논문 17-42-02-20 The Journal of Korean Institute of Communications and Information Sciences '17-02 Vol.42 No.02
https://doi.org/10.7840/kics.2017.42.2.432

MPEG-H 3D 오디오 표준 복호화기 구조 및 연산량 분석

문 현 기◆, 박 영 철*, 이 용 주**, 황 영 수°

MPEG-H 3D Audio Decoder Structure and Complexity Analysis

Hyeongi Moon◆, Young-cheol Park*, Yong Ju Lee**, Young-soo Whang°

요 약

MPEG-H 3D 오디오 표준은 UHDTV 등의 초고해상도 방송서비스에 대응하는 실감음향 서비스의 제공을 목표로 한다. 이를 위해 본 표준은 다채널 신호, 객체 신호, 장면 기반 신호의 부호화/복호화 기술과 다양한 재생 환경에서 3차원 오디오 제공을 위한 렌더링 기술, 후처리 기술 등 방대한 기술을 통합하였다. 본 표준의 참조 소프트웨어 복호화기는 여러 모듈들이 결합된 구조로 다양한 모드에서 동작이 가능하며, 각 모듈들이 독립된 실행파일로 순차적으로 실행되어 실시간 처리가 불가능하다. 본 논문에서는 MPEG-H 3D 오디오의 코어 복호화기, 포맷 변환기, 객체 렌더러, 바이노럴 렌더러의 각 함수를 동적 라이브러리화 및 통합하여 프레임 기반 복호화가 가능하도록 하였다. 또한 MPEG-H 3D 오디오의 각 모드별 연산량을 측정하여 다양한 하드웨어 플랫폼에서 적합한 모드를 선택하기 위한 참고 자료를 제공한다. 연산량 분석 결과, 한국 방송 표준에 포함된 저연산량 프로파일은 채널 신호로 렌더링을 할 경우 QMF 합성 연산의 2.8배에서 12.4배의 연산량을 가지며, 바이노럴 렌더링을 할 경우 QMF 합성 연산의 4.1배에서 15.3배의 연산량을 가진다.

Key Words : MPEG-H 3D Audio, Core decoder, Format converter, Object renderer, Binaural renderer, Complexity analysis

ABSTRACT

The primary goal of the MPEG-H 3D Audio standard is to provide immersive audio environments for high-resolution broadcasting services such as UHDTV. This standard incorporates a wide range of technologies such as encoding/decoding technology for multi-channel/object/scene-based signal, rendering technology for providing 3D audio in various playback environments, and post-processing technology. The reference software decoder of this standard is a structure combining several modules and can operate in various modes. Each module is composed of independent executable files and executed sequentially, real time decoding is impossible. In this paper, we make DLL library of the core decoder, format converter, object renderer, and binaural renderer of the standard and integrate them to enable frame-based decoding. In addition, by measuring the computation complexity of each mode of the MPEG-H 3D-Audio decoder, this paper also provides a reference for selecting the appropriate decoding mode for various hardware platforms. As a result of the computational complexity

※ 본 연구는 미래창조과학부의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구(No.B0101-16-0295,초고품질 콘텐츠 지원 UHD 실감방송/디지털시네마/사이니지 융합서비스 기술 개발)입니다.
◆ First Author : School of Electrical and Electronic Engineering, Yonsei University, orexis@dsp.yonsei.ac.kr, 학생회원
° Corresponding Author : Department of Electronic and Communication Engineering, Kwandong University, hysoo@cku.ac.kr, 정회원
* Computer and Telecommunications Engineering Division, Yonsei University, young00@yonsei.ac.kr, 종신회원
** Audio & Acoustics Research Section, Broadcasting·Media Research Laboratory, ETRI, draball@etri.re.kr, 정회원
논문번호 : KICS2016-07-166, Received July 25, 2016; Revised December 12, 2016; Accepted December 28, 2016

measurement, the low complexity profiles included in Korean broadcasting standard has a computation complexity of 2.8 times to 12.4 times that of the QMF synthesis operation in case of rendering as a channel signals, and it has a computation complexity of 4.1 times to 15.3 times of the QMF synthesis operation in case of rendering as a binaural signals.

Ⅰ. 서 론

최근 진행되고 있는 MPEG-H 3DA(3D-오디오)[1] 표준화의 목적은 고화질 영상신호에 대응하는 실감음향을 제공하기 위함이다. 몰입감과 현장감을 높이기 위하여 영상 기술은 해상도 측면에서 FullHD를 넘어서 4k, 8k UHD (Ultra high definition)가 사용되고 있으며, 시점 측면에서도 2D에서 3D, 다시점 영상[2] 등으로 발전하고 있다. UHDTV의 경우 일반적인 환경에서 시야각이 100도에 달하며, 이러한 고해상도 영상에 적합한 3D 오디오는 높은 음질과 함께 화면의 객체와 정확히 일치하는 3차원 음상정위를 지녀야 한다.

MPEG-H 3DA는 1.2Mbps~256kbps의 비트율에서 22.2채널 신호의 효율적인 부호화 및 재생을 목표로 2013년 1월 CfP (Call for Proposal)가 발표되었다. 2013년 5월 채널과 객체신호, HOA (High Order Ambisonics) 분야의 부호화 및 재생 기술이 제안되었다. 2013년 7월 참조 모델 0(Reference Model 0)이 선정되었으며 표준화 과정을 통하여 2014년 7월 DIS (Draft International Standard)가 승인되었다. 2016년 7월 13일 참조 소프트웨어 (Reference software) 버전 7이 배포되었다. 48kbps ~128kbps의 낮은 비트율에서 다채널 오디오 신호를 서비스하기 위한 MPEG-H 3DA Phase 2는 2016년 10월 AMD (Amendment) 발행을 앞두고 있다.

MPEG-H 3DA 기술은 아래 3가지의 입력 포맷을 지원한다. 높은 몰입감과 정확한 3차원 음상정위를 제공하기 위하여 높이 채널이 포함된 고채 다채널 신호를 지원한다. 객체 기반 오디오 신호(Object-based Audio)를 통해 사용자들이 오디오 장면을 자유롭게 조절할 수 있는 기능을 제공한다. HOA 기반 오디오를 통해 장면 기반 오디오 신호를 전송할 수 있도록 한다.

USAC 3D 코덱은 MPEG-H 3DA의 코어 코덱으로 최신 기술인 USAC (Unified Speech and Audio Coding[3])에 다채널 신호와 객체 신호, HOA 신호의 효율적인 부호화 및 복호화를 위하여 QCE (Quad Channel Element), IGF (Intelligent Gap Filling), MCT (Discrete Multi-Channel coding Tool), Hybrid residual MPS (MPEG Surround) 툴 등이 추가되었다. USAC 3D 코덱은 22.2채널 신호를 512kbps의 비트율로 렌더링 할 경우 MUSHRA (MUlti Stimulus test with Hidden Reference and Anchor[4]) 테스트에서 “Excellent” 점수를 받았다[5].

일반적인 가정에서는 다채널 스피커가 거의 사용되지 않으며, 최적의 위치에 스피커를 배치하기 힘든 경우가 많다. 또한 모바일 디바이스 사용의 확대로 바이노럴(Binaural) 재생환경에서의 콘텐츠 소비가 급격히 늘어나고 있다. MPEG-H 3DA 기술은 콘텐츠 제작자가 의도한 3D 오디오를 사용자에게 효과적으로 제공하기 위하여 다양한 단말의 재생환경에 대응하는 자유(Flexible) 렌더링/다운믹스 기능을 포함한다[6].

포맷 변환기(Format converter)는 입력 채널 신호와 재생단의 스피커 레이아웃이 일치하지 않더라도 효과적인 3차원 오디오를 제공하기 위하여 능동 다운믹스 기능을 제공한다. 이는 다운믹스 과정에서 발생 가능한 음질의 열화를 줄이는 기술을 포함한다. 객체 렌더러(Object renderer)는 단일 객체 신호를 채널 신호로 렌더링 하는 엔진이다. 다수의 스피커 채널의 이득 값을 조정함으로써 음상의 위치를 3차원 공간상에 정위시키는 3D VBAP[7] (Vector Base Amplitude Panning) 기술을 사용한다. 바이노럴 렌더러는 양이 방 충격 응답(Binaural Room Impulse Reponse)을 사용하여 채널 신호를 양이 신호로 변환하는 엔진으로 이어폰/헤드폰을 사용하여 3차원 오디오를 제공할 수 있는 기술이다.

MPEG-H 3DA 복호화기 참조 소프트웨어는 코어 복호화기와 HOA 복호화기, SAOC[8] (Spatial Audio Object Coding) 복호화기, 각종 렌더러 뿐만 아니라 MPEG-D 표준 DRC[9] (Dynamic Range Control), 리미터(Limiter) 등의 후처리기와 시간 및 주파수 영역 변환기 등 많은 모듈들이 결합되어 있다. 현재 참조 소프트웨어는 모듈간의 입력 및 출력에 오디오 및 메타데이터 파일을 사용하며, 각 모듈이 순차적으로 작동하여 실시간 처리가 불가능하다. 본 표준의 활용성을 높이기 위하여 각 모듈을 라이브러리화 하고 이를 통합하여, 프레임 기반의 복호화가 가능하도록 하는 것이 중요하다.

MPEG-H 3DA 복호화기는 3가지 입력 포맷에 대한 복호화와 렌더링, 후처리를 수행하기 위하여 높은 연산량을 가지며 다양한 모드에서 동작이 가능하다. PC, 모바일, TV 등의 다양한 플랫폼에서 본 표준을 사용하기 위해서는 제작하는 하드웨어 플랫폼에 적합한 모드가 무엇인지 판단해야 한다. 이를 위하여 각 모듈의 동작 모드별 연산량 측정이 필수적이다.

따라서 본 논문에서는 MPEG-H 3DA 참조 소프트웨어 버전 7의 코어 복호화기와 포맷 변환기, 객체 렌더러, 바이노럴 렌더러의 기능 및 구조를 분석한다. 그리고 각 모듈을 라이브러리화고 이를 통합하여 MPEG-H 3DA 실시간 복호화기를 구현한다. 구현한 복호화기를 사용하여 각 모듈의 연산량을 측정하고 다양한 복호화 모드에서의 연산량을 측정하였다.

본 논문의 구성은 다음과 같다. II장에서는 MPEG-H 3DA 에 사용된 기술들을 요약하고 III장에서는 참조 소프트웨어 버전 7 복호화기의 구조를 분석한다, IV장에서는 MPEG-H 3DA 실시간 복호화기를 구현하였으며, V장에서는 구현한 MPEG-H 3DA 실시간 복호화기를 사용하여 각 모듈과 복호화 모드별 연산량 및 연산 시간을 측정하였다. 마지막으로 VI장에서 결론을 맺는다.

Ⅱ. MPEG-H 3DA 기술 특징

2.1 코어 복호화기

코어 복호화기는 높이 채널을 포함한 다채널 부호화의 압축효율을 높이기 위하여 MPEG USAC 복호화기에 IGF, QCE, MCT, Hybrid residual MPS 툴이 추가되었다. IGF 툴은 MDCT (Modified Discrete Cosine Transform) 영역의 툴로, noise filling 툴의 성능을 개선한 것으로 인접한 시간-스펙트럼 타일의 정보를 이용하여 비어있는 시간-스펙트럼의 정보를 추정하는 툴이다. IGF 툴은 QMF 영역의 SBR (Spectral Band Replication) 툴을 대체하여 사용될 수 있다.

QCE 툴은 USAC 코덱의 복소수 예측 스테레오 (Complex prediction stereo) 툴과 MPS 212 툴을 결합하여 상하 좌우 4개의 채널을 스테레오 혹은 모노 채널로 부호화 및 복호화를 수행한다.

MCT 툴은 조인트 부호화툴로 수평과 수직방향으로 배치된 다채널 신호를 스테레오 신호 쌍을 사용하여 시간-주파수 영역에서 복호화 하는 툴이다. MCT 툴은 IGF 를 이용하여 스테레오 신호의 시간-주파수 영역에서 0으로 양자화된 신호를 복호화 하는 Stereo Filling (SF) 툴을 포함한다.

Hybrid residual 툴은 MPS212 툴을 개선한 툴으로, 잔여 신호(Residual)와 역상관기(Decorrelator)를 동시에 사용하여 스테레오 신호의 상관도를 조절한다. 각 툴에 대한 내용은 참고문헌[3,10-11]에 설명되어 있다.

2.2 포맷 변환기

포맷 변환기는 다운믹스 설정기(Downmix matrix decoder)와 포맷 변환기(Format converter)로 나뉜다. 다운믹스 설정기는 입/출력 스피커의 위치정보를 사용하여 초기 다운믹스 매트릭스를 생성한다. 전송단과 출력단의 채널 레이아웃이 모두 표준 레이아웃에 해당할 경우 "Converter rules matrix[1]"를 사용하여 다운믹스 매트릭스 매트릭스를 생성한다. 전송단과 출력단의 채널 레이아웃중 하나 이상이 표준 레이아웃에 해당하지 않을 경우 3D VBAP 기반의 다운믹스 매트릭스 생성 방법을 사용한다.

포맷 변환기는 주어진 다운믹스 매트릭스를 사용하여 채널 신호를 다운믹스한다. 다운믹스 과정은 수동 모드와 능동 모드로 나뉜다. 수동 모드는 초기 다운믹스 매트릭스를 그대로 적용하는 방법으로 간단하지만, 다운믹싱 과정에서 콤 필터링(Comb- filtering), 컬러레이션(Coloration), 위상 왜곡 등의 문제점이 발생할 수 있다[12]. 능동 모드는 이를 해결하기 위하여 매 프레임마다 채널간의 공분산을 분석하고 각 채널의 위상을 정렬한 다음 다운믹스 한다. 능동 모드는 기본적으로 하이브리드 QMF 영역에서 작동한다.

높이 채널이 없는 5.0, 5.1 채널 레이아웃에 높이 채널 신호를 렌더링하는 방법으로 "Immersive Loudspeaker Rendering[1]" 모드가 있으며 이는 높이 채널의 다운 믹싱에 머리 전달 함수 기반의 EQ를 사용하는 방법이다.

2.3 객체 렌더러

객체 렌더러는 객체 메타데이터 해석부(OAM Decoder)와 객체 렌더러부(Object renderer)로 구성되어 있다. 메타데이터 해석부는 MP4파일에 포함된 부호화된 객체 메타데이터를 복호화 하며, 객체 렌더러부는 객체 메타데이터, 출력단 채널 레이아웃을 사용하여 객체 신호를 채널 신호로 렌더링한다. 객체 메타데이터(OAM: Object Audio Metadata)는 객체의 위치정보, 객체의 이득 값 등을 포함한다.

메타데이터의 부호화에는 일정한 간격의 프레임마다 메타데이터를 모두 저장하는 인트라코딩(Intracoding)을 사용하며, 필요에 따라서 한 프레임 내에서 차분코딩(Differential decoding)을 사용하여 시간 해상도를

높일 수 있다.

객체 렌더러는 3D VBAP 기술을 사용하여 각 스피커에 대한 객체 오디오의 이득 값을 계산하며, 이득 값은 시간 영역일 경우 샘플 단위, QMF 영역일 경우 타임 슬롯 단위로 선형 보간되어 적용된다. 재생단의 스피커의 간격이 충분히 촘촘하지 않을 경우 가상의 스피커를 추가하는 가상 스피커(Imaginary speakers) 방법[1]을 사용할 수 있다. 또한 MDAP (Multiple-Direction Amplitude Panning[13]) 기술로 음상의 퍼짐(Spread)을 구현할 수 있다.

"Screen-Related Element Remapping[1]" 기능은 화면의 객체 위치에 음상을 렌더링하는 기능으로, 화면의 크기정보가 주어질 경우 사용할 수 있다. MPEG에서는 영상의 효율적은 전송을 위하여 고해상도 동영상에서 사용자가 관심 있는 영역만을 선별적으로 전송할 수 있다[14]. 객체 렌더러는 이에 대응하여 확대된 영상으로 변화된 객체의 위치를 보정하여 렌더링 할 수 있다.

2.4 바이노럴 렌더러

채널 신호에 해당 채널의 양이 방 충격 응답을 사용하여 양이 신호로 렌더링하는 바이노럴 렌더러는 시간 영역과 주파수 영역의 엔진이 있다. 채널 신호에 양이 방 충격 응답을 컨볼루션하는 방법은 연산량 측면에서 비효율적이기 때문에 MPEG-H 3DA 바이노럴 렌더러는 초기 반사음은 고속 컨볼루션을 사용하고 후기 잔향은 각 도메인에 적합한 인공 잔향기(Reverberator)를 사용한다. 바이노럴 렌더러는 양이 방 충격 응답의 초기 반사음과 후기 잔향을 나누고 이를 각 렌더러에 적합하게 전 처리하는 바이노럴 파라미터화부와 파라미터화 된 양이 방 충격 응답을 사용하여 채널 신호를 양이 신호로 렌더링하는 렌더러부로 나뉜다.

Ⅲ. MPEG-H 3DA 참조 소프트웨어 버전 7 복호화기 구조 및 실시간 구현

본 논문에서는 MPEG-H 3DA 참조 소프트웨어 버전 7을 기반으로 실시간 복호화기를 구현하였다. 실시간 복호화기에 포함된 모듈은 코어 복호화기, 포맷 변환기, 객체 렌더러, 바이노럴 렌더러이다. 먼저 참조 소프트웨어 복호화기의 전체 구조를 분석하고, 각 모듈에 해당하는 프로젝트의 상세 구조를 분석한다. 각 모듈에서 MPEG-H 3DA 복호화 및 렌더링 과정에서 발생하는 중복되거나 불필요한 연산을 제거하고, 프레임 기반 복호화 및 렌더링 연산을 수행하는 함수를 동적 라이브러리화 및 통합하여 MPEG-H 3DA 실시간 복호화기를 구성하였다.

3.1 MPEG-H 3DA 복호화기 구조

MPEG-H 3DA 참조 소프트웨어 버전 7 복호화기의 구조는 [그림 1]과 같다. 코어 복호화기와 각 렌더러 그리고 후처리기 실행 프로그램을 순차적으로 호출하여, MP4 파일을 복호화 하고 각 파일의 포맷에

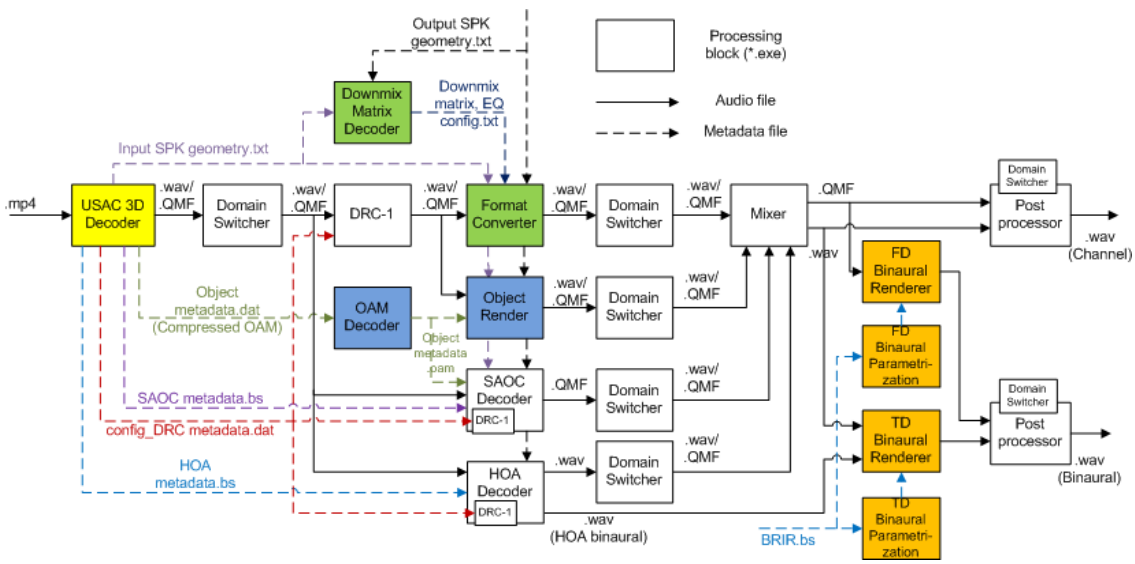

그림 1. MPEG-H 3D 오디오 코어 복호화기, 렌더러 및 후처리 모듈
Fig. 1. MPEG-H 3D Audio core decoder, renderer and post processing module

적합한 렌더링을 수행한다. 각 블록들은 독립된 실행 프로그램이며 점선은 메타데이터 파일, 실선은 시간 영역 혹은 QMF (Quadrature Mirror Filterbank) 영역의 오디오 파일이다.

코어 복호화기(USAC 3D Decoder)는 MP4 파일/비트스트림을 오디오 신호와 메타데이터 신호로 복호화한다. 오디오 신호는 채널 신호, 객체 신호, SAOC 신호, HOA 신호가 있다. 메타데이터 신호는 객체, DRC, SAOC, HOA 메타데이터와 입/출력단의 채널 레이아웃 정보 등이 있다.

영역 변환기(Domain switcher)는 시간 영역과 QMF 및 하이브리드(Hybrid) QMF[15] 영역 간의 변환을 수행하며 각 디코더 및 렌더러에 적합한 영역으로 신호를 변환한다.

복호화된 채널, 객체, SAOC, HOA 신호는 각각 포맷 변환기, 객체 렌더러, SAOC 3D 복호화기(SAOC 3D decoder), HOA 복호화기/렌더러(HOA decoder/renderer)를 통하여 출력단의 채널 신호로 렌더링된다. 메타데이터 해석부(OAM Decoder)는 압축된 객체 메타데이터를 복호화한다. DRC-1은 복호화된 각 채널과 객체신호의 다이나믹 레인지를 조절한다. 믹서는 각 렌더러의 출력을 더해준다. 바이노럴 환경에서는 각 영역에 적합한 바이노럴 렌더러가 사용된다. 후처리기는 DRC와 리미터 툴을 포함한다.

3.2 코어 복호화기 구조

코어 복호화기는 [그림 1]의 노란색 블록으로 MP4 파일에서 채널/객체/HOA 복호화기/렌더러 및 DRC와 관련된 메타데이터를 추출하는 "ASCparser Lib" 와 USAC 3D 복호화를 수행하는 "usacDec" 프로젝트로 나뉜다. [그림 2]는 "ASCparserLib" 와 "usacDec" 프로젝트의 종속성 구조이다. "ASCparser Lib" 프로젝트는 영상과 오디오가 포함된 MPEG4 파일의 구조체 정보와 핸들을 제공하는 "Libiso media" 프로젝트와 스피커 위치와 스피커 레이아웃 표준 인덱스간의 변환을 수행하는 "Cicp2geometry" 프로젝트를 포함한다.

"3DAudioCoreDeclib" 프로젝트는 USAC 3D 복호화를 위한 프로젝트들과 에러 방지 툴인 "libeptool", MPS를 수행하는 "spatiallib", QMF 합성 및 분석을 위한 "qmflib" 프로젝트를 포함한다.

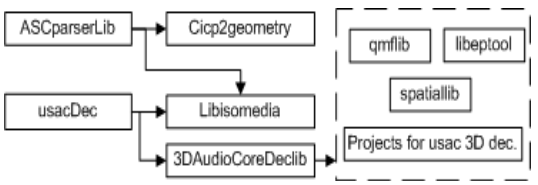

그림 2. ASCparserLib와 usacDec 프로젝트 종속성 구조
Fig. 2. Dependencies of the ASCparserLib and usacDec project

"objs_streamfile" 프로젝트는 오디오 MP4 파일을 프레임 단위로 스트리밍할 수 있는 핸들을 제공한다.

"3DAudioCoreDeclib" 프로젝트는 USAC 3D 복호화를 수행하는 USACDecodeFrame() 함수를 포함한다. [그림 3]은 USACDecodeFrame() 함수의 블록도이다. parseChannelElement() 함수는 산술 복호화 및 역양자화, NF (noise filling)을 수행한다. processingChannelElement() 함수는 IGF, TNS (Temporal Noise Shaping), M/S (Mid/Side), TW_MDCT (Time Warped MDCT) 툴을 포함한다.

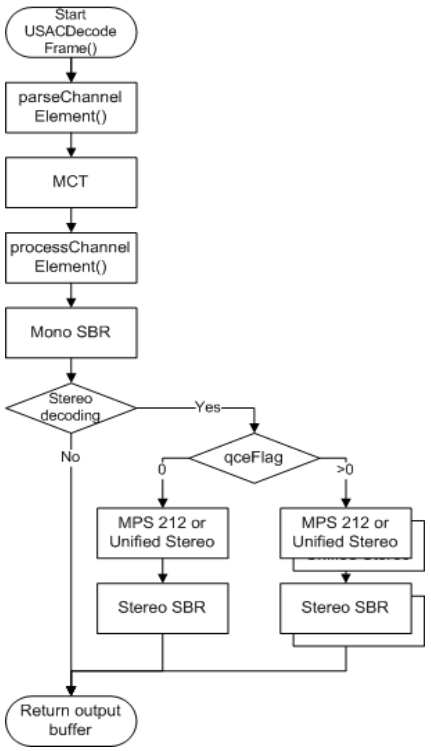

그림 3. USACDecodeFrame() 처리 블록도
Fig. 3. USACDecodeFrame() processing block diagram

3.3 포맷 변환기 구조

포맷 변환기는 "formatConverterCmdl" 프로젝트로 [그림 1]의 녹색 블록들에 해당하며, 그 종속성 구조는 [그림 4]와 같다.

"formatConverterLib" 프로젝트는 다운믹스 설정기와 시간 및 QMF 영역의 포맷 변환기 함수를 포함한다. 스피커가 표준 레이아웃이 아닐 경우 3D VBAP

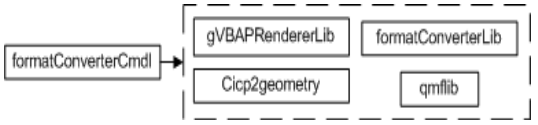

그림 4. formatConverterCmdl 프로젝트 종속성 구조
Fig. 4. Dependencies of the formatConverterCmel project

를 수행하는 “gVBAPRendererLib” 프로젝트를 호출하여 초기 다운믹스 매트릭스를 생성한다. “qmflib” 프로젝트는 능동 모드 포맷 변환기를 위한 하이브리드 QMF 분석 및 합성을 수행한다.

3.4 객체 렌더러 구조

객체 렌더러인 “gVBAPrendererCmel” 프로젝트와 부호화된 메타데이터를 복호화하는 “oamDecoderlib” 프로젝트는 [그림 1]의 파란색 블록들에 해당한다. “gVBAPRendererCmdl” 프로젝트의 종속성 구조는 [그림 5]와 같다. “gVBAPRendererLib” 프로젝트는 시간 및 주파수 영역에서 객체 렌더링을 수행하는 함수를 포함하며, “oamEncoderlib” 프로젝트는 복호화된 메타데이터 파일에서 메타데이터를 추출한다.

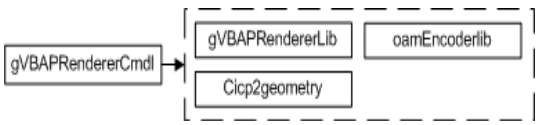

그림 5. gVBAPRendererCmdl 프로젝트 종속성 구조
Fig. 5. Dependencies of the gVBAPRendererCmdl project

3.5 바이노럴 렌더러 구조

바이노럴 렌더러는 [그림 1]의 오렌지색 블록들이다. 시간 영역과 주파수 영역 두 가지 렌더러가 있으며, 각각 양이 방 충격 응답 파라메터화부 프로젝트와 바이노럴 렌더러부 프로젝트로 나뉜다. 종속성 구조는 [그림 6]과 같다.

“binauralInterfaceLib” 프로젝트는 파라미터화된 양이 방 충격 응답을 읽어들이며 “BinauralRenderer Lib” 프로젝트는 바이노럴 렌더링을 수행한다. 바이노럴 렌더러는 초기 반사음의 연산에 고속 컨볼루션을 사용하기 때문에 고속 푸리에 변환을 수행하는 “fftlib” 프로젝트를 포함한다.

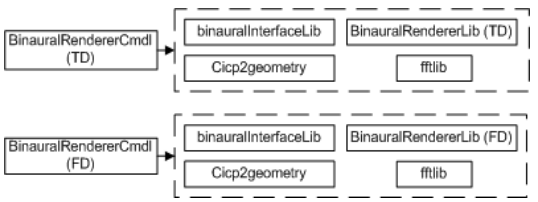

그림 6. BinauralRendererCndl 프로젝트 종속성 구조
Fig. 6. Dependencies of the BinauralRendererCndl project

3.6 MPEG-H 3DA 참조 소프트웨어 버전 7 실시간 복호화기 구현

코어 복호화기의 “ASCparserLib” 프로젝트와 “3DAudioCoreDeclib” 프로젝트, 포맷 변환기, 객체 렌더러, 시간 영역 바이노럴 렌더러, 주파수 영역 바이노럴 렌더러의 프레임 기반 복호화 및 렌더링을 수행하는 함수들을 동적 라이브러리로 구성하여 다양한 플랫폼에서 사용할 수 있도록 하였다. “fftlib”, “qmflib”, “Cicp2geometry”, “Libisomedia”, “binauralInterfaceLib” 와 같이 공통으로 사용되는 모듈은 각각 독립된 동적 라이브러리로 만들었다.

라이브러리를 통합하는 과정에서 중복되는 부분을 제거하였다. 참조 소프트웨어의 코어 복호화기의 SBR 툴과 MPS 툴은 QMF 합성을 항상 수행하였는데 렌더러가 QMF 영역에서 동작할 경우, 렌더러의 동작이 완료된 이후 QMF 합성을 수행하도록 수정하였다. 시간 영역의 바이노럴 렌더러는 입력 채널수와 상관없이 32개 채널 신호의 바이노럴 렌더링을 수행하였기 때문에 입력된 채널 신호만을 렌더링 하도록 수정하였다. [그림 7]은 구현한 MPEG-H 3DA 실시간 복호화기의 블록도이다.

MPEG-H_3DA_init() 함수는 사용자가 입력한 옵션과 MP4 파일의 메타데이터를 사용하여 복호화기 및 렌더러에서 사용되는 변수와 버퍼를 초기화 한다.

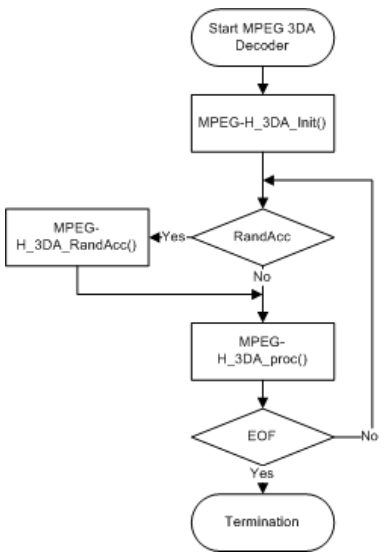

그림 7. MPEG-H 3DA 복호화기 블록도
Fig. 7. MPEG-H 3DA decoder block diagram

사용자가 선택 가능한 모드는 재생단의 스피커 위치정보, 포맷 변환기의 능동/수동 모드, 바이노럴 렌더러 사용여부 및 양이 방 충격 응답의 종류이다. 나머지 파라미터는 부호화기에서 결정된다.

MPEG-H_3DA_proc() 함수는 USAC 3D 복호화와 포맷 변환기, 객체 렌더러, 바이노럴 렌더러 연산을 수행한다. "objs_streamfile" 프로젝트에 포함된 StreamGetAccessUnit() 함수로 MP4 파일을 읽고 USACDecodeFrame() 함수로 USAC 3D 복호화를 수행한다. 코어 복호화기의 출력은 시간 영역 혹은 QMF 영역이며, 코어 복호화기의 출력 영역과 동일한 영역의 포맷 변환기 및 객체 렌더러, 바이노럴 렌더러가 동작한다. QMF 합성은 모든 렌더러의 동작이 완료된 다음 적용된다.

MPEG-H_3DA_RandAcc() 함수는 StreamGetAccessUnit() 함수를 사용하여 복호화 지점의 임의 접근 기능을 수행한다.

Ⅳ. MPEG-H 3DA 코어 복호화기 및 포맷 변환기, 객체 렌더러, 바이노럴 렌더러 연산량 분석

본 장에서는 MPEG-H 3DA 코어 복호화기, 포맷 변환기, 객체 렌더러의 연산량을 분석하고 실시간 연산 가능성을 살펴보았다. Microsoft사의 VS (Visual Studio) 2012의 프로파일러 기능을 사용하여 연산량을 측정하였다. VS 2012 프로파일러는 소스코드의 각 함수에 대하여 CPU의 종류에 영향을 받지 않는 연산량을 측정하는 기능인 "이식 가능한 이벤트"의 데이터를 수집할 수 있다. 이중 CPU에서 실제 수행된 upos (executable micro operation)를 측정하는 "Instruction retired counter" 를 사용하였다.

연산 시간은 VS 2012의 clock() 함수를 사용하여 측정하였다. [표 1]의 사양을 가지는 무부하 상태의 PC에서 싱글 스레드를 사용하여 초기화 및 파일 입출력 시간을 제외한 핵심 알고리즘의 연산 시간을 30번 측정한 평균값이다. 사용된 샘플은 5초 길이의 음원으로 스테레오, 5.1 채널, 7.1 채널, 10.2 채널, 22.2 채널

표 1. 테스트 PC 환경
Table 1. Test PC environment

CPU	Intel Ivy Bridge i5-3570 (3.4GHz, 64bit)
RAM	DDR3 16GB (PC3-128000 800MHz)
GPU	AMD Radeon R9 200 (940MHz, 3GB)
HDD	120GB SSD
OS	Windows 10, 64bit

표 2. QMF 분석 및 합성 연산량
Table 2. QMF analysis/synthesis computational complexity

Config.	Mono	22.2
QMF analysis	8.988 MIR	214.441 MIR
QMF synthesis	7.399 MIR	180.206 MIR

과 객체신호를 사용하였다.

본 논문에서는 연산량이 잘 알려진 64밴드 QMF 필터 뱅크의 연산량을 기준으로 다른 모듈의 연산량을 표시하였다. [표 2]는 1채널, 32 타임 슬롯 당 단일 채널과 22.2채널의 QMF 필터 뱅크의 연산량 측정 결과이다. 단일 채널의 QMF 합성에는 약 7.4MIR (Million instruction retired)의 연산량이 필요하였으며, 이를 QMR_MIR 으로 정의하였다.

4.1 MPEG-H 3DA 부호화 모드

복호화기의 연산량 측정을 위하여 MPEG-H 3DA 참조 소프트웨어 버전 7 부호화기를 사용하여 MP4

표 3. MPEG-H 3DA 부호화기 툴 및 명령어 스위치
Table 3. Tools and command switches of the MPEG-H 3DA encoder

Tool / module	Command switches and description	Profile	
		Lo	Hi
Mid/Side	-ms 1 (auto-detect)	O	O
	-cplx_pred (complex stereo prediction)	O	O
IGF	-enf	O	O
NF	-nf	O	O
MCT	-mctMode [int] 0 = Prediction 1 = Rotation 2 = Prediction + SF 3 = Rotation + SF	O	O
TNS	-tns_lc	O	O
	-tns	X	O
TW_MDCT	-usac_tw	X	O
SBR	-sbrRatioIndex [int] 1 = 4:1 SBR 2 = 8:3 SBR 3 = 2:1 SBR	X	O
Harmonic SBR	-hSBR	X	O
MPS 212	-mps_res (with residual)	X	O
	-mps_hybResidual 1 (hybrid residual mode)	X	O
	-ipd (with IPD coding)	X	O
QCE	-mps_qce (22.2 only)	X	O

파일을 생성하였다. 본 부호화기는 채널, 객체, HOA 신호의 부호화를 지원하며, 부호화에 사용 가능한 툴들은 [표 3]과 같다.

MPEG-H 3DA 표준의 프로파일(Profile)은 고 비트율에서 높은 음질을 제공할 수 있는 고연산량 프로파일(High profile)과 방송 및 스트리밍 상황에서 적합하도록 복호화 채널의 개수와 툴의 작동을 제한한 저연산량 프로파일(Low Complexity profile)이 있다. 저연산량 프로파일은 낮은 연산량을 위하여 SBR과 MPS 등의 QMF 영역의 툴을 사용하지 않으며, 채널과 객체, HOA 신호, 바이노럴 렌더러에 사용된 채널 개수를 기준으로 5단계로 나뉜다[1].

채널 신호의 부호화에는 모든 툴이 적용 가능하며, 객체 신호의 부호화에는 Mid/side, MCT, MPS, QCE와 같은 조인트 스테레오 툴을 제외한 툴이 적용 가능하다. 낮은 비트율로 객체 신호를 부호화해야 할 경우 MPEG-H 3D 오디오 표준의 SAOC 3D 모듈을 사용할 수 있다.

4.2 코어 복호화기 연산량

앞에서 생성한 MP4 파일을 사용하여 코어 복호화기의 연산복잡도 및 연산시간을 측정하였다. 채널당 32kbps의 비트율을 사용하였으며, 22.2채널의 경우 MPEG-H 3DA Phase 2의 청취실험에서 사용된 512kbps를 사용하였다. 코어 복호화기와 각 렌더러의 출력 신호의 프레임 크기는 2048이다.

저연산량 프로파일은 [표 3]의 사용가능한 모든 툴을 사용하여 부호화하였다. MCT 툴은 “Prediction” 모드와 “Rotation” 모드가 있으며, 전자는 스테레오 신호의 복호화에 소수점 첫째자리를 가지는 예측 계수를 사용하는 방법이고, 후자는 예측 계수를 테이블에서 찾아서 사용하는 방법이다. MCT 툴은 “Rotation+SF” 모드를 사용하였다. [그림 9]는 저연산량 프로파일의 연산량과 연산 소요 시간 측정 결과이다.

고연산량 프로파일의 연산량 및 연산 소요시간 측정 결과는 [그림 10]이다. 현재 버전의 참조 소프트웨어의 MPS는 22.2채널 신호의 부호화만을 지원하기 때문에 22.2채널 신호를 사용하여 MPS와 hybrid residual MPS, MCT 툴의 연산량을 비교하였다. SBR은 2:1 모드를 사용하였다. 고연산량 프로파일은 QMF 합성을 제외할 경우, 저연산량 프로파일과 비교

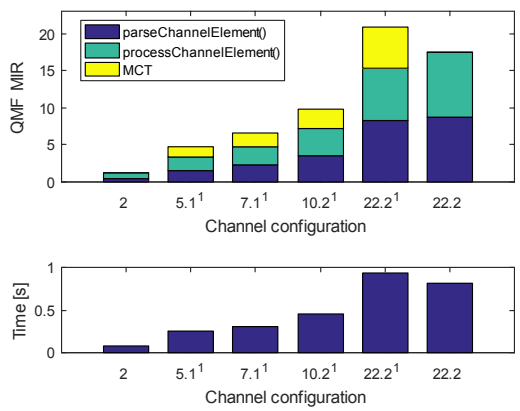

그림 8. 저연산량 프로파일 코어 복호화기 연산량 (위, 단위: QMF_MIR) 및 연산 소요 시간 (아래, 단위: 초). 1.MCT 사용
Fig. 8. LC profile core decoder computational complexity (upper, unit: QMF_MIR) and processing elapsed time (lower, unit: s) 1.with MCT

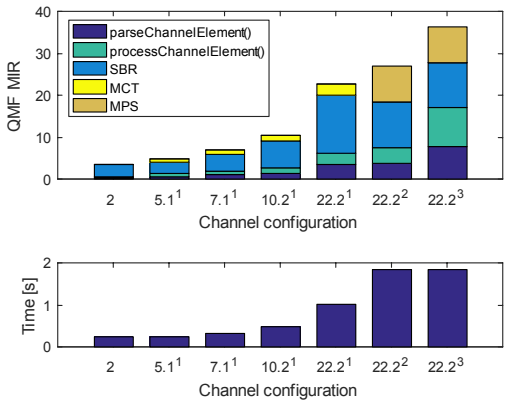

그림 9. 고연산량 프로파일 코어 복호화기 연산량 (위, 단위: QMF_MIR) 및 연산 소요 시간 (아래, 단위: 초). 1.MCT 사용, 2.MPS 사용 3.Hybrid residual MPS 사용
Fig. 9. High profile core decoder computational complexity (upper, unit: QMF_MIR) and processing elapsed time (lower, unit: s) 1.with MCT, 2.with MPS, 3.with Hybrid residual MPS

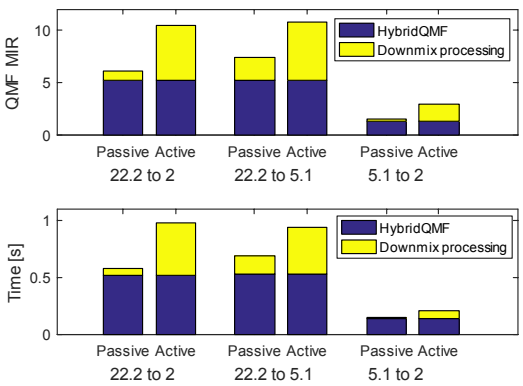

그림 10. 주파수 영역 포맷 변환기 연산량 (위, 단위: QMF_MIR) 및 연산 소요 시간 (아래, 단위: 초)
Fig. 10. FD format converter computational complexity (upper, unit: QMF_MIR) and processing elapsed time (lower, unit: s)

하여 최소 2배에서 최대 5배의 연산량을 가진다. 한 프레임에 복호화에 소요되는 평균 연산 시간과 최대 연산 시간의 비는 저연산량 프로파일의 경우 1.15 ~ 1.37배, 고연산량 프로파일의 경우 1.18 ~ 1.38배이다.

4.3 포맷 변환기 연산량

주파수 영역 포맷 변환기를 사용하여 22.2채널과 5.1채널 신호를 5.1채널, 스테레오 신호로 다운믹싱하는데 사용된 연산량과 연산시간을 측정하였다. 포맷 변환기는 77 밴드의 하이브리드 QMF 영역에서 동작하며 저연산량 프로파일일 경우 71 밴드의 하이브리드 QMF 영역에서 동작하기 때문에, 다운믹싱 이전과 이후에 하이브리드 QMF 분석과 합성 연산이 적용 된다. 한 프레임에 복호화에 소요되는 평균 연산 시간과 최대 연산 시간의 비는 3 ~ 7.1배이다.

수동 모드는 위상 정렬 과정이 없이 EQ와 다운믹스 매트릭스만을 적용하여 능동 모드에 비하여 연산량이 상대적으로 낮다. 시간 영역 포맷 변환기는 수동 모드로 작동되어, 하이브리드 QMF 연산을 제외한 수동 모드의 주파수 영역 포맷 변환기와 연산량과 연산 시간이 거의 동일하다.

4.4 객체 렌더러 연산량

시간 영역의 객체 렌더러를 사용하여 1, 2, 6개의 객체를 스테레오, 5.1채널, 22.2채널로 렌더링하는데 사용된 연산량과 연산 시간을 측정하였다. 메타데이터의 버전[1]은 2이며, 메타데이터의 프레임 크기는 1024이다. [그림 11]은 연산량 및 연산 시간 측정 결과이다.

동일한 재생 레이아웃일 때 연산량은 객체의 개수에 비례하며, 메타데이터의 프레임 크기에 반비례한다. 3D VBAP을 사용하여 객체에 대한 스피커의 이득을 계산하는데 필요한 부분은 전체 객체 렌더러 연

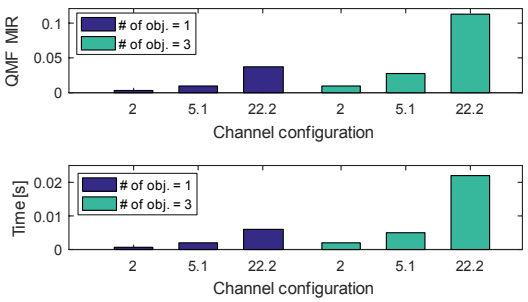

그림 11. 시간 영역 객체 렌더러 연산량 (위, 단위: QMF_MIR) 및 연산 소요 시간 (아래, 단위: 초)
Fig. 11. TD object renderer computational complexity (upper, unit: QMF_MIR) and processing elapsed time (lower, unit: s)

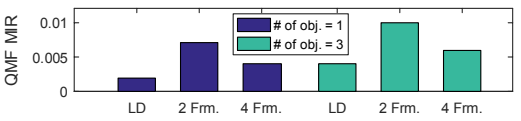

그림 12. 메타데이터 해석부 연산량 (단위: QMF_MIR)
Fig. 12. OAM decoder computational complexity (unit: QMF_MIR)

산량의 약 20%를 차지한다. 주파수 영역 객체 렌더러는 시간 영역 객체 렌더러를 QMF 신호의 실수와 허수부에 각각 적용한 것으로 그 연산량은 시간 영역 객체 렌더러의 2배이다.

객체 메타데이터의 부호화 방법으로는 저 지연모드(LD: Low delay) 혹은 n-프레임 간격으로 부호화 가 가능하다. 1개 혹은 3개의 객체 메타데이터를 저 지연 모드 부호화, 2 프레임 간격으로 부호화, 4 프레임 간격으로 부호화한 다음 이를 복호화 하는데 필요한 연산량은 [그림 12]와 같다.

객체 렌더러는 타 모듈에 비하여 연산량이 상대적으로 낮다. 24개의 객체를 22.2채널로 렌더링하는 경우 약 1 QMF_MIR의 연산량을 가진다.

4.5 바이노럴 렌더러 연산량

시간과 주파수 영역의 바이노럴 렌더러의 연산량과 연산 소요 시간 측정 결과는 [그림 13]과 같다. 사용된 양이 방 충격 응답의 평균 잔향 시간은 0.56초이다. 한 프레임에 복호화에 소요되는 평균 및 최대 연산 시간의 비는 시간 영역의 경우 1.04 ~ 1.15배, 주파수 영역의 경우 1.06 ~ 1.13배이다.

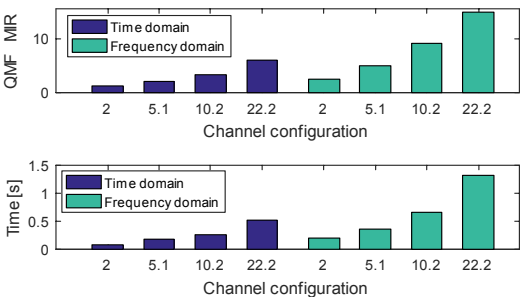

그림 13. 바이노럴 렌더러 연산량 (위, 단위: QMF_MIR) 및 연산 소요 시간 (아래, 단위: 초)
Fig. 13. Binaural renderer computational complexity (upper, unit: QMF_MIR) and processing elapsed time (lower, unit: s)

4.6 저연산량 프로파일 및 고연산량 프로파일 연산량 비교

본 절에서는 저연산량 프로파일 각 단계의 최대 채

널과 객체 신호를 복호화 및 렌더링 할 때의 연산량 및 연산시간을 측정하였다. 비교를 위하여 같은 개수의 채널과 객체 신호를 고연산량 프로파일로 복호화 및 렌더링 할 경우의 연산량 및 연산시간을 측정하였다. 1단계부터 4단계까지의 저연산량 프로파일의 최대 채널 신호 개수는 각각 스테레오, 5.1채널, 10.2채널, 22.2채널이며, 최대 객체 신호 개수는 각각 3개, 3개, 4개, 4개이다.

[그림 14]는 코어 복호화기, 포맷 변환기, 객체 렌더러를 사용하여 MP4 파일을 스테레오 신호로 렌더링하는데 소요된 시간이다. 저연산량 프로파일의 포맷 변환기는 시간 영역에서 수동 모드로 동작하며 고연산량 프로파일의 포맷 변환기는 QMF 영역에서 능동 모드로 동작한다. 저연산량 프로파일의 단계별 연산량은 각각 QMF 합성 연산의 2.8배, 6.6배, 12.4배, 23.9배이며 코어 복호화기가 대부분의 연산량 및 연산시간을 차지한다. 같은 채널 신호 객수와 객체 신호 개수를 사용한 고연산량 프로파일의 단계별 연산량은 각각 QMF 합성 연산의 11.3배, 16.6배, 26배, 62배이다.

[그림 15]는 동일한 MP4 파일을 코어 복호화기, 객체 렌더러, 바이노럴 렌더러를 사용하여 바이노럴 신호로 렌더링하는데 소요된 시간이다. 객체 렌더러는 객체 신호를 각각 스테레오, 5.1채널, 10.2채널, 22.2

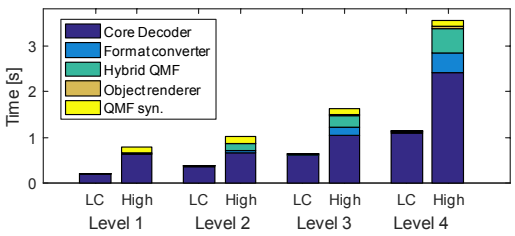

그림 14. MPEG-H 3DA 복호화기 채널 렌더링 소요 시간 (단위: 초)
Fig. 14. MPEG-H 3DA decoding and channel rendering elapsed time (unit: s)

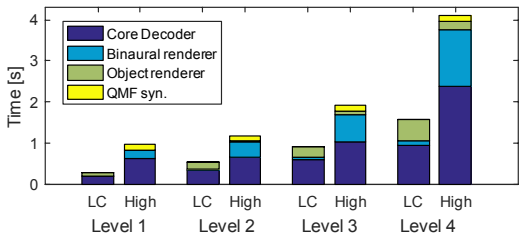

그림 15. MPEG-H 3A 복호화기 바이노럴 렌더링 소요 시간 (단위: 초)
Fig. 15. MPEG-H 3DA decoding and binaural rendering elapsed time (unit: s)

채널 신호로 렌더링 하였으며, 포맷 변환기는 사용하지 않았다. 저연산량 프로파일의 단계별 연산량은 각각 QMF 합성 연산의 4.1배, 8.4배, 15.3배, 29배이다. 같은 채널과 객체를 사용한 고연산량 프로파일의 단계별 연산량은 각각 QMF 합성 연산의 13.4배, 17.3배, 29배, 61배이다.

V. 결 론

MPEG-H 3DA 표준은 다채널 신호, 객체, HOA 신호의 효율적인 압축을 위해 최신 기술인 USAC에 다채널 신호를 위한 부호화/복호화 툴들이 추가되었으며, 다양한 재생환경에서 높은 몰입감과 현장감, 정확한 음상정위를 제공하기 위한 다운믹스 및 렌더링 기술과 후처리 기술을 포함한다.

MPEG-H 3DA 참조 소프트웨어 복호화기는 코어 복호화기와 각종 렌더러, 후처리 모듈들이 순차적으로 실행되는 구조로 실시간 처리가 불가능하다. 본 논문에서는 코어 복호화기, 포맷 변환기, 객체 렌더러, 바이노럴 렌더러의 함수의 구조를 분석하고 라이브러리화 하여 프레임 기반 복호화를 수행하는 MPEG-H 3DA 복호화기를 만들었으며, 복호화 및 렌더링에 필요 없는 부분을 제거하여 연산량을 줄였다. 또한 다양한 하드웨어 플랫폼에서 MEPG-H 3DA 복호화기를 사용하기 위한 참고자료를 제공하기 위하여 각 모듈의 동작 가능한 모드에 대한 연산량과 연산 시간을 측정하였으며 저연산량 프로파일과 고연산량 프로파일 간의 연산량 및 연산시간을 비교하였다.

싱글스레드를 사용한 일반적인 PC환경에서 본 논문에서 통합한 MPEG-H 3DA 복호화기를 사용하여 연산 시간을 측정한 결과 저연산량 프로파일과 고연산량 프로파일이 실시간으로 동작함을 확인하였다. 한국 방송 표준에 포함된 1단계에서 4단계까지의 저연산량 프로파일의 연산량을 측정한 결과, 채널 신호로 렌더링을 할 경우 QMF 합성 연산의 2.8배에서 12.4배의 연산량을 가지며, 바이노럴 렌더링을 할 경우 QMF 합성 연산의 4.1배에서 15.3배의 연산량을 가진다. 고연산량 프로파일의 연산량을 측정한 결과, 채널 신호로 렌더링을 할 경우 QMF 합성 연산의 11.3배에서 62배, 바이노럴 렌더링을 할 경우 13.4배에서 61배의 연산량을 갖는다.

References

[1] ISO/IEC 23008-3:2015, *Information technology*

— *High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D audio, AMENDMENT 2: MPEG-H 3D Audio File Format Support.*

[2] Y. S. Kim, H. Lee, E. D. Lee, and G. Lee, "A viewing zone analysis of a time-multiplex auto-stereoscopic multi-view 3D display," in *Proc. KICS Winter Conf.*, pp. 955-956, Korea, Jan. 2016.

[3] ISO/IEC 23003-3:2012, *Information technology - MPEG audio technologies - Part 3: Unified speech and audio coding.*

[4] ITU Recommendation BS.1534-3, *Method for the Subjective Assesment of Intermediate Quality Levels of Coding Systems.*

[5] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H 3D audio - The new standard for coding of immersive spatial audio," *IEEE J. Sel. Topics in Sign. Process.*, vol. 9, no. 5, pp. 770-779, Aug. 2015.

[6] J. Seo, K. Kang, and D. G. Jeong, "Overview of MPEG 3D audio standard activities for high-order multichannel realistic audio service," in *Proc. Korea Broadcast Eng.*, pp. 170-172, Korea, 2012.

[7] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, p. 456, Jun. 1997.

[8] ISO/IEC 23003-2:2010, *Information technology - MPEG audio technologies - Part 2: Spatial Audio Object Coding.*

[9] ISO/IEC 23004-4:2007, *Information technology — Multimedia Middleware — Part 4: Dynamic range control.*

[10] T. J. Lee, K. O. Kang, and W. W. Kim, "MPEG audio new standard: USAC technology," *J. Broadcast Eng.*, vol. 16, no. 5, pp. 693-704, Sept. 2011.

[11] M. Neuendorf, et al., "The ISO/MPEG unified speech and audio coding standard consistent high quality for all content types and at all bit rates," *J. Audio Eng. Soc.*, vol. 61, no. 12, pp. 956-977, Dec. 2013.

[12] S. K. Zielinski, F. Rumsey, and S. Bech, "Effects of down-mix algorithms on quality of surround sound," *J. Audio Eng. Soc.*, vol. 51, no. 9, pp. 780-798, Sept. 2003.

[13] V. Pulkki, "Generic panning tools for MAX/MSP," in *Proc. Int. Comput. Music Conf.*, pp. 304-307, Berlin, Germany, Aug-Sept. 2000.

[14] S. Y. Lim, J. M. Seok, and J. I. Seo, "Tiled panoramic video transmission system based on MPEG-DASH," in *Proc. KICS Int. Conf. Commun.*, pp. 804-805, Korea, Jun. 2015.

[15] ISO/IEC 23004-1:2007, *Information technology - Multimedia Middleware - Part 1: Architecture.*

문 현 기 (Hyeongi Moon)

2013년 2월 : 연세대학교 전기전자공학부 학사 졸업

2013년 3월~현재 : 연세대학교 전기전자공학부 석박 통합과정

<관심분야> 오디오 신호처리, 3D 오디오, 오디오 부호화

박 영 철 (Young-cheol Park)

1986년 2월 : 연세대학교 전자공학과 학사 졸업

1988년 2월 : 연세대학교 전자공학과 석사 졸업

1993년 2월 : 연세대학교 전자공학과 박사 졸업

현재 : 연세대학교 컴퓨터정보통신공학부 교수

<관심분야> 음성 신호처리, 적응 신호처리, 오디오 신호처리

이 용 주 (Yong Ju Lee)

1999년 2월 : 경북대학교 전자공학과 (공학사)
2001년 2월 : 경북대학교 전자공학과 (공학석사)
2001년 2월~현재 : ETRI 오디오연구실 책임연구원

<관심분야> 오디오 신호처리, 바이노럴 오디오 렌더링

황 영 수 (Young-soo Whang)

1982년 2월 : 연세대학교 전자공학과 학사 졸업
1994년 2월 : 연세대학교 전자공학과 석사 졸업
1990년 2월 : 연세대학교 전자공학과 박사 졸업
현재 : 가톨릭관동대학교 전자공학과 교수

<관심분야> 음향 및 음성 신호처리, 음향학, 멀티미디어