

MPEG에서의 객체기반 오디오 기술 표준화 현황

□ 서정일, 장인선, 강경옥 / 한국전자통신연구원 방송융합미디어연구부

I. 서론

최근 들어 한국과 프랑스 등에서는 청취자가 음악을 구성하는 각 요소(예로, 악기 별 연주 및 보컬 등)를 자유로이 제어하며 들을 수 있는 객체기반 음악이 서비스 되고 있다. 프로듀서에 의해 다양한 오디오 트랙이 믹싱되고 마스터된 후 최종적으로 하나의 트랙으로 사용자에게 제공되어 온 기존의 일반적인 디지털 음악 콘텐츠와는 달리, 객체기반 음악 서비스는 믹싱 이전의 각 오디오 트랙과 이들을 조합하기 위한 부가 정보를 제공하여 사용자가 자신의 취향에 맞도록 음악을 재조합하도록 한다. 따라서 기존의 일방적으로 수동적인 음악 서비스에 비해 사용자가 음악 콘텐츠 제작에 직접 참여하는 능동적인 서비스를 제공한다.

한국에서는 MUSIC 2.0이란 이름으로 객체기반 음악 CD의 온/오프 라인 판매가 이루어지고 있으며 이러한 상업적인 성공을 바탕으로 더 넓은 영역에서

의 객체기반 서비스가 계획되고 있다.

이와 같은 객체기반 서비스를 위한 기술의 표준 제정은 MPEG-4로부터 시작되었다. MPEG-4는 비디오, 오디오 등 멀티미디어 데이터를 객체별로 분리하여 효과적으로 저장, 전송 및 표현하기 위한 국제 표준으로 가장 다양한 사용자 상호 작용 및 풍부한 의미를 표현할 수 있는 멀티미디어 콘텐츠 표준이다. MPEG-4 표준에서는 객체기반의 오디오 신호를 압축하여 처리하기 위한 음성 및 음악신호 압축 기술과 다양한 오디오 객체들을 이용하여 3차원 음향장면을 구성하기 위한 음향장면기술(sound scene description) 방법에 대한 표준이 제정되었다.

객체기반 오디오 서비스의 가시화에 따라 MPEG에서는 다객체 오디오 신호를 효율적으로 압축하기 위한 부호화 기술과 다객체 오디오 콘텐츠를 상호운용성이 보장되도록 파일로 패키징하기 위한 파일포맷에 대한 표준화가 진행 중이다.

본 논문의 구성은 다음과 같다. 먼저 MPEG에서 표준화 작업이 진행 중인 다객체 오디오 부호화 기술을 2장에서 소개하고, 다객체 오디오 서비스 파일 포맷에 대한 기술을 3장에서 소개하며, 마지막으로 결론을 맺는다.

II. Spatial Audio Object Coding

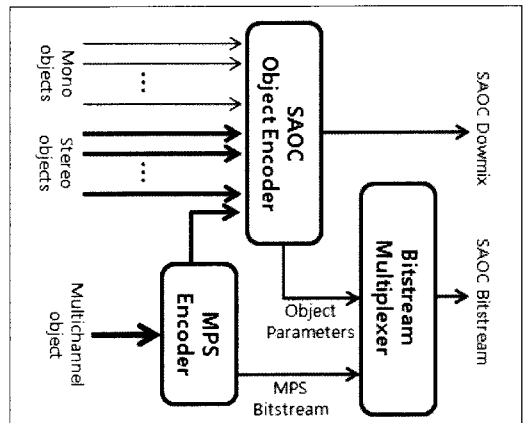
2006년말 MPEG Surround(MPS)의 표준화가 완료되어갈 즈음에 BCC Type I을 이용한 멀티객체 오디오 부호화 기술에 대한 표준화가 독일의 프라운호퍼 연구소에 의해 제안되었으며 ETRI, LG전자 등과 같은 여러 기관들의 공동작업을 통해 CFP(Call for Proposal)가 2007년 1월 발표되었다. 표준화 초기에 SAOC가 목표로 한 것은 MPS를 기본 디코더로 이용하고 SAOC 비트스트림을 MPS 비트스트림으로 변환하는 Transcoder를 표준화 하고자 하였다.

1. Application Scenarios

SAOC의 응용분야는 음원신호를 직접 효율적으로 부호화하는 것을 전제로 한 객체기반 대화형 오디오(interactive re-mix) 서비스 및 게임/리치미디어 분야와 음원으로 원격회의시스템의 음성을 다루는 Teleconferencing/telecommunication 이었다. Interactive re-mix와 Game/rich media 서비스에서는 입력된 음원신호들을 공간 오디오 부호화 기술을 이용하여 낮은 비트율로 압축하는 것을 목표로 하였으며, Teleconferencing 서비스에서는 원격회의에 참여하는 음성신호들을 효율적으로 결합 및 배분하는 것을 목표로 하였다.

2. SAOC Encoder

<그림 1>은 SAOC 인코더의 구조를 나타낸 블록도이다. SAOC 인코더는 입력 신호로써 모노, 스테레오 또는 멀티채널 오디오 객체를 입력받을 수 있다. 그러나 오디오 객체 파라미터를 추출하는 과정을 간략화하기 위하여 스테레오 오디오 객체는 모노 오디오 객체가 두 개 존재하는 것으로 처리하며, 멀티채널 오디오 객체의 경우에는 MPS 인코더를 통과한 스테레오 다운믹스 신호를 입력 오디오 객체로 처리하고 MPS 비트스트림을 객체 파라미터와 다중화하여 SAOC 비트스트림을 구성한다. 멀티채널 오디오 객체는 SAOC 트랜스코더나 디코더에 의해서 생성되는 음향장면에서 멀티채널 배경음을 구성할 뿐 청취자에 의하여 제어되지 않는 것을 가정하고 있기 때문이다.



<그림 1> SAOC 인코더 블록도

SAOC 인코더 과정 중 해석되는 객체 파라미터는 아래와 같다.



■ NRG(absolute object eNeRGy): 가장 큰 에너지를 가

진 객체의 에너지 값

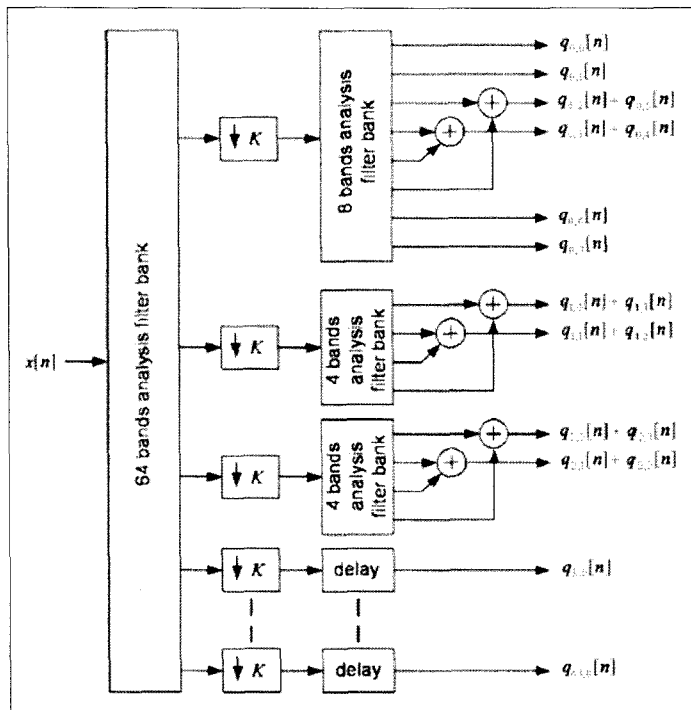
- OLD(Object Level Difference): 해당하는 객체와 가장 큰 에너지를 가진 객체와의 에너지비
- IOC(Inter-Object Correlation): 객체들 간의 유사도
- DMG(DownMix Gain): 객체들을 다운믹스할 때 취해지는 에너지 스케일값
- DCLD(Downmix Channel Level Difference): 스테레오 다운믹스 신호의 각 채널에 대한 객체들의 레벨 차이값
- PDG(Post Downmix Gain): SAOC 인코더에서 생성된 다운믹스 신호와 외부에서 입력된 다운믹스 신호사이의 레벨 차이값

3. T/F Transform

멀티객체 신호들로부터 객체 파라미터를 예측하

고 복원하는 과정은 주파수 대역별로 수행되며 입력된 시간축 신호를 주파수축으로 변환하기 위한 툴로써는 MPS와 동일한 Hybrid QMF를 이용한다.

낮은 연산량과 시간축 필터링을 이용할 수 있다는 장점을 지닌 QMF는 입력되는 시간축 신호를 고속 연산을 위해서 2^n 밴드로 균등하게 분배한다. MPS와 SAOC에서는 64밴드의 QMF 필터를 이용하게 되는데 고주파수 대역에서는 각 밴드의 분해능이 사람이 주파수를 분해하는 능력인 critical band보다 우수하기 때문에 문제가 되지 않으나 저주파수 대역에서는 하나의 QMF 밴드가 여러 개의 critical band에 맵핑됨으로 인해 적절한 객체 파라미터 해석이 불가능하다. 이러한 QMF 변환의 단점을 보완하기 위하여 저주파수 밴드를 다시 낮은 차수의 QMF 변환을



<그림 2> Hybrid QMF 해석필터 블록도

적용함으로써 critical band와 유사한 주파수 분해능을 가지도록 한 것이 Hybrid QMF 변환이다.

<그림 2>는 MPS와 SAOC에서 사용하는 71밴드 Hybrid QMF 필터의 블록도이다. 실제 객체 파라미터의 해석은 71개의 hybrid 밴드를 4개에서 28개의 파라미터 밴드로 묶어서 처리하게 된다. 또한, 파라미터 기반의 오디오 부호화 기법에서 발생할 수 있는 위상의 불일치를 보정하기 위하여 Hybrid QMF는 실수와 허수 부분을 분리하여 처리하게 되며, 연산량의 감소를 위하여 Low-Power SAOC 디코더의 일부 대역에 대해서는 실수 Hybrid QMF만을 수행한다.

또한, SBR에서도 동일한 64밴드 QMF 변환을 이용하므로 다운믹스 신호가 SBR로 처리되었을 경우에는 QMF 영역의 다운믹스 신호를 직접 입력받아 처리하는 것도 가능하다. 단, 다운믹스 신호의 결합 형태에 따라서 T/F 변환에 의한 지연(delay)시간이 달라지므로 이에 대한 보상이 필요하다.

4. SAOC Bitstream

SAOC 인코더를 통하여 부호화된 SAOC 비트스트림은 디코딩에 필요한 모든 정보를 포함하며 주요 syntactic element는 아래와 같다.



- SAOCSpecificConfig(): SAOC 디코더를 초기화하기 위한 헤더
- SAOCFrame(): Huffman 부호화된 SAOC 파라미터를 저장하는 프레임 데이터
- SAOCExtensionConfig(): residual 신호, 프리셋정보와 같이 부가적으로 추가될 수 있는 데이터를 전송하기 위한 container에 대한 헤더
- ObjectMetaData(): 객체에 대한 메타데이터 정보
- PresetMatrixData(): 사전에 정의된 렌더링 매트릭스

(Preset) 정보

- ResidualData(): 특정 객체의 음질을 보상하기 위한 residual signal 정보

상기 element들 가운데 PresetMatrixData()에는 SAOC 디코더가 복원된 객체 신호들을 어떠한 비율로 믹싱하여 출력신호를 생성하는가에 대한 정보인 렌더링 매트릭스를 담고 있으며 이를 프리셋(preset)이라 명명하였다.

5. SAOC Operation Mode

SAOC 비트스트림을 해석하여 출력 신호를 생성할 때 출력되는 오디오 신호의 채널 수에 따라서 SAOC는 디코더 형태를 표 1과 같이 구분하여 정의한다.

<# 1> SAOC의 동작모드

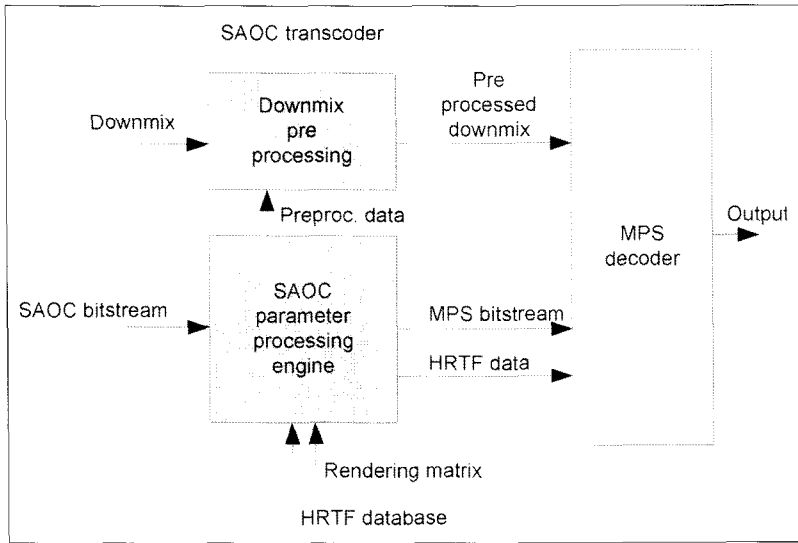
	SAOC Decoder	SAOC Transcoder
출력신호형태	mono/stereo/ binaural	multichannel
채널수	1 or 2	> 2
SAOC 모듈의 출력신호	PCM output	MPS bitstream, downmix signal
MPS 디코더	NO	YES

SAOC 디코더는 일반적인 오디오 디코더와 같이 다운믹스 신호와 SAOC 비트스트림을 입력받아 디코딩된 오디오 신호를 출력하지만, SAOC 트랜스코더는 SAOC 비트스트림을 MPS 비트스트림으로 변환(transcoding)하여 조정된 다운믹스 신호와 함께 MPS 디코더로 전달함으로써 최종 멀티채널 오디오 출력신호는 MPS 디코더가 생성하게 한다. 이런 이유는 표준화 과정에서 SAOC 디코더의 멀티채널 오디오 출력은 가정하지 않았으며, 멀티채널 복원은 MPS 디코더에 의해 처리되도록 결정하였기 때문이다.

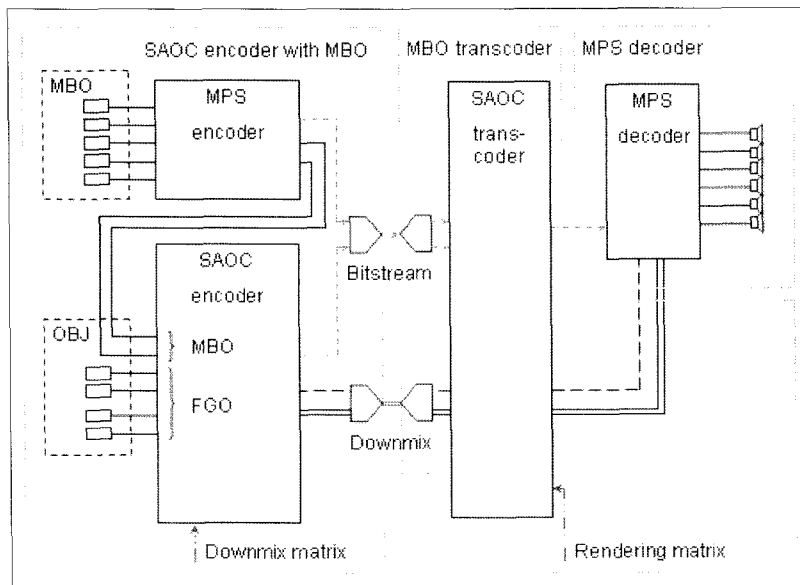
6. SAOC Transcoder Mode

<그림 3>에서와 같이 SAOC 트랜스코더는 다운믹스 신호를 조정하는 다운믹스 전처리부와 SAOC

파라미터 처리 엔진부로 구성된다. 다운믹스 전처리부에서는 입력된 다운믹스 신호에서 특정 객체를 삭제하거나 MPS 디코더에서 불가능한 위치로 객체를 이동시키기 위하여 좌우 채널신호를 변경하는 등의



<그림 3> SAOC Transcoder 블록도



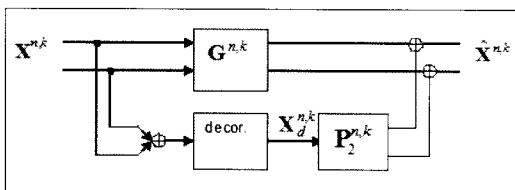
<그림 4> MBO 트랜스코딩 처리 과정도

전처리 과정을 수행한다. SAOC 파라미터 처리 엔진 부에서는 입력된 SAOC 비트스트림을 MPS 비트스트림으로 변환하는 과정을 수행하며, 이를 위하여 외부나 SAOC 비트스트림으로부터 전달받은 렌더링 매트릭스를 부가적으로 이용한다. <그림 3>에서 HRTF(Head Related Transfer Function) 데이터가 트랜스코딩 과정에서 변경되는 것으로 되어 있지만 바이노럴 출력 모드는 SAOC 디코더 모드에서 처리되는 것으로 표준이 변경되어 SAOC 트랜스코딩 모드에서는 필요하지 않게 되었다.

<그림 4>와 같이 멀티채널 배경음 객체(MBO: Multichannel Background Object)가 MPS 인코더로 부호화되어 SAOC 비트스트림으로 입력되었을 경우에는 MBO가 부호화된 MPS 비트스트림이 MPS 디코더로 직접 입력되게 되며, 객체 파라미터들로부터 변환된 MPS 비트스트림도 함께 MPS 디코더 입력되어 멀티채널 출력신호를 생성한다. 또한, MBO 객체의 재생 유무와 객체신호의 재생 유무에 따라서 MBO 비트스트림만이 전송되거나 제거되기도 한다.

7. SAOC Decoder Mode

앞서 설명한바와 같이 SAOC 디코더는 모노 또는 스테레오 다운믹스 신호를 입력받아 모노, 스테레오, 바이노럴 스테레오 신호를 출력한다. 바이노럴 스테레오 신호는 다운믹스 전처리부와 MPS 바이노



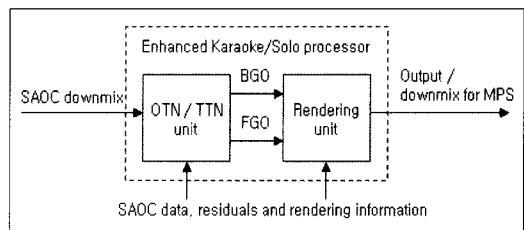
<그림 5> 스테레오 다운믹스 신호에 대한 바이노럴 디코더 처리 구조

럴 디코더가 결합된 형태를 의미한다. <그림 5>는 스테레오 다운믹스 신호가 입력되고 바이노럴 스테레오 신호를 출력하는 과정을 나타낸 그림이다. G 매트릭스는 객체 파라미터와 렌더링 매트릭스에 의해 결정되는 upmix 매트릭스이며 P 매트릭스는 G 매트릭스와 유사한 기능을 수행하지만 decorrelation된 다운믹스 신호를 입력받는다라는 차이가 있다.

8. Enhanced Karaoke/Solo Mode

실제로는 SAOC 인코더에서 다운믹스 과정을 통해 합쳐진 신호에서 객체 파라미터만으로 원래의 객체 신호들을 완벽하게 복원하는 것은 불가능하다. 따라서 SAOC에서는 ±12dB 이하로 객체의 레벨을 조절하는 것을 목표로 하였으며, 객체 신호를 모두 복원하는 과정은 고려하지 않았었다. 그러나 가라오케와 같이 특정 객체를 완벽하게 삭제하거나, 특정 객체만을 재생하는(Solo) 응용분야가 존재하기 때문에, 이와같은 상황에서도 만족할만한 음질을 제공하기 위하여 SAOC에서는 Enhanced Karaoke/Solo (EKS) 모드를 제공한다.

특정 객체 신호를 완벽하게 제거하거나 복원하기 위해서 원음과 객체 파라미터로 복원되는 음 사이의 잔차신호(residual signal)를 이용하며, 이는 MPS에서 사용하던 잔차신호 부호화 기법을 일부 수정하여 이



<그림 6> Enhanced Karaoke/Solo 모드 디코더의 처리구조

용한다. SAOC는 MPS의 기본 처리 블록인 OTT와 TTT를 사용하지 않지만 MPS의 잔차신호 부호화를 적용하기 위해서 EKS 모드에서도 이용한다. EKS 모드의 처리과정은 <그림 6>과 같으며, BGO (BackGround Object)는 객체 파라미터만으로 처리되는 객체를 뜻하며 FGO(ForeGround Object)는 잔차신호를 이용하여 특별히 처리되는 객체를 의미한다.

III. Interactive Music Application Format

대화형 음악 응용 포맷 (Interactive Music Application Format; IM AF)은 객체기반 음악 서비스를 위한 파일 포맷으로, 국제 표준화 기구인 MPEG에서 개발 중인 국제 표준 기술이다.

IM AF는 객체기반 음악 서비스를 위한 응용 포맷의 국제 표준화를 목적으로 2008년 4월 84차 MPEG 회의에서 ETRI와 오디오준이 공동 제안하여 MAF (Multimedia Application Format) under Development의 아이টে็ม으로 채택되었다[5]. 이후, 2008년 10월 86차 회의에서 정식 표준화 번호인 ISO/IEC 23000-12를 승인받았으며, 2009년 2월에 CD (Committee Draft) 단계를 거쳐 현재 study on CD 단계이다[6,7]. IM AF의 표준화에는 한국의 ETRI, 오디오준 및 프랑스의 iKlax Media 등이 참여하고 있다.

본 논문은 2009년 4월 88차 MPEG회의에서 승인된 IM AF의 study on CD를 기준으로 작성하였다.

1. IM AF의 개념

앞서 서론에서 언급했듯이, 현재 디지털 음악 콘

텐츠는 프로듀서 혹은 음악 엔지니어가 개별적으로 녹음된 보컬 및 피아노, 드럼 등 다양한 오디오 트랙을 믹싱 및 마스터링하여 제작하며, 최종적으로 하나의 트랙으로써 사용자에게 제공된다. 따라서 사용자는 프로듀서가 제작한 음악만을 수동적으로 청취하는 것이 일반적이었다.

객체기반 음악 서비스는 믹싱 이전의 각 오디오 트랙과 이들을 조합하기 위한 부가 정보를 제공하여 사용자가 자신의 취향에 맞도록 음악을 재조합하도록 한다. 따라서 기존의 일방적으로 수동적인 음악 서비스에 비해 사용자가 음악 콘텐츠 제작에 직접 참여하는 능동적인 서비스를 제공한다.

IM AF는 이러한 객체기반 음악 서비스를 위한 응용 파일 포맷으로써 여러 개의 오디오 트랙 등 미디어 데이터와 이들의 재조합에 관한 부가 정보 및 메타데이터를 ISO 파일 포맷(ISO-Based Media File Format; ISO-BMFF)으로 저장한다. ISO 파일 포맷은 멀티미디어의 렌더링, 편집, 관리 및 교환을 위해 시간 정보를 가지는 멀티미디어 데이터를 저장하기 위한 기본 포맷으로 정의되어 왔다.

세부적으로 IM AF 파일은 다음과 같이 구성된다.

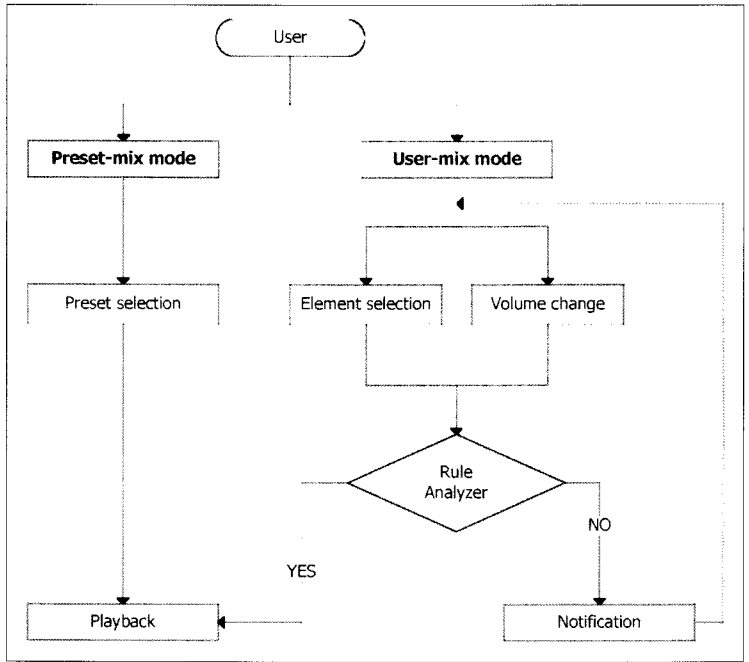


- 음악을 구성하는 여러 개의 오디오 트랙들
- 오디오 트랙간의 계층 구조를 정의한 그룹 데이터
- 프로듀서 등에 의해 사전 제작된 믹싱 정보 즉, 프리셋 데이터
- 사용자 인터렉션에 적용되는 규칙 즉, 룰 데이터
- 가사 등을 지원하기 위한 타임드 텍스트 데이터
- 앨범, 노래 등 관련 이미지
- 앨범, 노래 등 관련 메타데이터

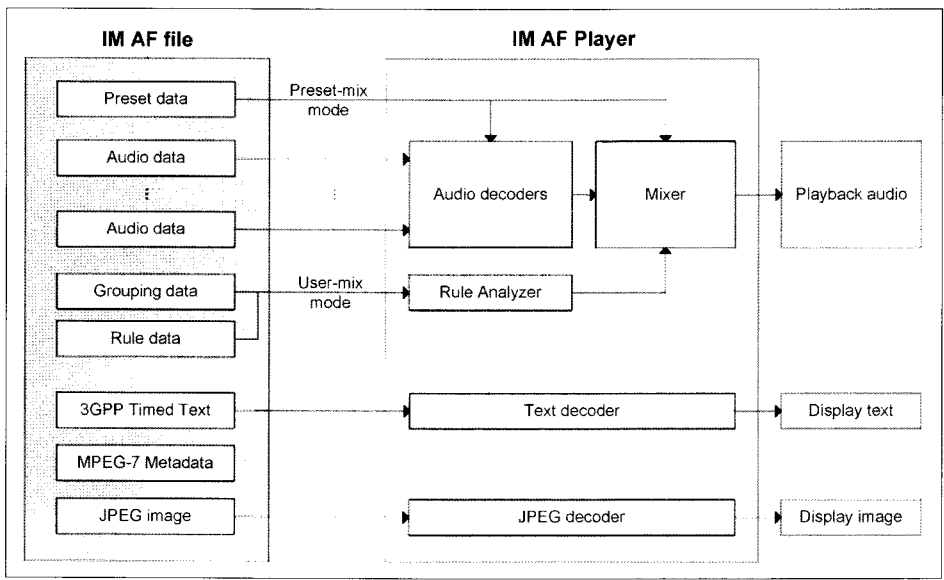
IM AF 기반의 객체기반 음악 서비스에서는 사용자 인터렉션 및 재생에 있어서 두 종류 즉, 프리셋 믹

싱 모드와 사용자 믹싱 모드의 믹싱 모드를 제공한다. <그림 7>은 두 믹싱 모드의 수행과정을 간략화

한 순서도이며, <그림 8>은 IM AF 플레이어의 기본 블록도이다.



<그림 7> IM AF 믹싱 모드 동작 순서도



<그림 8> IM AF 플레이어 기본 블록도

프리셋 믹싱모드에서는 사용자가 IM AF 파일 내에 저장되어 있는 프리셋 중 하나의 프리셋을 선택하고, 선택된 프리셋 정보를 기반으로 오디오 트랙들이 믹싱되어 재생된다.

사용자 믹싱모드에서는 사용자가 오디오 트랙 및 그룹을 선택하고 그들의 볼륨을 제어한다. 이 때, 플레이어 내의 규칙 해석 모듈(rule analyzer)에서는 요청된 사용자의 인터랙션이 프로듀서에 의해 미리 정의되어 있는 규칙을 준수하는지를 검증하고 그 결과에 따라 오디오 트랙들을 믹싱하고 재생한다.

2. IM AF의 콘텐츠 구성

앞 절에서 설명했듯이 IM AF 파일은 오디오 콘텐츠, 이미지 콘텐츠, 텍스트 콘텐츠, 메타데이터 등의 컴포넌트로 이루어진다. (표 2 참조)

<표 2> IM AF 컴포넌트

종류	요소명	약칭
Audio	MPEG-1 Audio Layer III	MP3
	MPEG-4 Audio AAC profile	AAC
	MPEG-D SAOC	SAOC
	PCM	PCM
Image	JPEG Image	JPEG
Text	3GPP Timed Text	3GPP Timed Text
Metadata	MPEG-7 Multimedia Description Scheme	MPEG-7 MDS

오디오 콘텐츠는 일반적으로 음악 서비스에서 많이 사용되고 있는 MP3와 AAC 오디오 스트림뿐만 아니라 SAOC 비트스트림, PCM 비트스트림을 수용한다. 따라서 IM AF는 모바일 기기에서의 객체기반 음악 서비스부터 고음질의 객체기반 음악 서비스까지 넓은 영역의 객체기반 음악 서비스를 망라할 수 있다.

이미지와 관련해서는 JPEG 이미지 타입을 지원하며, 텍스트는 3GPP timed text format을 지원한다. 따라서 음악 재생 시 동기화된 가사의 디스플레이가 가능하므로 가라오케 서비스 등에 응용할 수 있다. 또한, 메타데이터는 MPEG-7 MDS를 지원한다.

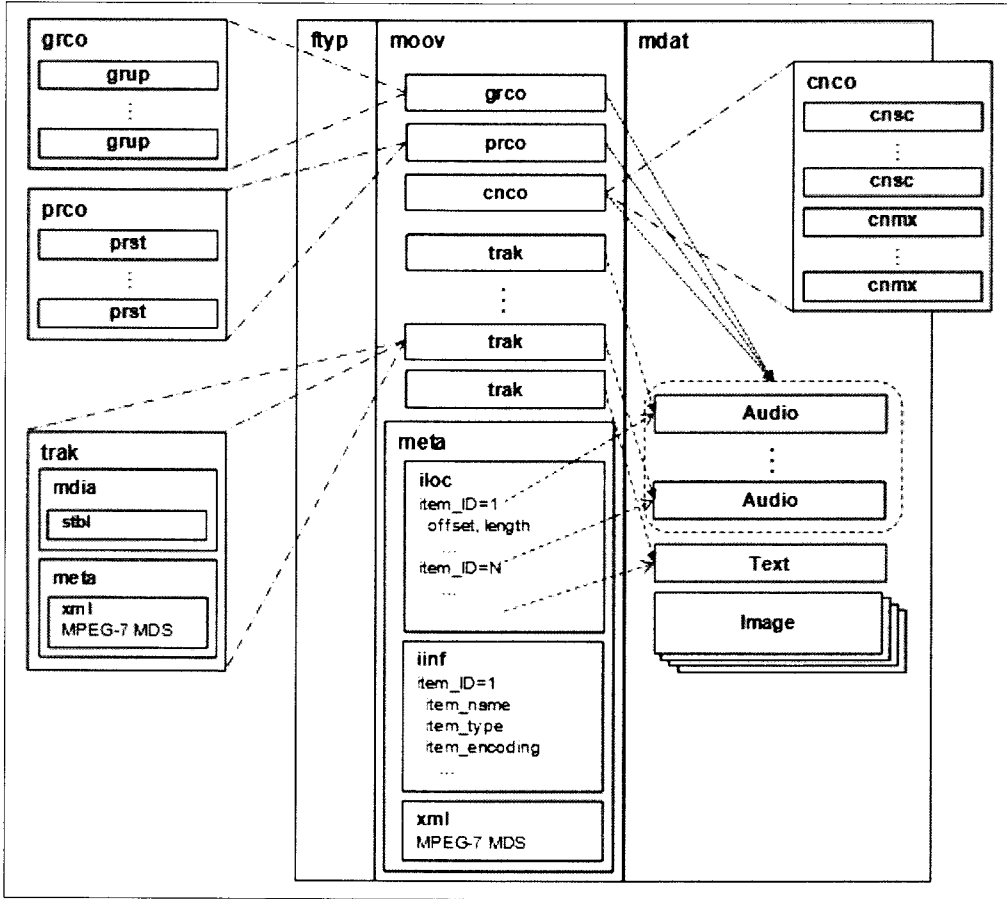
3. IM AF 파일 포맷 기술

IM AF는 ISO 파일 포맷을 기본 골격으로 한다. <그림 9>는 IM AF의 단일 형태의 파일 구조를 나타낸다. 단일 콘텐츠는 한 개의 노래에 대한 오디오 트랙들과 이와 관련된 부가 리소스 트랙들로 구성된다. 모든 미디어 리소스들은 Media Data Box('mdat')에 저장되며, 각 미디어 리소스에 대한 초기화 정보 및 동기화 정보는 Movie Box('moov') 내에 개별 미디어 트랙 즉, Track Box('trak') 단위로 저장된다. Meta Box('meta')에서는 각 미디어를 개별 소비 단위인 아이টে็ม으로 분리하여 관리하기 위한 정보가 포함되며, 아울러 각 미디어에 대한 상세정보 메타데이터가 포함될 수 있다.

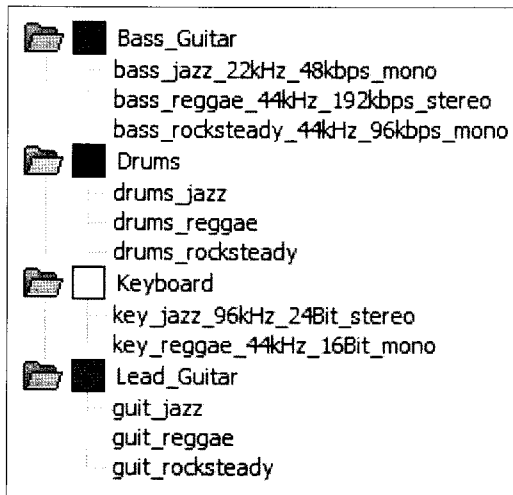
이외에도 IM AF 표준에서 특정하게 지정한 박스들과 그 기능들은 아래와 같다.

첫째, IM AF 파일 내 저장된 많은 오디오 트랙을 그룹지어 관리함으로써 사용자 인터랙션에 대한 규칙을 적용하거나 프리셋 정보를 적용하는데 편리성을 제공한다. <그림 10>은 네 개의 그룹 즉, Bass Guitar, Drum, Keyboard 및 Lead Guitar의 그룹으로 구성된 계층 구조의 예이다. 이와 같은, 오디오 트랙 및 그룹 간의 계층 정보는 Group Container Box('grco')에 개별적 Group Box('grup') 단위로 저장된다.

둘째, IM AF에서는 사용자가 음악을 쉽고 간편하



<그림 9> 단일형태의 IM AF 파일 구조

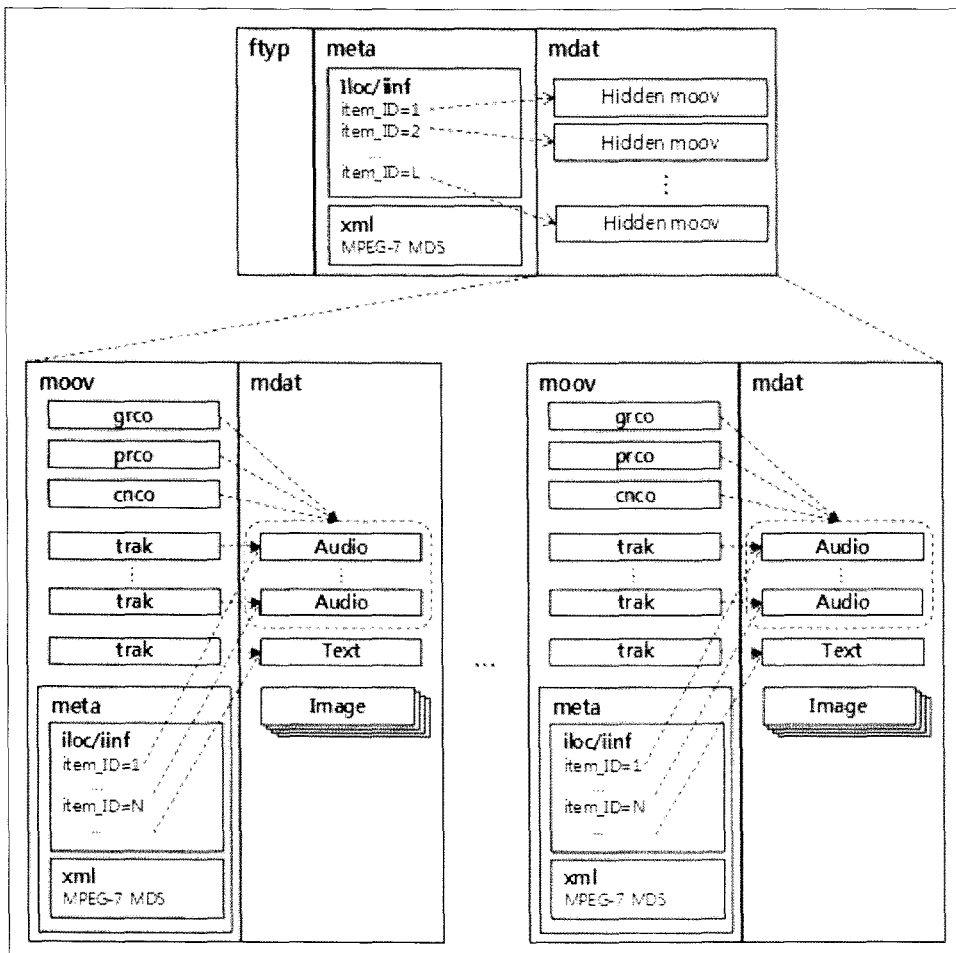


<그림 10> 여러 그룹으로 구성된 계층 구조의 예

게 재조합 할 수 있도록 IMAF 파일 제작 시 프로듀서 등이 프리셋 정보를 제작하여 저장한다. 이러한 프리셋 정보는 Preset Container Box('prco')에 개별적인 Preset Box('prst') 단위로 저장된다. 프리셋은 시간 변화에 따른 볼륨 값의 변화 및 볼륨 값의 적용 대상에 따라 정적/동적 트랙 볼륨 프리셋 및 정적/동적 객체 볼륨 프리셋의 네 가지의 타입으로 분류하여 저장할 수 있다.

셋째, Rule Container Box('ruco')에는 트랙 및 그

룹 선택에 대한 사용자 인터랙션 규칙이 개별적인 Selection Rule Box('rusc') 박스 단위로, 믹싱에 대한 사용자 인터랙션 규칙이 개별적인 Mixing Rule Box('rumx') 박스 단위로 저장된다. <그림 11>은 다중 형태의 파일 구조를 나타낸다. 다중 형태의 파일 구조는 앞서 설명한 단일 형태의 IMAF 파일들을 하나로 묶어 저장하는 것이다. 따라서 단일 음악에 대한 IMAF 파일들을 하나로 묶어 객체기반 음악 앨범을 구성하는 경우 다중 형태로 저장하면 된다. <그



<그림 11> 다중 형태의 IMAF 파일 구조

림 11>에서 보듯이 단일 형태의 파일 구조는 다중 형태의 IM AF 파일 내의 Media Data Box(‘mdat’) 내에 모두 저장되며 (즉, ‘hidden moov’), 단일 형태의 IM AF 단위의 아이템 식별 및 상세정보 메타데이터가 Meta Box (‘meta’)에 저장된다.

4. 메타데이터

IM AF는 파일 내에 음악 콘텐츠에 대한 상세 정보를 저장하여 전달함으로써 사용자가 음악 앨범, 노래 및 세부 오디오 트랙에 대한 정보를 쉽게 파악할 수 있도록 한다.

IM AF의 메타데이터는 <표 3>과 같이 앨범, 노래 그리고 트랙 레벨로 정의되며, 각 레벨의 메타데이터의 저장위치는 <표 4>와 같다.

<표 3> IM AF 메타데이터

Description	Level		
	Album	Song	Track
Title	o	o	o
Singer	o	o	-
Composer	-	o	-
Lyricist	-	o	-
performing musician	-	-	o
Genre	o	o	-
file date	o	o	o
CD track number of the song	-	o	-
Production	o	o	-
Publisher	o	o	-
copyright information	o	o	-
ISRC (International Standard Recording Code)	-	o	-
image	o	o	-
URL site address related to the music and the artist (e.g. album homepage, fan cafe, music video)	o	o	-

<표 4> IM AF에서의 각 레벨 메타데이터 저장 위치

Metadata	Location
track level	trak/meta box
song level	moov/meta box
album level	meta box of file

메타데이터의 핸들러는 ‘mp7t’이다. 트랙과 노래 레벨의 메타데이터는 MPEG-7 Creation-Information DS, MediaInformation DS and Semantics DS 를 이용한다. 앨범 레벨의 메타데이터는 앨범 내에 노래들이 순차적으로 정렬되어 있기 때문에 이에 대한 구조 정보 또한 가지므로, 앨범 레벨의 메타데이터는 MPEG-7 CreationInformation DS 뿐만 아니라 MPEG-7 ContentCollection DS를 이용하여 기술한다. 즉, 앨범 레벨의 메타 데이터를 기술하기 위한 ContentCollection DS 내의 콘텐츠 요소 하에 CreationInformation DS을 포함함으로써 메타 데이터를 결합시킨다.

IV. 결론

MUSIC 2.0과 같은 객체기반 음반서비스의 출현과 Teleconferencing 시스템에서의 효율적인 부호화를 위해 파라메트릭 부호화 기술을 이용하는 SAOC의 표준화가 완료단계에 있으며, 객체기반 음반 콘텐츠의 표준화된 파일포맷을 위한 IM AF 표준화도 현재 활발히 진행 중이다. SAOC는 새로운 개념의 대화형 오디오 서비스를 기존 오디오 서비스와 유사한 대역폭으로 제공하는 것이 가능하므로 빠른 시기에 가라오케, AOD(Audio on Demand) 등과 같은 응용분야에 적용될 것으로 예상된다. 또한 IM AF는 한국, 프랑스 등 일부국가에서만 제공되고 있

는 새로운 개념의 객체기반 음반 콘텐츠 서비스를
 국제 표준화함으로써 전 세계의 음반시장을 기술적
 으로 선도하고 새로운 시장을 개척하는 촉진제가 되
 리라 기대된다.

● 참고문헌 ●

- [1] ISO/IEC 14496-3:2005, MPEG-4 coding of audio-visual object, Part 3: Audio, 2005.
- [2] ISO/IEC 23003-1:2007, MPEG-D MPEG audio technologies, Part 1: MPEG Surround, 2007.
- [3] ISO/IEC FCD 23003-2, MPEG-D MPEG audio technologies, Part 2: Spatial Audio Object Coding (SAOC), 2009.
- [4] Christof Faller and Frank Baumgarte, "Binaural Cue Coding Applied to Audio Compression with Flexible Rendering," 113th AES Convention, Oct., 2002.
- [5] ISO/IEC JTC1/SC29/WG11/N9840, "MAF Overview," 2008.
- [6] ISO/IEC JTC1/SC29/WG11/N10468, "Text of ISO/IEC 23000-12 CD Interactive Music AF," 2009.
- [7] ISO/IEC JTC1/SC29/WG11/N10615, "Study of ISO/IEC CD 23000-12 Interactive Music AF," 2009.

필자소개



서정일

- 1994년 2월 : 경북대학교 전자공학과 학사
- 1996년 2월 : 경북대학교 전자공학과 석사
- 2005년 8월 : 경북대학교 전자공학과 박사
- 1998년 3월 ~ 2000년 10월 : LG반도체 주임연구원
- 2000년 11월 ~ 현재 : 한국전자통신연구원 선임연구원
- 주관심분야 : 오디오 부호화, 다채널 음장재현 시스템, 3차원 오디오, 디지털 방송 시스템, 객체기반 오디오 시스템



장인선

- 2001년 2월 : 충북대학교 전기전자공학부 정보통신공학 학사
- 2004년 2월 : 포항공과대학교 컴퓨터공학과 석사
- 2004년 8월 ~ 현재 : 한국전자통신연구원 연구원
- 주관심분야 : 객체기반 오디오 시스템, 음원분리, 오디오 부호화, 3차원 오디오 및 음향 신호처리



강경욱

- 1985년 2월 : 부산대학교 물리학과 학사
- 1988년 2월 : 부산대학교 대학원 물리학과 석사
- 2004년 2월 : 한국항공대학교 대학원 항공전자공학과 박사
- 1991년 2월 ~ 현재 : 한국전자통신연구원 팀장
- 2006년 4월 ~ 12월 : 영국 Southampton 대학 방문연구원
- 주관심분야 : 음향 신호처리, 3차원 오디오, 디지털 방송 시스템, 객체기반 오디오 시스템