간이 걸리지 않을 것으로 보인다. 몇 년 뒤 HMD를 저렴한 가격에 구매할 수 있고, 충분히 작고 가벼워지며, 고해상도의 시각 정보를 실시간으로 전달할 수 있고, 눈이 크게 피로하지 않아 HMD를 장시간 사용할 수 있게 된다면 정말로 현실과 메타버스 세상이 혼재된 모습을 일상에서 아무렇지 않게 보게 될 수 있을지도 모른다.

## 2. 메타버스를 구현하는 실시간 컴퓨터 그래픽스 기술

디스플레이 기술이 메타버스를 구성하는 핵심 하드웨어 기술이라고 한다면, 컴퓨터 그래픽스 기술은 메타버스의 핵심 소프트웨어 기술이라고 할 수 있다. 흔히 CG 기술이라고 부른다. 컴퓨터의 자원과 데이터 처리 규약 (정보를 원활하게 교환하기 위해 필요한 규칙의 집합)을 특수한 방법으로 정제해 실제 사람이 볼 수 있는 형태의 데이터로 가공하고, 이를 디스플레이와 같은 출력 장치로 송출하여 시각 자극을 제공한다.

예를 들어보면, 사용자에게 가상의 숲 공간을 HMD를 통해 보여주고 싶을 때, CG기술은 0과 1로 저장 되어 있는 나무, 돌, 풀과 같은 데이터를 공간 좌표상에 배치하고 이들이 본래 가지는 색과 빛의 영향을 고려하여 색을 결정한다. 또한 바람의 방향과 세기를 고려하여 나무, 돌, 풀 등의 애니메이션을 생성한 후 현재 송출하는 그 순간의 최종 위치를 결정하여 최종 장면을 생성한다. 그리고 이들을 HMD를 통해 어떻게 출력해야 할지를 결정하고 해당 패턴의 신호를 HMD에 전달한다.

결국 CG 기술은 컴퓨터에 0과 1로 저장된 데이터를 가공하여 사람이 이해할 수 있는 시각 정보로 보여주는 기술의 총체라고 할 수 있다. 메타버스는 고품질의 가상공간을 실시간으로 생성하고 상호작용이 가능해야 하므로 고퀄리티의 CG기술이 필수적이다.

CG에서는 물체를 표현할 때 점, 선, 면을 이용한다. 예를 들어 Stanford Bunny 라고 불리는 토끼를 정의하려면 수많은 Vertex(꼭짓점) 데이터의 위치값이 필요하다. 데이터의 가장 앞에 나오는 V는 Vertex 정보임을 나타내고, 그 뒤의 숫자들은 3차원 좌표 공간에서의 위치값으로 나타낸다.

Stanford Bunny 를 자세히 보면 Vertex (점) 이 보이기 보다는 삼각형 들의 군집이 더 잘 보일 것이다. 이들은 삼각형 폴리곤이라고 불리는 것으로 Vertex 데이터가 있을 때 특

정 규약에 따라서 하나의 면으로 만들어낸 것이다. 예를 들어 위에서부터 차례 대로 3개의 꼭지점(Vertex)이 존재할 때, 이들을 묶어서 하나의 폴리곤(다각형)이 되게 하겠다고 규약을 정했다면, 위에서부터 차례대로 3개의 Vertex를 묶어서 삼각형 폴리곤의 위치를 정의할 수 있게 되나. 폴리곤은 삼각형뿐 아니라 우리가 알고 있는 다양한 도형을 모두 사용할 수 있으나 주로 삼각형이나 사각형을 사용한다.

이렇게 폴리곤을 통해서 각 물체가 정의되었으면, 이 물체들을 이용해서 보여줘야 할 공간에 대해 정의해야 한다. 지형, 나무, 돌의 위치를 적절히 배치해서 원하는 장면의 기초가 되는 물체의 배열을 만든다. 그리고 해당 배치를 바라보는 카메라의 위치를 결정한다. 또한 카메라에서 해당 장면을 바라봤을 때 카메라에 이들의 모습이 어떻게 투영될지를 수학적으로 결정한다. 최종적으로 디스플레이에는 카메라에 투영된 모습이 송출된다.

배열이 완료되었다면, 그 다음에는 각 물체의 세부 특성을 고려해야 한다. 텍스처라고 불리는 이미지를 각 물체에 입혀서 색을 결정하고, 물체의 특성에 따라서 빛을 연산한다. 예를 들어 이파리 텍스처, 나무 밑동 텍스처, 나무 기둥 텍스처 등을 이용하여 나무가 초록색, 갈색과 각 무늬를 가지도록 만든다. 공간상에 존재하는 다양한 빛의 특성과 물체의 재질 특성뿐 아니라 반사, 굴절, 흡수량을 고려하여 물체의 최종 색을 결정한다. 그리고 그 모양에 따라 알맞은 형태의 그림자를 계산하여 장면에 넣어준다. 이러한 작업이 모두 끝나면 그림 공간에 대한 정의가 완료된다. 계산된 최종 데이터를 HMD와 같은 디스플레이 장치에 보내주면 사용자에게 해당 장면을 보여 줄 수 있다.

CG기술은 시각 데이터를 중요하게 다루는 게임이나 영화 산업에서 활발하게 사용되어 온 기술이다. 그래서 사실적인 장면을 더 빠르게 저비용으로 만드는 방향으로 기술이 발전했다. 하지만 메타버스의 등장으로 기존과는 조금은 다른 방향으로 CG기술이 발전하기 시작했다. 가상공간에서의 상호작용, 현실공간의 데이터 수집과 가상공간과의 통합, 가상공간과 물체 생성, 사람의 행동이나 표정 등을 가상 아바타에 매핑하기, 사람처럼 행동하는 NPC 등이 그것이다. 현실공간을 고려하여 가상공간을 생성한 후 그 가상공간에 사용자가 접속해 현실처럼 살아가려 함에 따라 필요해진 기술들로 인공지능 기술의 발전과 맞물려 최근 활발하게 연구되고 있다.

### 3. 실감 멀티모달 출력 기술들

시각 이외의 감각에 대한 출력을 시각과 동시에 해주는 장치를 멀티모달 출력 장치라고 한다. 시각 외의 감각 중 의미 있는 감각이 바로 청각과 촉각이다.

청각 디스플레이는 다른 감각에 비해 상대적으로 몰입형 디스플레이 장치를 만들기가 굉장히 쉽다. 이는 청각 신호와 사람의 청각 인지 기관의 특징에 기인한다. 우선 청각을 자극하는 신호는 시각과는 다르게 신호가 즉 한 줄로 차례대로 오는 1차원의 시계열 공기 압력 변화 데이터로 이루어진다. 따라서 다차원의 신호를 동시에 만들 필요가 없고, 많아야 귀 두쪽에 들리는 1차원 신호 2개만 만들어주면 되므로 신호를 만드는 액추에이터 구조가 매우 단순하다. 또한 청각 신호를 받아들이는 몸속의 센서 자체가 귀 안의 한 곳에 몰려 있으므로, 귀만 커버할 수 있으면 손쉽게 몰입형 출력이 가능하다는 장점이 있다.

이런 이유로 우리가 흔히 귀에 꽂는 헤드폰이나 이어폰 자체가 거의 완벽한 몰입형 청각 디스플레이가 될 수 있다. 즉 주위 환경 소리를 차단하고 100% 컴퓨터가 만들어내는 소리로 귀를 자극할 수 있는 장치인 것이다. 따라서 청각 관련 하드웨어 장치 기술은 이미 기

술적으로 충족되어 있으며, 그보다는 여러 가지 사실적인 소리를 다양하게 가상으로 조합해 낼 수 있는 소프트웨어 알고리즘이 훨씬 더 주목받는 핵심 기술이다.

그렇다면 우리가 일상적으로 듣고 있는 ‘소리’를 어떻게 가상으로 만들어 낼 수 있을 것인가? 크게 3가지에 대한 시뮬레이션이 필요하다. 첫 번째는 각기 다른 소리, 즉 공기의 고주파 떨림을 실시간으로 어떻게 합성할까? 다시 말해 소리 신호 자체를 어떻게 생성해 낼 것인가이다. 두 번째는 이렇게 만들어진 소리 소스를 사람의 두 귀로 어떻게 잘 전달해 주어야 사람이 소리만으로 공간감을 느낄 것인가이다. 마지막으로 해야 할 일은 가상환경의 구조적, 물리적 특징으로 인해 소리가 변형되는데 이를 시뮬레이션하여 좀 더 사실적인 소리를 만들어내는 것이다.

#### ① 소리는 어떻게 합성할까?

메타버스에서 유리창을 때리면 고음의 유리 소리가 나야 하고, 기차가 지나갈 때는 기차 엔진, 바퀴 등에서 나는 소리가 합쳐져서 복합적인 소리가 나야 한다. 물리적으로 보면 이러한 소리는 물체가 외부의 충격을 받거나 혹은 내부 에너지가 방출되면서 고주파로 떨게 되면 이 떨림이 공기의 떨림을 만들어내고, 이 에너지가 공기를 매개로 사람들의 귀에 도달하는 과정에서 발생한다. 이론적으로는 가상물체에 대해 이러한 미세한 떨림을 물리적으로 시뮬레이션해주면 소리를 만들어 낼 수 있다. 하지만 현실적으로는 물리적인 시뮬레이션으로 소리를 만들기가 힘들다. 고주파 떨림 자체가 시뮬레이션 계산량이 매우 크며, 이러한 소리는 나의 움직임과 주위의 변화에 실시간으로 반응해서 만들어지기 때문이다. 아주 빠르게 변화에 반응하는 실시간 시뮬레이션이 필요한데, 이러한 시뮬레이션에 대한 계산이 사람이 느끼기에 지연되지 않고 계산되려면 아주 빠른 고성능의 시스템이 필요하다.

실제 가상현실에서 소리는 좀 더 단순한 방법으로 만들어진다. 만들기 쉬우면서 효과도 그럴듯한 방법인데, 실제 음향을 녹음해서 음향콘텐츠를 만들고 이것을 가상환경에서 재생하는 것이다. 물론 당연히 단점이 있다. 단순 녹음 재생의 단점은 미세한 변화를 만들어내기가 불가능하다는 것이다. 실제 소리는 사람의 상호작용에 따라 변화하는데, 이를 반영하기가 힘들다. 다른 상황의 소리를 일일이 녹음해서 상황에 맞게 들어주는 것도 한계가 있다. 이러한 문제를 해결하기 위해 녹음된 소리끼리 음향 합성을 하거나, 혹은 녹음된 소리를 실시간 변형하는 방법을 사용하기도 한다. 모든 경우가 아닌 대표적인 소리만 녹음해 소리 데이터베이스를 만들고, 실시간 상호작용에 따라 소리를 합성하는 것이다. 그렇게 되면 이제 일일이 다 녹음하지 않더라도 다양한 소리를 사실적으로 만들어 낼 수 있다.

물체들에서 나는 소리 외에 중요하나 소리가 하나 더 있다. 바로 사람들의 목소리이다. 사람의 목소리를 합성하는 것은 가상현실과 메타버스가 생겨나기 전에도 그 유용성 때문에 여러 분야에서 TTS (Text to Speech, 텍스트 음성변환) 라는 이름으로 많이 연구된 중요한 분야다. 지금은 기술의 발전으로 실제로 사람이 내는 목소리와 구별하는 것이 거의 불가능할 정도다,

#### ② 어떻게 소리로 공간을 느낄 수 있을까? (공간 사운드)

사람에게는 두 개의 귀가 있다. 사람은 양쪽 귀에 들리는 소리의 차이를 이용해서 소리가 어느 쪽에서 나는지를 파악한다. 이를 ‘사운드 위치 인지’라고 하는데, 크게 3가지 정보를

사용해 사운드 소스의 위치를 파악한다. 사운드 소스가 자신의 오른쪽에 있다면 같은 소리라도 오른쪽 귀에 들리는 사운드가 더 크게 들릴 것이고, 오른쪽 귀에서 해당 소리가 약간 더 빨리 들리게 된다. 이 두 가지 정보를 바탕으로 우리는 평면상에서 소리가 어느 방향에서 오는지 알게 된다. 그렇다면 소리의 상하 위치는 어떻게 인지할까? 우리 귓볼의 복잡한 모양으로 인해서 위에서 오는 소리와 아래에서 오는 소리의 파장이 약간 변형된다. 우리는 이를 인지할 수 있고 이미 학습으로 파악하고 있어서 이를 이용하여 소리의 수직 위치도 파악할 수 있다. 가상현실에서도 사운드 위치 인지를 위한 신호들을 제공해야 사실적인 사운드가 만들어질 수 있다. 단순히 양쪽 귀에 다른 소리가 들리는 스테레오 사운드가 아니라, 실시간으로 변하는 사용자의 머리 방향에 맞추어 이러한 사인이 정확하게 제공되어야 한다. 그러면 실시간으로 변하는 사용자의 머리 방향에 따라 양 귀에 들리는 소리 변화를 어떻게 시뮬레이션해야 할까? 가장 단순하면서 확실한 방법이 머리전달함수를 만들어 활용하는 것이다. 머리 전달 함수란 동일한 소리를 전방위에서 발생시켜 방향에 따른 주파수 반응을 측정하여 3차원 함수로 정리한 것을 말한다. 가상현실 시스템은 보통 아직 찌그러지기 전인 원래 소리 파형을 가지고 있다. 이 소리를 가상환경의 특정 위치에서 생성된 듯한 사운드로 변화시키기 위해 모델링한 소리전달 함수를 원래 소리 파형에 적용하여 변형된 소리 파형을 만들어낸다. 이를 사용자에게 재생해주면 사람들은 이 음향이 목표한 위치에서 나오는 것이라고 느끼게 된다.

사용자가 바뀔 때마다 머리 전달 함수를 만들어야 하는 것은 당연히 아니다. 보통 대부분의 3D사운드 소프트웨어 API에서는 미리 만들어진 머리전달 함수를 가지고 있고, 파라미터만 조정하면 최적의 공간 사운드를 손쉽게 만들어낼 수 있는 툴이 제공된다. 가상환경에서의 사운드 소스 위치만 입력하면 자동으로 결과 음향 파형을 만들어주는 수준으로 간단하게 구현할 수 있다.

### ③ 환경 사운드

마지막으로 사실적인 사운드를 위해 고려해야 할 부분이 환경 사운드다. 사용자 주위의 환경에 따라서도 소리가 바뀌는 효과를 가상 환경에서도 제공해줘야 한다, 그 대표적인 예가 동굴이다 같은 소리라도 동굴 속에 있을 때와 개활지에 있을 때 소리의 느낌이 완전 다르다. 여러 이유가 있겠지만 대표적인 원인은 벽에 반사되는 소리 때문이다. 이처럼 가상현실에서 내가 밖에 있다가 갑자기 방안으로 들어갔을 때 같은 소리라도 다르게 들리게 만들어줘야 한다. 그러면 이런 효과는 어떻게 만들까? 벽에 대한 반사도 물리 시뮬레이션을 통해 해결할 수 있다.

환경 사운드의 또 다른 예로 도플러 효과를 들 수 있다. 빠른 속도로 움직이는 사운드 소스가 있다면 이 사운드는 나와 멀어질 때와 나와 가까워질 때의 소리 진동수가 달라진다. 가상환경에서 내 앞으로 소방차가 썹 하고 지나갈 때 소방차 사이렌 소리가 올 때와 갈 때 달라져야 하는 것이다. 이러한 효과도 시뮬레이션을 통해서 지원될 수 있다.

이렇게 음향합성, 공간사운드, 환경사운드를 시뮬레이션함으로써 가상환경 내에서 사실적인 음향을 제공할 수 있다. 요즘 개발되는 음향 API들은 이러한 기능들 대부분이 구현되어 있고, 메타버스 시스템에 손쉽게 적용될 수 있도록 개발자를 위한 인터페이스를 제공하고 있다.



#### 4. 초실감이 가능한 상호작용 기술

메타버스가 가상으로 만들어진 3D 세상인 만큼 우리는 그 세계와 현실처럼 상호작용하고 싶어 한다. 가상의 세계를 두 발로 걸어 다니면서 탐험하고 가상에 존재하는 물체를 만지고, 밀고, 들어 올릴 수 있기를 원한다. 이를 위해서는 현실에 존재하는 사람의 행동을 인식하고 이를 데이터화해 메타버스 세계에 반영하기 위한 하드웨어와 소프트웨어 기술이 필수적이다. 그렇다면 메타버스에서 서로 상호작용 하기 위해 어떻게 메타버스 세계에서 걸을 것인가? 과거에 PC나 모바일 기기를 통해 사용자에게 가상세계를 탐험하는 경험을 제공하는 것은 그다지 어려운 일이 아니었다. 가상세계에 사용자를 대신하는 가상 아바타를 제공하고 나서, 사용자에게 키보드나 터치스크린과 같은 기기를 통해 아바타를 조종하라고 하면 사용자는 충분히 가상세계를 탐험한다는 느낌을 받았다. 하지만 HMD를 착용하고 메타버스를 이용하는 사용자들에게는 이러한 방식을 사용할 수가 없게 되었다. 우선 사용자들이 기대하는 몰입의 수준이 달라졌다. 단순히 자신을 대신하는 아바타가 가상세계를 탐험하는 경험이 아닌 자신이 직접 가상세계를 탐험하는 느낌을 경험하고 싶어 하게 된 것이다. 하지만 키보드나 컨트롤러 같은 외부 입력 기기를 통해 가상 아바타를 움직이는 것은 메타버스 세계에 대한 몰입감과 현존감을 심각하게 떨어뜨린다는 연구 결과가 나와 있다. 이는 외부 기기를 통한 조작이 사용자들에게는 자신이 직접 움직인다고 보다는 그저 자신을 대신하는 무엇인가를 외부에서 조종하는 느낌을 받기 때문이다.

외부 입력 기기를 통한 메타버스 세계 탐험에는 또다른 문제가 존재한다. 그것은 바로 사용자가 가상 멀미를 겪을 확률이 높아진다는 것이다. 가상 멀미가 발생하는 원인에 대해서는 여러 견해가 있는데, 그 중 대표적인 이론이 감각 갈등 이론이다. 감각 갈등 이론은 인간의 시각과 전정감각이 일치하지 않을 때 멀미를 유발한다고 가정한다. 인간은 보통 균형을 잡기 위해 시각과 전정감각을 통합한다. 시각은 우리 신체의 위치와 주변 사물에 대한 상대적인 위치 정보를 제공하며, 전정 감각은 신체의 움직임에 대한 정보를 제공한다.

하지만 때때로 두 개의 감각이 서로 다른 정보를 제공할 때가 있다. 배나 차를 탈 때 우리의 시각은 우리가 움직이고 있다는 정보를 제공하지만 전정 감각은 앉아 있거나 서 있다는 정보를 제공한다. 그러면 두 가지 감각이 전달하는 정보가 일치하지 않게 되고, 이로 인해 멀미가 발생한다. 외부 입력 기기를 통해 메타버스 세계를 탐험하다 보면 이처럼 감각이 충돌하는 상황이 발생할 수 있다.

사용자가 의자에 앉아서 외부 컨트롤러를 통해 자신의 아바타를 걷게 했다고 가정해보자. 그렇다면 사용자의 HMD는 앞으로 걷고 있는 영상을 송출할 것이므로 사용자의 시각 정보는 앞으로 걷고 있는 상태에 대한 정보가 될 것이다. 하지만 실제 세계에 존재하는 사용자의 몸은 의자에 앉아있는 상태이므로 전정 정보는 앉아 있는 상태에 대한 정보가 될 것이다. 즉 사용자가 컨트롤러를 통해 아바타를 움직이는 순간부터 감각 충돌이 일어나게 되고, 이러한 현상이 지속될수록 가상 멀리는 점차 심화 될 것이다.

이러한 이유로 메타버스 세상에서의 이동은 현실세계로의 실제 이동을 사실 세계의 이동과 일치시켜 달성하는 것이 권장된다. AR/MR 환경은 실제 공간에 사용자가 존재하고, 실제 공간의 일부를 가상의 물체로 대체하거나 가상 물체를 실제 공간 위에 더하는 형태로 동작하기 때문에 현실의 사용자 이동과 AR/MR 환경에서의 사용자 이동이 완전히 일치한다. 하지만 VR HMD와 같이 닫힌 디스플레이 기반 HMD를 사용하게 되면 현실 공간에서의

이동을 별개의 센서를 통해 인식하고, 사용자가 이동한 거리만큼을 가상세계에 반영하여 가상세계에 존재하는 사용자 아바타의 이동에 반영해야 하는 절차들이 필요해진다. 그러나 일반적으로 VR세계는 현실세계보다 크거나 다양하고 격렬한 형태의 움직임이 필요한 경우가 많아 단순히 현실세계의 사용자 움직임을 가상세계의 움직임에 반영하는 데는 한계가 있다. 최근 이런 한계를 극복하기 위해 트레드밀을 응용한 기기 장치 등으로 사용자의 감각을 속여 실제로는 좁은 공간을 움직이고 있지만 가상세계에서는 광활한 세계를 탐험하는 것처럼 느끼도록 하는 기술들이 연구되고 있다.

## 5. 혼합현실과 다중 사용자 메타버스 기술

현실과 같은 경험을 제공하는 것이 메타버사이므로, 메타버스 세계도 타인과의 교류와 상호작용을 지원해야 한다. 하지만 안타깝게도 수많은 메타버스 기술과 콘텐츠는 단일 사용자 중심으로 개발되었다. 이러한 현상이 일어난 가장 큰 이유는 기술과 비용의 한계 때문이다. 단일 사용자만을 대상으로 하더라도 충분히 높은 품질의 경험을 제공하는 것이 힘든 게 메타버스 관련 기술의 현 상황이다. 이런 상황에서 다중 사용자를 위한 인프라를 확보하면서 관련 기술을 개발할 여력을 가진 곳은 몇 곳 밖에 되지 않는다. 하지만 연구계와 산업계에서는 다중 사용자를 지원하는 메타버스 세계에 언젠가는 반드시 도달해야 한다는 것을 인식하고 있다. 이에 관련 기술을 확보하기 위한 노력이 느리지만 천천히 이루어지고 있다. 2016년 마이크로소프트사는 ‘홀로포테이션 : 실시간 가상 3D 텔레포트’라는 제목으로 관련 기술을 발표해 전 세계의 이목을 집중시켰다. 홀로포테이션은 다수의 고성능 3D 이미지를 만들고, 이를 다른 사용자가 착용 중인 마이크로소프트사의 홀로렌즈를 전송하는 기술을 일컫는다. 그러면 홀로렌즈를 착용 중인 사용자는 자신이 존재하는 공간에 원거리 사용자의 3D 모습이 더해져 제공되므로 마치 자신의 눈앞에 있는 것처럼 보고 대화를 나누는 혼합현실 경험을 할 수 있게 된다. 다수의 고성능 3D 카메라, 카메라를 요구사항대로 설치하기 위한 충분히 넓은 공간, 실시간 이미지 변환을 위한 고성능 CPU, 홀로렌즈까지 요구되는 인프라 비용이 엄청나기 때문에 일반 사용자들이 사용할 수 있는 기술은 아니다. 하지만 마이크로소프트사는 이날 수많은 대중에게 XR기술이 가져올 미래를 꿈꿀 수 있는 구체적인 장면을 시연했다. 이는 최근의 메타버스 붐을 일으키는 계기가 되었다.

최근 마이크로소프트사는 자사의 클라우드 컴퓨팅 플랫폼인 애저 클라우드(Azure Cloud)에서 혼합현실 기능을 제공하는 협업 플랫폼 메시(Mesh)를 발표했다. 단순히 기술을 시연하는 데 불과했던 홀로포테이션을 넘어 서로 다른 공간에 있는 여러 명의 작업자가 다양한 기기에서 홀로그램 경험을 공유하며 협업하는 플랫폼을 제시했다. 마이크로소프트사는 다중 사용자를 지원하는 메타버스 환경을 이용해 원격 협업 환경을 구축하려는 것이 목표다. 만약 메시에서 온라인 회의를 진행한다면, 지구 반대편에 있는 여러 사용자가 동시에 한 공간으로 모여 마치 바로 옆에 있는 것처럼 회의와 업무를 진행할 수 있다. 또한 필요에 따라서 디지털 데이터를 시각화하여 보여줄 수 있으며, 회의 공간이나 회의실의 레이아웃을 실시간으로 바꿀 수 있다. 실제로 마이크로소프트사는 ‘이크나이트2021’이라는 글로벌 콘퍼런스 행사를 가상현실에서 진행한 적이 있다. 이때 메시 플랫폼을 활용하여 현실과 가상현실을 아우르는 공간에서 다수의 사용자가 협업을 통해 업무에서 다양하고 긍정적인 변화를 가져올 수 있음을 보여주었다.

코로나 19 사태로 다중 사용자 접속 환경에 대한 수요가 폭증하게 되었고, 관련 기술과 플랫폼이 폭발적으로 성장했다. 이러한 성장에도 불구하고 여전히 대다수 플랫폼과 기술들은 기존의 한계를 극복하지 못한 상태이다.

## 6. 실제 세계를 디지털 데이터화하는 기술

실제 세계의 정보를 가상세계에 투영하는 데 필요한 핵심 기술 중 하나가 3D 센싱/ 스캐닝과 관련된 기술이다. 물체의 부피, 모양, 방향 등 형상과 관련된 정보를 얻기 위한 기술은 크게 4가지로 분류된다.

### ① ToF (Time of-Flight) 기법

ToF기술은 스캐닝 하고자 하는 물체를 향해 입자나 파장을 발사하고 다시 돌아올때 까지 걸리는 시간을 이용하여 해당 물체의 외형적인 특징을 추정하는 기술이다. ToF 카메라에 부착된 이미터(Emitter)에서 스캐닝 하고자 하는 목표 물체를 향해 특정 파장을 지닌 빛을 쏘았다고 가정했을 때 해당 물체에 도달한 빛은 반사되어 다시 돌아올 것이고, 이렇게 돌아오는 빛은 ToF카메라에 부착된 센서를 통해 데이터화할 것이다. 그러면 데이터 분석을 통해 이미터에서 몇 초전에 발사한 빛이 어떤 위치에서 도달한 것인지를 분석하고, 이를 통해 ToF카메라로부터 물체까지의 거리 정보를 얻어낼 수 있다. 만약 여러 다발의 빛을 일정 거리를 두고 쏘아내서 받아낼 경우, 특정 공간을 차지하는 물체까지의 거리를 나타내는 깊이 맵(Depth Map)을 얻을 수 있고, 이를 기반으로 해당 물체의 외형 정보를 추정할 수 있다.

자율주행차를 구현하려면 주변 상화에 대한 정보를 실시간으로 얻어야 하는데, 이를 위해 ToF 기술을 사용하기도 한다. 자율 주행차를 구현하는 데 주로 쓰이는 라이다(LiDAR)센서는 주변부로 여러 다발의 레이저를 쏘고 다시 돌아오기까지 걸린 시간을 측정해 자동차 주변의 환경이 현재 어떤 형상인지를 얻어낸다. ToF 기술은 세밀하게 물체를 구별해내는 거리분해능력이 뛰어나다. 따라서 고해상도의 데이터를 만들기 수월하고, 짧은 시간에 물체를 정확하게 감지할 수 있어서 각종 실시간 거리가 필요한 곳에 적용하기 쉽다는 장점이 있다. 다른 기술보다 가격이 저렴하고, 레이저 기반의 센서를 사용하면 꽤 넓은 공간의 정보를 얻을 수 있다. 그러나 태양광의 간섭이 심한 환경에서는 돌아오는 빛과 태양광을 구분하기 힘들어 오류가 발생하는 단점이 있다. 또한 고해상도 데이터를 얻기 위해서 빛을 받아들이는 칩이 많이 필요하므로 다른 센서에 비해 크기가 크다는 단점이 있다.

### ②레이저 삼각 측량 기법

레이저 삼각 측량기법은 타깃이 되는 특정 입사각으로 레이저 빔을 쏘았을 때 타깃 대상인 물체의 표면에서 산란되는 레이저를 특정 위치에서 포커싱하여 이미지를 얻는 기법이다. 레이저의 발사 위치와 포커싱을 위한 센서 사이의 각도를 알고 있으면 삼각 측량 기법을 통해 물체까지의 거리를 측정할 수 있다. 레이저 삼각 측량 기법의 장점은 다른 기술보다 저렴한 비용으로 고정밀 측정이 가능하다는 점이다. 게다가 측정 범위도 넓어서 광범위한 앱의 요구사항을 충족할 수 있다. 그러나 레이저의 경로에 따라서 정확도가 영향을 받기 때문에 먼지나 기타 이물질이 많은 공간에서는 정확도가 떨어지고, 민감한 전자부품을 사

용하는 경우가 많아 사용할 수 있는 온도가 다소 제한된다.

### ③구조화 광 패턴 기법

구조화 광 패턴 기법은 특정한 패턴을 가진 빛을 물체에 비출 때 물체 표면에 생기는 패턴을 분석하여 거리를 측정하는 방식이다. 일반적으로 작선형 무늬나 점 패턴을 물체에 투사하고 물체의 굴곡에 따라 패턴이 굽어져 보이는 것을 분석한다. 물체 표면의 이미지를 분석하여 거리값을 얻을 수 있으므로 다른 기법들에 비해 저비용으로 고해상도의 데이터를 얻을 수 있으며, 한번에 넓은 면적의 3차원 데이터를 획득할 수 있다는 장점이 있다. 하지만 물체 표면에 맺힌 패턴 인식을 방해하는 요소들에 취약하다는 단점이 있다.

### ④스테레오 비전 기법

스테레오 비전 기법은 카메라 두 대를 통해 얻은 이미지 정보를 이용하여 거리 정보를 획득하는 기법이다. 두 대의 카메라 사이의 위치 관계를 알고 있을 때, 한쪽 카메라의 이미지 평면에 맺히는 데이터와 반대쪽 카메라의 이미지 평면에 맺히는 데이터를 삼각법을 이용하여 분석하는 것이다. 이는 사람이 두 눈을 통해 물체를 볼 때 양쪽 눈에 비치는 이미지가 얼마나 다른가를 분석하여 거리감을 느끼는 것과 같은 원리다. 대표적인 장점은 2개의 영상 데이터와 깊이 데이터를 동시에 얻을 수 있어 다양한 처리를 병행 할 수 있다는 점이다. 그러나 영상을 얻기 힘든 조도가 낮은 환경이나 너무 밝은 환경에서는 데이터와 정확도가 떨어지며, 2개 이미지를 처리하는 알고리즘의 복잡도가 높아 고성능 프로세서를 필요로 한다는 점이 대표적인 단점이다.

4가지 기술 모두 저마다의 장단점을 가지고 있다. 따라서 목적과 사용 환경에 따라 최적의 기술을 사용하면 된다. 4가지 기술에는 공통적인 한계가 존재한다. 첫째, 모든 환경에서 완벽하게 동작하는 센싱 기법이 존재하지 않는다. 둘째, 해상도를 높이기 위해서는 센서 자체의 성능을 끌어올려야 하는데, 일반 사용자들이 감당하기에는 큰비용이다. 셋째, 센서들이 민감한 경우 통제되지 않은 상황에서 어떤 오작동을 일으킬지 보장되지 않는 경우가 많다. 이런 한계로 인해 일반 사용자를 위한 메타버스 환경에 바로 접목하기에는 아직 무리다.