

ETRI 미디어부호화연구실 주요 연구 소개

김종호, 강정원, 정세윤, 백승권, 강경욱, 정영호, 방건, 성종모, 최진수, 이태진 / ETRI 미디어부호화연구실

한국전자통신연구원(ETRI)은 지난 1976년 설립된 이후 45년간 대한민국 정보통신산업의 성장을 이끌어온 과학기술정보통신부 산하 글로벌 정보통신기술(ICT) 국책연구기관으로, ‘미래 사회를 만들어 가는 국가지능화 종합 연구기관’이라는 비전과 함께 ‘AI 강국 코리아’실현을 위해 힘찬 도약을 준비 중이다.

ETRI가 현재 수행하고 있는 주요 연구개발 분야로는 ▲빅데이터, 클라우드, 자율주행차 등 초지능 SW·콘텐츠 인공지능 분야 ▲차세대 방송서비스와 5G, 위성 및 전파 관련 통신미디어 기술 ▲도시, 에너지, 국방, 복지 등 사회 전반에서 상호작용하는 초연결 통신 및 지능화 융합연구 ▲국가 산업 기반 고부가가치 ICT 소재부품 창의연구 등이 있다.

ETRI는 박사급 연구원 1천여 명을 포함해 인력 규모가 총 2천 5백여 명에 달하며 연간 R&D 예산이 6천 2백억 원 수준이다. 명실상부 대한민국 최고 싱크탱크(Think Tank)로서 ICT R&D를 책임지고 있는 정부출연연구기관이다.

ETRI는 4개의 전문연구소(인공지능연구소, 통신미디어연구소, 지능융합연구소, ICT창의연구소)로 구성되어 있으며 미디어부호화연구실은 통신미디어연구소 산하 미디어연구본부에 속해있다.

미디어부호화연구실은 ETRI에서 30여 년간 비디오/오디오 신호처리 기술을 개발해 온 전문연구실로, 여러 조직명으로 비디오/오디오 관련 독립 조직을 이루어 오다가 '17년부터 비디오/오디오 연구실을 하나의 그룹으로 통합하였고, '19년부터 미디어부호화연구실로 조직명을 변경하였다.



<그림 1> ETRI 본원 전경

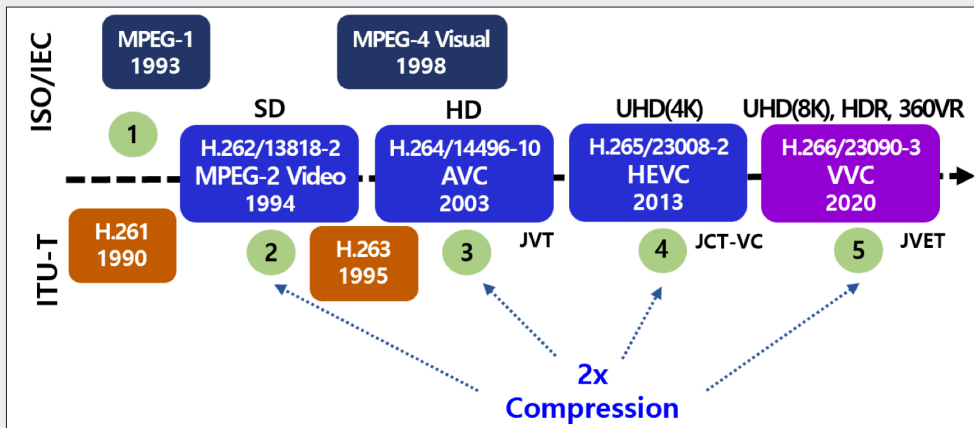


<그림 2> ETRI 미디어부호화연구실 history

미디어부호화연구실에서는 미디어(비디오/오디오) 부호화 기술 연구개발 및 표준화와 함께 비디오 품질 측정 및 향상 기술, 오디오 획득, 재현 및 응용 기술에 대해 30여 명의 연구원이 연구를 수행하고 있다. ETRI 미디어부호화연구실에서 수행하고 있는 주요 연구/개발 주제는 다음과 같다.

1. 2D 평면 비디오 부호화 기술

일반적으로, 2D 평면 비디오 부호화 기술은 비디오 신호에 포함된 시간과 공간 중복성을 제거하고 남은 차분 신호에 대해 주파수 영역으로 변환한 후 양자화와 엔트로피 코딩을 하는 방법으로 하이브리드 비디오 부호화 방식 위주로 발전해왔다. 이러한 하이브리드 비디오 부호화 방식에 기반한 비디오 부호화 표준은 국제표준화기구인 ISO/IEC JTC 1 SC 29 MPEG(Moving Pictures Experts Group)과 ITU-T SG



<그림 3> 비디오 부호화 기술 표준화 동향

16 VCEG(Video Coding Experts Group)에서 진행되어 왔으며, MPEG에서 CD와 같은 저장 매체를 위해 제정한 1세대 비디오 부호화 표준인 MPEG-1을 제외하면, 디지털 방송의 비디오 압축방식으로 사용된 2세대 비디오 부호화 표준부터 8K 고화질 영상과 HDR(High Dynamic Range), 360 VR(Virtual Reality)을 모두 지원하는 5세대 비디오 부호화 표준인 VVC(Versatile Video Coding)까지 다양한 비디오 부호화 표준을 MPEG과 VCEG이 공동으로 개발하였다.

2D 평면 비디오 부호화 기술개발과 표준화는 본 연구실에서 수행해 온 대표적인 연구 주제 중 하나로 '08년부터 HEVC(High Efficiency Video Coding) 표준화를 위한 핵심기술 개발로부터 시작되었다. HEVC 표준화가 본격적으로 시작된 '10년부터 표준화에 참여하여 다수의 지식재산권을 확보하였으며 해당 특허들의 표준 특허 풀 진입을 통해 대외적으로 본 연구실에서 연구한 기술에 대한 가치를 본격적으로 인정받기 시작하였다. 또한, 4K UHD 영상을 초당 60프레임의 속도로 실시간 압축하는 HEVC 기반 비디오 인코더 국산화 개발에도 성공함으로써 상용화 단계의 기술력도 갖추 수 있었다. 이후, HEVC 부호화 표준화의 경험을 발판삼아 '16년부터는 HEVC 다음 단계의 부호화 기술 개발을 시작할 수 있었고 '18년 CfP(Call for Proposal)를 시작으로 VVC 표준화에 본격적으로 참여하여 8K, 360도 비디오 등을 위한 초고압축 비디오 부호화를 위한 핵심기술들을 표준에 반영시켜 한 단계 발전된 연구실의 위상을 알릴 수 있는 계기가 되었다. '20년 7월 VVC 표준 제정 완료 후 JVET(Joint Video Expert Team)에서는 Post-VVC, 즉 VVC 이후 차세대 압축 기술 표준화를 위한 준비 작업들이 시작되었다. 그 일환으로 '20년 10월 회의에서 NNVC(Neural Network-based Video Coding)이라는 이름으로 신경망 기반 비디오 압축 기술에 대한 논의를 본격적으로 시작하였으며, '21년 4월 회의에서는 'Enhanced compression beyond VVC capability'라는 이름으로 기존의 신호처리 알고리즘 기반 비디오 부호화 기술에 대한 논의를 시작하고 기술 탐색 및 검증을 위한 참조 소프트웨어 ECM(Enhanced Compression Model)을 배포하였다. 이러한 흐름에 앞서 본 연구실은 '17년도부터 '초실감 테라미디어를 위한 AV부호화 및 LF미디어 원천기술 개발' 과제를 통해 Post-VVC 핵심 요소기술 개발 및 표준화 활동들을 해오고 있으며 아래와 같이 3가지 기술 개발 측면에서 연구를 진행해 오고 있다.

1) 전통적인 신호처리 기반 비디오 부호화

기존 비디오 압축 기술 표준화에서 주로 다루지는 기술 개발 방식으로 신호처리 기반 알고리즘을 이용하여 압축 툴/모듈들의 성능을 높이는 방식을 말한다. 본 연구실에서는 다양한 기술 개발 분야 중, 주로 움직임 예측에서의 후보 벡터 결정/생성 기술, 움직임 벡터 보정 기술 등의 화면 간 예측 기술들과 기하학적 분할 방식과 복수 참조 샘플 병합 기반 방식을 이용한 화면 내 예측 기술들에 관한 연구를 진행하고 있다.

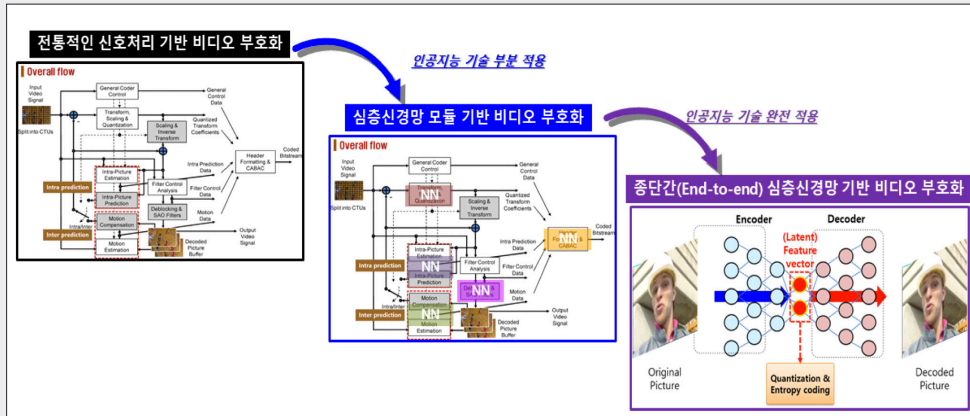
2) 심층 신경망 모듈 기반 비디오 부호화

심층 신경망 모듈 기반 비디오 부호화 기술은 압축과정에서 기존의 압축 툴/모듈들의 일부 또는 전체

를 최근 주목받고 있는 신경망 기술로 대체/확장하거나 이를 이용하여 새로운 압축 톨을 만들어 압축 성능을 높이는 방식을 말한다. 특히 본 연구실에서는 심층 신경망 기술에 기반한 화면 내 예측 블록 생성 기술, 양-예측 기술, 보간 기술 및 인-루프 기술 등 다양한 기술에 대한 연구를 진행하였고 관련 핵심 지식재산권 확보 및 기술 기고를 진행하였다.

3) 종단간(End-to-End) 심층신경망 기반 비디오 부호화

종단간(End-to-End) 심층신경망 기반 비디오 부호화 기술은 전통적인 압축 알고리즘에 의한 영상 압축 방식과는 완전히 다른 새로운 방식으로써, 단일신경망 오토인코더의 병목(bottleneck) 계층에서 추출한 은닉벡터(Latent vector)를 양자화 및 엔트로피 코딩하여 영상 압축효율을 높이는 기술을 말한다. 비디오 압축 기술 표준화 과정에서는 해당 기술이 아직 부각되고 있지 않지만, 본 연구실에서는 핵심 원천기술 확보 차원에서 선행 연구를 진행하고 있다. 관련하여 정지 영상 부호화 관련 국제 경진대회인 CLIC(Challenge on Learned Image Compression)에서 최근 몇 년간 최상위권 입상과 더불어 탑-티어 (Top-tier) 학회 논문발표 등 여러 방면에서 다양한 성과를 내고 있다.



<그림 4> 비디오 부호화 기술 개발 방식

2. 6DoF 입체공간 비디오 부호화 기술

미래의 미디어 서비스는 가상과 현실의 경계를 허무는 초실감 기술을 통해 사용자가 3차원 가상공간을 자유롭게 이동할 수 있는 서비스를 추구할 것으로 전망된다. 이에 입체공간 미디어 서비스의 장에 요소 가운데 중 하나인 압축률 한계를 극복하기 위하여, 미디어부호화연구실은 라이트필드, 포인트클라우드, 디지털 홀로그램과 같은 입체공간 비디오에 특화된 새로운 초고압축 부호화 기술을 개발하고 MPEG Immersive Video, Point Cloud Compression 등 표준화를 진행하고자 '19년부터 10년간

‘6DoF 입체 공간 비디오의 초고압축 부호화 기술개발 및 표준화 과제’를 수행하고 있다. 본 과제에서는 ‘24년까지 6DoF 라이트필드 비디오 부호화 기술을 집중적으로 개발하고, ‘25년부터 ‘28년까지 개발한 6DoF 라이트필드 비디오 부호화 고도화 기술 개발 및 디지털 홀로그램 부호화 기술 개발을 진행할 예정이다. 현재 개발이 진행 중인 6DoF 라이트필드 비디오 부호화 기술은 다음과 같은 세부 기술들을 포함한다. (<그림 5> 참고)

1) 3차원 기하변환 및 표현 기술

다수의 카메라 배열들로부터 획득한 다시점의 영상들을 3차원 공간상의 동일좌표계로 변환하여 동일 좌표계상에서 시점별 속성(색상 정보) 표현이 가능하도록 하는 기술들로, 플레넵틱 포인트클라우드 및 메쉬 변환을 통해 공간 내 시점별 속성을 표현하는 기술, 다중평면 이미지(MPI, Multi-Plane Image)를 이용하여 공간 내 시점별 속성을 표현하는 기술, 딥러닝 네트워크를 이용하여 시점별로 서로 다른 장면 정보를 생성하거나 표현하는 기술 등을 주제로 연구하고 있다.

2) 시점 종속 성분과 시점 독립 성분 분해를 통한 시점간 속성 중복성 제거 기술

동일좌표계에서 표현된 시점별 속성(색상 정보)들 간의 중복성을 제거하기 위한 기술들로, DCT와 같이 전통적인 주파수 분해 및 해석을 통해 중복성을 제거하는 방법, 가우시안 또는 BRDF(Bidirectional Reflectance Distribution Function)와 같이 광학계 물리 모델을 기반으로 중복성을 제거하는 방법, 딥러닝을 이용하여 시점 종속/독립 성분을 데이터 압축 관점에서 유리하게 분해하여 중복성을 제거하는 방법들에 대한 연구를 진행 중에 있다.

3) 3차원 기하 정보 및 속성 정보의 시공간 압축 기술

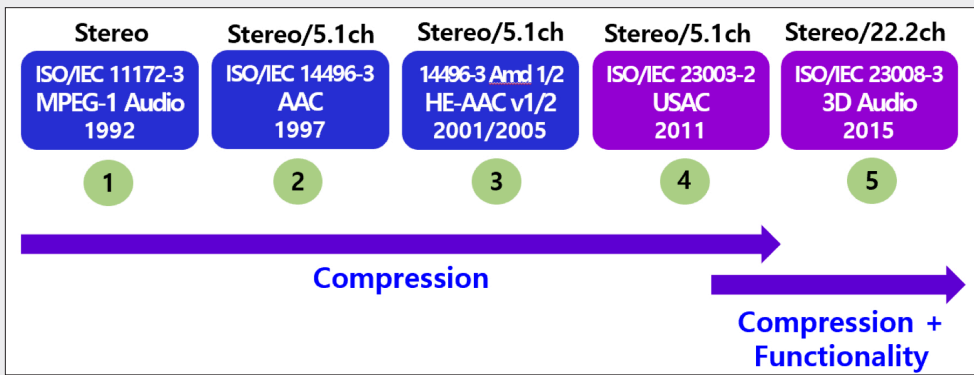
3차원 기하변환에서 생성된 기하정보와 시점 간 중복성이 제거된 속성정보를 시공간으로 압축하는 기술로, HEVC, VVC와 같은 기존의 평면 비디오 부호화 표준 기반에서 기하정보와 속성정보들의 상호 특성을 반영하여 압축 효율을 높이기 위한 기술 개발을 진행 중이다.



<그림 5> 6DoF 입체공간 비디오 부호화 개념도

3. 오디오 부호화 기술

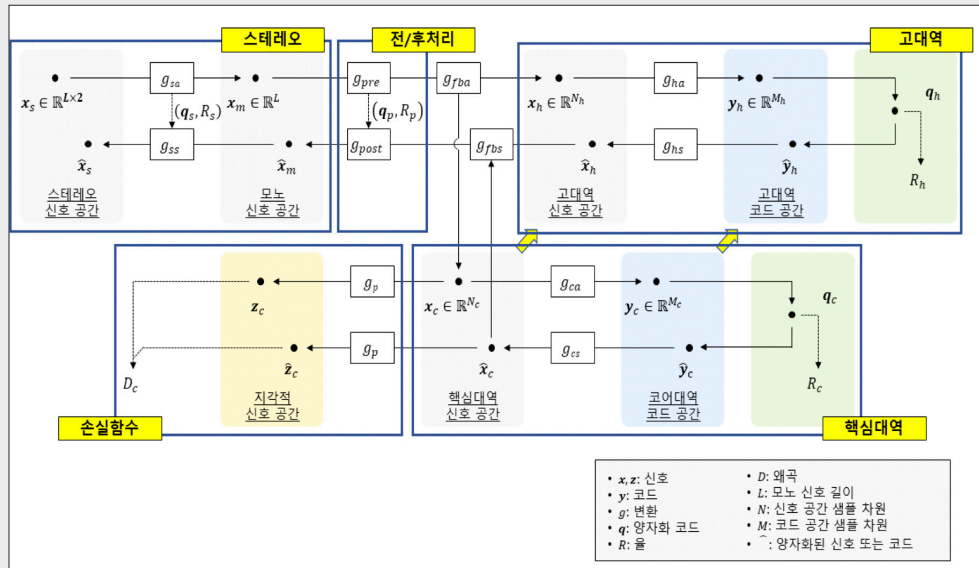
오디오 부호화 기술은 심리음향 모델을 이용하여 오디오 신호를 인간이 들을 수 있는 신호와 들을 수 없는 신호로 구분하고, 들을 수 있는 신호 위주로 비트를 할당하는 방법으로 압축을 수행한다. 대표적으로, MPEG에서는 <그림 6>과 같이 1세대 오디오 부호화 기술인 MPEG-1 오디오 부호화 표준을 시작으로 5세대 오디오 부호화 기술인 MPEG-H 3D Audio에 이르기까지 압축률 향상과 사용자 상호작용 및 렌더링 등 추가적인 기능을 제공하는 것을 목표로 다양한 오디오 부호화 기술에 대한 개발 및 표준화를 진행하였다.



<그림 6> 세대별 MPEG 오디오 부호화 기술 표준화

최근 많은 관심을 받고 있는 초실감 메타버스 서비스를 위해서는 가상공간에서 사용자에게 실시간 소통과 고품질의 몰입감 및 체험을 제공할 수 있는 오디오 부호화 기술이 필요하다. 그러나, 현재 존재하는 음성 부호화 기술과 오디오 부호화 기술은 실시간성과 품질면에서 서로 상충되는 특성을 갖는다. 즉, 기존 음성 부호화 기술은 짧은 알고리즘 지연으로 실시간 서비스를 제공하는데 적합한 반면, 고품질 서비스를 제공하기 어렵고, 오디오 부호화 기술은 상대적으로 긴 분석 프레임을 통해 신호에 내재된 여러 중복성을 효과적으로 제거할 수 있어 고품질의 서비스를 제공할 수 있지만, 긴 분석 프레임으로 인해 지연 시간이 증가하므로 실시간 서비스에는 적합하지 않다. 따라서 미래 초실감 메타버스 서비스를 위해서는 저지연 특성과 고품질 특성을 동시에 만족하는 새로운 오디오 부호화 기술 개발이 필요하다. 이러한 요구에 대응하여 본 연구실에서는 저지연과 고품질을 함께 제공하는 오디오 부호화 기술 개발을 위해 기존 신호처리 방식을 개선하는 연구뿐만 아니라 AI 기술을 적용하는 방식에 대한 연구를 수행하고 있다. <그림 7>은 AI를 활용한 오디오 부호화 프레임워크를 나타낸 것으로 일반적으로 효율적인 오디오 부호

화를 위해 스테레오 신호를 다운믹스한 모노 신호와 채널 상관도 정보로 표현하고, 모노 신호를 다시 핵심대역 신호와 고대역 신호로 분리하여 처리한다. 음질에 가장 민감한 핵심대역 신호는 오토인코더에 기반한 AI 기술을 활용하여 압축을 수행하고, 고대역 신호는 압축률 향상을 위해서 핵심대역 신호와 고대역 신호에 대한 약간의 부가 정보로부터 AI 기반 예측을 통해서 복원한다.



<그림 7> AI를 활용한 오디오 부호화 프레임워크

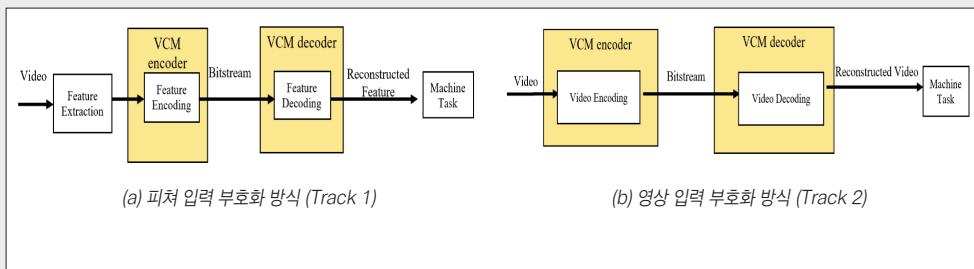
AI를 이용한 오디오 신호 부호화는 오토인코더를 기본 구조로 하여 기존 음성 부호화를 위한 CELP(Code-Excited Linear Prediction)에 기반한 신경망 CELP와 정규화 플로우(Normalizing Flow) 및 GAN(Generative Adversarial Network) 등 생성모델에 기반한 오디오 압축 방법에 대한 연구를 수행하고 있다. 신경망 CELP는 기존 CELP와 유사하게 선형예측 분석 결과로 얻어지는 신호의 주기적인 성분과 잡음 성분을 나누어서 처리하는 구조로 설계되었으며, 주기적인 성분 모델링을 위해 시계열 데이터 분석에 효과적인 순환신경망을 채용하고 잡음 성분은 일반적인 합성곱으로 이루어진 순방향 신경망을 이용하여 모델링한다. 정규화 플로우에 기반한 오디오 압축 방식에서는 Gaussian 분포를 갖는 잠재 변수로부터 입력 신호의 분포를 추정하기 위한 비선형 가역 합성곱과 선형 커플링으로 이루어진 플로우 생성모델을 이용한 신경망 보코더인 WaveGlow에 조건 벡터 대신에 잠재 신호를 직접 양자화하는 과정을 추가함으로써 오디오 신호를 부호화 및 복호화한다. 점진적 신경망 오디오 부호화 방식은 복수의 오토인코더를 잔차 연결(Residual Connection)을 통해 다단계로 연결된 구조를 가지고 있으며, 각 단계에서 목표로 하는 신호 대역을 달리하기 때문에 신호 대역폭, 비트율 및 품질에 대한 확장성을 갖는 것을 특징으로 한다.

일반적으로 AI 기반 오디오 부호화 모델에서는 사람이 느끼는 지각적 품질과 신호-레벨 복원 품질 간의 불일치 문제가 발생한다. 이러한 문제를 해결하기 위해 다양한 지각적 척도를 바탕으로 심리음향 모델 기반 지각적 손실함수, GAN 손실함수 및 특징맵 손실함수 등에 대한 연구도 함께 진행하고 있다.

오디오 신호의 고효율 부호화를 위해서는 대역폭 확장 및 스테레오 부호화가 필수적이다. 이를 위해 부호화단에서 신경망을 통해 추출한 고대역 부가 정보로부터 고대역 신호를 복원하는 AI 기반 대역폭 확장 기술과 오토인코더 구조의 잠재신호 영역에서 M/S(Mid/Side) 스테레오 부호화를 적용한 AI 기반 스테레오 기술을 개발 중에 있다. 또한, 기존 HE-AAC 오디오 코덱의 부호화단에서 신경망을 통해 추출한 부가 정보를 이용하여 복호화단에서 후처리 과정을 수행함으로써 복호화 품질을 향상하는 연구를 수행하였다.

4. 기계를 위한 비디오 부호화 기술

기계를 위한 비디오 부호화 기술(VCM: Video Coding for Machine)은 사람을 위한 비디오 부호화 기술이 인지적 화질을 유지하면서 비트율을 줄이기 위한 율-왜곡 최적화(Rate-Distortion Optimization) 기반의 부호화 기술인 반면에, 기계 임무 성능 측면에서 중요한 정보를 유지하면 비트율을 줄이기 위한 율-임무 성능 최적화(Rate-Performance Optimization)에 초점을 둔 비디오 부호화 기술이다. MPEG에서는 '19년 7월부터 VCM 표준화를 위한 준비작업을 진행하고 있으며, '22년부터 본격적인 표준화 진행이 예상된다. MPEG VCM에서는 피쳐 입력 부호화 방식(Track 1)과 영상 입력 부호화 방식(Track 2)으로 분야를 나눠 기술 탐색(Exploration)을 진행하고 있고, 본 연구실에서는 2가지 방식의 기술을 모두 연구하고 있다.



<그림 8> MPEG VCM의 2가지 부호화 방식

5. MPEG-I Immersive Audio 기술

공간음향 기술은 VR(Virtual Reality)/AR(Augmented Reality)/MR(Mixed Reality) 등 미래의 가상

현실 응용에 활용할 수 있는 실감음향 기술을 일컫는 용어라고 할 수 있다. 기존의 기술이 움직임이 거의 없는 청취자를 중심으로 Sweet Spot의 청각 경험을 수동적으로 제공하는 기술이라고 한다면, 공간음향 기술은 사용자의 움직임이 있을 때 그에 따라 변하는 음향의 방향감, 거리감, 공간감 등 현실에서 경험하는 청각 경험을 제공함으로써 마치 공간에 실재하는 듯한 체험을 가능하게 하는 기술이라고 할 수 있다.

머리의 회전을 추적(Head Tracking)하여 공간적으로 동기되도록 하는 기술을 3DoF 공간 음향이라고 하며, 머리의 회전과 함께 청취자의 움직임까지 추적(Motion Tracking)하여 가상공간 임의 위치에서의 실감 음향을 제공하는 기술을 6DoF 공간 음향이라고 한다. 이러한 추세에 따라 SC29/WG6(MPEG Audio Coding)에서는 VR/AR/MR 응용을 위하여 MPEG-I Immersive Audio의 표준화를 추진하고 있으며, '21년 4월 MPEG-I Part 4 Immersive Audio 표준화를 위한 CfP를 공표하였고, '22년 1월 WG6 회의에서 후보기술(RM0)을 선정하는 것을 시작으로 '23년 10월 FDIS(Final Draft International Standard) 문서 발간을 계획하고 있다. MPEG-I Immersive Audio의 CfP 제안기술의 평가를 위해 개발된 평가 플랫폼은 기존과는 달리 복수의 렌더러를 가상현실 환경에서 실시간 동작시키면서 성능을 직접 비교하여 평가 방법을 사용하고 있다.

ETRI에서는 객체 기반 오디오 기술을 통하여 MPEG-A IMAF(Interactive Music Application Format) 기술을 '10년도에 표준화하였고, 이후 채널 오디오와 객체 오디오가 혼합된 하이브리드 오디오 기술 개발 경험을 기반으로 공간 음향 분석 및 재현 기술 개발을 수행하고 있다. 최근에는 ETRI에서는 미래의 초실감 음향 기술로서 공간 음향 기술을 개발하고 있으며, MPEG-I Immersive Audio 표준화에도 참여하여 '21년 11월 ETRI의 공간 음향 기술을 CfP에 대응하여 제안하였다.

현재 연구가 진행되고 있는 ETRI의 공간 음향 기술은 가상체험과 같이 가상현실 콘텐츠의 재생에 의한 가상공간에서의 가상 이벤트의 체험을 목적으로 하는 공간 음향 콘텐츠 서비스 기술을 1단계로 개발하는 것이 목표이며, 2단계로는 양방향 통신에 의한 공유 공간 체험에 의한 협업 및 가상 원격회의에 활용될 수 있는 저지연 공간 음향 통신 기술을 개발하는 것을 목표로 하고 있다. 공간 음향 기술은 공간 음향 콘텐츠를 렌더링하기 위한 모든 정보를 비트스트림으로 생성하고 다시 파싱하여 복원하는 코덱부를 중심으로 공간 음향 장면 모델링 기능과 공간 음향 렌더링 기능으로 구분할 수 있다. 공간 음향 장면 모델링 기능은 가상현실 콘텐츠의 고해상도 시각적 공간 모델과는 별개로 음향효과에 유의미한 부분으로 단순화한 음향적 공간 모델을 기반으로 공간 음향 렌더링에 필요한 정보를 분석 및 추출하는 기능으로 요약할 수 있다. 공간 음향 렌더링은 단말에서 움직이는 청취자의 위치 및 머리 방향에 따른 공간 음향 콘텐츠의 다양한 음향효과를 생성하는 기능을 수행한다.

공간 음향 기술에서는 움직이는 청취자에 의해 공간 음향 콘텐츠의 임의 위치의 음향효과를 생성할 수 있어야 하는데, 이를 위하여 음원 모델이 전통적인 점 음원에서 부피 음원(Volume Source) 혹은 지향성 음원으로 확장되고, 공간을 구성하는 다양한 음향 차폐물들에 의한 반사, 산란, 투과, 회절 음향효과를 청취자의 위치에 기반하여 실시간으로 생성하여 주어야 하며, 청취자 움직임과 상대적인 음원의 움직임을 고려한 도플러 효과를 생성하여야 한다. 2단계의 저지연 공간 음향 통신 기술에서는 상기 공

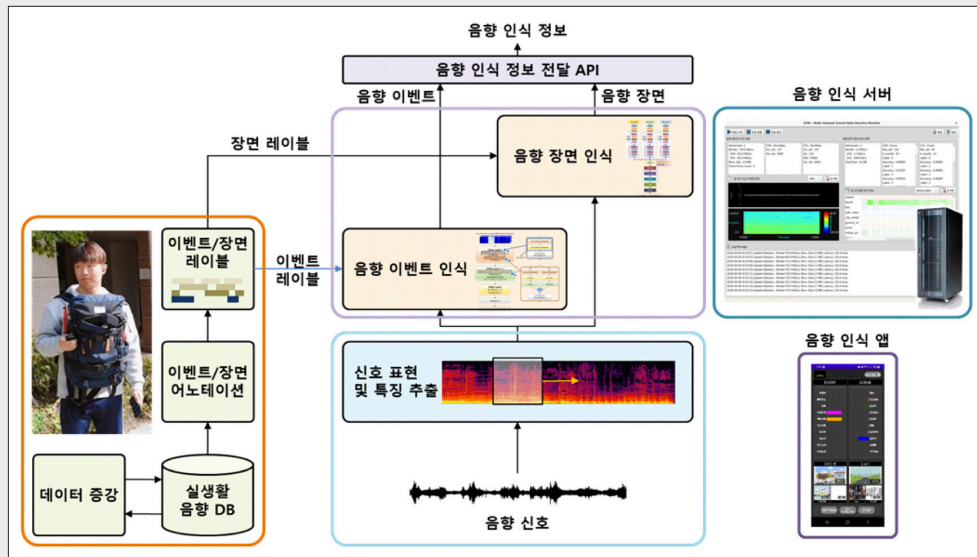
간음향 기술에 더하여 공간 음향 획득 기술이 추가되어야 하며, 실시간 통화가 가능하도록 저지연 처리의 요구사항이 추가된다.

6. 음향 사건 및 장면 인식 기술

음향 인식 기술은 인식하고자 하는 대상에 따라 음향 이벤트 및 음향 장면 인식 기술로 구분할 수 있으며, 화재 경보, 비명 소리, 차량 경적 소리 등과 같이 주변에서 발생하는 개별 음원을 음향 이벤트로, 공항, 쇼핑몰, 레스토랑 등과 같이 개별 음원들의 조합에 의해 형성된 공간적 음향 특성을 갖는 장소를 음향 장면으로 정의한다.

본 연구실은 AI 기반 음향 이벤트 및 장면 인식 기술 개발을 위해, '17년도부터 음향 데이터를 수집하기 시작하여 현재 국내 최대 규모의 실생활 음향 DB(음향 이벤트(63종) 인스턴스 개수 : 총 61,501개, 음향 장면(20종) 녹음 시간 : 총 3,205분)를 구축하였다. 이를 바탕으로 시계열 데이터 특성 파악에 강인한 Advanced GCRNN 음향 이벤트 인식 모델 및 비균등 주파수 대역 분할을 특징으로 하는 Advanced Trident-ResCNN 음향 장면 인식 모델을 개발하였다. 또한, 세계적인 음향 인식 기술 경진대회인 DCASE2020 챌린지의 복수 단말 대상 음향 장면 인식 과제에 참가하여 전체 28개 팀 순위에서 1위를, 팀 별 최대 4개까지 제출한 총 92개 시스템에 대한 평가에서 1, 2위를 차지하는 성과를 거두었다.

AI 기반 음향 이벤트 및 장면 인식 기술은 노약자 및 청각 장애인 등을 대상으로 하는 위험 상황 알림, 시설 보안 감시, 미디어 자동 태깅, 환경 소음 및 장비 상태 모니터링 등의 다양한 응용 분야에 활용될 수 있다.



<그림 9> 음향 이벤트 및 장면 인식 기술 구성도

7. 맺음말

가상과 현실의 경계를 허무는 미래 초실감 미디어 서비스는 초대용량의 비디오/오디오를 기본으로 하고 있으므로, 초대용량 미디어에 대한 고품질/고효율 압축 기술인 미디어 부호화 기술은 미래 미디어 서비스를 위한 핵심기술이다. ICT 강국 대한민국에서 미래 초실감 미디어 서비스를 위한 핵심 기술인 미디어 부호화 기술의 한계에 도전하여, 우리의 손으로 새로운 미디어 부호화 기술의 미래를 이끌어 가기 위해 ETRI 미디어부호화연구실은 오늘도 불철주야 연구개발에 매진하고 있다.



<그림 10>