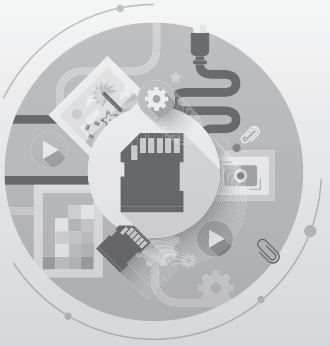
UHDTV 방송 서비스를 위한 MPEG-H 3D 오디오 기술



이미숙 ETRI 방송미디어연구소 오디오연구실 책임연구원
백승권 ETRI 방송미디어연구소 오디오연구실 선임연구원
이태진 ETRI 방송미디어연구소 오디오연구실 책임연구원/실장

1. 머리말

TTA의 '지상파 UHDTV 방송 송수신 정합' 표준[1]은 북미 차세대 방송 표준 규격인 ATSC 3.0을 기반으로 하고 있다. 다양한 오디오 신호를 압축하여 전송하고 이를 다시 복원하여 재현하는 오디오 코덱의 경우, ATSC 3.0에서는 AC-4와 MPEG-H 3D 오디오 기술을 잠정표준으로 채택하고 있다. 그러나 TTA 표준에서는 압축 성능, 오디오 품질 그리고 다양한 서비스 제공 가능성 등을 검토하여 MPEG-H 3D 오디오를 단일 오디오 코덱 표준으로 채택하였다.

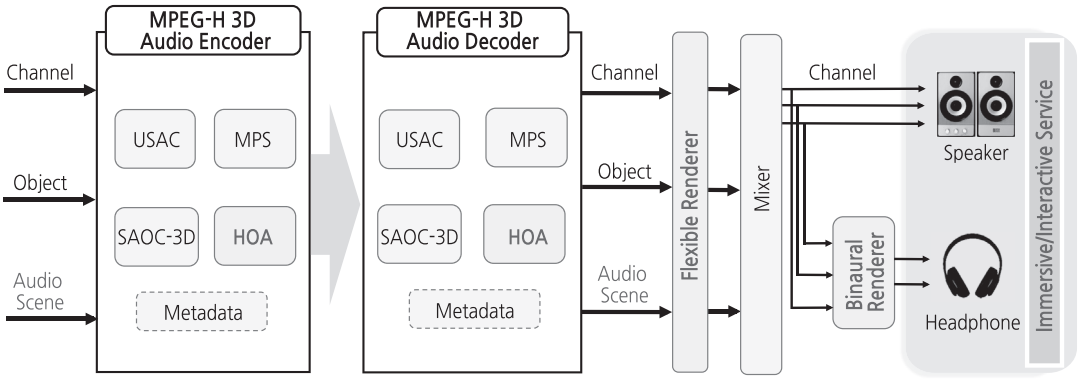
멀티미디어 국제 표준화 그룹인 MPEG에서는 MPEG-H 3D 오디오[2] 표준화를 두 단계로 나누어 진행하였다. 첫 번째 단계(Phase 1)에서는 UHD 방송 표준에 대응하기 위한 주요 기능들을 중심으로 표준화를 진행하였고, 두 번째 단계(Phase 2)에서는 태블릿 PC나 스마트폰과 같은 모바일 환경을 고려한 기술을 표준화하였다. 그리고 두 단계를 걸쳐 개발된 기술과 추가의 기술을 포함한 MPEG-H 3D 오디오 기술 표준을 2016년 하반기에 승인할 예정에 있다.

MPEG-H 3D 오디오는 몰입감/현장감 극대화와 개인 맞춤형 오디오 재현을 위해 전통적인 오디오 신호인 채널(Channel)뿐만 아니라 객체(Object)와 오디오 장면(Audio scene) 신호를 입력으로 수용하고 있다. 또한, 다양한 스피커 배치 환경과 헤드폰에서 최적화된 3D 오디오를 재현하기 위한 여러 가지 렌더링(Rendering) 기술을 표준에 포함하고 있다. 따라서 MPEG-H 3D 오디오 기술을 통해 기존의 HDTV에서는 경험하지 못했던 다양한 서비스를 시청자에게 제공할 수 있을 것으로 기대된다.

본고에서는 '지상파 UHDTV 방송 송수신 정합'에서 표준 오디오 코덱으로 선정한 MPEG-H 3D 오디오 기술에 대해 설명하고 이 코덱을 이용하여 제공 가능한 서비스에 대해 간략히 살펴보고자 한다.

2. MPEG-H 3D 오디오 기술

대화면 고해상도로 대변되는 UHDTV 방송 서비스를 고려하여 MPEG에서는 몰입감과 현장감을 극대화한 개인 맞춤형 오디오 서비스를 제공



[그림 1] MPEG-H 3D 오디오 기술 개요

하기 위해서 채널, 객체, 그리고 오디오 장면 신호까지 처리할 수 있는 MPEG-H 3D 오디오 기술에 대한 표준화를 진행하고 있다. MPEG에서는 이렇게 다양한 형태의 입력신호를 처리하기 위해 [그림 1]과 같이 기존의 MPEG 오디오 표준 기술인 USAC(Unified Speech & Audio Coding)[3], MPS(MPEG-Surround)[4] 그리고 SAOC(Spatial Audio Object Coding)[5] 기술 등을 활용하고 오디오 장면 처리를 위한 고차 앰비소닉스(HOA, High Order Ambisonics)와 다양한 재현환경에 최적화된 오디오 재생을 위한 렌더링 기술을 새로 개발하여 MPEG-H 3D 오디오 기술을 표준화하였다.

2.1 MPEG-H 3D 오디오 핵심 기술

일반적으로 오디오 콘텐츠는 채널 신호로 전송되어 정해진 위치의 스피커를 통해 재생된다. 현재 국내 HDTV 방송에서는 오디오 신호를 대부분 스테레오 채널로 전송하고 일부 음악 방송에서만 5.1 채널을 사용하고 있지만, NHK에서는 22.2 채널 오디오 시스템을 개발하여 일본 UHDTV 방송 표준에 적용하였다. 이렇게 고차 다채널에 대한 요구가 생기면서 MPEG-H 3D 오디오 코덱에서는 22.2 채널까지 지원하고 있다.

오디오 코덱에서 고차 다채널을 지원하기 위해서는 특히 압축 성능이 중요하므로 MPEG에서는 음성과 음악 신호 모두에서 고른 성능을 나타내고 압축 효율이 뛰어난 USAC를 MPEG-H 3D 오디오의 핵심 기술로 채택하였다. USAC에서는 스테레오 신호 코딩을 위해 성능이 개선된 MPS(MPEG Surround) 기술을 사용하고 있는데, MPEG-H 3D 오디오에서는 이 기술을 스테레오 쌍으로 확장하여 고차의 다채널 오디오 신호를 처리한다.

객체신호는 오디오 장면을 구성하는 각각의 음원들을 의미한다. 예를 들어, 스포츠 중계에서 관중석의 응원 소리와 아나운서의 해설을 별도의 객체신호로 볼 수 있다. 기존 채널신호에서는 관중석의 소리와 아나운서의 목소리가 혼합된 오디오를 재생단의 스피커에 1:1로 매칭되는 채널이라는 단위로 묶어서 처리한다. 그러나 객체 단위로 신호를 압축하여 전송하면 시청자의 스피커 구성 환경을 고려하여 신호를 재현할 수 있다. 객체신호는 채널신호와는 달리 객체신호가 재생되는 스피커 위치를 제어할 수 있으며, 사용자의 선택에 따라 인터렉티브(Interactive) 서비스 제공이 가능하다. 이러한 객체

신호는 채널신호로 간주하여 USAC로 압축하거나 SAOC-3D[6]로 압축하여 전송할 수 있다. 그러나 객체신호의 수가 많거나 전송 비트율이 낮은 경우에는 많은 채널과 동적 객체신호에 대해 압축효율이 좋은 SAOC-3D 기술을 적용한다.

오디오 장면 신호는 MPEG-H 3D 오디오에 새롭게 소개된 신호형태로 기존의 MPEG 오디오 표준 기술로는 이 신호를 처리할 수 없기 때문에 HOA 기술을 새로 개발하였다. 특수 장비를 통해 획득한 다채널 오디오 신호는 HOA 기술을 통해 메자닌 포맷(Mezzanine format)이라 불리는 6개의 PCM 신호와 메타데이터(Metadata)로 표현된다. 이때 전송 비트율을 낮추기 위해 PCM 신호는 다른 채널신호와 마찬가지로 USAC로 압축하여 전송한다.

또한, MPEG-H 3D 오디오 코덱은 시청자가 채널 또는 콘텐츠별로 인지하는 음량이 달라서 TV 시청 시에 오디오 볼륨을 상시 조절해야 하는 번거로움을 배제하기 위한 DRC(Dynamic Range Control) 기술을 포함하고 있다. 이렇게 오디오 코덱에서 DRC를 통해 음량을 제어하여 오디오 신호를 출력하면 시청자의 볼륨조절에 대한 번거로움을 최소화 시킬 수 있다.

2.2 MPEG-H 3D 오디오 렌더링 기술

MPEG-H 3D 오디오 기술을 통해 22.2 채널 콘텐츠를 제공하더라도 모든 시청자가 22.2 채널의 스피커를 정해진 표준 위치에 설치하는 것이 아니기 때문에 스테레오 스피커나 다른 위치에 배치된 스피커 또는 헤드폰 사용자에게도 원 제작자의 의도에 충실한 음향 장면을 제공할 수 있어야 한다. MPEG-H 3D 오디오의 주요 특징 중 하나는 바로 다양한 스피커 배치환경과 헤드폰에서도 최적화된 3D 오디오를 재현할 수 있는 렌더링 기술이라고 할 수 있다.

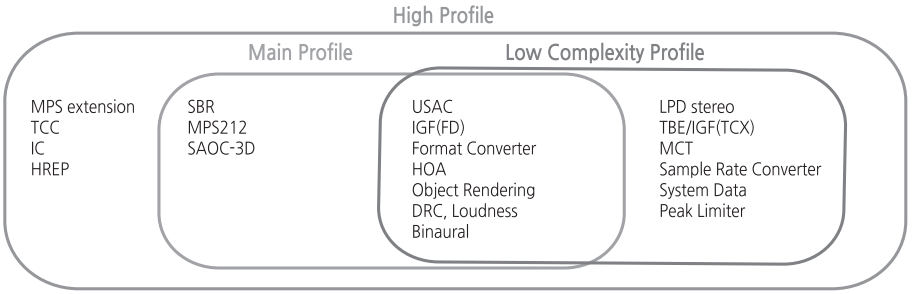
[그림 1]의 플렉서블 렌더러(Flexible Renderer)에는 채널, 객체, 그리고 오디오 장면신호 각각을 렌더링하는 모듈이 포함되어 있다. 포맷변환기(Format Converter)는 입력 채널과 출력 채널 구성 간의 변환장치로 원래 콘텐츠의 효과를 최대한 반영하기 위한 능동적 믹싱을 지원한다. 예를 들어, 22.2 채널을 단순하게 다운믹싱하여 스테레오 채널 신호를 생성하는 것이 아니라 채널 신호 간의 상관관계나 위치 정보 등을 고려하여 다운믹싱함으로써 원래 콘텐츠의 음향 장면에 가까운 스테레오 신호를 생성한다.

객체신호는 객체 렌더러를 통해 특정 스피커 재생 환경에 맞게 렌더링 되고, 오디오 장면신호는 디코딩된 PCM 신호와 HOA 메타데이터를 사용하여 HOA 렌더러를 통해서 특정 스피커 재생 환경에 맞게 생성된다.

스피커가 아닌 헤드폰을 통해 오디오 신호를 재현할 경우에는 바이노럴 렌더러(Binaural Renderer)를 통해 원 콘텐츠의 효과를 최대한 반영하는 스테레오 신호를 생성한다. 바이노럴 렌더러는 디코딩된 고차의 다채널 신호를 스테레오 신호로 변환하는 과정에서 공간상의 스피커 위치에서 발생하는 전달함수를 적용하여 헤드폰을 통해 3D 오디오를 경험할 수 있도록 하는 기술이다.

3. MPEG-H 3D 오디오 프로파일

MPEG 표준은 RM(Reference Model) 선정 이후에도 성능향상을 위해 다양한 툴들을 추가하기 때문에 활용분야에 따라 채널 수, 비트율, 사용 툴 등에 대해 프로파일로 정의한다. 각 프로파일은 서비스 목적에 따라 제공하는 기능 및 처리 가능한 채널 수를 명시하고 있다. MPEG-H 3D 오디오는 [그림 2]와 같이 세 개의 프로파일을 지원하며 각각의 특징은 다음과 같다[7].

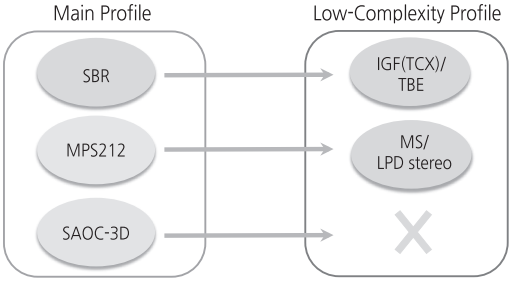


[그림 2] MPEG-H 3D 오디오 프로파일

- **메인 프로파일(Main Profile):** 메인 프로파일은 Phase 1에서 개발된 기술들에 대한 프로파일로, 복잡도에 대한 제약사항 없이 높은 몰입감과 여러 기능을 지원하기 위해 최대한 필요한 기술들을 모두 포함하고 있다. 가능한 채널 수 및 객체 수도 제약을 두지 않고 있다.
- **저-복잡도 프로파일(Low-Complexity(LC) profile):** MPEG-H 3D 오디오가 높은 몰입감을 제공하고 다양한 기능을 지원함으로써 새로운 오디오 코덱 형상을 제시한 것은 사실이나, 높은 복잡도로 인해 메인 프로파일을 실제 서비스 단말에서 지원하기는 다소 부담이 될 수 있다. 이를 해결하기 위해서 Phase 2 기술 개발 기간 동안에 복잡도가 낮은 기술, 또는 기존 기술의 복잡도를 낮춘 기술들을 표준에 반영하였다. 저-복잡도 프로파일은 이러한 기술들을 조합하여 만든 프로파일이지만 몰입감의 수준과 제공하는 기능은 메인 프로파일과 크게 다르지 않다.
- **상위 프로파일(High profile):** 상위 프로파일은 메인 프로파일과 저-복잡도 프로파일을 확대한 프로파일이다. 저-복잡도 프로파일이 메인 프로파일의 복잡도를 낮추기 위한 프로파일 이기는 하지만 메인 프로파일에는 기술되지 않은 새로운 기술들을 포함하게 되었다. 이에 따라 MPEG에서는 Phase 1과 Phase 2에서 개발한 모든 기술을 포함하는 상위 프로파일을 새롭게 정의하였다.

3.1 저-복잡도 프로파일

ATSC 3.0에서는 MPEG-H 3D 오디오 코덱 메인 프로파일과 저-복잡도 프로파일을 잠정표준으로 채택하고 있으나, TTA 표준에서는 저-복잡도 프로파일만 표준으로 선정하였다. 저-복잡도 프로파일에서는 복잡도를 고려하여 [그림 3]에 나타낸 바와 같이 메인 프로파일에서 사용하는 일부 기술 대신 다



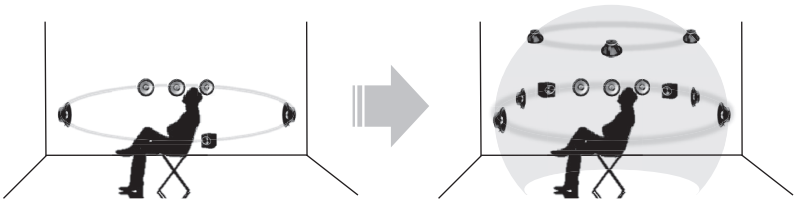
[그림 3] 저-복잡도 프로파일에서 사용하는 기술

른 기술을 사용하였다. 주요 변경내용을 보면, 메인 프로파일에서 채널신호 처리 시 사용하던 대역폭 확장기술인 SBR(Spectral Bandwidth Replication) 대신 TCX(Transform CodeeXcitation)영역의 IGF(Intelligent Gap Filling)와 ACELP(Algebraic Code-Excited Linear Prediction) 영역의 TBE(Time domain Bandwidth Extension) 기술을 사용하였다. 또한, 메인 프로파일에서 고차의 다채널 신호 처리를 위해 사용하는 MPS212 대신 저-복잡도 프로파일에서는 MS(Mid-Side)와 LPD(Linear Prediction Domain) 스테레오 기술을 사용한다. 객체신호는 SAOC-3D 기술을 사용하지 않고 각 객체를 개별적인 채널신호로 간주하여 처리한다.

〈표 1〉은 저-복잡도 프로파일에서 레벨에 따른 채널과 객체의 최대 수 그리고 HOA의 최대 차수를 나타내고 있다. TTA의 지상파 UHDTV 방송표준방식에서는 MPEG-H 3D 오디오 코덱의 저-복잡도 프로파일 레벨 1, 2, 그리고 3을 지원한다.

<표 1> 지상파 UHDTV 방송의 비디오 신호 규격

레벨	최대 표분화 주파수	최대 채널 수 (비트스트림/ 디코더)	최대 스피커 구성(예)	최대 디코더 객체 수	채널+객체 구성(예)	최대 HOA 차수	최대 HOA 차수+ 객체 구성(예)
1	48000	10/5	2/2.0	5	2채널+3객체	2	2차+3객체
2	48000	18/9	8/7.1	9	6채널+3객체	4	4차+3객체
3	48000	32[1]/16	12/11.1	16	12채널+4객체	6	6차+4객체



a) 5.1 채널 스피커 구성　　b) 10.2 채널 스피커 구성

[그림 4] 몰입형 다채널 오디오 서비스

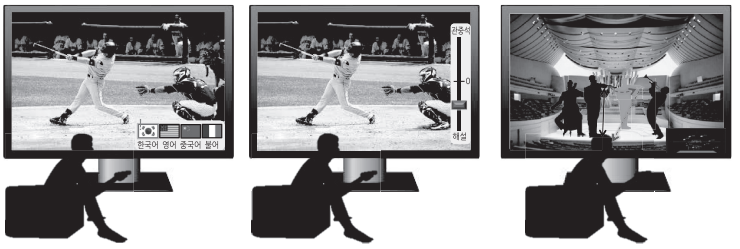
4. MPEG-H 3D 오디오 서비스

[그림 4]는 5.1 채널과 지상파 UHDTV 방송 표준 방식에서 지원하는 채널 형식 중 하나인 10.2 채널 스피커의 배치를 보여주고 있다. 기존의 HDTV 서비스에서 제공하는 스테레오 또는 5.1 채널 오디오는 수평방면에서의 공간감을 제공했다면, MPEG-H 3D 오디오에서는 수평뿐만 아니라 수직방면에서의 공간감을 제공함으로써 사용자를 중심으로 3차원의 모든 방향에서 오디오가 둘러싸는 입체적 공간감을 제공할 수 있다. 기존 HDTV 시청환경에 비해 UHDTV는 화면이 크고 시청거리가 상대적으로 가깝기 때문에 화면에 표출되는 영상의 위치와 일치하는 오디오를 재생하기 위해서는 좌우뿐만 아니라 높낮이에서도 차별된 소리를 제공할 수 있어야 하므로 수직방면의 공간감이 매우 중요하다. 따라서 이러한 고차 다채널 신호와 함께 동적 객체와 오디오 장면 신호 처리 기술을 이용하면 몰입감과 현장감이 극대화된 오디오 서비스를 제공할 수 있을 것이다.

또한, MPEG-H 3D 오디오의 객체신호 처리 기술을 사용하면 시청자의 취향이나 선택에 따라 특정 오디오 신호를 제어하는 서비스를 제공할 수 있다. 예를 들어, 특정 객체신호의 음량 또는 재생되는 소리의 위치의 변경이 가능하다. [그림 5]는 객체신호를 이용한 서비스의 예를 보여주고 있다.

MPEG-H 3D 오디오 코덱 기술을 사용하면 기존에 제공하던 다중언어와 화면해설 서비스를 효율적으로 제공할 수 있을 뿐만 아니라 새로운 서비스 제공이 가능하다. [그림 5] b)처럼 스포츠 중계 시 관중석의 생생한 현장 소리와 아나운서의 해설을 객체 신호로 처리함으로써 사용자의 취향에 따라 두

1) TTA의 '지상파 UHDTV 방송 송수신 정합' 표준에서는 최대 채널 수를 16채널로 제한하고 있다.



a) 언어 선택 b) 음량 제어 c) 악기신호 위치 제어

[그림 5] 객체 오디오 기반 인터렉티브 서비스 예

소리의 크기를 달리하여 재생할 수 있다. 즉, 현장의 생생함을 느끼고 싶을 경우에는 관중석 소리를 키우고, 명료한 해설을 듣고 싶을 때는 관중석 소리를 줄이고 아나운서의 목소리를 크게 조정할 수 있다.

또한, [그림 5] c)와 같이 특정 객체신호의 재생 위치를 제어할 수도 있다. 예를 들어, 음악을 들을 때 악기에서 발생하는 소리의 위치를 조절하여 시청자가 연주자의 앞이 아니라 옆에서 듣는 것과 같은 느낌을 갖도록 할 수도 있다.

5. 맺음말

본고에서는 TTA의 '지상파 UHDTV 송수신 정합 표준'에서 오디오 코덱 표준으로 선정한 MPEG-H 3D 오디오 기술 및 서비스에 대해 간략히 살펴보았다.

MPEG-H 3D 오디오 기술은 다양한 형식의 입력신호를 코딩하고, 제작된 콘텐츠의 채널 구성과 다른 재현 환경에서도 최적화된 오디오 신호를 재생할 수 있는 기술이다. 따라서 대화면의 고해상도 UHDTV 비디오에 부합하는 고도의 몰입감과 더불어 새로운 인터렉티브 서비스를 시청자에게 제공할 수 있을 것으로 기대된다.

물론 아직까지는 대부분의 방송 프로그램이 스테레오로 제작되고 있기 때문에 5.1 채널 이상의 다채널 서비스와 객체기반의 인터렉티브 서비스가 보편화될 때 까지는 시간이 필요할 것이다. 그러나 청취자가 특정 채널을 선택한다든지 오디오 신호의 볼륨을 제어하는 비교적 간단한 인터렉티브 서비스를 먼저 시작함으로써 UHDTV 방송을 통해 시청자는 HDTV와는 차별화된 서비스를 경험할 수 있을 것이다. 그리고 가정 내에서 다채널 오디오를 편리하게 재현할 수 있는 사운드바와 같은 장치가 보급되면 MPEG-H 3D 오디오의 다양한 기술을 통해 시청자에게 몰입감과 현장감이 극대화된 오디오 서비스를 제공할 수 있을 것으로 기대된다.

※본 원고는 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임. [No.B0101-16-0295,초고품질 콘텐츠 지원 UHD 실감방송/디지털시네마/사이니지 융합서비스 기술 개발]

[참고문헌]

[1] TTAK.KO-07.0127, '지상파 UHDTV 방송 송수신 정합', 2016. 6.

[2] ISO/IEC 23008-3:2015, 'Information Technology–High efficiency coding and media delivery in heterogeneous environments-Part 3: 3D Audio, Amendment 3'

[3] ISO/IEC 23003-3, 'Information technology-MPEG audio technologies-Part 3: Unified speech and audio coding'.

[4] ISO/IEC 23003-1:2007, 'Information technology-MPEG audio technologies- Part 1: MPEG Surround'.

[5] ISO/IEC 23003-2, 'Information technology-MPEG audio technologies-Part 2: Spatial Audio Object Coding (SAOC)'

[6] A. Murtaza, J. Herre, J. Paulus, L. Terentiv, H. Fuchs, S. Disch:

'ISO/MPEG-H 3D Audio: SAOC 3D Decoding and Rendering', 139th AES Convention, New York, USA, 2015.

[7] 이태진외 'UHDTV를 위한 차세대 오디오 표준: MPEG-H 3D Audio,' 방송과 미디 어, 제 21권 3호, 2016. 7.

정보통신 **용어 해설** • 웹: http://terms.tta.or.kr • 모바일: http://terms.tta.or.kr/mobile/main.do • 홈페이지: http://www.tta.or.kr

프레임 간 예측 부호화 interframe predictive coding

영상 부호화 방법의 하나로, 시간 축 방향으로 나열된 화면 간에 매우 높은 상관성을 이용하여 현재 화면의 신호와 이전 화면에서 예측된 신호의 차이만을 부호화하는 방식.

일반적으로 동영상은 시간축 방향으로 인접 화면 간 진폭이 급격하게 변화하지 않기 때문에 매우 높은 상관성을 갖게 된다. 따라서, 프레임 간 예측 부호화를 사용하면 시간적 중복성을 제거할 수 있게 되어, 매우 높은 부호화 효율을 얻을 수 있다. 프레임 간 예측 부호화는 크게 움직임 예측(motion prediction), 움직임 보상(motion compensation) 및 예측 오차, 예측 오차의 양자화 및 부호화 단계를 거친다. 프레임 간 예측 부호화는 동영상 압축 방식의 핵심 기술로 엠펙(MPEG) 등 거의 대부분 표준에서 사용되고 있다.