대화형 3차원 오디오 방송단말 구현

박기윤, 이태진, 강경옥
한국전자통신연구원 디지털방송연구단 방송미디어연구그룹
gypark@etri.re.kr, tjlee@etri.re.kr, kokang@etri.re.kr

An Implementation of Interactive 3D Audio Broadcasting Terminal

Gi Yoon Park, Taejin Lee, Kyeongok Kang
Broadcasting Media Research Group at ETRI

Abstract

본 논문에서는 사용자의 입력에 따라 3차원 오디오 장면을 재구성하여 전달할 수 있는 대화형 오디오 방송단말의 구현 예를 제시한다. MPEG-4 AudioBIFS 규격에 따라 계층적으로 표현한 오디오 장면의 속성을 사용자의 입력에 따라 갱신하고, 주어진 속성을 참조하여 오디오 데이터를 3차원 공간 상에 재합성하는 방식을 취한다. 속성을 갱신하는 모듈은 MPEG-4 Audio 프로파일을 지원하게 하되 AudioBIFS 노드 유형에 따른 사용자 인터페이스를 미리 정의하여 단말 측에 저장해 두고 이용함으로써 대화형 방송 서비스를 구현했다. 3차원 오디오 데이터를 재생하는 기능은 사용자의 입력에 대한 피드백을 풍부하게 하여 대화형 방송의 효과를 극대화하고, 사실감을 제고하는 데 중요한 역할을 담당한다. 요소 기술로 음상의 위치, 지향성, 모양, 잔향특성 등을 구현하기 위한 3차원 오디오 기술에 대해 소개한다. 또한 대화형 3 차원 오디오 방송단말을 이용한 서비스의 예로 대화형 합주 및 합창 프로그램을 소개한다.

1. 개요

시청각 콘텐츠를 일방적으로 다수의 시청자에게 송출하던 기존 방송시스템은 다양해진 시청자의 욕구를 충족하지 못하여 대화형 방송의 필요성이 제기되었다. 근래에 방송시스템의 전송 모듈이 확장하여 인터넷 등의 양방향 채널을 포함하는 소위 방송통신융합 현상으로 인해 대화형 방송을 위한 인프라가 확충되었다.

더불어 MPEG-4 를 중심으로 활발하게 표준화가 진행되고 있는 객체지향 혹은 객체기반 시스템 기술을 방송분야에 도입하여 대화형 방송시스템을 구현하려는 노력이 세계 각국에서 이루어졌다. 본 연구팀에서 MPEG-4 시스템에 기반을 둔 방송시스템에 대한 연구를 수행한 바 있으며[1], 일본 NHK 에서도 대화형 방송서비스로서 유아교육용 숨바꼭질 프로그램 등을 선보이기도 했다. 이들 시스템은 사용자의 의도에 따라 콘텐츠를 갱신해 줌으로써 콘텐츠 내에서 시청자의 자유도를 증대하는 효과를 거두었으며, 특히 오디오 콘텐츠의 경우 1960 년대부터 꾸준히 연구가 진행되어 온 입체음향 기술과 상승작용을 일으켜 현실감의 극대화에도 공헌하였다. 3D 오디오 기술은 사용자의 입력에 따라 오디오 장면의 공간적 구조를 재합성할 수 있는 수단을 제공함으로써 사용자 제어의 효과를 제고한다.

대화형 서비스를 구성하는 요소기술로서 MPEG-4 AudioBIFS 규격과 3차원 오디오 기술에 대해 2장과 3장에서 각각 소개하고, 방송단말의 구조 및 사용자 인터페이스에 대해 4장에서 설명한다. 대화형 3차원 오디오 방송 단말을 이용한 서비스의 예를 5장에서 소개하고 6장에서 결론을 맺는다.

2. MPEG-4 AudioBIFS

MPEG-4 시스템은 자연/합성 AV 스트림에 객체기술자(object descriptor)를 붙여 객체화하고, AV 객체가 장면(scene)을 구성하는 구조를 별도로 기술하게 한다. 부호화하고자 하는 대상을 객체 단위로 분할하고 객체의 속성에 부호화기를 선정함으로써 압축률을 높임과 동시에, 사용자의 입력을 참조하여 객체 별로 서로 다른 효과를 줄 수 있다.

객체를 물리적 시공간에 배치하기 위해서는 객체 자체에 대한 정보 이외에 별도로 장면정보가 필요하다. 장면정보는 그래프로 표현되는 계층구조를 가진다. 각 노드(node)에는 유형에 따라 미리 정의한 필드가[2] 하위 구조로서 속하며, 특정 유형의 필드는 노드를 값으로 취할 수 있어 한 노드가 다른 노드에 하위구조로 포함되는 관계를 에지(edge)로 표현할 수 있으며, 특정 유형의 필드를 이용해 객체기술자를 참조함으로써 해당 객체와 장면 그래프를 연결할 수 있다.

장면그래프 중 오디오 노드로 이루어진 오디오 서브그래프(sub-graph)는 비디오 객체의 처리와 별도로 오디오 객체를 처리하는 과정을 SFG (signal-flow graph) 스타일로 표현할 수 있게 하였다 (그림 1 참조).

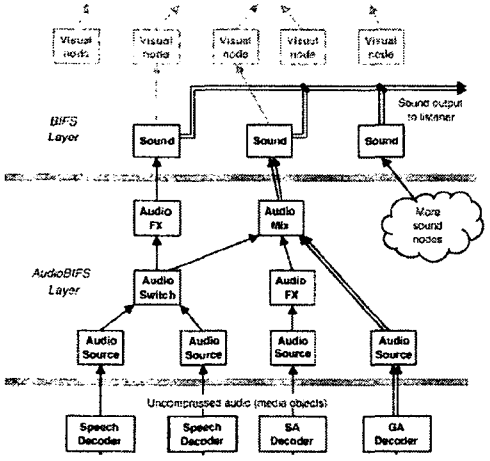

그림 1. MPEG-4 오디오 장면 합성의 예

오디오 서브그래프는 delay, mixing, switch 등 물리적 의미와 상관없이 산술적인 연산을 표현하는 부분과 가상현실 효과를 나타내는 부분으로 나뉜다. 본 논문의 구현대상으로 후자에 속하는 노드 인터페이스를 나열하면 표 1과 같다.

표 1 오디오 가상현실 효과를 나타내는 노드 인터페이스

Node type	Purpose
Sound	Sound spatialization, elliptic directivity model,
DirectiveSound	Spatialization, linear filter directivity model, reverberation
WideSound	Spatialization, sound shape, reverberation
AcousticScene	Physical reverberation model
PerceptualParameters	Perceptual reverberation model

Sound 노드는 음상을 정위하기 위한 인터페이스를 제공하며, DirectiveSound 노드와 WideSound 노드는 각각 오디오 객체의 방향성과 모양을 정의하기 위해 확장한 인터페이스를 제공한다. 더욱이, AcousticScene 노드 인터페이스에 따른 잔향모델 외에도 PerceptualParameters 노드를 하위구조로 포함하여 소리의 밝기, 온도 등 인지적 특성에 따른 잔향 특성을 객체 별로 설정할 수 있다.

3. 3차원 오디오 기술

AudioBIFS 노드 인터페이스에 따른 오디오 랜더링 기능을 지원하기 위해 사용한 기술에 대해 소개한다.

3.1. 음상정위

음원부터 청취자의 귀까지 이어지는 채널을 방위각(azimuth angle)에 따른 선형 시스템으로 모델링한 HRTF (head-related transfer function)를 이용해 음상의 방향을 합성하는 방법을[3] 취했다. 소리가 청취자에게 전달되는 중에 일어나는 반사, 회절 등의 효과를 모델링하기 위해 가슴과 머리, 귓바퀴를 포함하는 상반신 모형을 이용해 채널의 충격응답을 측정했다.

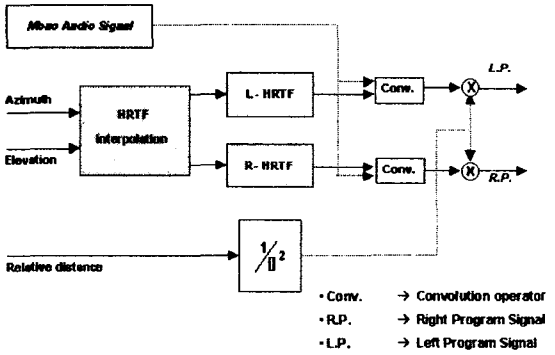

그림 2 음상정위 시스템의 블록 다이어그램

HRTF를 입력 모노 신호에 입힌 후, 음원과 청취자 사이의 거리에 따라 음압이 감쇄하는 현상을 역제곱의 법칙(inverse-square law)으로 모델링하여 음압을 조절함으로써 음상을 정위하였다 (그림 2 참조).

3.2. 음원확장

점음원이 모여 면적이나 체적을 지닌 음원을 이루는 현상을 재현하기 위한 기술이다. 주어진 음원을 서로 인접한 위치에 합성하여 재생하되 소리의 확장감을 제고하기 위해 음상정위 이전에 각 음원 사이의 상관도(correlation)를 떨어뜨린다[4].

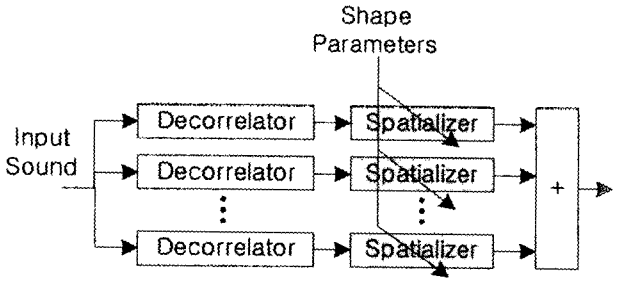

그림 3 음원모양 재현 시스템

입방체, 타원기둥 등 음원의 모양에 따라 모양 파라메터를 음상정위 모듈에(3.1 참조) 인가하고 점음원을 그림 3의 시스템으로 처리함으로써 선/면/체적 음원을 합성할 수 있다.

4. 대화형 오디오 방송단말

사용자 인터페이스를 포함한 오디오 방송단말을 구현한 절차에 대해 소개한다.

4.1. 구조

객체지향 프로그래밍 언어로 단말 시스템을 구현하기 위한 데이터 구조와 처리흐름에 대해서 설명한다.

장면 그래프의 각 노드와 객체기술자는 각각 MPEG-4 시스템 규격에 명시한 노드 필드와 비트 필드로 이루어진 구조체로 선언하여 BIFS/OD 스트림을 파싱(parsing)하기 위한 틀(templates)의 기능을 수행하며, 오디오 장면을 합성하기 위한 노드 및 객체기술자 별 신호처리 모듈을 포함한다.

재생단말은 콘텐츠로부터 오디오 스트림과 장면 그래프 등을 추출하기 위한 DEMUX, 그리고 장면을 합성 및 출력하기 위한 Presenter를 생성하여 재생작업을 준비한다 (

그림 4 참조).

IOD 복호기가 DEMUX로부터 IOD 스트림을 넘겨받아 BIFS, OD 복호기를 초기화하면, BIFS 복호기는 장면 그래프를 복원하여 Presenter에게 전달하고, OD 복호기는 AAC 복호기를 통해 raw data를 생성하고 장면 그래프 상의 오디오 입력 노드에 연결한다. Audio renderer는 Presenter로부터 받은 audio subgraph를 참조하여 AAC decoders의 출력을 처리하여 오디오 장면을 합성한다.

4.2. 사용자 인터페이스

장면그래프를 사용자의 입력에 따라 갱신함으로써 대화형 오디오 재생환경을 구현할 수 있다.

MPEG-4 시스템 규격에 따르면 스크립트, 센서 노드, Route 메커니즘 등을 이용해 사용자 인터페이스를 장면 그래프에 포함할 수 있지만, 인터페이스 데이터를 복호하기 위한 계산 상의 부담과 함께 복합적인 인터페이스를 표현함에 한계가 있다. 이에 본 재생단말에서는 버튼, 리스트 박스 등을 통한 GUI (graphical user interface)를 AudioBIFS의 각 노드에 따라 정의해 두고 호출하여 사용하는 방법을 취했다. 또한 방송환경에 적용하기 위해 입력도구로서 TV 리모콘 등 간단하고 직관적인 형태를 고려하였다.

BIFS 스트림을 복호할 때 Sound, DirectiveSound,

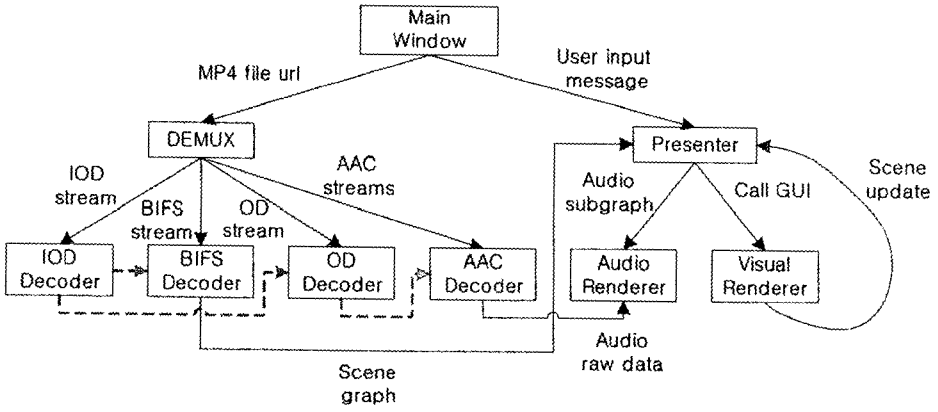

그림 4 재생단말의 데이터 흐름

WideSound 등 AudioBIFS 최상위 노드의 목록을 저장해 둔다. 사용자가 객체 목록을 호출하고 방향키 등을 이용해 객체를 선택하면, 목록을 지우고 노드 타입에 따른 GUI를 화면에 생성하여 사용자의 입력을 받아들여 장면 그래프를 갱신한다.

예를 들어 WideSound 노드 타입의 객체를 선택한 경우 Presenter 를 통해 그림 5 와 같은 GUI 를 호출하여 화면에 출력한다. 사용자가 리스트 박스를 통해 오디오 객체의 모양으로 shuck, box, ellipsoid, cylinder 중 하나를 선택하면 해당 필드 값을 갱신하고, 그에 따른 체적음원의 모양을 우측에 나타낸다.

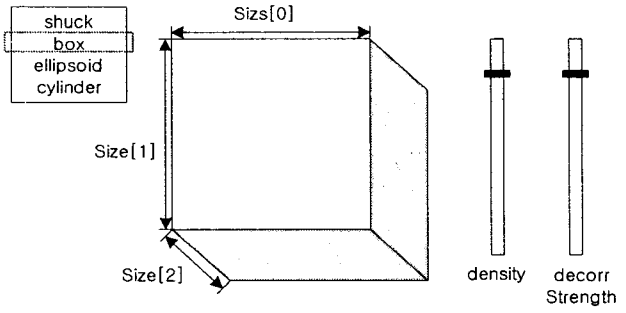

그림 5 WideSound 노드의 GUI

체적음원의 크기, 농도 등 실수 타입의 필드는 슬라이드 바 등을 이용해 갱신할 수 있게 하고, 화면에 나타난 GUI 에 실시간으로 반영한다.

5. 서비스 예

대화형 합주 및 합창 방송 서비스 시나리오를 소개한다. 합주에 참여한 각 악기소리에 물리적 특성, 연주자에 대한 정보 등의 기술자(descriptor)을 부가하여 오디오 객체를 구성한다.

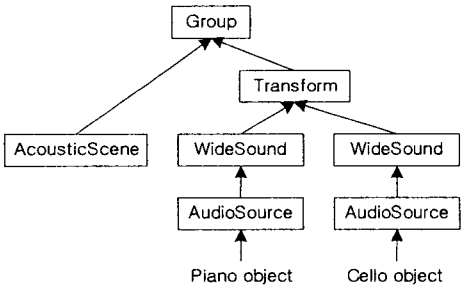

그림 6 합주장면 그래프

주어진 악기 소리의 음감을 확장하여 연주자의 수를 늘리는 효과를 기대하여 WideSound 노드를 통해 악기 객체를 장면그래프에 입력한다. 또한 공연장의 크기, 벽면의 재질 등에 따른 잔향 특성을 제어할 수 있도록 AcousticScene 노드를 장면그래프에 포함하여 (그림 6 참조) 콘텐츠를 제작한다. 재생단말에서는 장면그래프를 복원하여 초기 장면을 구성하고, 사용자의 입력에 따라 연주자 또는 가수의 수자 및 배치와 공연장의 잔향특성 등을 갱신함으로써 대화형 합주 또는 합창 장면을 재현할 수 있다.

6. 결론

MPEG-4 시스템 규격을 참조하여 대화형 오디오 재생 시스템을 구현하되, 방송 환경에 활용하기 용이하도록 사용자 인터페이스를 재생 단말에서 미리 정의하여 두고 사용자의 입력에 따라 호출하여 활용하게 하였다.

또한 대화형 오디오 방송의 효과를 제고하기 위한 3 차원 오디오 기술로 음상정위, 음감확장 기술 등에 대해 소개하였다.

7. 감사의 글

본 논문은 정보통신부의 연구사업인 "지능형 통합정보방송 기술개발" 과제의 일환으로 수행한 결과로서 정보통신부 당담자 및 관련 연구원들의 노력에 감사를 드립니다.

참고문헌

[1] Jeongil Seo, Gi Yoon Park, Dae-Young Jang, and Kyeongok Kang, "Implementation of Interactive 3D Audio Using MPEG-4 Multimedia Standards," *Proc. 115th AES Convention*, October 2003.

[2] ISO/IEC 14496-1, *"Information technology – Coding of audio-visual objects Part 1: Systems,"* 2002.

[3] Richard O. Duda, "Modeling Head Related Transfer Functions," *Proc. Conf. Signals, Systems and Computers*, vol. 2, pp. 996-1000, 1993.

[4] Gary S. Kendall, "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery," *Computer Music Journal*, vol. 19, no. 4, pp, 71-87, Winter, 1995.

[5] ISO/IEC 14472-1, "Virtual Reality Modeling Language," 1997.