한국음향학회지 제22권 제7호 pp. 592~602 (2003)

객체기반 3차원 오디오 방송 시스템 설계

The Design of Object-based 3D Audio Broadcasting System

강 경 옥*, 장 대 영*, 서 정 일*, 정 대 권**
(Kyeong-Ok Kang*, Dae-Young Jang*, Jeongil Seo*, Dae-Gwon Jeong**)

*한국전자통신연구원 방송미디어연구부, **한국항공대학교 항공전자공학과
(접수일자: 2003년 7월 8일; 채택일자: 2003년 9월 18일)

본 논문은 오디오를 동반한 방송 서비스에 있어서 기존의 단순 청취형의 오디오 서비스에서 탈피한 대화형의 객체 기반 3차원 오디오 방송 시스템의 구조를 설명한다. 객체기반 3차원 오디오 방송 시스템은 3차원 오디오 입력부, 3차원 오디오 편집/제작부, 3차원 오디오 부호화부, 3차원 오디오 복호화부, 3차원 오디오 장면합성부 및 3차원 오디오 재생부로 구성된다. 오디오 입력부에서는 3차원 배경음 객체와 독립적인 오디오 객체들을 획득한다. 편집/제작부에서는 오디오 객체들에 대한 3차원 음상정위 및 오디오 이미지 생성을 위한 파라미터를 설정하고, 이들을 조합하여 3차원 오디오 장면을 편집/제작한다. 부호화부에서는 장면정보와 오디오 객체들을 부호화하고, 복호화부에서는 오디오 객체들을 복원하고 장면정보를 획득한다. 장면 합성부에서는 장면정보와 오디오 객체들을 이용하여 오디오 장면을 구성한다. 3차원 오디오 재생부에서는 3차원 오디오 객체들재생하고, 사용자의 제어신호를 이용한 대화형 기능을 구현한다.
핵심용어: 오디오 방송 시스템, 객체기반, 3차원 오디오, MPEG-4, 오디오 장면 기술자, 대화형 방송
투고분야: 뉴미디어 분야 (13.1)

This paper aims to describe the basic structure of novel object-based 3D audio broadcasting system. To overcome current uni-directional audio broadcasting services, the object-based 3D audio broadcasting system is designed for providing the ability to interact with important audio objects as well as realistic 3D effects based on the MPEG-4 standard. The system is composed of 6 sub-modules. The audio input module collects the background sound object, which is recored by 3D microphone, and audio objects, which are recorded by monaural microphone or extracted through source separation method. The sound scene authoring module edits the 3D information of audio objects such as acoustical characteristics, location, directivity and etc. It also defines the final sound scene with a 3D background sound, which is intended to be delievered to a receiving terminal by producer. The encoder module encodes scene descriptors and audio objects for effective transmission. The decoder module extracts scene descriptors and audio objects from decoding received bistreams. The sound scene composition module reconstructs the 3D sound scene with scene descriptors and audio objects. The 3D sound renderer module maximizes the 3D sound effects through adapting the final sound to the listner's acoustical environments. It also receives the user's controls on audio objects and sends them to the scene composition module for changing the sound scene.
Keywords: *Audio broadcasting system, Object oriented, 3D audio, MPEG-4, User interactivity, Sound scene descriptor*
ASK subject classification: *New media (13.1)*

I. 서 론

지난 십여 년간 MPEG (Moving Picture Experts Group)에서는 디지털 멀티미디어의 부호화를 위한 표준 제정에 심혈을 기울여 왔고, 이를 통하여 방송 및 통신 분야에서의 디지털 멀티미디어 서비스의 괄목할 만한 성장이 이루어졌다.

방송 매체의 디지털화 및 고품질화에 의해 고선명 TV 방송의 상용화가 이루어지고 있으며, 라디오 방송 또한 디지털 오디오 압축 기술에 의해 보다 향상된 품질과 서비스로서 유럽을 중심으로 DAB (Digital Audio Broadcasting)에 의해 상용화되어 점차 그 서비스 영역을 늘려가고 있다.

고선명 TV 방송에서는 MPEG-2 표준에 의해 5.1채널

책임저자: 강경옥 (kokang@etri.re.kr)
305-350 대전광역시 유성구 가정동 161
한국전자통신연구원 방송미디어연구부 방송컨텐츠응용연구팀
(전화: 042-860-5521; 팩스: 042-860-5479)

의 오디오 서비스가 가능하여 졌으며, 또한 부가 데이터의 전송에 의한 데이터 방송, 양방향 채널 혹은 대용량 저장 장치에 의한 대화형 방송, 메타데이터와 아카이빙 기술을 이용한 사용자 맞춤형 방송이 TV-Anytime 규격 등을 통하여 상용화가 급속히 진행되고 있다[1].

MPEG에서는 이러한 방송 기술의 발전 추세에 따라 대화형 방송 및 대화형 멀티미디어 매체의 효과를 최대한 실현하기 위해 1999년 MPEG-4 표준을 제정하였다. MPEG-4에서는 객체기반의 오디오 및 비디오 등 멀티미디어 데이터의 부호화 도구 및 이들 멀티미디어 데이터를 객체기반으로 장면 합성을 하기 위한 BIFS (BInary Format for Scene)라는 장면 기술언어를 표준화 하였다[2]. 이러한 객체기반의 멀티미디어 데이터를 활용할 경우 원격지에서 서로 각 객체에 대한 접근 및 변환이 자유로우며, 이에 의해 서로간의 데이터 및 정보 공유의 느낌을 극대화할 수 있다.

또한 MPEG-4에서는 3차원 멀티미디어 데이터를 처리하기 위해 3차원 오디오 및 비디오의 표현 방법을 BIFS에서 표준화 하였다. 3차원 오디오에 있어서는 모노 채널의 음원 객체에 대해 3차원 공간 상에서의 위치, 공간의 물리적/청각적 특성 등을 기술하여 수신단에서 이를 이용하여 3차원 오디오 장면을 합성하여 재생할 수 있도록 하였다[2].

그러나 MPEG-4의 기본 개념이 실제 장면을 모든 객체로 세분화하는 것이 그 목적이므로 실제 환경에서의 오디오에 있어서 음원 모델이나 음장의 모델을 객체화하는 것에 한계가 있다.

3차원 오디오는 인간의 3차원 오디오 지각 능력을 활용하여 임의 공간의 음향적인 분위기를 가상적으로 체험할 수 있게 하며, 또한 각 오디오 객체 신호들이 가상적인 3차원 공간 상에 배치되므로 실제와 유사한 오디오 객체들에 대한 분별력을 가질 수 있게 한다.

3차원 오디오를 녹음 및 재생하는 기술은 오래 전부터 미국, 유럽, 일본 등에서 개발되어 왔으나, 재생 환경의 제약 때문에 상용화 단계에까지는 이르지는 못하고 있다. 그러나 최근 수년간의 DVD와 고선명 TV방송의 상용화로 인하여 멀티채널 오디오의 급속히 보급되고 있고, 3차원 오디오를 녹음 및 재생하는 기술의 지속적인 상용화 시도로 3차원 오디오는 새로운 오디오 방송 기술의 하나로 대두되고 있다.

본 논문에서는 3차원 오디오 처리 기술과 MPEG-4를 기반으로 하는 객체기반 멀티미디어 처리 기술을 이용하여, 객체기반 3차원 오디오 방송을 위한 시스템의 구조 및 서비스 구현을 위한 기반 기술들에 대하여 설명한다.

본 논문의 구성은 다음과 같다. 2장에서는 3차원 오디오 처리 기술에 대하여 설명하고, 3장에서는 MPEG-4 에서의 장면표현 기술과 오디오 부호화 기술에 대해서 설명하고, 4장에서는 객체기반 3차원 오디오 방송 시스템의 구조 및 특징에 대하여 기술하며, 끝으로 5장에서는 향후 과제 및 결론을 기술한다.

II. 3차원 오디오 처리 기술

3차원 오디오란 음원의 방향뿐만 아니라 녹음되는 장소의 음향학적 특성을 동일하게 녹음하고 재생하는 기술이다. 궁극적인 3차원 오디오를 재현하기 위해서는 3차원 오디오를 녹음하는 마이크로폰 기술로부터 스피커를 통하여 재생하는 기술까지 포함하는 통합적인 시스템을 구현하여야 한다[3]. 본 절에서는 3차원 오디오를 재현하기 위하여 필수적인 기술들을 3차원 오디오 획득 기술, 부호화 기술, 재생기술로 구분하여 설명한다.

2.1. 3차원 오디오 획득 기술

지난 수십년간 3차원 오디오를 획득하기 위하여 여러 가지 형태의 마이크로폰 구조가 제안되었다. 가장 대표적인 방법은 사람의 머리를 모델링한 더미헤드 마이크로폰과, 지향성 마이크로폰과 무지향성 마이크로폰의 조합을 통하여 3차원 오디오를 녹음하는 엠비소닉 (Ambisonic) 마이크로폰을 들 수 있다[4,5].

2.1.1. 더미헤드를 이용한 3차원 오디오 획득기술

더미헤드는 3차원 음장의 수음, 재생 및 평가에 이용되고 있다. 더미헤드는 사람의 머리 형상을 한 마네킹의 귀 위치에 마이크를 장착한 것이며, 1983년에 IEC에서는 보청기 측정을 위한 더미헤드를 심의하여, 몸통을 포함한 계측용 더미 헤드는 HATS (Head and Torso Simulator)라고 명명하였다.

더미헤드는 사람의 머리 형상을 모델링하였기 때문에, 더미헤드가 놓여진 곳의 3차원 오디오 음장 환경을 적절하게 녹음할 수 있다. 그러나 스피커를 이용하여 재생할 경우 좌우 스피커간의 간섭 (cross-talk)에 의하여 3차원 음장을 정확하게 표현하기 힘들다. 스피커 배치에 따라서 발생하는 크로스토크 효과를 줄이기 위하여, 스테레오 다이폴과 같은 특수한 스피커 배치 방법[6,7]이나 재생 환경을 모델링한 역필터를 설계하여 크로스토크 효과를

줄여주는 알고리즘들이 제안되었다[8,9].

2.1.2. 구형 더미헤드 마이크로폰

더미헤드는 사람의 머리 모양을 모델링하여 제작되기 때문에 형태가 크고 시각적으로 편하지 않으므로 실제 라이브 환경에서 사용하기에는 적절하지 않다. 따라서 사람의 머리형태를 구형으로 간략화시킨 구형 더미헤드가 많이 이용되고 있다.

일반적으로 구형 더미헤드 모델은 모든 주파수 영역에 대해 인간의 머리를 묘사할 수 없다고 알려져 있다. 이는 고주파에서 파형의 길이가 상대적으로 작으므로 몸통, 머리, 귀바퀴 등의 영향이 고주파 성분을 인식하는데 중요한 역할을 하기 때문이다.

음원의 방향은 좌우 귀에 도달하는 음파의 시간 차이 (Interaural Time Difference, ITD)와 세기 차이 (Interaural Level Difference, ILD)를 단서로 하여 지각한다. 일반적으로 저주파 대역에서는 ITD가 고주파 대역에서는 ILD 음원의 위치를 파악하는데 중요한 역할을 한다고 알려져 있다. 1.5 kHz 이하에서 측정된 구형 모델에서의 ITD 특성은 같은 지름을 가지는 인간의 머리와 비슷한 결과를 보이므로 실제 인간의 머리를 적절하게 묘사한다고 볼 수 있으나, 1.5 kHz 이상의 주파수 영역부터는 귀바퀴에 의한 소리의 스펙트럼 변경이 고려되지 못하므로 전후방 구별에 혼동을 미치고, 4~16 kHz 까지의 고주파 성분의 위치 결정에도 영향을 미친다[7].

라이브 공연 같은 생방송 환경에 사용하기 위하여 고안된 구형 더미헤드 마이크로폰은 저주파수 대역에서는 더미헤드와 비슷한 성능을 보이지만 고주파수 대역에서는 음원의 위치를 적절히 표현할 수 없으므로 이를 개선하기 위한 구형 더미헤드 마이크로폰 구조가 제안되고 있다.

2.1.3. 엠비소닉 마이크로폰

엠비소닉 마이크로폰은 파면 합성 기법 (Wave Field Synthesis)을 이용한 3차원 녹음 기법으로써, 마이크로폰 위치에서 녹음한 음장을 스피커로 동일하게 재생해낼 수 있으므로 현장감 있는 입체음향을 얻을 수 있다[10]. 엠비소닉 마이크로폰은 한 개의 무지향성 마이크로폰 (W)과 세 개의 지향성 마이크로폰 (X, Y, Z)을 조합하여 원 포인트 기법으로 3차원 음장을 녹음한다. 원 포인트 녹음 기법은 네 개의 마이크포폰이 한 지점에 배치되어야 하지만 실질적으로 불가능하기 때문에 이를 보정회로를 이용하여 해결한 마이크로폰을 음장 마이크로폰 (sound field microphone)이라고 부른다[11]. 음장 마이크폰 유닛의 구성과 처리과정을 그림 1에 나타내었다.

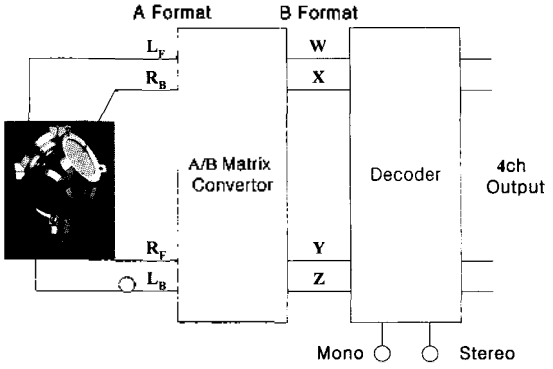

그림 1. 음장 마이크로폰과 A포맷, B포맷의 변환
Fig. 1. The conceptual model of soundfield microphone and A/B format.

음장 마이크로폰으로부터 얻어진 신호를 A포맷 신호라고 하고, A포맷 신호를 A/B 매트릭스 변환기로 W, X, Y, Z의 신호로 합성한 구성을 B포맷으로 하고 있으며 A포맷 신호와 B포맷 신호의 관계를 수식으로 나타내면 (1)과 같다.

$$
\begin{aligned}
W &= L_F + L_B + R_F + R_B \\
X &= L_F - L_B + R_F - R_B \\
Y &= L_F + L_B - R_F - R_B \\
Z &= L_F - L_B - R_F + R_B
\end{aligned} \tag{1}
$$

엠비소닉 마이크로폰은 세계적으로 수요가 적어 소량 생산되고 있고, 특성이 동일한 마이크로폰을 사용해야 하므로 가격이 비싼 단점이 있다. 그러나 마이크로폰 위치의 조절없이 인코더 상에서 거리조절, 음폭, 깊이감, 높이감 및 지향특성 등의 조절이 가능하다는 장점이 있다.

2.2. 3차원 오디오 부호화 기술

3차원 오디오 부호화 기술은 3차원 오디오 획득 기술을 이용하여 녹음된 오디오 신호를 객체별로 분리하는 기술, 분리된 객체를 압축하여 부호화하는 기술, 객체들을 조합하여 3차원 오디오 장면을 표현하는 기술로 나눌 수 있다.

2.2.1. 3차원 오디오 객체 분리 기술

오디오 부호화에서 사용되는 오디오 객체들이 녹음 과정에서 분리가 어려운 상태로 서로 혼합되어 있을 때는 별도의 분리과정이 필요하다. 예를 들면, 칵테일 파티장에는 여러 사람들의 대화소리와 음악 등이 섞여 있으며

경우에 따라 이들 소리들은 하나의 오디오 객체로 간주되기 보다는 분리한 후 각각을 독립된 오디오 객체들로 간주하여 다룰 필요가 있다. 또한 피할 수 없는 잡음들에 의해 음질이 심히 저하되는 경우 잡음들로부터 목표 오디오 객체를 효과적으로 분리해낼 필요도 있다.

3차원 오디오 객체분리는 시간적으로 혼합된 여러 개의 오디오 객체를 분리하는 것으로 혼합되기 전의 오디오 신호에 대한 정보없이 분리하므로 블라인드 신호 분리(blind source separation)라고 부른다[12,13,14]. 블라인드 신호 분리에 대한 개념은 그림 2를 통해 알 수 있다.

독립된 오디오 객체 신호 s는 혼합환경 $\boldsymbol{A}$에 의해 혼합되어 혼합신호 $\boldsymbol{x}=\boldsymbol{A}s$를 만들어낸다. 혼합된 신호 $\boldsymbol{x}$로부터 s를 분리해내기 위해서는 $u=Wx=A^{-1}x$를 만족하는 분리회로 $\boldsymbol{W}$가 필요하다. 여기서 독립된 오디오 객체 신호 s와 혼합회로 $\boldsymbol{A}$는 알려져 있지 않으므로 오직 관찰 가능한 혼합신호 $\boldsymbol{x}$만으로부터 분리회로를 예측해야만 한다. 이때 사용되는 가장 중요한 사실은 서로 다른 오디오 객체들이 서로 통계적으로 독립적 (statistically independent)이라는 점이며 이러한 이유로 신호분리 기술은 독립성분분석기술 (independent component analysis)의 하나로 분류된다.

오디오 객체 분리 기술은 오디오 객체의 혼합형태에 따라 크게 순시혼합 (instant mix)과 지연혼합 (delayed and convolved mix)된 경우로 구분된다. 예를 들어 칵테일 파티장에서 각 사람들의 대화 소리와 악기 음과 같은 소리들은 파티장의 벽면이나 다른 물체들에 닿은 뒤 반사되어 크기가 변한 다음 혼합되어 귀에 들리게 된다. 이 경우 지연혼합에 해당한다. 반면에 무향실 (anechoic chamber)과 같은 특수한 환경에서의 오디오 혼합은 순시혼합에 해당한다.

한편 오디오 객체 분리는 통계적 독립성을 활용하는 방법에 따라 고차 통계적 (higher order statistics, HOS) 방식, 정보량 최대화 방식, 중복 상관도 제거 (multiple decorrelation) 방식 등으로 구분된다. 이들 방식은 모두 순시혼합과 지연혼합에 적용 가능하다.

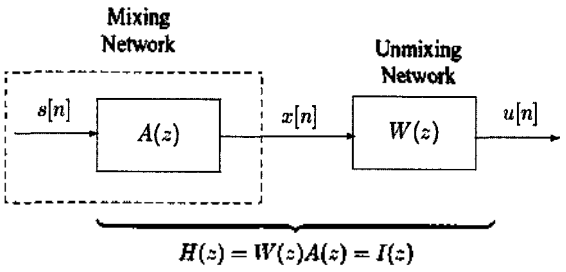

그림 2. 블라인드 신호 분리
Fig. 2. Blind source separation.

2.2.2. 오디오 객체 압축 기술

다수의 마이크로폰을 이용하여 객체별로 녹음하거나 2.2.1절에 설명한 방법으로 분리된 오디오 객체들을 보다 효율적으로 전송 또는 저장하기 위하여 디지털로 변환하여 압축하게 된다.

디지털 오디오 압축방식에는 인간의 지각특성을 이용하는 지각부호화 방식과 반복되는 데이터를 제거하는 엔트로피 부호화 방식이 있다. 국제 표준화 기구인 MPEG (Moving Picture Experts Group)에서는 상기의 두 가지 방식을 혼합한 MPEG-1, MPEG-2 BC, MPEG-2 AAC (Advanced Audio Coding), MPEG-4 GA (General Audio Coding) 기술을 표준화 하였고[15,16,17], Dolby, AT&T, Sony 등에서도 이와 유사한 디지털 오디오 압축방식을 사용하고 있다[19].

2.2.3. 3차원 오디오 장면표현 기술

오디오 비쥬얼 객체들을 시공간상에 배치되어 이루는 가상공간을 장면 (scene)이라고 부른다. 이러한 장면을 구조화된 언어로 기술 (description)한 것을 장면기술 (scene description)이라고 하며, MPEG-4와 같이 오디오 비쥬얼 데이터가 객체별로 부호화되어 단말에 전송될 때는 장면기술 정보에 근거하여 비쥬얼 데이터를 디스플레이하고 오디오 데이터를 재생하게 된다. 그러므로 객체기반의 멀티미디어 플랫폼에서는 각 객체별 압축/부호화 방식뿐만 아니라 이들을 조합한 인터렉티브 (interactive)한 장면을 어떻게 기술하는가에 대한 내용도 포함하고 있어야 한다. 위와 같은 3차원 대화형의 장면을 기술하기 위하여 국제 표준으로 채택되었거나 상용화된 방법은 VRML, X3D, SMIL, QuickTime, MPEG-4 BIFS, MPEG-4 XMT 등이 있다[20,21].

2.3. 3차원 오디오 재생 기술

3차원 오디오 재생 기술은 3차원 오디오 렌더링 기술, 3차원 오디오 출력 제어 기술 및 사용자 상호작용 기술 등의 요소기술로서 구분할 수 있다.

2.3.1. 3차원 오디오 렌더링 기술

객체기반 3차원 오디오 렌더링 기술은 각각 객체로서 전달된 오디오 신호들을 함께 전달된 속성에 의해 정확히 3차원 공간 상에서 표현될 수 있도록 3차원 오디오 효과

처리를 하는 기술을 의미한다.

3차원 오디오 효과로서는 음원의 3차원 공간상에서의 방향을 처리하는 음상 정위와 공간의 반사 및 잔향 특성을 모사하는 공간감 제어로서 구분할 수 있다. 임의 공간에서의 거리감은 주로 직접음대 잔향음 비율에 의해 정해지므로 공간감 특성이 잘 모사되면 거리감도 함께 제어할 수 있다.

음상 정위 기술로서는 머리전달함수를 이용하여 심리음향적인 음원의 방향을 제어하는 방법과 멀티채널 스피커에 의해 임의 방향으로부터의 음 혹은 음장을 재현하여 음원의 방향을 제어하는 물리음향적인 방법으로 구분할 수 있다. 현재 주로 사용되고 있는 것은 5.1 채널 스피커를 통해 전방 및 측, 후방의 음을 재현하는 전통적인 방법이 많이 사용되고 있다. 현재 잘 사용되고 있지는 않지만 머리전달함수를 이용하는 방법과 엠비소닉의 B포맷 신호에 의한 음상 정위 방법을 고려할 수 있는데, 머리전달함수를 이용하는 방법은 2개의 스피커만으로 대략적인 3차원 음향을 재현할 수 있으며, 엠비소닉에 의해 4채널 이상의 스피커로 재생할 경우 좀 더 완전한 3차원 오디오를 재생할 수 있으나, 스피커 배치를 위한 공간상의 제약이 따르고, 속삭임 등 근접 신호에 대해서는 표현이 어렵다는 단점이 있다.

공간감 제어 기술은 임의 공간의 구조 및 크기에 따른 공간 특성을 각 음상 정위된 음원 객체에 적용함으로써, 3차원 음상 정위 및 거리감 성능을 향상시키며, 자연스럽게 모든 오디오 객체들을 믹싱하기 위해 사용된다. 현재 대부분의 응용에서는 공간의 크기 및 음원의 거리에 따른 잔향 시간을 조절하는 것을 가장 중요한 요인으로서 처리하고 있지만, 공간의 구조 및 벽면의 재질 등도 중요한 요인이 된다. 또한 공기의 흡음효과 방해물에 의한 회절 및 차단 효과 등도 공간감 제어를 위해 제외될 수 없는 요인들이다.

2.3.2. 3차원 오디오 출력 제어 기술

3차원 오디오 출력 제어 기술은 3차원 음상 정위 및 공간감 제어 처리가 완료된 3차원 오디오 신호를 사용자에게 정확히 청취할 수 있도록 처리하는 모든 처리를 포함한다. 물리적인 처리로서는 청취실의 흡음 처리 및 스피커 배치를 포함하며, 전기음향적인 처리로서는 크로스토크 제거 및 재생환경 특성 등화 기술을 포함한다.

청취실의 흡음 처리는 벽면에는 흡음재 혹은 커튼 등으로 처리할 수 있으며, 바닥은 카펫으로 처리할 수 있다. 물리적인 처리는 시공 사양에 따라 비용이 많이 차이가 나며, 대부분의 가정에서는 이러한 투자를 하기 어려운 실정이다. 그러므로 일반적으로 가장 중요한 것은 스피커를 통하여 출력할 때 스피커의 배치를 정확히 하는 것이다. 이때 벽면 혹은 바닥면의 반사를 고려하여 청취실의 특성에 적합하게 스피커를 배치를 하는 것도 중요하다. 필요에 따라서는 불가피한 변칙 스피커 배치에 있어 이를 사용자가 보정할 수 있는 기능도 여러 곳에서 연구가 진행되고 있다.

바이노럴 재생에 있어 오른 쪽 신호는 오른쪽 귀에 왼쪽 신호는 왼쪽 귀에 정확히 전달되어야 하는데, 스피커 재생에 있어서는 공기를 통해 서로 반대편에 전달되므로 이를 제거하는 크로스토크 제거 필터가 필요하게 된다. 크로스토크 제거 기술은 크로스토크 발생경로 및 전달 특성을 측정하여 이 특성과 역의 특성을 가지는 필터를 전단에 통과시킴으로써, 청취자의 귀에 입력된 신호를 그대로 전달한다.

일반적으로 충실하게 오디오 신호를 재생하기 위해서는 음의 반사 및 잔향이 전혀 없는 무향실에서 재생하여야 하지만, 일반적으로는 어느 정도의 반사와 잔향을 포함하는 청취실에서 재생하게 된다. 이때 재생공간의 특성에 따른 반사, 잔향 특성과 재생 시스템의 특성을 등화하여 전기적인 신호가 음향적인 신호로 사용자의 귀에 전달될 때까지 신호의 변형이 없도록 하여야 한다. 크로스토크 제거 기술과 같이 필터링 기술로서 이 문제를 해결할 수 있다.

2.3.3. 사용자 상호작용 기술

사용자 상호작용 기술은 객체기반 3차원 오디오 장면을 사용자가 임의로 제어할 수 있도록 사용자 인터페이스 및 오디오 장면 제어를 하는 것이다. 장면제어에는 음원 객체의 위치제어 및 행위제어로 나눌 수 있으며, 사용자 인터페이스는 그래픽에 의한 GUI (Graphic User Interface) 및 청각에 의한 AUI (Auditory User Interface)로 구분할 수 있다.

음원 객체의 위치 제어에 있어서는 객체기반 3차원 오디오 렌더링 이전에 전송 혹은 저장된 음원 객체의 장면 정보를 변경함으로써 이루어질 수 있다. 또한 음원 객체의 행위 제어는 음원의 운동 제어, 선 음원, 면 음원 및 체적 음원의 방향 제어를 포함하며, 체적 음원의 경우 청취자를 중심으로 하는 멀티뷰 형식에서 화면을 회전시키는 제어를 포함한다.

사용자 인터페이스는 일반적으로 GUI를 통해 할 수 있지만, 3차원 오디오 지각에 있어 지각과 실제의 오차가

존재하므로, 실제 사용자가 원하는 제어를 하려면 청각에 의해 음원의 위치 및 행위를 제어하는 AUI가 더욱 실용적이라고 할 수 있다. 특히 AUI는 GUI에 의한 접근이 불가능한 시각장애인을 위해 매우 유용한 도구가 될 것이다.

III. MPEG-4

3.1. MPEG-4

MPEG-4에서는 기존의 MPEG-1과 MPEG-2에서 다루던 영상, 오디오 신호의 압축 부호화에 추가해서 정지영상, 컴퓨터 그래픽스, 분석 합성계와 음성 부호화, MIDI (Musical Instrument Data Interface) 등에 의한 합성 오디오 및 텍스트도 포함하는 종합 멀티미디어 부호화 규격을 목적으로 하고 있다[22]. 따라서 MPEG-4에서는 부호화 대상을 3차원 공간 내의 AV (Audio Visual) 객체가 구성하는 장면으로 하며, 이 장면을 재생측의 2차원 디스플레이에 표시하기 위한 정보의 부호화 포맷을 규정하고 있다. 장면을 구성하는 각 객체는 개별적으로 최적의 부호화 방식이 적용되므로, 높은 부호화 효율을 실현함과 동시에 재생측에서 사용자가 객체 조작 (User Interface)을 가능하게 하고 있다.

3.2. MPEG-4 BIFS

MPEG-4에서 부호화 대상이 되는 다양한 멀티미디어 콘텐츠들은 객체별로 디지털로 표현/저장/전송된다. MPEG-4 BIFS (BInary Format for Scene)는 MPEG-4 객체들을 구성 및 혼합하고 상호작용을 하는데 사용되는 장면표현 (Scene Description) 기술이다[2].

MPEG-4 BIFS 구조가 대부분의 VRML (Virtual Reality Modeling Language)로부터 가져온 것이기는 하지만 비트스트림 표현은 완전히 다르다. 또한 MPEG-4는 VRML에 비하여 몇가지 구별되는 메커니즘 (데이터 스트리밍, 장면 갱신, 압축)을 추가하였다. MPEG-4는 클라이언트-서버 모델을 사용하다. MPEG-4 클라이언트 (또는 플레이어, 브라우저)는 MPEG-4 서버에 접속하여 콘텐츠를 요청한 후 수신하고 표현한다. 콘텐츠는 비디오, 오디오, 정지 영상, 합성 2D/3D 데이터 등으로 구성될 수 있으며, 이러한 각기 다른 데이터들을 취합하여 사용자의 화면이나 스피커에 출력하는 방법은 MPEG-4 BIFS를 이용한 장면기술에 의해 결정된다.

콘텐츠와 장면기술정보는 스트림형태로 전송된다. 즉 클라이언트는 데이터를 받아서 필요할 때 재생한다. 예를 들면, 비디오는 서버로부터 도착했을 때 플레이될 수 있다. 이것은 VRML과 명백히 반대되는 것으로 VRML에서는 먼저 서버로부터 장면에 관한 모든 정보를 전송 받은 후, 랜더링하며 이로 인하여 긴 대기시간을 필요로 한다.

일단 장면이 생성되면, 서버는 BIFS-command 라는 MPEG-4 장면 갱신 (update) 메커니즘을 이용하여 장면을 수정할 수 있으므로, 갱신과 스트리밍 메커니즘을 결합하여 BIFS는 장면 요소들을 움직일 수 있다. 또한 거의 모든 장면 요소를 여러 단계의 산술 코더 스트림으로 구성하는 BIFS-animation 메커니즘을 이용하면 BIFS-command와 동일한 장면 갱신 능력과 높은 코딩 효율을 얻을 수 있다.

위의 그림 3은 MPEG-4 BIFS 구조를 이용하여 표현한 장면의 한 예이다.

3.3. MPEG-4 AudioBIFS

VRML에서 제공하는 오디오 노드의 기능은 한정되어 있었으므로 MPEG-4에서는 이를 확장한 AudioBIFS 노드들을 정의하여 MPEG-4 BIFS에 포함시켰다.

AudioBIFS 노드들은 오디오 객체들을 계층적이고 객체 지향적인 합성을 통한 2차원 또는 3차원 오디오 장면을 구성할 수 있게 하며, 또한 오디오 서브트리란 오디오 노드들만의 트리를 제공함으로써 응용프로그램을 구성할 때 오디오를 분리하여 처리할 수 있도록 하였다.

그림 4는 MPEG-4 오디오의 장면 합성의 흐름을 나타낸다. 객체 기반으로 부호화된 오디오 객체들은 AudioBIFS의

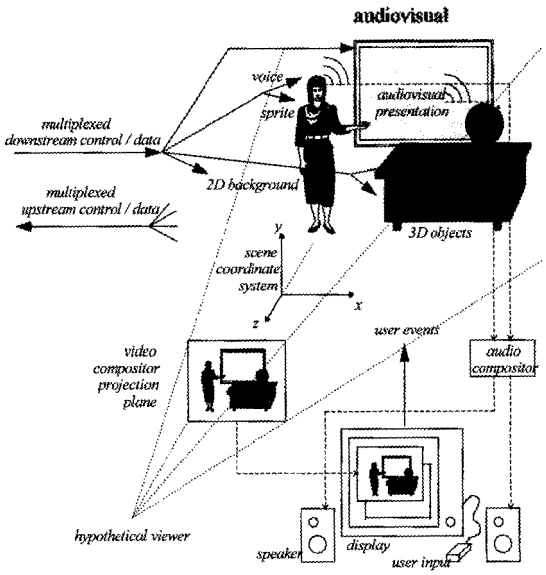

그림 3. 객체기반 멀티미디어 장면의 예
Fig. 3. Example of object-oriented multi-media scene.

장면 표현에 따라 처리된 다음 믹싱되어 출력되게 된다.

AudioBIFS는 하나의 오디오 장면에 합성오디오를 포함한 다양한 형태의 오디오 데이터를 결합하여 오디오 장면을 구성할 수 있도록 하였으며, 지연 노드를 통한 동기 기능, 표본화 주파수 변경을 포함하는 고품질 재생기능 등을 제공한다. 또한 합성 오디오를을 처리하기 위한 AudioFX 노드를 별도로 제공함으로써 신호처리 및 필터링 기능을 유연하게 처리할 수 있도록 하였다.

AudioBIFS는 사용자의 다양한 제어를 받아들일 수 있도록 AudioBIFS 노드 또는 다른 BIFS 노드들과의 유연한 상호작용 (interaction) 기능을 제공한다. AudioBIFS를 구성하는 7개의 오디오 전용 노드 (버전 1)들은 아래와 같다.

- **AudioBuffer**: 사용자 제어에 따른 연속적인 오디오 신호의 재생을 지원하기 위하여 오디오 클립 (짧은 길이의 오디오 데이터)을 버퍼에 저장하는 노드
- **AudioDelay**: 영상신호나 다른 오디오 신호 등과 동기를 맞추기 위하여 오디오 신호를 지연시키는 노드
- **AudioFX**: 합성오디오 신호와 같은 효과음을 처리하거나 필터링 효과를 주기 위하여 신호처리 알고리즘을 적용시키는 노드
- **AudioMix**: N개의 채널을 가지는 입력 오디오 신호를 M개의 채널로 믹싱하는 노드
- **AudioSwitch**: 입력 채널 중에서 M개의 채널을 선택하는 노드
- **Sound**: 임의의 오디오 신호를 3차원 공간상에 위치시키는 노드

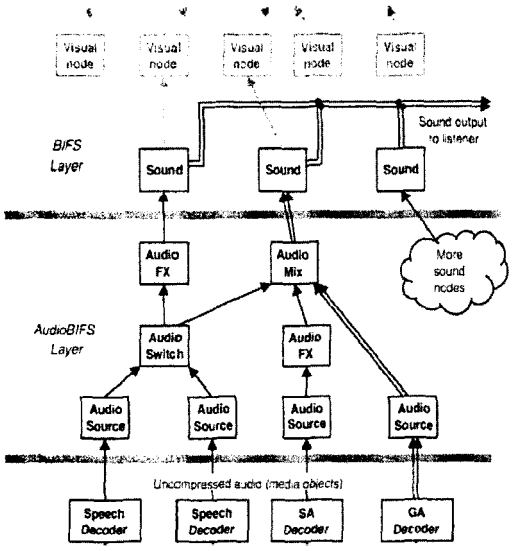

그림 4. MPEG-4 오디오 장면 합성의 예
Fig. 4. Example of MPEG-4 audio scene synthesis.

- **Sound2D**: 임의의 오디오 신호를 2차원 공간상에 위치시키는 노드

MPEG-4 시스템 규격이 버전2로 확장함에 따라 AudioBIFS의 기능을 확장한 것이 어드밴스드 AudioBIFS이다. 어드밴스드 AudioBIFS는 3차원 오디오 환경을 보다 적절히 모델링하기 위하여 다음과 같은 4개의 노드들을 새로이 정의하였으며, 이런 노드들을 이용하여 보다 실제적인 3차원 오디오 재생환경을 모델링해 줌으로써 현장감 (realistic) 있는 3차원 오디오 환경을 구성할 수 있게 한다.

- **AcousticScene**: AcousticScene 노드의 부모 그룹 노드의 하위 노드로 구성되는 오디오 서브그래프에서의 음향환경 (주로 잔향특성)을 정의
- **AcousticMaterial**: Material 노드의 서브셋으로 구성되며 IndexedFaceSet에 의하여 구성되는 음향공간상에서의 반사 및 투과 특성을 정의
- **DirectiveSound**: 3차원 오디오 신호의 주파수 의존적인 지향특성을 추가하여 **Sound** 노드의 기능을 확장한 것
- **PerceptualParameters**: 지각적인 음향환경 파라미터에 기반하여 3차원 오디오를 표현하는 노드

3.4. MPEG-4 Audio

MPEG-4에서는 오디오의 부호화를 위하여 객체기반의 오디오 부호화 도구들을 표준화하고 있으며, 이들 도구들에는 AAC (Advanced Audio Coding)를 기반으로 한 고품질 오디오 부호화, 저 비트율 오디오 부호화, 음성 부호화 등의 자연 음향 부호화 도구 (Natural Audio Tools)와 합성음 부호화 도구 (Synthetic Audio Tools) 및 문자 음성 변환 도구 (Text-To-Speech Interface, TTSI)가 포함된다[18].

그림 5에 MPEG-4 오디오 부호화 도구들의 종류 및 분류를 나타내고 있으며, 3차원 오디오 표현에 대해서는 합성 오디오 도구에 함께 표현되어 있다.

IV. 객체기반 3차원 오디오 방송 시스템의 설계

객체기반 3차원 오디오 방송 시스템은 DAB (Digital Audio Broadcasting) 및 디지털 방송에서, MPEG-4 표준을 기반으로 하여 객체기반의 사용자 제어가 가능한 오디

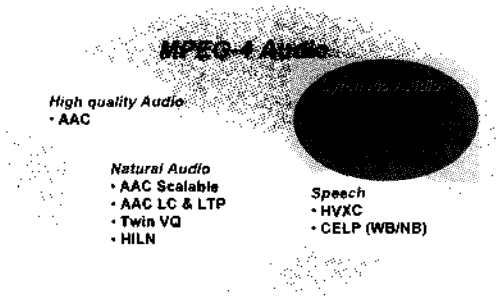

그림 5. MPEG-4 오디오 부호화 도구의 분류
Fig. 5. Conceptual view of MPEG-4 audio codectools.

오 컨텐츠를 제작하고 부호화하여 전송하며, 수신된 컨텐츠를 3차원 공간상에서 장면을 구성하여 재생하며, 사용자 제어에 의해 오디오 객체들의 위치 및 특성에 관련된 속성들을 제어하는 기능들을 제공하는 시스템이다. 본 시스템은 3차원 오디오, 단말에서의 대화형 컨텐츠, 관련 이미지 혹은 영상 데이터를 부가하여 향상된 멀티미디어 서비스를 제공하기 위한 것이다.

그림 6은 객체기반 3차원 오디오 방송시스템의 블록도를 나타낸다. 그림에서와 같이 본 시스템은 송신 서브 시스템과 단말 서브 시스템으로 구분할 수 있으며, 송신 서브 시스템은 객체기반 오디오 입력부, 객체기반 3차원 오디오 편집/제작부 및 객체기반 3차원 오디오 부호화부로 구성되며, 단말 서브 시스템은 객체기반 3차원 오디오 복호화부, 객체기반 3차원 오디오 장면 합성부, 사용자 제어부 및 3차원 오디오 재생부로 구성된다. 이하의 절에 객체기반 3차원 오디오 방송 시스템의 각 부별 구조, 기능 및 특징에 대해서 기술한다.

4.1. 객체기반 3차원 오디오 입력부

객체기반 3차원 오디오 입력부는 단일 채널, 스테레오, 더미헤드 마이크로폰 혹은 멀티채널 3차원 마이크로폰과 같은 여러가지 형태의 마이크로폰으로부터 오디오 신호를 입력받아 객체기반 3차원 오디오 객체들을 편집/제작부에 전달한다.

멀티채널 3차원 마이크로폰의 경우 물리적인 마이크로폰 장비 이외에 획득된 멀티채널 신호들을 단말에서의 스피커환경에 맞도록 처리하는 모듈을 포함한다.

또한 오디오 객체 분리기술을 이용하여 녹음 당시에 믹싱되어 있는 음원을 객체별로 분리하거나 추출하는 기능도 제공한다.

단말에서의 실시간 처리를 위하여 객체기반 3차원 오디오 입력부의 전체 오디오 채널 (객체)의 수는 8개로 제한하며 이용 가능한 출력 객체의 조합은 표 1과 같다.

본 시스템에서는 점음원뿐만 아니라 다차원의 멀티채널 음원을 모델링하고 있다. 이들 음원은 점음원 이외에 그림 7과 같은 선 형태의 1차원 음원, 면 형태의 2차원 음원, 부피를 가지는 3차원 음원으로 구분할 수 있다.

인간의 청각은 형체를 가진 음원의 경우 그 형체를 잘 식별하지는 못하지만 일정 각도 이상의 수직, 수평 방향의 넓이감은 비교적 정확하게 인지할 수 있다. 선형태의 1차원 음원 모델은 이러한 음원의 넓이감을 표현하는데 사용될 수 있다[23,24].

2차원 음원 및 3차원 음원 모델은 주변을 둘러싸는 음장을 표현하기 위해 사용될 수 있으며, 2차원 음원은 특정 방향에서 도래하는 음장을 표현할 수도 있다. 이러한 모델에 의해 가상 음장 안에서의 청취 위치를 제어할 수 있다.

2차원 및 3차원 음원 모델은 또한 여러 가지 음원을 가지는 객체 즉, 자동차와 같은 형상의 객체를 가상 전시장을 통해 원격지의 사용자에게 대화형 컨텐츠로서 제시할 때 사용할 수 있다.

4.2. 객체기반 3차원 오디오 편집/제작부

객체기반 3차원 오디오 편집/제작부는 객체기반 오디오 입력부로부터 입력된 오디오 신호 중 객체 오디오의 3차원 정보를 편집하며, 필요한 경우 공간의 3차원 정보

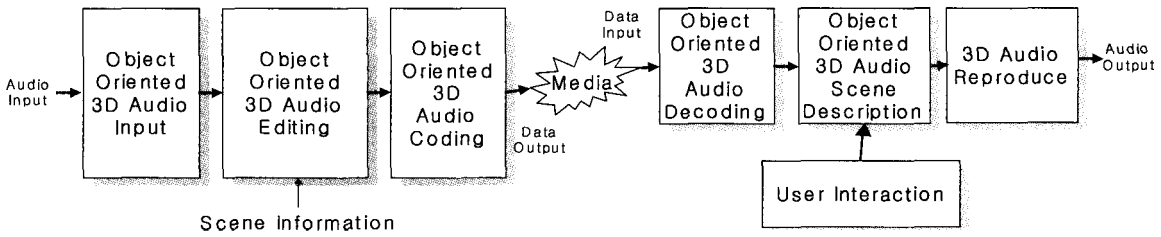

그림 6. 객체기반 3차원 오디오 방송시스템의 구조
Fig. 6. The structure of object-oriented 3D audio broadcasting system.

표 1. 오디오 입력부에서의 가능한 출력 객체 조합
Table 1. Possible configuration of audio objects.

Composition Type	Object Configuration	Application
Background Sound Object	8ch Background Sound Object	- Multi-view type interaction - Virtual car exhibition - Orchestra
Background Sound Object + Sound Source Object	Background Sound Object (2~5.1ch) + Sound Source Object (2~6)	- Small concert - Interaction audio game
Sound Source Object	Sound Source below 8 objects (point, 1D, 2D, 3D)	- Studio music recording - Interaction audio game - Educational contents

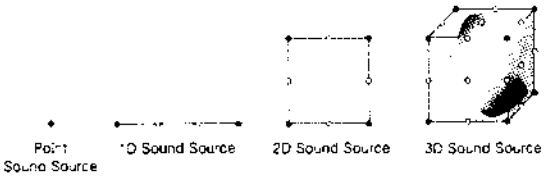

그림 7. 형체를 가지는 음원 모델
Fig. 7. Various sound source models according to its shape.

를 입력하여 단말에서 정확한 3차원 오디오 장면을 구성하여 재생할 수 있도록 한다.

제작자 또는 편집자는 청취점과의 상대적인 위치 (수평 각도, 수직 각도, 거리), 지향성 등으로 표현되는 음원 객체의 3차원 위치정보와 녹음 공간의 형태, 크기, 벽면의 반사율, 잔향시간 등으로 표현되는 녹음공간의 3차원 정보를 입력하여 편집할 수 있다.

음원객체 및 공간의 3차원 정보는 텍스트 포맷인 MPEG-4 XMT-A 형식의 3차원 오디오 장면 정보로 기술되며 장면 편집/합성부의 최종 출력이 된다.

4.3. 객체기반 3차원 오디오 부호화부

객체기반 3차원 오디오 부호화부는 오디오 객체 신호들을 MPEG-4 오디오 부호화 도구를 사용하여 부호화 한다. 오디오 부호화 방식으로는 MPEG-4 AAC 알고리즘을 이용한다.

객체기반 3차원 오디오 편집/제작부에서 전달된 XMT-A 형식의 3차원 오디오 장면 정보는 바이너리 포맷인 BIFS로 변환되며, 제작자나 편집자로부터 입력받은 데이터를 이용하여 IOD (Initial Object Descriptor)와 OD (Object Descriptor)를 생성하고, 부호화된 비트스트림들을 MPEG-4 SL (Sync Layer) 패킷화한다.

마지막으로 SL 패킷 및 IOD 데이터를 DAB 멀티미디어 전송 규격 또는 인터넷 스트리밍 전송 규격에 따라 다중화하여 전송한다.

4.4. 객체기반 3차원 오디오 복호화부

객체기반 3차원 오디오 복호화부는 다중화되어 전송된 데이터를 수신하여 역다중화하고, SL 패킷을 해석하며, MPEG-4 오디오 복호화 도구를 사용하여 오디오 객체 신호들을 복원하며, IOD, OD 및 BIFS를 추출한다.

4.5. 객체기반 3차원 오디오 장면합성부

객체기반 3차원 오디오 장면합성부는 입력된 오디오 객체 중 음원객체를 3차원 장면 정보에 따라 3차원 공간상에 렌더링하며, 공간의 3차원 장면 정보에 따라 공간 특성을 처리하여 공간감을 생성하고 렌더링하며, 공간감 처리가 수행된 음원 객체를 배경음 객체와 믹싱한다. 이때 3차원 장면 정보는 사용자 (청취자)에 의해 변경될 수 있으며, 변경된 장면 정보에 의해 렌더링이 수행될 수 있다.

또한 청취공간에 설치된 스피커의 구성을 사용자로로부터 입력받아 스피커 구성에 따라 2채널 혹은 멀티채널에 의한 3차원 렌더링을 선택하여 수행한다.

4.6. 3차원 오디오 재생부

3차원 오디오 재생부는 청취환경 및 재생 스피커 구성에 따라 합성된 오디오 장면이 적절히 재생될 수 있도록 처리하는 부분으로서 크로스토크 제거 및 실내 음향 특성의 제어를 수행하는 모듈이다.

재생 블록은 3차원 오디오 신호의 재생형태 즉 헤드폰, 스테레오 스피커, 스테레오 다이폴, 멀티채널 스피커 등의 정보를 입력받아 크로스토크 제거 필터링을 수행하고 청취자의 실내 음향 특성을 제거하기 위한 역필터링을 수행한다.

또한 사용자의 선호도에 따라 특정 음향환경 (콘서트홀, 운동장등)으로 변환하기 위한 필터링[25]도 수행할 수 있다.

4.7. 응용분야

본 시스템의 방송에의 응용 분야로서는 3차원 오디오와 대화형 오디오 서비스를 접목한 서비스라고 할 수 있으며, 3차원 방송, 멀티뷰 방송의 오디오, 가상 전시장의 대화형 객체, AUI 등에 응용할 수 있다.

3차원 방송은 스포츠 중계, 영화, 자연 다큐멘터리 등의 방송에 있어 파노라믹 영상과 함께 3차원 음향을 제공함으로써 현실감을 높일 수 있으며, 또한 공연장의 느낌을 가정에서 그대로 느낄 수 있도록 3차원 음악 방송을 제공할 수 있다. 현재 유럽에서는 DAB 시스템을 이용하여 오디오뿐만 아니라 이미지, 영상, 부가 데이터를 포함하는 멀티미디어 서비스의 단말로서 구상되고 있으며, 이 채널을 이용하여 가정에서 감상할 수 있는 객체기반의 3차원 음악 방송이 가능하다.

멀티뷰 방송은 현재의 방송이 단일 시청각을 제공하는 데 대해 다각도의 이미지 혹은 비디오와 더불어 3차원 오디오를 제공함으로써 현장감을 향상시킬 수 있는 기술이다. 현재 MPEG의 3DAV (3 Dimensional Audio Visual) 그룹에서 표준화 단계에 있으며[26], 2차원 혹은 3차원 음원을 사용하여 임의 각도의 오디오 신호를 생성할 수 있다.

가상 전시장의 대화형 객체는 자동차와 같이 각 부분에서의 소리가 다르며 크기가 충분히 큰 객체에 대해서는 온라인 전시장에 전시되어 있는 자동차를 원격지의 사용자가 직접 회전시키거나 확대시키면서 각 부분의 소리를 자세히 청취하도록 할 수 있다. 이러한 서비스는 상품의 소개 및 홍보를 위하여 사용될 수 있으며, 평소에 접근이 어려운 물체에 대한 청각적인 경험을 할 수 있도록 한다.

AUI는 3차원 오디오에 의해 다수의 아이템 혹은 임의 구조의 멀티미디어 데이터를 사용자에게 효율적으로 제시하고, 사용자는 이들 데이터 사이에서 3차원 오디오 인터페이스에 의해 자유롭게 움직이면서 필요한 컨텐츠를 검색하는 기능을 수행할 수 있는 기능이다. 이러한 기능은 긴 복도에 많은 방이 있고 문이 열려 있어 이 복도를 지나가면서 각 방의 컨텐츠의 핵심내용을 오디오로서 제시하면서 사용자가 찾는 컨텐츠를 검색할 수 있도록 하는 오디오 통로 (Audio Hallway)와 같은 개념을 사용함으로서 많은 채널로 구성되어 있는 TV 방송의 채널 안내 등을 할 수 있도록 한다[27]. 또한 이러한 기능을 응용하여 각종 오디오 게임을 제공할 수 있다.

V. 결 론

본 논문에서는 MPEG-4 표준에 기반한 객체기반 3차원 오디오 방송 서비스를 제안하고 가능한 시스템의 구조 및 응용 분야를 살펴 보았다.

제안된 시스템은 2채널 혹은 멀티채널 스피커 환경하에서 실감있는 3차원 오디오를 제공할 뿐만 아니라, 청취자에게 일방적으로 음악만을 들려주던 종래의 서비스에서 탈피하여, 사용자가 주어진 장면정보를 이용하여 음원을 객체별로 제어할 수 있는 대화형 오디오 방송 서비스를 제공하는 것을 목표로 한다. 이러한 서비스를 제공하기 위해서는 3차원 오디오를 획득, 처리, 재생하는 기술과 3차원 오디오 장면을 편집하고 표현하는 기술과 사용자 제어를 처리하는 기술이 필요하다. MPEG-4 표준은 객체기반 대화형 서비스 및 3차원 오디오에 기반한 매우 융통성 있는 서비스의 가능성을 시사하였다. 그러나 현실적인 서비스로서 실현하기 위해서는 객체 모델의 최적화가 필요하며, 이를 위해 음원객체의 수를 최소한으로 감소시키고, 배경음 객체를 도입하는 것을 제안하였다. 또한 다양한 서비스를 위해 점음원뿐만 아니라 1차원, 2차원 및 3차원 음원 모델을 제시하였고 이들을 이용한 응용 서비스의 몇가지 예를 제시하였다. 제안한 시스템은 객체기반 3차원 오디오 방송뿐만 아니라 가상현실, 게임, 인터넷 쇼핑등의 분야에도 응용할 수 있다.

본 시스템은 향후 시스템의 구현 및 시험을 통해 실효성을 입증하게 될 것이다.

감사의 글

본 논문은 정보통신부의 연구사업인 "지능형 통합 정보 방송 기술 개발" 과제의 일환으로 수행한 결과로서 정보통신부 담당자 및 관련 연구원들의 노력에 감사를 드립니다.

참 고 문 헌

1. A. Puri and T. Chen, *Multimedia Systems, Standards, and Networks*, Marcel Dekker, 2000.
2. ISO/IEC 14496-1, "Information technology – Coding of audio-visual objects, Part 1: Systems," International Standard, 1999.
3. 장대영, 서정일, 이태진, 강경옥, "객체기반 3차원 오디오 방송 시스템," 2002년도 한국음향학회 하계학술발표대회.

4. 강성훈, 강경옥, 입체음향, 기전연구사, 1997.
5. J. Blauert, *Spatial Hearing*, MIT Press, Cambridge Mass., 1983.
6. O. Kirkeby, P. A. Nelson, and H. Hamada, "Local sound field reproduction using two closely spaced loudspeakers," *J. Acoust. Soc. Am.*, **104** (4), 1973–1981, Oct. 1998.
7. Y. Kahana, P. A. Nelson, O. Krikeby and H. Hamada, "A multiple microphone recording technique for the generation of virtual acoustic images," *J. Acoust. Soc. Am.*, **105** (3), Crosstalk cancellation 1, March 1999.
8. P. A. Nelson, H. Hamada, and S. J. Elliott, "Adaptive inverse filters for streophonic sound reproduction," *IEEE Transactions on Signal Processing*, **40**, 1621–1632, 1992.
9. O. Kirbey, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularisation," *IEEE Transactions on Speech and Audio Processing*, 189–194, 1998.
10. 이신렬, 이준수, 성광모, "입체방송을 위한 입체음향 제작기법 연구," 한국음향학회 하계 학술발표대회, 2001.
11. www.soundfield.com
12. J. F. Cardoso, "Blind signal separation: Statistical Principles," *Proceedings of the IEEE*, **86** (10), Oct., 1998.
13. L. Tong, R. W. Liu, V. C. Soon, and Y. F. Huang, "Indeterminancy and identificability of blind indentification, " *IEEE Transactions on Circuits and Systems*, **38** (5), May 1991.
14. L. Tong, Y. Inouye, and R. W. Liu, "Waveform-preserving blind estimation of multiple independent sources," *IEEE Transactions on Signal Processing*, **41** (7), July, 1993.
15. ISO/IEC 11172-3, "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbis/s, Part 3: Audio," *International Standard*, 1993.
16. ISO/IEC 13818-3, "Generic Coding of Moving Pictures and Audio, Part 3: Audio," *International Standard*, 1997.
17. ISO/IEC 13818-3, "Generic Coding of Moving Pictures and Audio, Part 7: Advanced Audio Coding (AAC)," *International Standard*, 1997.
18. ISO/IEC 14496-3, "Information technology – Coding of audio-visual objects, Part 3: Audio," *International Standard*, 1999.
19. T. Kientzle, *A Programmer's Guide to Sound*, Addison Wesley, 1998.
20. ISO/IEC 12272-1:1997, "Virtual Reality Modeling Language (VRML)," *International Standard*, 1997.
21. W3C Recommendation 07, "Synchronized Multimedia Integration Language (SMIL 2.0)," 2001.
22. Eric D. Scheirer, Ritta Vaananen, and Jyri Huopanienmi, "AudioBIFS: Describing Audio Scenes with the MPEG-4 Mutlimedia Standard," *IEEE Trans. on Multimedia*, 1 (3), Sep., 1999.
23. G. Potard, J. Spille, "Study of sound source shape and wideness in virtual and real auditory displays," *114th AES Convention*, March, 2003.
24. ISO/IEC JTC1/WG11, M9457, "Report on the third MPEG-4 AudioBIFS Sound Source Wideness Core Experiment," May, 2003.
25. D.-S. Lee and K.-M. Sung, "Sound field processor for creating virtual concert hall impression in home," *IEEE Transactions on Consumer Electronics*, **41** (2), May 1995.
26. ISO/IEC JTC1/WG11, N5539, "Applications and Requirements for 3DAV," May, 2003.
27. C. Schmandt, "Audio Hallway: a Virtual Acoustic Environment for Browsing," *UIST '98 Proc.*, 163–170.

저자 약력

● 강 경 옥 (Kyeong-Ok Kang)

1985년 2월: 부산대학교 물리학과 학사
1988년 2월: 부산대학교 대학원 물리학과 석사
1993년 3월 ~ 현재: 한국항공대학교 대학원 전자공학과 박사과정
1991년 ~ 현재: 한국전자통신연구원 방송미디어연구부 방송컨텐츠응용연구팀장 (선임연구원)
※ 주관심분야: 디지털 방송, 메타데이터 신호처리, 음향신호처리, 객체기반 3차원 오디오 신호처리 기술

● 장 대 영 (Dae-Young Jang)

1991년 2월: 부경대학교 전자공학과 졸업 (공학사)
2000년 2월: 배재대학교 대학원 컴퓨터공학과 졸업 (공학석사)
2000년 3월 ~ 현재: 배재대학교 대학원 컴퓨터공학과 박사과정
1991년 1월 ~ 현재: 한국전자통신연구원 방송미디어연구부 선임연구원
※ 주관심분야: 음향 신호처리, 멀티미디어 응용, 객체기반 3차원 오디오 신호처리

● 서 정 일 (Jeongil Seo)

1994년 2월: 경북대학교 전자공학과 공학사
1996년 2월: 경북대학교 대학원 전자공학과 공학석사
1996년 3월 ~ 현재: 경북대학교 대학원 전자공학과 박사과정
1998년 3월 ~ 1999년 5월: LG반도체 근무
1999년 6월 ~ 2000년 10월: 현대전자 근무
2000년 11월 ~ 현재: 한국전자통신연구원 방송미디어연구부 연구원
※ 주관심분야: 음향신호처리, 음성신호처리, 객체기반 3차원 오디오 신호처리, MPEG, DMB

● 정 대 권 (Dae-Gwon Jeong)

한국음향학회지 제20권 제3호 참조