

# 360VR 오디오 기술 및 표준화 동향

□ 이용주, 이태진, 김휘용 / 한국전자통신연구원

## 요약

최근 영화, 스포츠 중계, 게임 등 다양한 종류의 콘텐츠가 VR 형태로 서비스되고 있는데, 이러한 360도 영상 서비스 환경에서 영상과 오디오를 더욱 실감나게 제공하는 기술에 대한 연구도 많이 이루어지고 있다. 오디오의 경우 스테레오, 5.1 채널 오디오 등과 같은 채널 기반 오디오가 방송, 영화 등에서 주로 사용되었는데, 최근에는 객체 기반 오디오, 장면 기반 오디오 기술의 적용이 시도되고 있다. VR 서비스에서도 더욱 실감 있는 오디오를 제공하기 위하여 이러한 최신 오디오 기술의 적용이 시도되고 있다. 본 고에서는 증강 현실 및 가상 현실에 활용될 수 있는 최신 오디오 기술에 대해 간략하게 살펴보고, 최신 오디오 기술이 어떻게 VR 서비스에 적용이 되고 있는지에 대해 살펴본다. 그리고, VR 서비스를 위한 규격 제정 측면에서 오디오 기술은 어떻게 진행이 되고 있는지에 대하여 간략하게 살펴본다.

## I. 서 론

최근 영화, 드라마, 스포츠 중계, 게임, 가상 체험 등 다양한 종류의 콘텐츠가 VR(Virtual Reality) 형태로 서비스되고 있다. 많은 VR 서비스에서는 360도 영상을 제공하여 사용자가 높은 현실감을 느낄 수 있도록 하고 있는데, 현실감의 향상을 위하여 고해상도, 고품질의 영상을 제공하는 방법에 대한 연구가 이루어지고 있다.

고품질의 VR 영상을 제공하기 위한 연구와 함께 VR 환경에서 실감 있는 오디오를 제공하기 위한 연구도 많이 이루어지고 있다. 기존 오디오 시스템의 경우 체험자가 움직이지 않는 환경에서 최적의 음향을 재현하는 것을 기반으로 하고 있어, 체험자의 시선 또는 공간 상의 위치가 변경될 수 있는 VR 환

\* 본 연구는 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [18ZR1100, 초실감 공간미디어 원천기술 개발]

경에서 최적의 실감 오디오를 제공하는 것에는 제한이 있을 수 있다. 이러한 제한을 해결할 수 있는 오디오 기술로서 객체 기반 오디오와 장면 기반 오디오 기술 등이 제안되고 있는데, 이를 적용하면 VR 환경에서 더욱 실감 있는 오디오의 재현이 가능할 수 있다. 객체 기반 오디오 및 장면 기반 오디오를 적용한 콘텐츠의 제작 및 재현을 가능하게 하는 기기들이 이미 제품으로 출시가 되고 있으며, 일부의 VR 콘텐츠에는 객체 기반 오디오와 장면 기반 오디오 기술이 적용되고 있는 상태이다.

본 고에서는 VR 서비스에 활용될 수 있는 최신 오디오 기술에 대해 간략하게 살펴보고, 이러한 기술을 제공할 수 있는 제품들을 통하여 최신 오디오 기술이 어떻게 VR 서비스에 적용이 되고 있는지를 살펴본다. 그리고, 여러 국제 표준화 기구에서 수행되고 있는 VR 서비스 관련 규격 제정에서 오디오 관련 내용이 어떻게 진행되는 지에 대해서도 간략하게 살펴본다.

## II. VR을 위한 오디오 기술

방송, 영화, 라디오 등과 같은 멀티미디어 서비스에서 영상과 함께 오디오는 빠질 수 없는 요소이다. 오디오 서비스는 하나의 재현 채널만을 제공하는 모노 오디오 서비스에서, 스테레오, 5.1채널, 7.1 채널 오디오 등과 같이 재현 채널 수가 증가되어 왔는데, 이는 더욱 공간감 있는 오디오를 제공하기 위함이다.

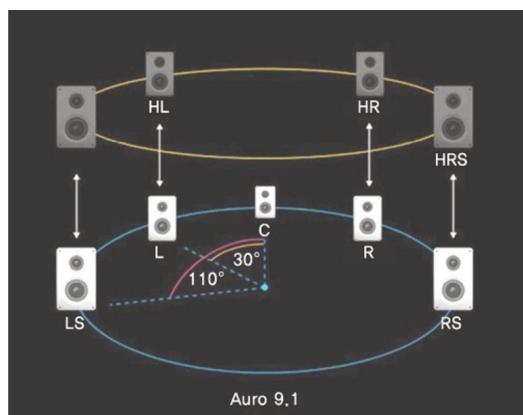
기존의 오디오 서비스에서는 위와 같은 채널 기반의 오디오 서비스를 주로 제공하였는데, 최근 표준화가 완료된 MPEG-H 3D audio와 AC-4와 같은 오디오 부복호화 기술에서는 채널 기반의 오디오와 함께 객체 기반 오디오와 고차원 앰비소닉

(HOA, Higher Order Ambisonics)을 활용한 장면 기반 오디오를 포함하기도 한다[1][2].

본 장에서는 기존 오디오 시스템에서 많이 사용되어왔던 채널 기반 오디오와 객체 기반 오디오, 장면 기반 오디오에 대해서 간략하게 살펴본다.

### 1. 채널 기반 오디오

채널 기반 오디오는 재현되는 오디오의 채널 수와 재현되는 스피커의 위치가 미리 정해진 오디오 재현 환경에서 최적의 오디오가 재현되도록 저작된 오디오 파일 형식인데, 대표적으로 스테레오, 5.1채널, 7.1채널 오디오 등이 있다. 아날로그 방송 시대의 TV나 라디오의 경우 대부분 스테레오 오디오 서비스를 제공하였으며, 디지털 방송으로 전환한 이후에도 일부 영화, 음악 프로그램을 제외한 대부분의 방송 프로그램이 스테레오 오디오 형태로 서비스되고 있다. 5.1채널 또는 7.1채널 오디오의 경우 전면 부에만 스피커가 배치되어있는 스테레오 오디오와 달리 측면 또는 후면에 스피커가 배치되어 스테레오 오디오 대비 더욱 다양한 방향의 소리를 표현할 수



〈그림 1〉 9.1채널 스피커 재현 환경

있어, 영화, 공연 실황 등에서 많이 사용되고 있다.

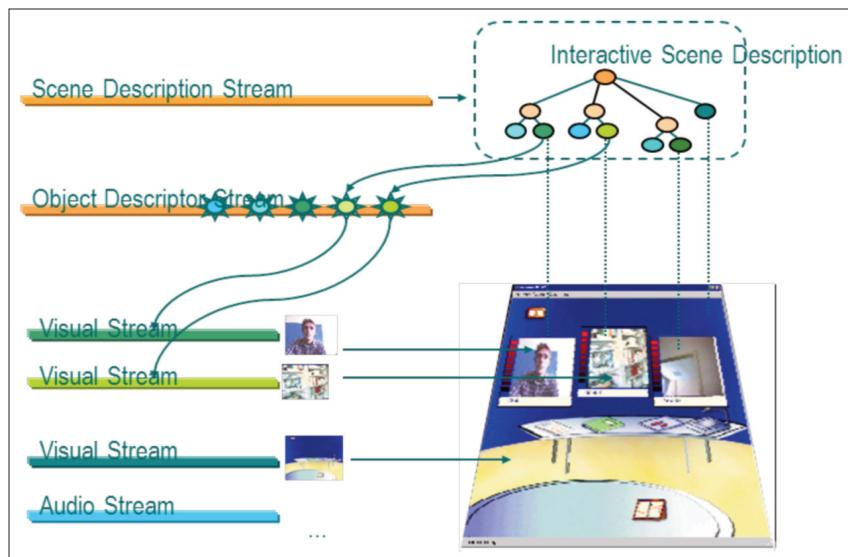
최근에는 9.1채널, 10.2채널, 11.1채널, 22.2채널 등과 같이 수평 방향뿐 아니라 수직 방향에도 스피커를 배치하여 수직 방향의 공간감을 증가시키는 오디오 재현 시스템이 제안되고 상용화가 되고 있다. <그림 1>은 수직 방향에 스피커를 포함하는 채널 기반 오디오 시스템의 한 예인 9.1 채널 오디오에 대한 스피커 재현 환경을 나타낸다. MPEG-H 3D audio, AC-4 등의 표준에서도 수직 방향의 스피커를 포함하는 오디오 재현 환경을 고려하고 있으며, 2016년도에 표준화가 완료된 국내 UHDTV 표준에서도 10.2채널 오디오 재현 시스템을 표준으로 채택하였다[3].

## 2. 객체 기반 오디오

객체 기반 오디오 기술은 20세기 후반의 MPEG-4로부터 본격적으로 출발되었는데, MPEG-4는 비

디오와 오디오를 화면(frame) 단위 또는 채널 단위가 아니라, 여러 개의 객체로 구분하고, 객체 별로 부복호화, 전송, 재현하는 방법에 대한 표준 기술이다[4]. <그림 2>는 이러한 MPEG-4 기술에 대한 개념을 나타낸 것이다.

채널 기반 오디오의 경우 다양한 오디오 신호들이 저작 단계에서 채널 기반 오디오로 렌더링(rendering)되고, 이를 전송, 재생하는 방식인데, 객체 기반 오디오는 이와 달리 오디오 신호를 개별적인 객체 단위의 음원들과 각 음원 객체의 위치, 크기 등과 같은 정보를 포함하는 메타 데이터로 나누어 기술하는 방법으로 오디오 신호를 재현하는 단말에서는 음원 객체와 메타 데이터를 이용하여 스피커의 개수나 배치 등 재현 환경에 맞게 최적화된 렌더링을 하는 방식이다. 또한 사용자가 특정 음원을 선택하여 음량, 스피커 위치 등을 조절할 수 있어 상호 작용(interactive) 서비스가 가능하다. 예를 들어, 스포츠 중계 장면에서 관중석의 응원 소리



<그림 2> MPEG-4 시스템의 객체 기반 멀티미디어 서비스 개요

를 낮추고 아나운서 해설 소리를 크게 조정하거나, 음악을 들을 때 앞에서 나는 악기 소리의 위치를 옆으로 옮길 수 있다[5].

실제로 객체 기반 오디오 기술을 이용한 음악 서비스가 ‘Music 2.0’이라는 기술로 사업화가 되었는데, 이 음악 서비스에서는 각 악기에 대한 소리 크기의 제어가 가능하다. <그림 3>은 객체 기반 오디오 기술을 이용하여 발표된 음반의 예이다.



<그림 3> 객체 기반 오디오를 활용한 음악 서비스의 예

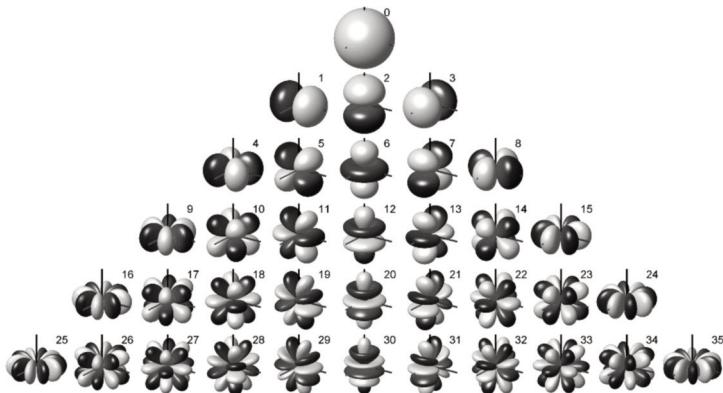
객체 기반 오디오 기술은 일부 서비스에 활용되기는 하였으나 크게 활성화되지는 못하고 있었으

나, 최근에 표준화가 완료된 오디오 부복호화 표준인 MPEG-H 3D audio와 AC-4에 포함이 됨에 따라 UHDTV 방송 등에서 활용이 될 전망이다.

### 3. 장면 기반 오디오

장면 기반 오디오는 공간 상에서의 다양한 지점에서의 음압 값으로 오디오 장면을 표현하는 기술이다. 장면 기반 오디오는 일차 앰비소닉(FOA, First Order Ambisonics)과 고차원 앰비소닉(HOA, Higher Order Ambisonics)을 모두 포함하는데, 앰비소닉을 활용한 음장의 표현은 오디오 장면의 등방성 표현을 초래하는 직교 구면 고조파 기저 함수(SH, Spherical Harmonics)를 사용하여 이루어진다. <그림 4>는 차수에 따른 직교 구면 고조파를 나타낸다.

오디오 콘텐츠 제작 또는 배포 시 고차원 앰비소닉 기술을 적용하면, 사용자는 고차원 앰비소닉 포맷의 오디오 신호를 헤드폰, 스테레오(stereo), 5.1 채널 오디오 재현 시스템 등 오디오 재생 환경에 맞춰 렌더링하여 최적화된 입체 음향을 재현할 수



<그림 4> 고차원 앰비소닉의 차수에 따른 직교 구면 고조파

있다. 또한 고차원 앰비소닉 포맷의 오디오는 음장(sound field)의 회전, 확대/축소, 이동 등이 구현 가능하여 가상 현실(VR) 오디오 제작에도 이용된다[5].

### III. VR 오디오의 획득 및 재현

현재 서비스되고 있는 대부분의 VR 콘텐츠는 채널 기반 오디오로 제작이 되어 있는데, 일부 콘텐츠에서는 객체 기반 오디오와 장면 기반 오디오가 사용되고 있다. 채널 기반 오디오의 경우 기존의 방식에 따라 획득이 되고 있으나, 객체 기반 오디오와 장면 기반 오디오의 경우 기존의 채널 기반 방식과는 다른 방식의 획득 방법이 요구되기도 한다. 객체 기반 오디오와 장면 기반 오디오의 경우 재현 시스템도 기존의 채널 오디오 신호를 단순히 출력하는 채널 기반 오디오 방식과는 다르게 동작하는데, 본 장의 아래에서 VR 오디오를 위한 획득과 재현에 대해서 조금 더 상세하게 알아본다.

#### 1. VR 오디오 획득 및 저작 기술

오디오 콘텐츠의 획득은 일반적으로 마이크로폰을 통해 이루어진다. 기존의 채널 기반 오디오에서는 하나 이상의 모노 또는 스테레오 마이크를 활용하여 오디오를 획득하고, 이를 믹싱하여 목적하는 오디오 재현 환경(예, 5.1채널 오디오)에서의 오디오 신호를 제작하는 방법이 사용된다. 스포츠 중계와 같이 현장감을 중요하게 여기는 장르에서는 다채널 마이크로폰을 활용하여 현장음을 다채널로 획득하는 방식이 사용되기도 한다.

VR 오디오에서는 360도 영상에 대응하는 공간감 있는 오디오를 획득하는 것이 이슈가 되고 있는데,

여전히 많은 경우에서 기존과 같은 채널 기반 오디오 획득 방법이 사용되고 있으나, VR 오디오를 위한 새로운 획득 방법도 시도되고 있다. 일차 앰비소닉 마이크로폰 또는 고차원 앰비소닉 마이크로폰을 활용한 방법이나, 바이노럴(binaural) 마이크로폰 등이 대표적이라 할 수 있다.

일차 앰비소닉 마이크로폰 또는 고차원 앰비소닉 마이크로폰은 <그림 5>와 같이 장면 기반 오디오(Scene based Audio)를 획득하기 위한 마이크로폰인데, 최근에는 VR 오디오 획득을 위한 소형 마이크가 출시되고 있다.



<그림 5> 장면 기반 오디오 획득을 위한 마이크로폰의 예

고차원 앰비소닉은 앞서 기술한 것과 같이 해당 공간에서의 다양한 방향에 대한 오디오를 획득하고 재현할 수 있는 기술로서, 현장감 있는 오디오를 획득하는 것에 장점이 있으며, 위와 같은 하나의 마이크로폰을 활용하여 획득이 가능하므로, 실시간 방송과 같은 서비스에서 잘 활용될 수 있다.

바이노럴 마이크로폰은 사람이 귀로 듣는 것과 유사한 소리를 획득하기 위한 마이크로폰으로서 사람 머리 모양이나 귀 모형 안에 마이크로폰을 설치하여 수음을 하는 방식이다. 바이노럴 마이크로폰은 사람의 머리 전달 함수 등을 획득하는데 많이 사용되었는데, 최근 360VR 오디오의 획득에도 사용되고 있다. 머리 전달 함수의 측정 등에서는 두 개의 마이크



&lt;그림 6&gt; 바이노럴 오디오 획득을 위한 마이크로폰의 예

로폰을 활용하는 방식이 일반적이었는데, 사람의 머리 움직임에 대응하기 위해 여러 개의 귀를 가진 마이크로폰이 등장하였고, 오클러스에서 만든 바이노럴 마이크폰이 잘 알려져 있다. <그림 6>은 제품으로 출시된 바이노럴 마이크로폰의 예를 나타낸다.

최근에는 VR 콘텐츠를 하나의 장치에서 획득하기 위하여 카메라와 마이크를 일체형으로 한 장치를 선보이고 있는데, 대표적인 것이 5개의 카메라와 64개의 마이크를 하나의 장치로 개발한 visisonics 사의 5/64 Audio/Visual Camera이다.

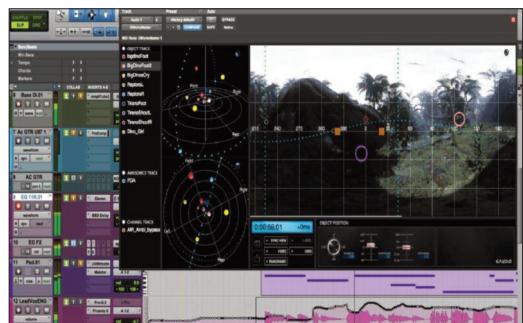


&lt;그림 7&gt; 360영상과 오디오를 함께 획득하는 장치의 예

고차원 앰비소닉 오디오나 바이노럴 오디오의 경우 획득한 신호를 믹싱 과정을 거치지 않고 전송하는 것이 가능하나, 대부분의 오디오는 오프라인에서의 제작 단계를 거친 후 전송된다. 기존의 오

디오는 스테레오, 5.1채널, 7.1채널 등과 같이 채널 기반으로 저작이 되었으며, 현재의 VR 콘텐츠의 경우에도 대부분 이와 같은 채널 기반으로 제작이 되어있다. 이러한 채널 기반의 오디오 제작 기법은 360VR 콘텐츠가 아닌 일반 콘텐츠용 오디오를 저작하기 위하여 사용이 되어온 것으로 360VR 콘텐츠의 제작에는 다소의 불편함이 있을 수 있다.

최근에는 VR용 오디오를 더욱 쉽게 저작할 수 있는 도구가 플러그인(Plug-in) 형태로 출시되고 있는데, 대표적인 것이 국내 업체인 가우디오(GAudio)에서 개발한 ‘SOL VR 360 Works’이다[6]. <그림 8>은 가우디오의 저작도구 ‘SOL VR 360 Works’의 실행 화면인데, 가우디오의 SOL VR 360 Works는 VR용 오디오를 제작할 수 있는 프로툴즈(Pro tools)



&lt;그림 8&gt; 가우디오의 저작도구 SOL VR 360 Works 실행 화면

용 플러그인으로서 영상의 특정 위치에 오디오 음원을 위치시키고, 시간에 따라 음원의 3차원 공간상의 위치를 이동하는 것이 가능하도록 하여, VR-용 오디오를 쉽게 저작할 수 있도록 하였다. 저작된 오디오는 가우디오의 독자 포맷인 GA5 포맷으로 저장이



〈그림 9〉 AudioEase의 ‘360pan suite’ 실행 화면

되는데, 장면 기반 오디오와 객체 기반 오디오, 채널 기반 오디오를 모두 지원한다.

이와 유사한 플러그인으로 AudioEase의 ‘360pan suite’가 있다[7]. 〈그림 9〉는 AudioEase 사의 ‘360pan suite’ 실행화면인데, 360VR 콘텐츠의 오디오를 제작하는데 특화된 프로툴즈용 플러그인으로서 영상 장면의 특정 위치에 오디오 객체를 위치시키고, 시간에 음원의 3차원 공간상의 위치를 이동하는 것이 가능하다.

이외에도 360VR 오디오 저작을 위한 전용 플러그인이 출시가 되고 있으며, 향후 더욱 많은 플러그인이 출시가 될 것으로 전망된다.

한편, 360VR 전용 플러그인으로 볼 수는 없으나, 기존의 채널 기반 저작 도구와 달리 객체 오디오를 지원하는 플러그인이 있는데, 이러한 플러그인도 360VR 오디오의 제작에 활용될 수 있다. 대



〈그림 10〉 채널 기반 및 객체 기반 오디오 저작이 가능한 플러그인의 예

표적인 것으로는 Dolby, DTS, Auro3D의 플러그인이 있으며, 국내에서는 한국전자통신연구원에서 채널 기반 오디오와 객체 기반 오디오를 지원하는 프로토콜즈용 플러그인을 개발하고 있다.

## 2. VR 오디오 재현 기술

VR 콘텐츠의 오디오 재현 측면에서 중요한 것은 체험자의 머리 움직임에 따른 오디오의 상호 작용이라 할 수 있다. VR 콘텐츠를 시청하면서 머리를 다른 방향으로 이동하는 경우, 해당 방향의 영상이 HMD와 같은 디스플레이에 재현이 된다. 이때, 머리 움직임에 따라 소리도 방향이 변경되어야 한다. 소리를 스피커로 재현하는 환경에서는 체험자의 시선 방향이 변경되면 자연스럽게 사람의 머리와 스피커의 각도가 변경되기 때문에 별도의 보상 처리를 해 줄 필요가 없으나, 헤드폰으로 오디오를 청취하는 경우에는 체험자의 머리 움직임에 따라 소리의 방향도 적절하게 보상해 주어야 한다.

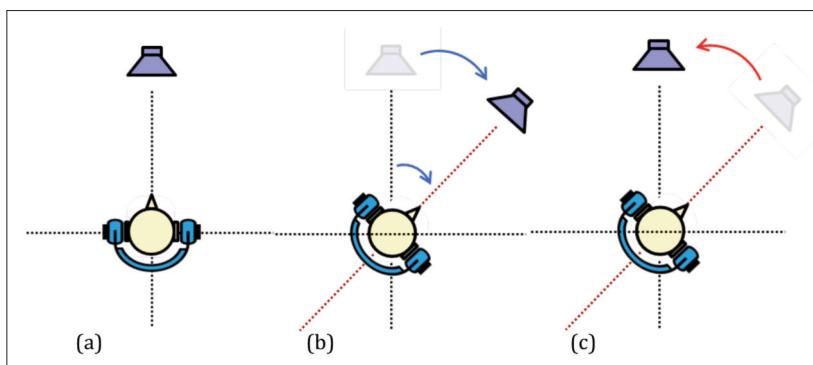
채널 기반 오디오 재현 단말에서는 다양한 각도에 대한 바이노럴 필터를 활용하여, 이러한 머리 움직임에 대한 연동을 지원할 수 있다. <그림 11>은

채널 기반 오디오를 헤드폰 청취 환경에서 머리 움직임을 보상하는 방법에 대한 것이다.

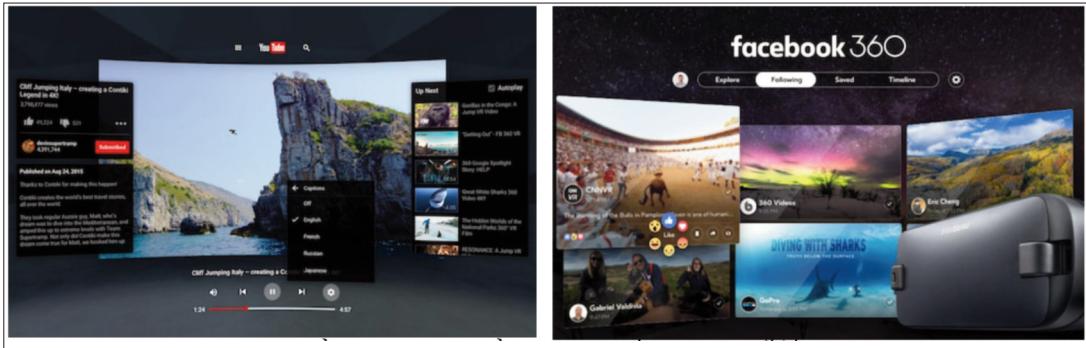
머리 움직임이 없을 때에 정면에 있던 스피커에서 재생되는 소리는 체험자가 (b)와 같이 머리를 오른쪽 방향으로 움직이면, 체험자의 시선을 기준으로 왼쪽 방향에서 소리가 나야 한다. 그러나, 머리 움직임에 대한 스피커의 위치 보정을 하지 않을 경우 (b)와 같이 체험자의 시선과 함께 소리의 위치도 함께 시선의 정면으로 이동하게 되어 소리가 원래의 위치에 남아 있지 않게 된다. 이때, 적절한 신호 처리를 통하여, (c)와 같이 소리가 사람의 왼쪽 방향에서 나도록 보정해 주면, 머리는 움직이더라도 소리는 3차원 공간상에서 움직이지 않은 것과 같은 효과를 줄 수 있다.

고차원 앰비소닉 오디오는 앞서 기술한 것과 같이 음장의 회전이 가능하여, 헤드폰 청취 환경에서도 머리 움직임에 따른 오디오 신호의 위치 보정이 가능하다.

객체 기반 오디오 재현 단말에서도 객체 오디오를 다채널 오디오로 변환하는 과정에서 머리 움직임에 따른 위치 보정을 해 주거나, 변환된 다채널 오디오에 적용하는 바이노럴 필터의 각도를 머리의 움직임에 따라 변경하는 방법을 활용하여 머리 움직임에 대한 오디오의 위치 보정을 헤드폰 청취 환



<그림 11> 채널 기반 오디오의 머리 움직임 보상 방법



(그림 12) YouTube와 Facebook의 360 VR 체험

경에서도 제공할 수 있다.

우리가 쉽게 접할 수 있는 YouTube, Facebook VR 플레이어는 고차원 앰비소닉 포맷을 지원하며, 머리 움직임에 따른 오디오 신호의 위치 보정도 제공된다. 그러나 대부분의 VR 재현 단말은 채널 기반의 오디오만을 지원하고 있으며, 머리 움직임에 따른 오디오의 위치 보정 기능을 포함하고 있지 않은 것으로 파악된다.

국내에서는 SBS, 가우디오에서 공동 개발한 안드로이드용 VR 재생 앱에서 채널 기반 오디오, 장면 기반 오디오, 객체 기반 오디오 등으로 구성된 오디오를 재현하고 있는데, 머리 움직임에 따른 오디오 신호의 보정 기능도 포함하고 있다. 한가지 아쉬운 점은 다양한 형태의 오디오를 포함하는 VR 관련 오디오 표준이 없어서 독자적인 규격을 사용하고 있는 상태로, 타 콘텐츠와의 호환성 측면에서는 어려움이 있는 상태이다.

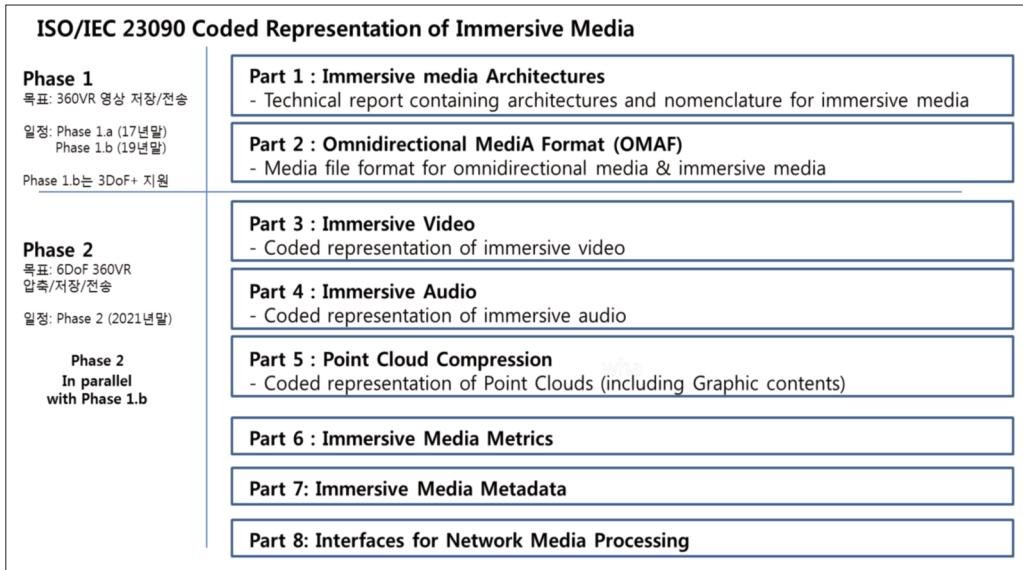
#### IV. VR 관련 오디오 표준화 동향

현재 방송, 게임, 가상 체험 등과 같은 많은 플랫

폼에서 VR 서비스가 제공되고 있으나, 대부분 자체적인 포맷을 사용하고 있어, 콘텐츠 간 호환이 되지 않는 경우가 많이 있다. 이는 VR 콘텐츠와 관련해서는 국제적인 표준화 기구에서 적절한 규격이 제정된 예가 없기 때문이라 할 수 있는데, 최근에는 MPEG(Moving Picture Experts Group), DVB(Digital Video Broadcasting), 3GPP(3rd Generation Partnership Project) 등과 같은 표준화 기구에서 VR 서비스를 위한 표준화를 준비 중이거나 진행 중에 있다.

##### 1. MPEG 표준화 동향

MPEG에서는 비디오/오디오의 압축, 저장, 전송과 관련한 많은 표준을 제정하여 왔는데, 오디오 부복호화 관련하여 가장 최근에 표준화가 완료된 표준은 2016년에 표준이 제정된 MPEG-H 3D audio이다. 기존의 채널 기반 오디오 기술을 중심적으로 표준화하였던 것과 달리 MPEG-H 3D audio에서는 객체 기반 오디오와 장면 기반 오디오를 포함하고 있다. 객체 기반 오디오와 장면 기반 오디오 기술을 포함함에 따라 MPEG-H 3D audio



&lt;그림 13&gt; MPEG-I 단계별 Part 구성

표준을 이용하면 3DoF(Degree of Freedom)에서의 오디오는 충분히 표현이 가능하다고 볼 수 있다.

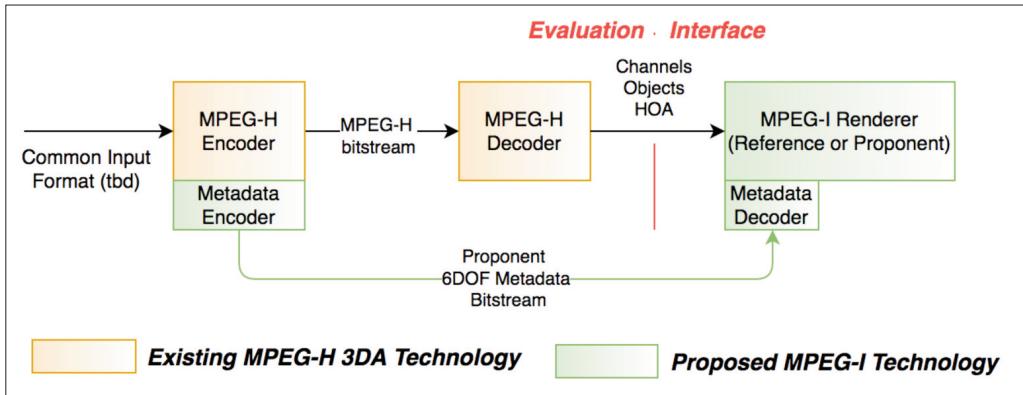
최근에는 몰입형 미디어에 대한 차세대 프로젝트인 MPEG-I (ISO/IEC 23090 Coded Representation of Immersive Media) 표준화가 진행이 되고 있는데, MPEG-I는 비디오, 오디오에서 시스템 요소에 이르기까지 전체 콘텐츠 에코시스템을 대상으로 하는 표준 기술로 구성된 프로젝트이며, 360 VR비디오와 MR(Mixture Reality) 및 6DoF 관련 콘텐츠를 포함한 몰입형 미디어 전체를 고려 대상으로 한다.

<그림 13>은 MPEG-I의 단계별 파트 구성을 나타낸 것인데, 총 2 단계의 Phase와 8개의 Part로 정의되어 있다. 파트 4에서 몰입형 오디오를 위한 규격을 정의하는데, 6DoF 콘텐츠 표현을 위해 필요한 특정 오디오 압축 및 재생 관련 표준이 포함된다.

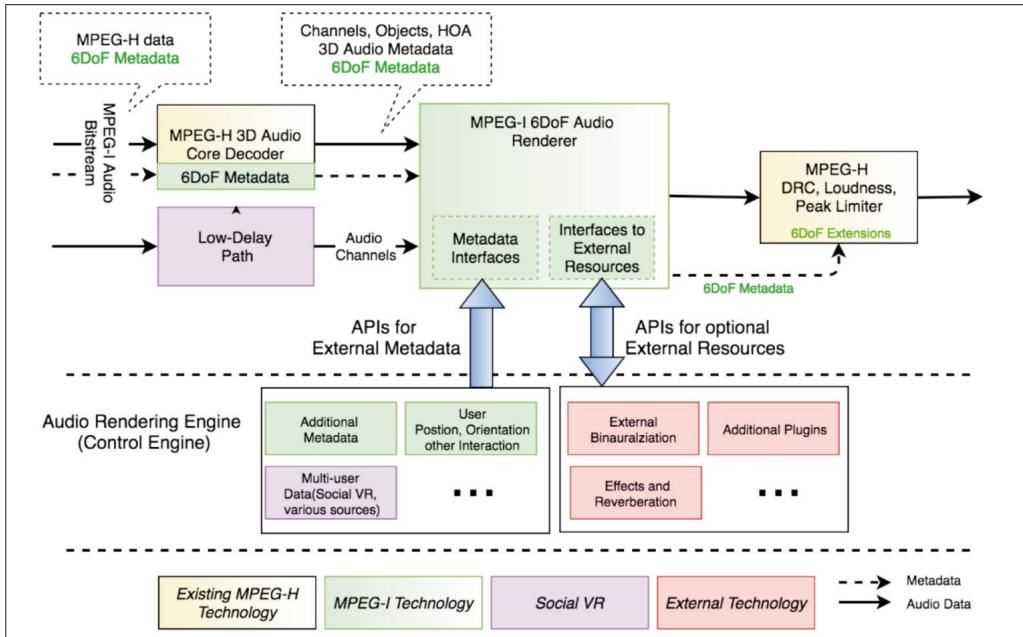
MPEG의 오디오 표준화 그룹에서는 AR/VR 표준화를 진행하고 있는 MPEG-I, OMAF 그룹과 오디오 관련 요구사항을 논의하고 있으며, 오디오 전문가 그룹에서 AR/VR 활용 사례를 다루기 위해 필요한 오디오 기술 등을 정리한 “Thoughts on AR/VR and 3D Audio”문서를 118차 MPEG 회의에서 발행하였다.

MPEG 119차 회의에서 오디오 서브그룹에서는 “MPEG-I Audio Architecture and Evaluation Procedures for 6DoF”문서를 발행하였다[8]. 해당 문서에서는 기존의 MPEG audio 기술과 MPEG-I에서 표준화가 되어야 하는 부분을 구분하고 있다. 이를 나타낸 것이 <그림 14>인데, MPEG-H 3D Audio를 기반으로 하며, 6DoF를 위한 메타 데이터와 이를 렌더링 하는 모듈을 추가하는 개념적인 구조로 되어 있다.

MPEG 122차 회의에서는 “Thought on MPEG-I audio Requirements” 문서를 발행하였다[9]. 해당



&lt;그림 14&gt; MPEG-I audio의 개념적인 구조도



&lt;그림 15&gt; MPEG-I 오디오 기준 시스템 구조

문서에서는 MPEG-I audio의 요구사항에 대한 다양한 의견을 포함하고 있으며, 이전에 제시하였던 것 보다 좀더 구체화 된 MPEG-I Audio 시스템 구조를 제시하고 있다. <그림 15>는 해당 문서에 기술된 MPEG-I 오디오 기준 시스템 구조를 나타낸다.

2018년 10월 기준으로 MPEG에서는 MPEG-I 오디오를 위한 다양한 요구사항(requirements)에 대해 논의하고 정의하는 작업이 여전히 진행되고 있으며, 아직은 6DoF VR 오디오를 위한 표준화의 시작 단계로 볼 수 있다.

## 2. DVB 표준화 동향

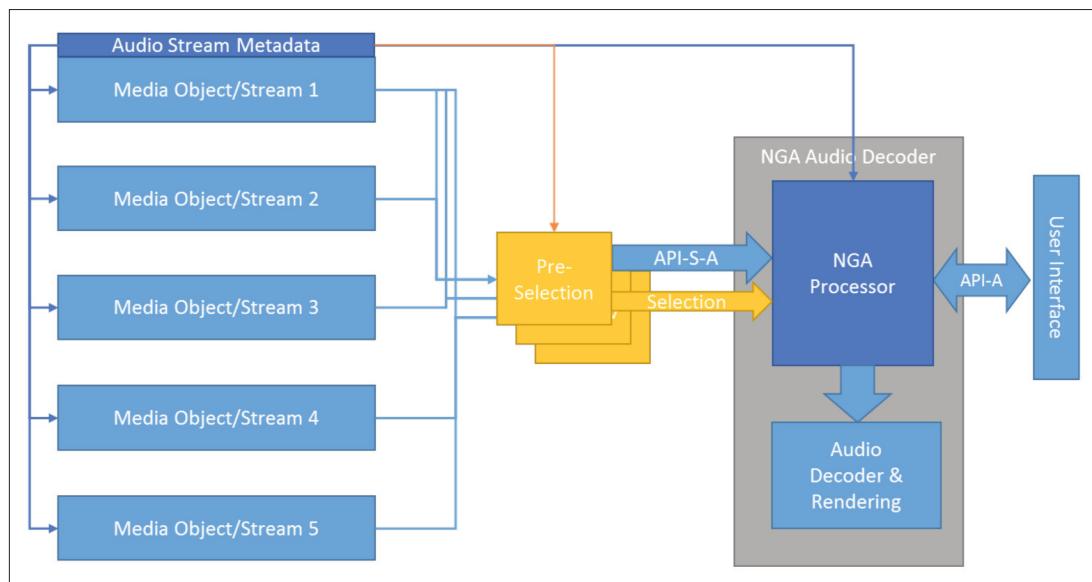
DVB는 시청자가 화면 내 시점을 자유롭게 선택할 수 있고 주시뷰(RoI: Region of Interest) 기반 고화질 공간 해상도를 제공함과 동시에 최대 6DoF 움직임을 제공하여 몰입감 및 사실감을 극대화 할 수 있는 VR 몰입형 미디어 방송기술 표준화 계획을 제시하고 있다. 이를 위하여 DVB는 2015년 10월부터 방송에서의 VR 미디어 적용 및 하이브리드망 기반의 전송 기술에 대한 표준화를 위하여 CM (Commercial Module) 산하에 VR(Virtual Reality) 그룹(이하 DVB-CM-VR-SMG(DVB Commercial Module VR Study Mission Group))을 새롭게 신설했었다.

DVB-CM-VR-SMG의 Audio Subgroup에서는 표준화에 앞서 VR 오디오를 위한 기술 분석을 진행하였으며, ‘Audio and Virtual Reality’라는 문서를 2016년 7월에 발간하였다. 해당 내용은 이후

2016년 11월에 발간된 DVB-CM-VR-SMG의 문서인 ‘Report of the DVB CM Study Mission on Virtual Reality (CM1697)’에 포함되었다[10].

해당 문서에는 VR을 위한 오디오의 요구사항과 VR에 활용될 수 있는 오디오 기술에 대해 분석한 내용이 기술되어 있는데, DVB의 UHDTV를 위한 표준인 ‘ETSI TS 101 154’[11]에서 사용되었던 NGA(Next Generation Audio) 기술을 이용한 VR 서비스에 대한 검토를 포함하고 있다. <그림 16>은 해당 문서에 포함된 NGA 전송 모델을 나타낸 것이다.

해당 문서에는 기술 분석에 따른 결론 및 DVB를 위한 권고 사항을 포함하고 있다. 오디오와 관련된 내용도 포함되어 있는데, 객체 기반 오디오와 장면 기반 오디오를 지원하는 것, 머리 움직임에 따른 오디오의 처리, 오디오의 품질에 대한 내용 등이 해당된다. 이러한 내용이 향후 DVB에서 VR 관련 표준화를 추진 시에 반영될 것으로 예측되는데, 현재로



<그림 16> DVB의 NGA 전송 모델

서는 본격적인 VR 서비스 관련 표준화가 진행되지 않는 않고 있다고 볼 수 있으며, 향후 MPEG, 3GPP 등과 같은 타 기관과의 협력을 통한 규격 제정이 예상된다.

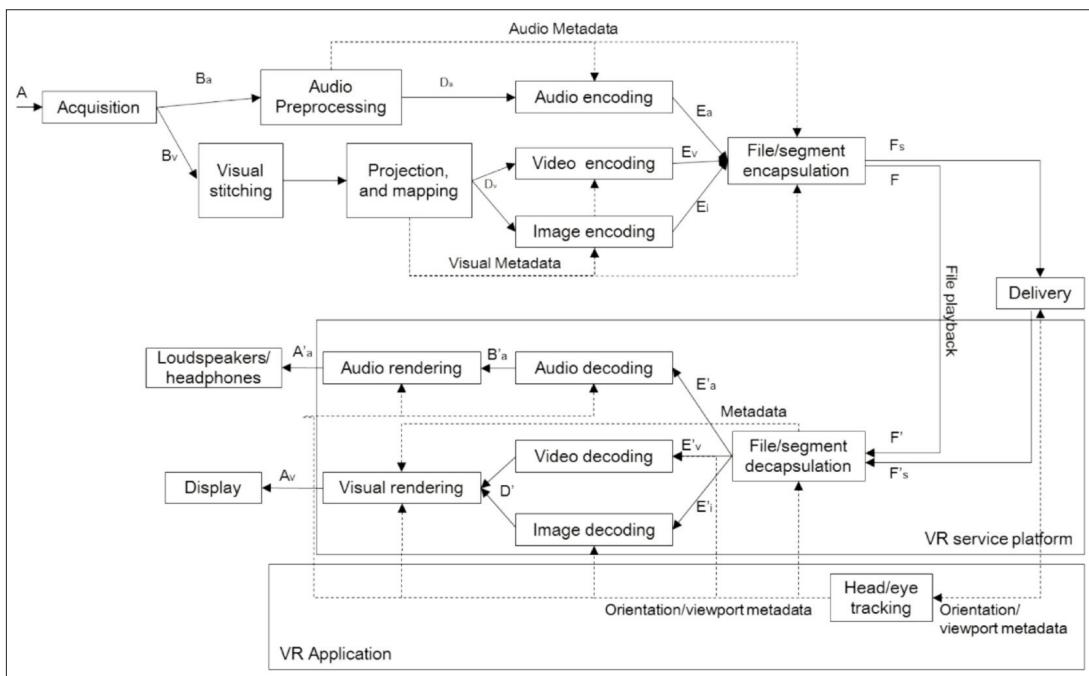
### 3. 3GPP 표준화 동향

3GPP에서는 이동통신 규격 하에서 VR 서비스를 제공하기 위한 표준화 작업을 진행 중에 있다. 규격의 제정에 앞서 3DoF 기반 VR 서비스를 기준으로 한 몰입형 비디오 및 오디오 기술에 대한 기술보고서를 2016년도부터 작성하고 있다[12]. 기술 보고서에는 VR 비디오 시스템, 오디오 시스템, VR의 사용 사례, 오디오 및 비디오의 품질 평가 방법, 지연과 동기화 문제 등과 같이 VR 서비스를 다양한 각도에서 조

사하고 분석하고 있다. 또한 이를 기반으로 기존의 3GPP 규격에서 서비스가 가능한 부분과 추가적으로 정의되거나 새로이 도입되어야 할 기술 등에 대해서도 기술하고 있다. <그림 17>은 기술 보고서에 제시된 VR 스트리밍 서비스를 위한 시스템의 구조의 예를 나타낸 것으로, A, B, D, E, F는 시스템간의 인터페이스를 구분하기 위하여 사용되었다.

3GPP의 기술 보고서에는 VR 서비스에서의 오디오 기술에 대해서도 상당히 자세하게 기술을 하고 있는데, 오디오 획득 시스템, 콘텐츠 제작 워크 플로우, 오디오 포맷, 오디오 렌더링 시스템, VR 전송 시스템에서의 오디오 성능 측정 등의 내용을 포함하고 있다. 또한, VR에서 요구될 것으로 예측되는 오디오 기술의 프로파일에 대해서도 분석을 하고 있다.

기술 보고서의 마지막 장에서는 VR 서비스를 위



<그림 17> VR 스트리밍 서비스를 위한 구조의 예

한 규격 제정을 위하여 추가적으로 수행되어야 할 부분에 대해서 기술하고 있는데, 오디오 관련 내용으로는 앰비소닉 오디오를 적용하면서도 좋은 성능의 오디오를 제공하는 방법에 대한 연구와 VR 오디오 시스템의 성능을 측정하는 방법에 대해서는 추가적인 연구가 필요하다는 의견을 제시하고 있다.

## V. 결 론

본 고에서는 VR 서비스에 활용될 수 있는 최신 오디오 기술과 최신 오디오 기술이 어떻게 VR 서비스에 적용이 되고 있는지를 간략하게 살펴보고, 여러 국제 표준화 기구의 표준화 현황에 대해서도 살펴보았다.

최신 오디오 기술로서, 기존에 많이 사용되는 채널 기반의 오디오와 함께 향후 VR 오디오에 적용될 것으로 전망이 되는 객체 기반 오디오와 장면 기반 오디오 기술에 대해 살펴보았다. 그리고 객체 기반

오디오와 장면 기반 오디오의 획득, 제작 및 재현을 위한 기술 및 제품에 대한 조사를 통하여 이미 이러한 최신 오디오 기술이 VR 콘텐츠에 활용이 되고 있음을 알 수 있었다.

그러나 VR 콘텐츠를 위한 표준이 제정되지 않은 상황에서 콘텐츠 간의 호환이 되지 않는 문제가 있으며, 객체 기반 오디오와 장면 기반 오디오의 제작 및 재현을 위한 기기 및 제품들이 아직은 초기 단계로 채널 기반 오디오 관련 제품에 비해 많지 않음을 알 수 있었다.

현재 MPEG, DVB, 3GPP 등과 같은 단체에서 VR 서비스 관련 표준 제정을 위한 노력이 진행되고 있는데, 이러한 표준화가 완료되면 콘텐츠 호환성에 대한 문제가 해결이 될 것으로 전망이 되며, 기존의 채널 기반 오디오 만을 지원하는 제품들도 객체 기반 오디오와 장면 기반 오디오를 지원하는 형태로 발전되고 있어, 향후에는 최신 오디오 기술을 포함하는 VR 콘텐츠의 제작 및 재현이 더욱 쉬워질 것으로 전망된다.

### 참고 문헌

- [1] ISO/IEC-23008-3, "Information technology - High efficiency coding and media delivery in heterogeneous environments Part 3: 3D audio", 2016. 1.
- [2] ETSI TS 103 190, "Digital Audio Compression (AC-4) Standard", 2014. 04.
- [3] TTAK.KO-07.0127/R1, "지상파 UHDTV 방송 송수신 정합", 2016. 12.
- [4] ISO/IEC 14496-1, "Information technology coding of audio-visual objects Part 1: Systems", 1998. 05.
- [5] 한국정보통신기술협회, 정보통신용어사전
- [6] <https://www.gaudiolab.com/sol-vr360-sdk>
- [7] <https://www.audioease.com/360>
- [8] ISO/IEC JTC1/SC29/WG11 MPEG2017/N17038, "MPEG-I Audio Architecture and Evaluation Procedures for 6DoF", 2017. 07.
- [9] ISO/IEC JTC1/SC29/WG11 MPEG2018/N17647, "Thoughts on MPEG-I Audio Requirements", 2018. 04.
- [10] DVB CM-VR-SMG, "Report of the DVB CM Study Mission on Virtual Reality (CM1697)", 2016.11.
- [11] ETSI TS 101 154, "Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications", 2018. 2.
- [12] 3GPP TR 26.918, "Virtual Reality (VR) media services over 3GPP", 2017. 7.

## 필자소개



### 이용주

- 1999년 : 경북대학교 전자공학과 공학사
- 2001년 : 경북대학교 대학원 전자공학과 공학석사
- 2013년 : 경북대학교 전자공학과 박사 수료
- 2001년 ~ 현재 : 한국전자통신연구원 책임연구원
- 주관심분야 : 디지털 방송기술, 디지털 신호처리, 3차원 오디오 신호처리



### 이태진

- 1996년 : 전북대학교 전자공학과 공학사
- 1998년 : 전북대학교 대학원 전자공학과 공학석사
- 2014년 : 충남대학교 전자전파정보통신공학과 공학박사
- 1998년 2월 ~ 2000년 5월 : Mobens, Co., Ltd. 영상음성기술연구팀 연구원
- 2002년 10월 ~ 2003년 9월 : 일본 Tokyo Denki University, 방문연구원
- 2000년 ~ 현재 : 한국전자통신연구원 책임연구원
- 주관심분야 : 오디오 부호화, 실감음향, 오디오 신호처리



### 김희용

- 1994년 : 한국과학기술원 전기및전자공학과 공학사
- 1998년 : 한국과학기술원 전기및전자공학과 공학석사
- 2004년 : 한국과학기술원 전기및전자공학과 공학박사
- 2003년 ~ 2005년 : (주)에드파크놀러지 기술연구소 멀티미디어 팀장
- 2006년 ~ 2010년 : 과학기술연합대학원대학교(UST) 이동통신 및 디지털 방송공학과 겸임교수
- 2013년 ~ 2014년 : Univ. of Southern California(USC) 멀티미디어통신연구실 방문연구원
- 2005년 ~ 현재 : 한국전자통신연구원 실감AV연구그룹장
- 주관심분야 : 영상처리/압축, 컴퓨터비전, UHD/HDR/3D/VR, MPEG