

논문 2010-47SP-5-4

사용자 기반 실감 객체 오디오 파일 포맷 및 오디오 장면 묘사 기법

(An User Controllable Object Audio File Format and Audio Scene Description)

조 충 상*, 김 제 우*

(Choong Sang Cho and Je Woo Kim)

요 약

최근의 오디오 기술은 사용자 중심으로 변화하고 있으며, 사용자의 환경과 의도에 따라 능동적으로 서비스가 이루어지는 대화형 오디오 서비스 시대로 변화하고 있다. 이에 맞추어 시장에서는 고품질 오디오 서비스를 위한 무손실 오디오 기술을 탑재한 멀티미디어 기기와 사용자가 선택적으로 악기를 조절할 수 있는 객체 오디오 음원 서비스가 이루어지고 있다. 본 논문에서는 사용자가 컨트롤 할 수 있는 실감 객체 오디오 파일 포맷을 설계하고 실감 정보를 낮은 복잡도에서 저장 매체와 전송 매체에 적용 가능한 오디오 장면 묘사 방법을 제안한다. 설계된 실감 객체 오디오 파일 포맷은 MPEG-4 파일 포맷을 기반으로 설계되었다. MPEG-4 파일 포맷은 MPEG-4에 속하는 고성능 오디오 코덱을 오디오 코덱 객체 번호를 통해서 쉽게 적용가능하다. 또한 오디오 객체 개수의 변화에 따라 파일 포맷의 트랙을 변화 시켜 사용하면 되므로 객체 오디오를 포함하기에 적절하다. 본 연구에서 개발된 파일 포맷은 실감 객체오디오 생성시 MPEG-4 오디오 코덱으로 압축된 객체 오디오, 실감 객체 오디오를 위한 오디오 장면 묘사 데이터를 독립적인 트랙으로 포함하고 있다. 포함된 오디오 장면 묘사 기법은 저장 매체를 위해 전체 오디오 장면에 적용되는 오디오 묘사 기법과 각각의 오디오 객체에 적용되는 오디오 묘사 기법을 노드 구조로 설계 하였으며, 전송 매체를 위해서 기본적인 객체 오디오 동작을 하기위한 필수 정보와 오디오 세부장면 묘사를 위한 정보로 분할하여 설계하였다. 이를 바탕으로 본 연구에서는 실감 객체 오디오 시뮬레이터를 개발하였다. 개발된 시뮬레이터는 객체 음원과 오디오 장면 묘사 정보를 부호화하여 MPEG-4 파일 포맷에 저장하며, 생성된 실감 객체 오디오 파일은 재생 모듈에서 오디오 객체에 입력 받은 사용자 정보와 오디오 장면 묘사 정보가 적용되어 사용자에게 몰입감이 높은 실감 오디오 서비스를 제공한다.

Abstract

Multi-media service has been changed into user based audio services, which service supports actively user's preference and interaction with the users. In the market, multi-media products which can support the highest audio-quality by using lossless audio technology have been released and object audio music which user can select the objects has been serviced. In this paper, we design user's preference information based object audio file format and audio scene description for storage and transmission media. The designed file format is designed based on MPEG-4 file format because high-quality audio codecs in MPEG-4 audio can be easily used and the track of file format can be flexibly controlled depend on the number of the instrument in music. The encoded audio data of each objects and encoded audio scene description by binary encoding that has independent track are packed in a file. The scene description for storage media is consist of full and object scene description, the scene description for transmission media has an essential description for object audio operation and a specific description for real audio sound. The designed file format based simulator is developed and it generates an object audio file with several scene descriptions. Also, the real audio sound is serviced by the interaction with user and the unpacked scene description.

Keywords : Object Audio, Scene Description, 객체 오디오, 실감 오디오, MPEG-4 파일 포맷

* 정희원, 전자부품연구원 멀티미디어IP 연구센터
(Korea Electronics Technology Institute, Multimedia IP research center)
접수일자: 2010년7월5일, 수정완료일: 2010년8월9일

I. 서론

최근 멀티미디어 서비스는 HDTV(high definition TV), IPTV(internet protocol TV)등을 통해 사용자에게 실감 고품질 콘텐츠를 제공되고, 사용자의 의도에 따라 능동적으로 서비스가 이루어지는 대화형 멀티미디어 서비스로 발전해 가고 있다^[1]. 또한 정보 전달 속도의 발전과 함께 유무선 인터넷 속도의 증가에 따라 언제 어디서나 사용자가 자신이 원하는 서비스를 선택적으로 즐기며, 자신의 요구를 표현하여 반영된 서비스를 받는 유비쿼터스형 양방향 멀티미디어 서비스로 진화하고 있다^[2]. 이러한 서비스 변화에 따라 사용자가 오디오 기술에서 참여할 수 있는 공간을 제공하고 사용자의 주변 환경에 적응적으로 실감을 제공하는 기술들이 연구되고 있다. 적응적 실감 오디오 기술은 오디오 음원이 생성과 물리적 공간이 다른 청취자에게 물리적 공간의 차이를 극복하고 실감을 제공하기 위한 음향학적인 기술과 사용자의 취향, 신체적, 심리적 조건, 사용자 청취 환경 등에 따라 최고의 실감 오디오 서비스가 이루어지도록 하는 사용자 중심의 오디오 기술이다.

이와 관련하여 오디오 분야에서는 멀티미디어 기기에서 고품질에 대한 요구가 증가하고 있으며, 이에 맞

추어 일부 휴대용 멀티미디어 기기에서는 무손실 오디오 코덱을 탑재하여 사용자가 원음을 감상할 수 있는 기회를 제공하고 있다^[1]. 대화형 오디오 기술로써는 능동적인 멀티미디어 기기 사용자를 위하여 사용자가 자신이 듣고자하는 음악의 악기 구성과 볼륨을 조절할 수 있는 객체 오디오의 초기 버전들이 출시되고 있다^[3-6]. 실감 입체 오디오 효과 기술들은 아직 많은 부분이 연구가 진행되고 있다. 연구되는 분야로 3차원 음향 가공, 음향정보 추출, 사람의 청각적 특성 분석, 공간에서의 사운드 특성 분석, 멀티채널 스피커 사용에 따른 다채널 스피커 기술들이 연구되고 있다^[7-9].

앞에서 언급된 실감 오디오 연구들을 바탕으로 사용자와 교감을 이루는 실감오디오 기술을 대중화하기 위하여 실감오디오를 담을 수 있는 파일 포맷이 요구된다. 본 연구에서는 그림 1과 같이 가정용 홈 서버 및 네트워크 기반 스트리밍 서비스가 가능한 대화형 실감 객체오디오 파일포맷을 설계한다.

설계된 파일 포맷은 객체오디오를 위한 정보 뿐 만 아니라 실감 정보를 위하여 3차원 오디오 지원이 가능해야 한다. 또한 설계된 포맷은 기존의 파일 포맷과 융합되어 기존의 포맷에 풍부한 실감 객체 오디오 지원이 가능해야 하며, 설계된 파일 포맷 단독으로 실감 객체

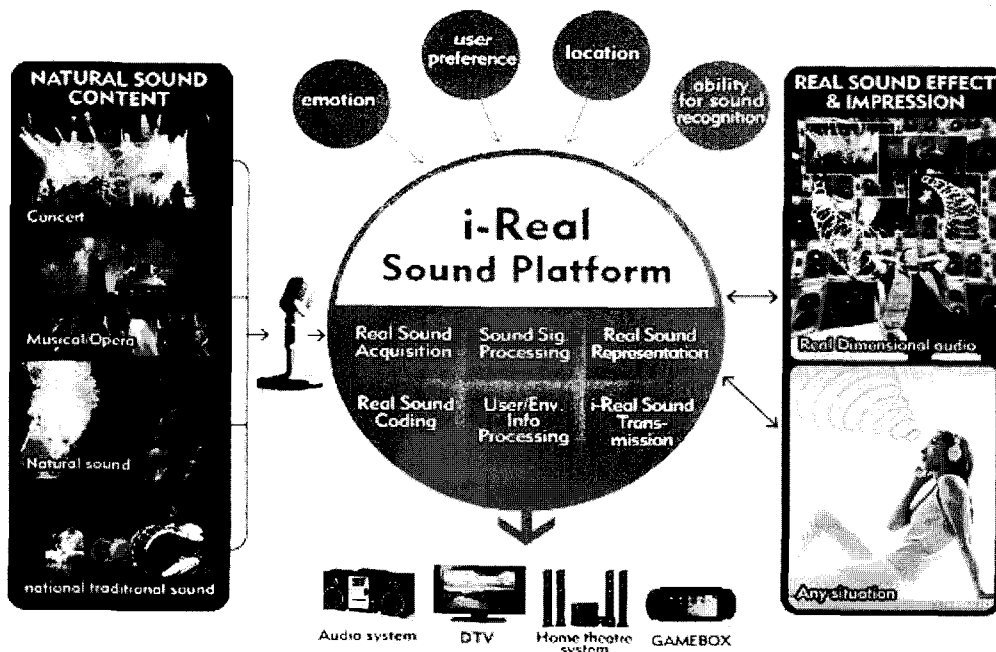


그림 1. 사용자 정보 기반 실감 객체 오디오 재생 플랫폼 개념도
 Fig. 1. Structure of real-object audio system based on user's information.

오디오 지원이 가능하도록 설계한다. 복잡도 측면에서는 대중적인 사용이 가능하도록 휴대용 멀티미디어 기기에서 동작 가능한 낮은 복잡도를 보여야 한다.

본 논문의 구성은 다음과 같다. 서론에 이어 제 II장에서는 실감 오디오 기술 동향을 분석하고, III장에서 사용자 정보기반 실감오디오에 대해 설명한다. 제IV장에서는 실감 객체오디오 파일 포맷 설계에 대해서 설명하며, 제 V장에서는 설계된 오디오 장면 묘사 기법을 설명한다. 제 VI장에서는 설계된 포맷을 이용한 시뮬레이터를 설명하고 마지막으로 제 VII장에서 결론을 맺는다.

II. 실감 오디오 기술 동향

세계 디지털 음악시장은 DVD(Digital Video Disc)와 blue ray 저장 매체를 통한 음원의 출시에 따라 기존의 스테레오 음원에서 5.1, 7.1채널 오디오 음원이 점점 대중화 되고 있다. 이러한 오디오 시장의 변화에 따라 실감 오디오 기술은 오디오 실감과 밀접한 관계가 있는 다채널 오디오와 대화형 서비스가 가능한 객체 오디오 기술로 부각되고 있다.

다채널 오디오는 채널수의 증가에 의해 사용자에게 좀 더 공간감을 강하게 느낄 수 있게 한다. 따라서 오디오의 실감을 증대하기 위해 오디오는 채널수가 증가하고 있으며, 현재 스테레오와 5.1 채널이 휴대용 멀티미디어 기기와 TV에서 주류를 이루고 있다. 하지만 보다 높은 실감을 위하여 sony digital sound는 7.1을 사용하고 있으며, 5.1채널 기반으로 10.2 채널 구조가 THX labs. 과 university of southern california에 의해 개발되었다^[10]. 10.2채널은 기존의 멀티채널과 달리 2개의 높이를 갖는 스피커를 포함하고 있으므로 더욱 공간감을 증대 시킨다. 일본의 NHK경우는 UHDV(ultra-high definition video)를 위하여 22.2채널 다채널 오디오 구성을 개발하였다^[11]. 22.2 채널 오디오는 3개의 레이어로 구성되어 최하단의 레이어는 5개의 스피커, 중간 레이어에는 10개의 스피커, 최상의 레이어는 9개의 스피커로 구성되어 멀티레이어, 다채널을 통한 공간감을 제공한다.

객체 오디오를 통한 실감 오디오분야는 MPEG에서 표준화된 SAOC(spatial audio object coding)가 있다. SAOC는 낮은 데이터레이트에서 객체오디오 구성이 가능하도록 다운 믹스 구조를 갖고 있다^[12~13]. 다운믹

스된 모노, 스테레오신호는 디코더에서 공간파라미터와 여기신호와 조합되어 각 객체별 오디오 신호를 복원하며, 완벽한 복원을 위해서는 여기신호가 요구되는 단점을 갖고 있다. 또한 SAOC의 다운 믹스된 신호는 SAOC 비트스트림을 변환하여 MPEG surround의 비트스트림으로 변환하여 MPEG surround와 호환 가능하다.

국내의 객체오디오 기술은 ETRI와 (주)오디즌에 의해 개발된 뮤직2.0은 객체오디오 음원의 첫 상용화 사례이다^[3]. 뮤직 2.0은 음악을 구성하는 악기를 사용자가 선택하여 음악을 구성하고, 각 악기(객체)의 볼륨을 사용자가 조절함으로써 사용자가 자신이 원하는 음악을 구성할 수 있게 하였다. 현재 국내에서 여러 가수의 음반이 객체 오디오로 제작되어 서비스가 이루어지고 있다. 국외의 객체오디오 기술은 프랑스 iklax사가 개발한 것으로 보컬, 기타, 드럼 등 각각의 독립적인 오디오 음원을 wav, mp3, ogg vorbis으로 압축하고, iklax사에 의해 개발된 포맷에 각 객체를 독립적으로 저장하여 사용자가 음악에서 구성하고자 하는 악기를 선택할 수 있으며, 각 악기의 볼륨을 조절할 수 있다^[4].

객체 오디오를 위한 파일 포맷은 2010년에 MPEG에서 interactive music application format (IMAF)이라는 이름으로 표준화가 완료되었다. 표준화된 포맷은 ISO 베이스 미디어 파일 포맷을 기반으로 설계되었으며, 객체 오디오를 위한 추가적인 box들을 정의하였다. 하지만 표준화된 포맷은 객체의 기본 정보와 볼륨을 컨트롤 정보만을 갖고 있으며, 3차원 오디오 효과를 위한 정보를 위한 box를 갖고 있지 않다^[5].

III. 사용자 정보 기반 실감 오디오 기술

사용자 정보기반 실감 오디오란 그림 1에 보이는 바와 같이 다양한 환경에서 취득된 고품질 콘텐츠를 사용자 특성, 취향, 환경 등을 기반으로 한 실감 사운드 처리를 통해, 청취자에게 실제 환경에서 느끼는 감동을 제공할 수 있는 능동적인 미래의 오디오 사운드 시스템이다. 이러한 시스템을 구축하기 위해서는 공간음향 기술을 이용한 실감 오디오 획득/저장 및 부호화 기술, 청취자의 취향이나 감성정보를 바탕으로 한 맞춤형 실감 사운드 재생 기술, 청취자의 청각특성과 환경정보 기반 실감 사운드 전송 및 표현 기술이 필요하다.

실감 오디오 획득/저장 및 부호화 부분에서는 획득된

오디오 데이터 레이트를 감소시키기 위하여 높은 오디오 품질을 제공하는 오디오 코덱으로 압축한다. 압축된 오디오 데이터와 오디오 신호에서 추출된 오디오 신호 관련 정보, 오디오 획득 환경에 대한 정보를 실감 오디오 파일 포맷에 포함시키는 작업을 수행한다.

실감 오디오 재생부에서는 실감 파일 포맷에 포함된 오디오 정보와, 재생부에서 수집된 청취자의 정보를 기반으로 사용자에게 최적의 실감오디오를 제공하기 위하여 재생될 오디오 악기 및 볼륨 컨트롤, 3D 사운드 효과 등이 적용되어 음악을 청취하는 사람의 환경에서 가장 실감나는 오디오 사운드를 제공한다.

IV. MPEG-4 파일포맷 기반의 실감 객체 오디오 파일 포맷

본 장에서는 MPEG-4 파일 포맷을 기반으로 실감 객체 오디오 데이터를 저장하기 위하여 설계된 파일 포맷의 구조를 설명한다.

1. MPEG-4 파일 포맷 기반 객체 오디오 포맷

MPEG-4 파일 포맷은 시간을 기준으로 정렬되는 미디어 데이터를 포함하기 위해 유동적이며, 확장 가능한 파일 포맷이다^[14]. 또한 MPEG-4 파일 포맷은 미디어 데이터를 위한 표현과 미디어에 대한 설명 정보를 포함하는 등 미디어에 대한 포괄적인 정보를 담을 수 있으며, MPEG-4에 속하는 다양한 오디오 코덱을 오디오 데이터 압축에 사용하고, 사용된 코덱의 객체 번호만을

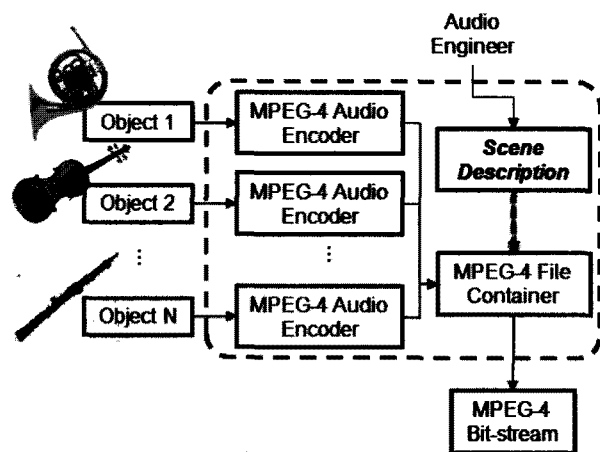


그림 2. MPEG-4 파일 포맷 기반의 실감 객체 오디오 파일 생성 구조도
Fig. 2. File generation structure of real object audio based on MPEG-4 file format.

표기하여 쉽게 다양한 코덱 사용할 수 있는 장점이 있다^[15~16].

이러한 장점을 갖는 MPEG-4 파일 포맷을 이용하여, 실감 객체오디오를 구성하는 객체를 MPEG-4에 속하는 오디오 코덱을 이용하여 독립적으로 부호화한다. 그림 2는 MPEG-4 파일포맷 기반으로 각 악기를 MPEG-4 오디오에 속하는 오디오 코덱으로 독립적으로 부호화 한다. 동시에 음악 녹음 엔지니어에 의해 생성된 정보와 녹음 환경/조건의 정보 및 3D 오디오 효과를 적용하여 실감 효과를 만들고자 하는 정보 등이 장면 묘사 정보에 포함되어 하나의 트랙으로 실감 객체 오디오 파일 포맷에 저장된다.

저장된 실감 객체 오디오 파일 포맷은 그림 3과 같은 실감 객체 오디오 재생 구조를 갖고 있다^[6]. 여러 오디오 객체와 여러 오디오 장면 묘사 정보로 구성된 파일 포맷에서 장면 묘사정보와 오디오 데이터를 분리한다. 분리된 오디오 데이터는 MPEG-4 오디오 복호화기를 통해서 복호화가 이루어지며, 오디오 장면 묘사 정보 또한 복호화 된다. 재생단에서는 음악 서비스를 이용하는 사용자의 주변 환경과 사용자가 선택한 정보들이 입력되고, 오디오 믹서부에서는 복호화된 오디오 장면 묘사 정보와 입력된 사용자 정보가 적용되어 실감 객체 오디오가 재생된다.

MPEG-4 파일 포맷 기반으로 개발된 실감 객체 오디오 파일 포맷의 전체 구성은 그림 4와 같다. 하나의 무비(mov) 데이터는 무비 헤더(mvhd), 객체 묘사 초기화(iods)와 트랙(trak)으로 구성된다. 객체 묘사 초기

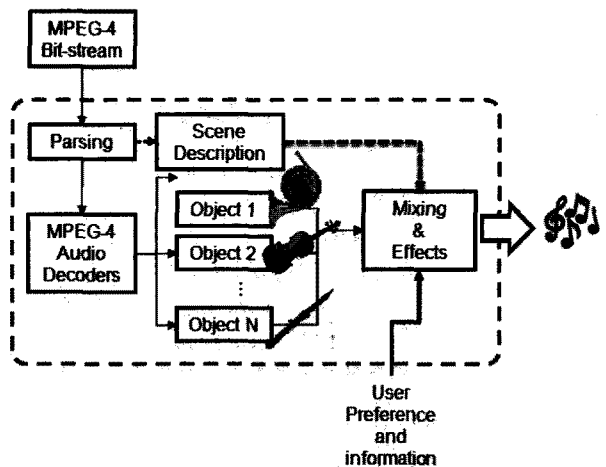


그림 3. MPEG-4 파일 포맷 기반의 실감 객체 오디오 파일 재생 구조도
Fig. 3. Audio player structure of real object audio file based on MPEG-4 file format.

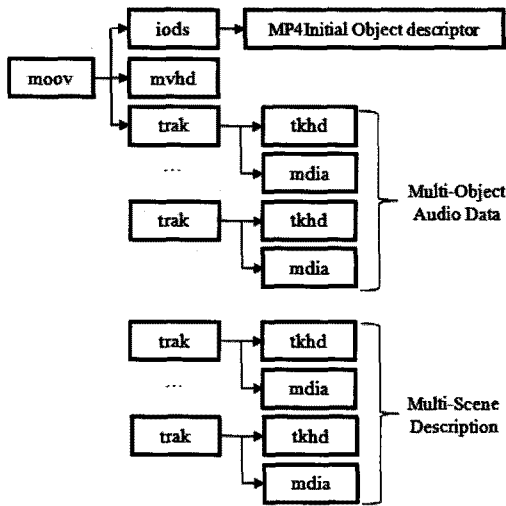


그림 4. MPEG-4 파일 포맷 기반의 실감 객체 오디오 파일 포맷 구조
 Fig. 4. Real object audio file format structure based on MPEG-4 file format.

화 노드에서는 사용된 MPEG-4 오디오 코덱에 대한 초기화 및 상세 정보가 포함되며, 트랙에는 트랙에 저장된 오디오 비트스트림의 위치 및 관련 정보가 저장된다.

각각 부호화된 오디오 객체는 그림 4와 같이 멀티 트랙으로 구성되어 MPEG-4 파일 포맷의 각 트랙에 저장된다. MPEG-4 파일 포맷의 트랙의 개수는 제안되어 있지 않으므로 음악을 구성하는 객체의 개수에 따라 유연하게 조절될 수 있다. 또한 오디오 객체 압축을 위해 사용되는 오디오 코덱은 MPEG-4에 속하는 고품질 고성능 오디오 코덱으로, 낮은 비트레이트의 오디오 코덱뿐만 아니라 무손실 오디오 코덱을 사용하여 최상의 오디오 품질의 제공이 가능하다. 실감 객체 오디오의 장면묘사 정보는 바이너리 부호화가 이루어진 후 트랙에 독립적으로 저장되며, 장면묘사의 종류에 따라 여러개의 장면묘사 트랙을 포함한다. 각 장면 묘사 비트스트림은 독립적인 트랙에 저장되어 사용자가 선택한 장면 묘사 정보에 따라 객체 오디오에 실감 효과를 적용하게 된다.

V. 사용자 정보기반 실감 오디오 장면 묘사 기법

본 장에서는 제안된 실감 객체오디오 신호를 위한 장면 묘사 기법에 대해서 설명한다. 제안된 방법은 객체 오디오를 지원할 수 있도록 구성되어 있으며, 휴대용 멀티미디어 기기에 적용 가능하도록 가볍게 설계되었

```

<ELEMENT SD (SDID, Duration, Num_Obj, Num_SE, Effect*,
  Object-, Mic_Config*, Record_Env*, Reserved*) >
  <ELEMENT SDID (#PCDATA) >
  <ELEMENT Duration (#PCDATA) >
  <ELEMENT Num_Obj (#PCDATA) >
  <ELEMENT Num_SE (#PCDATA) >
  <ELEMENT Effect
    (Effect_ID, Effect_Name, Effect_StartTime,
    Effect_EndTime, Effect_Info) >
  <ELEMENT Effect_ID (#PCDATA) >
  <ELEMENT Effect_Name (#PCDATA) >
  <ELEMENT Effect_StartTime (#PCDATA) >
  <ELEMENT Effect_EndTime (#PCDATA) >
  <ELEMENT Effect_Info (Effect_SL|Effect_VR|Effect_Ext|Effect_BS) >
  <ELEMENT Effect_SL (SL_NumofChannel, SL_Azimuth,
    SL_Distance, SL_Elevation, SLSpkAngle) >
  ...
  <ELEMENT Effect_VR (VR_PreDefined_Enable, (VR_RoomIdx|VR_MicPos),
    (VR_ReflectCoeff|VR_RoomSize), VR_SourcePos?,
    VR_ReflectOrder?, VR_ReflectCoeff?) >
  ...
  <ELEMENT Effect_Ext (Ext_NumofChannel,
    Ext_Angle, Ext_Distance, Ext_Spkangle) >
  ...
  <ELEMENT Effect_BS (BS_File0.3, BS_File2, BS_File3, BS_File4, BG_Index) >
  <ELEMENT Object (Obj_ID, Obj_File, Obj_Name, Obj_Seg, Obj_StartTime+,
    Obj_EndTime+, Obj_NumEffect, Obj_MixRatio, Effect*) >
  <ELEMENT Obj_ID (#PCDATA) >
  <ELEMENT Obj_File (#PCDATA) >
  <ELEMENT Obj_Name (#PCDATA) >
  <ELEMENT Obj_Seg (#PCDATA) >
  <ELEMENT Obj_StartTime (#PCDATA) >
  <ELEMENT Obj_EndTime (#PCDATA) >
  <ELEMENT Obj_NumEffect (#PCDATA) >
  <ELEMENT Obj_MixRatio (#PCDATA) >
  <ELEMENT Mic_Config (Layer, Layer_1*, Layer_2*, Layer_3*) >
  ...
  <ELEMENT Record_Env (Ref_Coeff, Space) >
  ...
  <ELEMENT Reserved (#PCDATA) >
    
```

그림 5. 제안된 오디오 장면 묘사를 위한 XML 형식
 Fig. 5. XML structure for proposed audio scene description.

다. 또한 사용자 정보 기반 오디오 효과 제공이 가능하도록 노드 구조로 설계되었다. 설계된 실감 객체오디오 장면 묘사의 용의함을 위해 그림 5와 같이 XML (extensible markup language) 형식으로 작성 되었다.

1. 저장 파일 포맷을 위한 장면 묘사 기법

실감 객체 오디오 기술이 기존의 휴대 멀티미디어 플레이어에 적용되기 위하여, 저장 파일 포맷을 위한 장면 묘사가 가능하다. 저장 장치를 위하여 장면 묘사 정보가 매번 트랙의 샘플에 포함 되어 있지 않고, 트랙의 첫 번째 샘플에 장면 묘사 정보가 모두 포함된다.

최상의 노드, 장면 묘사 노드(scene description node)는 복수개의 SD(scene description)가 가능한 구조를 갖고 있다. 그림 6에 보이는 바와 같아 각 SD를 식별하고 복수개의 SD가 가능하도록 하기 위하여 SD ID (identification) 필드를 갖고 있다. 이를 통해 현재 포함된 SD 개수 및 적용되고 있는 SD의 종류등을 분석하는데 사용된다. 새로운 SD가 추가 될 경우 ID가 부여 되고, 파일 포맷의 새로운 트랙에 저장된다. 또한 SD에는 SD의 지속 시간(scene duration) 필드가 포함되어 있어서 SD의 총 길이를 알 수 있으며, 현재 SD에 포함되어 있는 객체 수를 나타내는 Num_Obj 필드, SD에 적

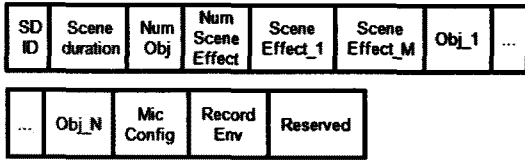


그림 6. 장면 묘사 노드 상세 구조도
Fig. 6. Structure of full scene description.



그림 7. 객체 묘사 노드 상세 구조도
Fig. 7. Structure of object scene description.

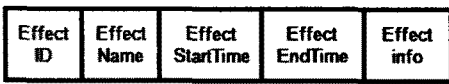


그림 8. 오디오 효과 묘사 노드 상세 구조도
Fig. 8. Structure of audio effect description.

용되는 오디오 효과의 개수 Num_SE 필드, 선언된 오디오 효과 개수 만큼 오디오 효과를 표현하는 Scene Effect 필드, SD에 포함된 객체 수만큼 오디오 객체 정보를 포함하는 Obj 필드, 녹음에 사용된 마이크 구성을 나타내는 Mic Config 필드, 녹음환경 정보를 포함하는 Record Env 필드로 구성된다.

두 번째 노드는 각 객체에 대한 묘사를 담은 객체 묘사 노드는 SD 노드에서 언급한 개수의 필드를 갖는다. 객체 노드는 그림 7과 같이 각 객체를 나타내는 Obj ID 필드, 객체의 이름을 나타내는 Obj Name 필드, 객체를 구성하는 세그먼트의 개수를 나타내는 Obj Seg 필드, 객체의 재생 시간을 나타내는 Obj Time 필드, 객체에 적용되는 효과의 개수 Obj NumEffect 필드, 오디오 믹싱 비율을 포함하는 Obj MixRatio 필드, 그리고 객체에 적용될 오디오 효과 개수 만큼의 오디오 효과 정보 필드로 구성된다.

마지막 노드인 오디오 효과 노드는 그림 8과 같이 구성된다. 오디오에 적용되는 효과의 종류를 구분하기 위한 오디오 Effect ID 필드, 적용되는 효과의 이름을 포함하는 Effect Name 필드, 효과의 시작 시간을 나타내는 Effect StartTime 필드, 효과의 종료를 나타내는 Effect EndTime, 그리고 각 효과의 세부 정보를 포함하는 Effect Info 필드로 구성된다.

2. 전송용 오디오를 위한 장면 묘사 기법

실감 객체 오디오를 IPTV와 같은 실시간 방송에 적용하기 위하여 논 논문에서는 실시간 전송을 위한 실감

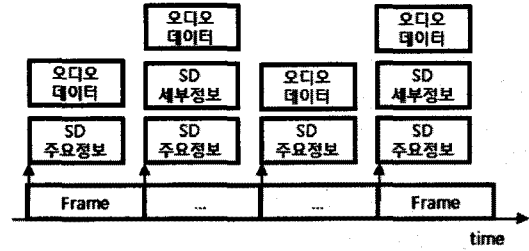


그림 9. 전송용 장면 묘사를 위한 기본 구조
Fig. 9. Scene description structure for transmission media.

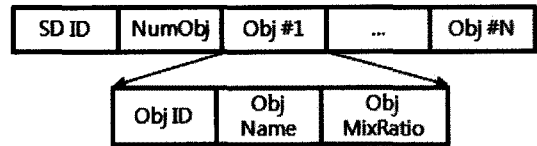


그림 10. 전송용 SD 주요 정보 노드 상세 구조도
Fig. 10. Structure of main scene description for transmission media.

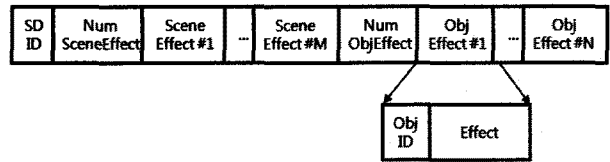


그림 11. 전송용 SD 세부 정보 노드 상세 구조도
Fig. 11. Structure of specific scene description for transmission media.

객체 오디오 장면 묘사 기법을 제안한다. 제안된 방법은 장면 묘사를 위해 매 프레임 마다 전송되는 데이터 레이트를 감소시키기 위하여 그림 9와 같이 매번 전송되어야 하는 SD 주요 묘사 정보와 필요한 경우에만 전송되는 SD 세부 묘사 정보로 구분하여 설계하였다.

최상위 노드는 SD 주요 정보 노드으로써 그림 10과 같은 구조를 갖는다. SD 주요 정보는 복수개의 SD를 가질 수 있으며 SD ID에 의해 구분된다. 또한 SD 주요 정보는 객체의 개수, 객체 개수만큼의 객체에 대한 기본 정보 및 객체의 믹싱 정보를 포함하여, 주요 정보 노드만으로 객체 오디오의 기본적인 동작이 가능하다. SD 세부 묘사 정보는 오디오에 실감을 제공하기 위해 요구되는 정보들로 그림 11과 같이 설계되었다. SD 세부 묘사 정보 노드는 SD ID를 통해 여러 개의 SD를 구분하며, 현재 프레임의 장면에 적용될 효과의 개수와 효과 개수만큼의 효과정보를 포함한 효과 필드를 갖고 있다. 효과 필드는 그림 8과 같은 구조를 사용한다. 또한 SD 세부 묘사 정보에는 각 객체에 적용될 효과의 수 필드와 효과 개수만큼의 객체 효과 필드로 구성된

다. 객체 효과 필드는 객체 ID를 갖고 있으므로 적용될 객체를 구분할 수 있으며, 효과 필드는 그림 8과 같은 구조를 갖고 있다.

VI. 시뮬레이터 개발

그림 12는 실감 객체오디오 파일 포맷을 사용하는 실감 객체오디오 인코더 시뮬레이터이다. 실감 객체오디오를 위한 오디오 장면 묘사 정보를 포함한 XML 파일과 객체 오디오 음원 정보를 갖고 있는 객체오디오 파일(*.oas)이 입력되고, 실감 객체오디오 파일은 MPEG-4 파일 포맷으로 출력된다. 실감 객체 오디오 파일 생성 시 장면묘사 정보 XML은 분석되어 바이너리 부호화되며, 객체오디오 파일의 하나의 트랙으로 저장된다. 또한 실감 객체오디오 시뮬레이터는 사용자에게 원음 그대로의 최상위 오디오 품질을 제공하기 위하여 MPEG-4 ALS을 이용하여 각 오디오 객체를 부호화한다.

그림 13는 실감 객체오디오 파일 포맷 기반의 실감 객체오디오 재생 시뮬레이터이다. 재생 시뮬레이터에서는 MPEG-4 파일에서 각 객체의 오디오 비트스트림과 장면 묘사 정보를 분리하며, MPEG-4 파일의 객체 초기화 부분 정보를 통해서 현재 오디오 비트스트림의 트랙 개수와 압축에 사용된 코덱 종류, 샘플링 주파수, 채널수 정보를 획득 한다. 오디오 비트스트림은 본 시뮬레이터가 탑재하고 있는 MPEG-4 ALS에 의해 복호화되며, 장면 묘사 데이터는 바이너리 복호화가 이루어

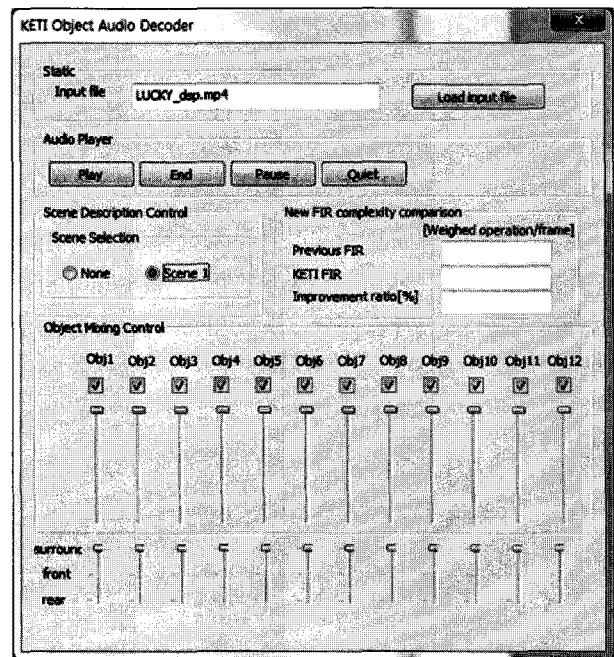


그림 13. 실감 객체오디오 재생기
Fig. 13. Audio player for real object audio.

진다. 본 시뮬레이터는 사용자가 객체를 선택적으로 청취할 수 있으며, 선택된 객체의 믹싱 비율을 조절할 수 있다. 또한 오디오 장면 묘사 정보를 사용자의 선택에 따라 선택적으로 적용할 수 있다.

VII. 결 론

본 논문에서 실감 객체오디오를 위한 파일 포맷을 설계하였으며, 실감형 오디오를 위한 사용자 정보기반의 오디오 장면 묘사 기술을 제안하였다. 제안된 객체오디오 파일 포맷은 여러개의 트랙 구성이 용이하고, 다양한 고품질 MPEG-4 오디오 코덱 사용이 편리한 MPEG-4 파일 포맷을 기반으로 설계되었다. 설계된 파일 포맷은 기존의 오디오 재생 시스템이 가능할 뿐만 아니라, 사용자의 요구 및 사용자의 환경에 따라 객체 선택, 객체 믹싱 조절, 3차원 오디오 효과를 통한 실감 증대 등이 가능하여 오디오에 대한 몰입이 높다. 또한 제안된 장면 묘사 기술은 저장용을 위한 방법과 전송용을 위한 두 가지 방법이 모두 개발되어 IPTV 및 휴대용 멀티미디어 기기에서 사용이 가능하다. 설계된 실감 객체오디오 파일 포맷 및 장면 묘사 기법을 검증하기 위하여 실감 객체 오디오 시뮬레이터를 개발하였다. 개발된 시뮬레이터는 장면 묘사 정보와 여러 객체를 부호화하여 하나의 MPEG-4 파일로 생성하며,

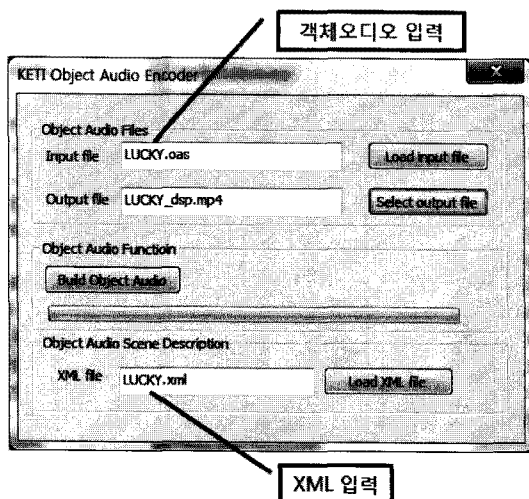


그림 12. 실감 객체 오디오 파일 포맷 생성기
Fig. 12. File format generator for real object audio.

생성된 MPEG-4 파일은 실감 객체오디오 재생 시플레이터를 통해 각 객체가 포함하고 있는 오디오 기본 정보와 오디오 장면 묘사가 복호화 되어 실감 객체 오디오를 재생한다. 또한 사용자의 객체 선택 및 오디오 믹싱 컨트롤에 의하여 실감 객체 오디오를 사용자의 요구에 따라 재구성 한다. 설계된 장면 묘사 방법은 노드 구조로 가벼우며, 저장용과 전송용 모두 설계가 이루어졌다. 그러므로 설계된 포맷 기술은 IPTV와 같은 최신 방송 서비스, 휴대용 멀티미디어 기기 등 다양한 분야에 쉽게 적용가능하다.

참 고 문 헌

- [1] T. Ricker, Zune vs. iPod specification smackdown, available at <http://www.engadget.com/2007/10/03/zune-vs-ipod-specification-smackdown/>, Oct. 2007.
- [2] 차지훈, 이인재, 박상택, 문경애, 홍진우, “모바일 리치미디어 방송 기술,” *전자통신동향분석*, 제 23권, 제3호, pp. 96-105, 2008년 6월.
- [3] Audizen.com, “*Music 2.0 demonstration*,” available at <http://www.audizen.co.kr/product/product2.htm>, 2008.
- [4] iKlax Media “*Audio Format*,” available at <http://www.iklaxmedia.com/article-26-iklax-technology.html>, May 2010.
- [5] ISO/IEC 23000-12, Information technology-Multimedia application format (MPEG-A)- Part 12, July, 2010.
- [6] C. S. Cho, J. W. Kim, H. S. Shin, B. H. Choi, “Implementation of an object audio system based on MPEG-4 audio lossless coding on DSP,” in *proc. of Broadband Multimedia Systems and Broadcasting(BMSB) 2010 IEEE International Symposium*, Mar. 2010.
- [7] Simeonov, Aleksandar, et al., “Advanced 3D Audio Algorithms by a Flexible, Low Level Application Programming Interface,” in *Proc. of 116st AES Convention*, May 2004.
- [8] M. Custer, “Multichannels room impulse response rendering on the basis of underdetermined data,” in *proc. of J. Audio Eng. Soc.*, Vol. 57, No. 6, June 2009.
- [9] G. Wersenyi, “Effect of emulated head-tracking for reducing localization error in virtual audio simulator,” in *proc. of IEEE Tran. on Audio, Speech and Language Processing*, Vol. 17, No. 2, Feb. 2009.
- [10] Paul A., “*Audyssey DSX 10.2 Surround Sound Overview*,” available at <http://www.audioholics.com/education/surround-sound/audyssey-dsx-10.2>, July 2009.
- [11] ITU-R BS., “*Multichannel sound technology in home and broadcasting applicaiton*,” June 2009.
- [12] ISO/IEC WD 23003-2, Information technology-coding of audio-visual object, Part1:MPEG Surround Spatial Audio Object Coding, Oct. 2007.
- [13] Jonas E., Barbara R., Cornelia F., et al., “Spatial Audio Object Coding(SAOC)-The Upcoming MPEG Standard on Parametric Object Based Audio Coding,” in *Proc. of 124st AES Convention*, May 2008.
- [14] ISO/IEC 14496-12, Information technology-Coding of audio-visual objects, Part12: ISO base media file format, Oct. 2008.
- [15] ISO/IEC 14496-3:2005/Amr 2 : 2006, Information technology-Coding of audio-visual object, Part3: Audio Lossless Coding(ALS), new audio profiles and BSAC extensions, Mar. 2006.
- [16] ISO/IEC 14496-3:2005/Amr 2 : 2006, Information technology-Coding of audio-visual object, Part3:Scalable Lossless Coding(SLS), June 2006.

저 자 소 개



조 충 상(정회원)-교신저자
 2006년 수원대학교 전자공학과
 학사.
 2008년 광주과학기술원 정보통신
 공학과 석사.
 2010년 현재 전자부품연구원 멀티
 미디어IP연구센터
 전임연구원

<주관심분야 : 오디오/음성 압축, 다채널 오디오
 신호처리, 오디오 신호처리>



김 제 우(정회원)
 1997년 서울시립대학교 제어계측
 공학과 학사.
 1999년 서울시립대학교 제어계측
 공학과 석사.
 2010년 현재 전자부품연구원 멀티
 미디어IP연구센터
 책임연구원

<주관심분야 : A/V 신호처리, 멀티미디어 코덱,
 코덱 알고리즘 및 최적화>