

실감방송을 위한 3차원 오디오 기술

Three Dimensional Audio Technologies for Realistic Broadcasting

장대영(D.Y. Jang)	3D미디어연구팀 선임연구원
서정일(J.I. Seo)	3D미디어연구팀 선임연구원
이태진(T.J. Lee)	3D미디어연구팀 연구원
박기윤(G.Y. Park)	3D미디어연구팀 연구원
강경옥(K.O. Kang)	3D미디어연구팀 책임연구원, 팀장

차세대 방송 서비스는 입체감있는 3차원 AV 콘텐츠와 자연스럽게 사용자와 인터랙션하는 대화형 콘텐츠를 기반으로 하는 실감방송으로 변화되어 갈 것으로 예상된다. 이러한 실감방송 서비스에서는 현장감을 효율적으로 나타낼 수 있는 음상 정위 및 음장 재현 등 3차원 오디오 기술과 사용자 인터랙션을 위한 객체기반 오디오 처리 기술들이 필요하다. 본 고에서는 이러한 현장감과 사용자 인터랙션을 통하여 가상현실에 근접한 서비스를 제공하기 위한 대표적인 3차원 오디오 기술의 개발 동향을 살펴본다. 우선 3차원 오디오 기술의 기본 개념 및 개요를 기술하며, 이러한 3차원 오디오 기술에 기반한 대화형 3차원 오디오 기술 개발에 대한 최근 동향을 살펴보고, 국내에서 개발하고 있는 객체기반 3차원 오디오 기술에 대하여 간략히 설명한다.

I. 서론

방송에 있어 오디오 기술은 필수적이며, 또한 중요한 위치를 차지하고 있다고 할 수 있다. 하지만, 일반적으로 오디오는 비디오에 종속되는 부가적인 매체로 생각하는 경우가 많다. 이것은 일상 생활에 있어 귀보다 눈에 의존하는 경향이 많은 현실에 기인한 것이다. 그러나 오디오에 의해서 대부분의 분위기나 정서적인 것이 전달된다는 사실은 이미 알려진 바이며, CD나 영화, 방송 등에서는 오디오 제작의 완성도를 매우 중요하게 생각하고 있다.

방송은 1920년 미국 펜실베이니아의 웨스팅하우스사에서 세계 최초로 라디오 방송을 송출한 이래 80년의 역사를 통하여 다양한 변천의 과정을 겪었다. 대표적인 방송 역사상의 이정표를 살펴 보면 흑백 TV 방송, 컬러 TV 방송, 인터넷 방송, 디지털 방송 등을 꼽을 수 있다.

현재 태동기에 있는 인터넷 방송 및 디지털 방송

의 특징을 살펴 보면 인터넷 방송은 멀티미디어 방송, 대화형 방송을 그 특징으로 하고 있고, 디지털 방송은 멀티미디어 방송, 고품질 방송 및 대화형 방송이 그 특징이라고 할 수 있다. 이러한 특징은 방송이 기존의 일방적인 서비스 전달 매체로서의 역할에서 벗어나 사용자들에게 다양한 정보를 맞춤형으로 서비스를 해야 한다는 것을 반영하고 있다.

이러한 방송의 발전은 멀티미디어 부호화 기술, 전송 기술 등의 발전에 기반을 둔다고 할 수 있으며, 최근에는 이러한 기술 이외에도 방송 통신 융합에 의해 사용자가 서비스 제공자와의 인터랙션을 통해 원하는 정보를 제공 받을 수 있는 대화형 서비스에 대한 관심이 높아지고 있다. 방송과 통신 서비스가 융합되면서 방송의 고품질 콘텐츠와 통신의 양방향 서비스가 상승 작용을 일으켜 향후 방송통신 융합망에서는 전혀 새로운 서비스들이 가능할 것으로 기대한다. 이러한 새로운 서비스 중 콘텐츠의 현장감을 그대로 전달하며, 사용자에게 몰입감을 극대화해 줄

수 있는 실감방송도 하나의 중요한 응용 분야로 인식되고 있다.

현재 디지털 방송에서는 ITU에서 규정하는 5.1 채널의 서라운드 재생방식을 채택하고 있다. 5.1 채널 오디오는 현재 주로 DVD 영화 및 음악용으로 사용되고 있으며, 앞으로는 가정에서 TV 방송을 통하여 누구나 손쉽게 서비스를 제공 받을 수 있게 된다. 이러한 5.1채널 오디오 서비스에 의해 화면에 표시된 장면뿐만 아니라 화면에 표시되지 않은 주변의 분위기도 현장감 있게 즐길 수 있다. 그러나 주로 채널간 레벨차 및 위상차이에 의해 음상을 제어하므로 3차원 오디오의 충실한 재생이 어려운 것이 사실이며, 사용자와의 인터랙션에 의한 대화형 서비스를 제공할 수 없다.

본 고에서는 보다 현실감 있는 방송, 통신 서비스를 위한 3차원 오디오 기술 및 개발 동향을 소개하고 향후 실감방송의 오디오 기술로서 객체기반 3차원 오디오 방송 기술 전망에 대하여 살펴보고자 한다. II장에서는 3차원 오디오 기술의 개요에 대하여 기술하고, III장에서는 객체기반 3차원 오디오의 개념 및 주요기술들을 설명하며, IV장에서는 객체기반 3차원 오디오 기술 개발 동향에 대하여 소개하고, 마지막으로 V장에서는 결론과 향후 필요한 기술 개발에 대하여 간략히 기술하고자 한다.

II. 3차원 오디오 획득 및 재생 기술

3차원 오디오 기술의 목적은 청취자에게 실제 음원을 획득한 장소와 같은 느낌을 주는 것이다. 즉 청취자가 실제 음원을 획득한 장소에 실제로 있는 것처럼 현장감, 거리감, 방향감 등을 주는 것이다. 이러한 효과를 주기 위해서는 음원을 얻기 위한 3차원 오디오 획득 기술 및 다양한 재생환경에서 재생하기 위한 3차원 오디오 재생 기술이 필요하다. 3차원 오디오 획득 기술은 크게, 더미헤드 기술과 음장(sound field) 녹음 기술로 나눌 수 있고 3차원 오디오 재생 기술은 헤드폰, 스테레오, 멀티채널 재생 기술 등으로 나눌 수 있다.

1. 3차원 오디오 획득 기술

가. 더미헤드 마이크로폰

3차원 오디오 획득을 위해 실제 사람의 귀에 마이크를 장착하는 것도 가능하지만, 고성능의 마이크를 사람의 귀에 장착하기가 힘들고, 잡음 및 움직임의 영향 때문에, 사람의 형태를 한 더미헤드에 마이크를 장착하여 3차원 오디오 신호를 획득한다. 더미헤드를 이용하여 획득한 신호를 바이노럴 신호라고 하는데, 헤드폰을 통해 청취할 경우 입체감을 느낄 수 있다[1]. 더미헤드는 머리부분만을 마네킹으로 모사한 것이고, 더미헤드에 몸통까지 포함한 것을 HATS(Head And Torso Simulator)라 한다. 현재 여러 기업에서 HATS와 더미헤드를 제작하여 판매하고 있는데, 대표적인 더미헤드의 형상을 (그림 1)에 나타낸다.



(위쪽부터 시계 반대방향으로 KU-80i, KEMAR, Aachener Kopf)

(그림 1) 대표적인 더미헤드 형상

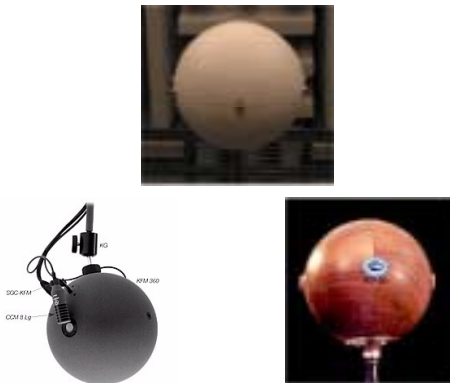
나. 구체형 더미헤드

구체형 더미헤드는 기존 사람의 머리 형태를 갖는 기존의 더미헤드를 구체형으로 간략화한 것이다. 더미헤드는 실제 사람의 머리 형태를 가지기 때문에 사람의 청각 특성과 유사하지만, 형태가 크고 보기에 편하지 않기 때문에, 실제 라이브 환경에서 사

용되기에는 어려운 점이 있다. 따라서 사람의 머리 형태를 구체형으로 간략화한 구체형 더미헤드가 많이 이용되고 있다.

구체를 이용하여 획득한 3차원 오디오 신호는 스피커를 통해 재생하는 경우, 더미헤드 획득 신호에서 발생하는 double equalization(녹음과 재생 시 귀의 영향을 두 번 받는 현상)이 없고, 형태의 단순성 때문에 3차원 오디오 획득에 많이 이용되고 있다.

ETRI에서는 구체 위에 다수의 마이크를 장착하여 3차원 오디오 신호를 획득하고 재생하는 시스템을 개발하였다. ETRI의 구체 마이크는 5개의 마이크를 구체 위에 장착하여 3차원 오디오 신호를 획득하고 후처리를 통해 헤드폰, 스테레오, 스테레오 다이폴, 4채널, 5채널 등의 환경에서 재생 가능한 3차원 오디오 신호를 생성한다[2]. (그림 2)는 현재 가장 대표적인 구체형 더미헤드들을 보여주고 있다 [3],[4].



(위쪽부터 시계 반대방향으로 ETRI 3D Multichannel Microphone, KFM360, BS-3D)

(그림 2) 대표적인 구체형 더미헤드 형상

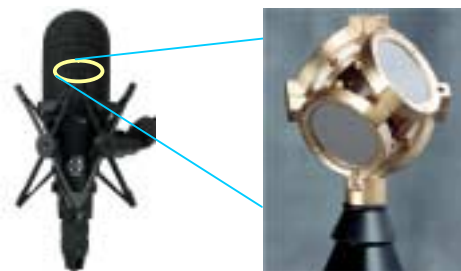
다. Ambisonic 기술

Ambisonic은 음장에 대한 정보를 획득하는 방법과 일정한 방식의 스피커를 통해 재생하는 시스템으로, 실제적인 3차원 오디오 이미지를 생성하는 시스템이다.

1920년대 말 Alan Blumlein이 개발한 기술은 8자 지향성을 갖는 마이크 쌍을 가능한 가깝게 위치시키고, 각각 전방에서 45도 각도를 향하게 하는 방식이었다. 이 방식은 훌륭한 스테레오 이미지를 제공하지만 8자 지향성 때문에 후방에서 오는 소리 또한 픽업되고, 스피커 쌍을 통해 재생될 때 이 소리가 접혀 전방에 매핑되어 과도한 잔향이 있는 것처럼 들리게 된다.

Blumlein 기술의 단순성과 정확성을 좋아하는 순수한 레코딩 엔지니어들은 이러한 문제점을 해결하기 위해 8자 지향성을 갖는 마이크를 크로바형 지향성 마이크로 교체하고, 이들 사이의 각도를 변경하여 원하는 사운드 스테이지만을 포함하도록 개선하였다. 또한 크로바형 지향성 마이크의 후방으로부터 도래하는 소리에 대해 응답이 감소하도록 하여 후방으로부터 도래하는 잔향 신호가 전방 사운드 스테이지에 매핑되는 것을 줄였다.

현재는 Alan Blumlein의 원래의 생각으로 다시 돌아가서, 단지 무지향 마이크를 8자 지향성 마이크 쌍에 추가함으로써, 가능한 모든 음장 정보를 획득할 수 있도록 개선되었다. 물론 이것은 캡슐을 완전히 일치하게 배치하였다는 것을 가정한다. 즉 모든 3개의 캡슐이 음장 내에서 음향학적으로 동일한 장소에 위치하여야 하지만 현실적으로 불가능하다. 이러한 문제는 sound field 마이크에 의해 해결되었는데, 이 마이크는 입력신호를 얻기 위해 (그림 3)의 오른쪽 그림과 같이 가상의 구의 표면에 4개의 작은 캡슐을 배치함으로써 문제를 해결하였다.



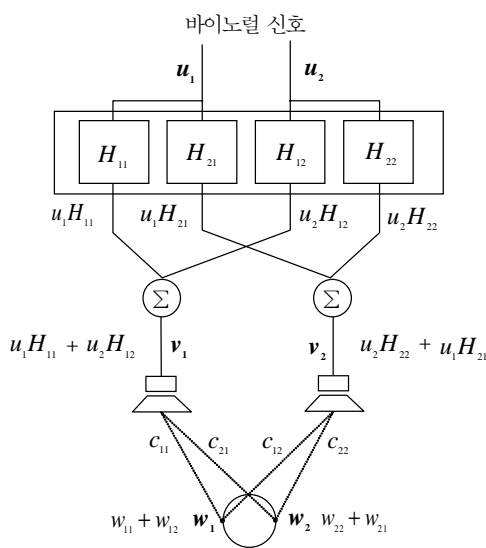
(그림 3) Ambisonic 마이크로폰 형상

2. 3차원 오디오 재생 기술

3차원 오디오 재생 기술은 그 방법에 따라 바이노럴 및 멀티채널로 구분할 수 있다. 바이노럴은 2채널의 스피커 혹은 헤드폰에 의해 3차원 오디오를 출력하는 방식으로서, 헤드폰 혹은 스피커의 특성을 보정하고 스피커의 경우 공기 중의 전달 특성을 보정하기 위해 역필터링을 위한 필터 매트릭스를 적용하여야 한다.

(그림 4)는 2채널 재생인 경우의 크로스토크 제거 과정을 보인다. 입력으로 바이노럴 신호 u_1, u_2 가 들어올 때, 이를 바로 두 개의 스피커를 통해 재생하는 경우, 헤드폰을 이용한 재생과 달리, 반대편 스피커의 출력까지 귀에 들리게 된다. 이 문제를 해결하기 위해, 스피커에서 귀까지의 plant $C(z)$ 를 측정 한 후, 이를 보상해 주는 역필터 $H(z)$ 를 설계한다. 이러한 역필터를 통과한 신호 v_1, v_2 가 스피커의 입력 신호가 된다. v_1, v_2 는 스피커에서 청취자까지 전달함수 plant $C(z)$ 의 특성을 이미 보상한 신호이기 때문에, 역필터를 잘 설계하면, w_1, w_2 는 입력신호 u_1, u_2 와 동일하게 된다. 따라서 크로스토크가 제거된 신호를 청취자의 귀에서 재생할 수 있다.

(그림 5)는 스테레오 재생인 경우 역필터와 재생

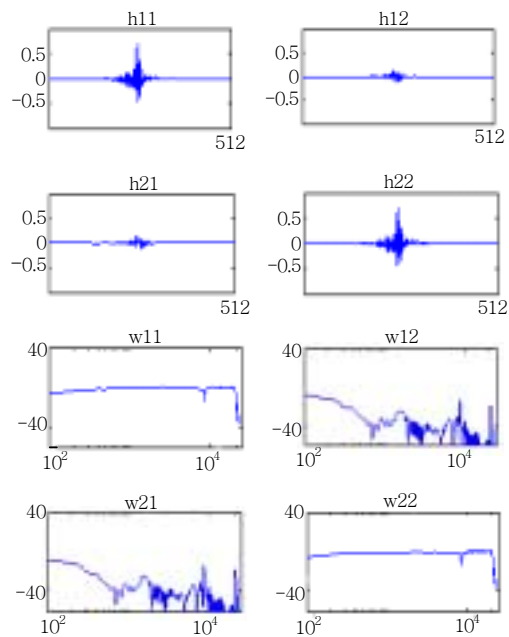


(그림 4) 2채널 크로스토크 제거 필터

신호를 보인다. 그림에서 h 가 역필터를 의미하고 w 가 재생신호를 의미한다. 그림은 각 재생신호의 주파수 특성의 예를 보인 것인데, w_{11}, w_{22} 가 좌측 스피커에서 좌측 귀로, 우측 스피커에서 우측귀로 전달되는 신호이고, w_{12}, w_{21} 은 크로스토크 성분을 나타낸다. 그림에서 크로스토크 성분은 제거되고, 각각의 스피커에서 나온 신호가 각각의 귀에만 전달됨을 알 수 있다.

멀티채널의 경우 스피커의 배치 및 지향성 등을 잘 조절하여 채널간의 간섭을 최대한 줄여야 하지만 주로 가정에서는 이에 대한 고려는 하지 못하고 있다. 다만 ambisonic 기술의 경우 재생 스피커의 개수 및 배치에 따라 출력 신호를 제어하여 최대한 원 음장을 재현할 수 있다.

바이노럴 기술은 대부분 60도의 각도로 배치된 두 개의 스피커에서 재생하는 것을 기반으로 하여 구현되고 있지만, 청취실의 음향 특성에 따라 그 성능이 매우 저하되며, 유효 청취 영역도 매우 좁다. 영국의 사우쓰햄프턴 대학의 필립 교수와 동경전기 대학의 하마다 교수는 기본 개념을 뒤엎고 10도의 스피커 배치에 의해 유효 청취영역을 넓히는 스테레



(그림 5) 2채널 역필터의 특성



(그림 6) 스테레오 다이폴 스피커 시제품

오 다이폴의 연구에 성공하였으며, DiMAGIC이라는 음향 회사를 설립하여 스테레오 다이폴을 실용화하는 한편 홈씨어터 및 휴대용 음향기기에의 적용을 추진하고 있다[5]. (그림 6)은 PS2와 DVD에서 스테레오 다이폴을 적용한 제품들을 보여주고 있다.

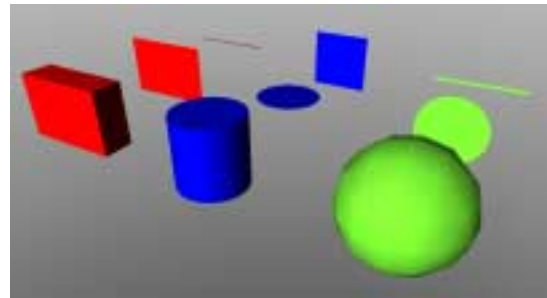
III. 객체기반 3차원 오디오 기술

객체기반 3차원 오디오는 오디오 신호들을 객체 단위로 구분하여 부호화 및 전송하고, 재생부에서 이들 객체 신호들을 조합하여 재생하며, 필요에 따라 사용자가 객체 단위로 제어하게 함으로써 정보 전달 및 이용을 극대화하기 위한 수단이다. 객체기반 3차원 오디오는 3차원 오디오에 의한 임의 현장의 가상 체험 효과와 객체기반 처리에 의한 사용자 상호작용을 통하여, 정보의 이용을 극대화할 뿐 아니라 실제적인 현실감을 전달하기 위한 기술이다.

객체기반으로 3차원 오디오를 처리하기 위해서는 3차원 오디오 장면을 구성하는 객체를 정의하고, 객체들을 이용하여 3차원 오디오 장면을 구조화된 언어로 기술하고 재생할 수 있는 도구들이 필요하며, 또한, 혼합된 신호로부터 중요한 오디오 객체를 분리하기 위한 기술도 요구된다.

1. 오디오 객체의 종류

오디오 객체들은 속성제어가 가능한 음원 객체와 전체적인 음향 분위기를 구성하는 배경음 객체로 구분할 수 있다[6]. 객체 별로 녹음되었거나 혼합음으로부터 분리된 음원 객체는 3차원 오디오 장면을 구성하거나 사용자의 제어에 의하여 속성(위치, 거리,



(그림 7) 다양한 입체 음원 모델

<표 1> 객체기반 3차원 오디오 객체의 종류

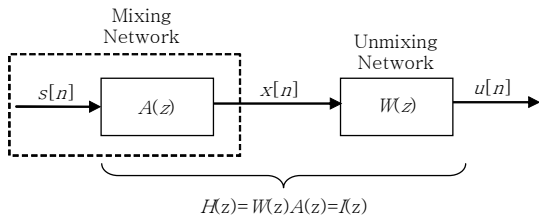
오디오객체	객체 구분	비고
음원객체	점음원	음원 소스가 한 점임
	선음원	선으로 이루어진 음원 소스 (기타의 현)
	면음원	면으로 이루어진 음원 소스 (벽, 면)
	체적음원	부피를 가지는 음원 소스 (근접 자동차 소리)
배경음객체	바이노럴	더미헤드로 입력한 신호의 형태
	멀티채널	스테레오~멀티채널
	엠비소닉	B-포맷 신호

크기, 모양)을 변경할 수 있으나, 배경음 객체는 두 개 이상의 객체들이 혼합된 것이므로 속성 변경에 제한을 가지는 것이 특징이다. 또한, 점음원으로 국한된 음원 모델을 선, 면, 체적의 형태까지 지원하기 위한 방법이 최근 MPEG에서 WideSound 노드란 이름으로 표준화가 진행중이다[7]. (그림 7)은 WideSound 노드에서 표현 가능한 음원의 형태를 나타낸 것이다.

객체기반 3차원 오디오 장면을 구성하는 오디오 객체들의 종류를 정리하면 <표 1>과 같다.

2. 음원 분리 기술

오디오 부호화에서 사용되는 오디오 객체들은 녹음 과정에서 분리가 어려운 상태로 서로 혼합되어 있어 별도의 분리과정이 필요하다. 예를 들면, 카페 일 파티장에는 여러 사람들의 대화소리와 음악 등이 섞여 있으며, 경우에 따라 이들 소리들은 하나의 오디오 객체로 간주되기 보다는 분리한 후 각각을 독



(그림 8) 블라인드 신호 분리

립된 오디오 객체들로 간주하여 다룰 필요가 있다. 또한 피할 수 없는 원치 않은 잡음들에 의해 음질이 심히 저하되는 경우, 원치 않은 잡음들로부터 목표 오디오 객체를 효과적으로 분리해낼 필요도 있다.

3차원 오디오 객체분리는 시간적으로 혼합된 여러 개의 오디오 객체를 분리하는 것으로 혼합되기 전의 오디오 신호에 대한 정보 없이 분리하므로 블라인드 신호 분리(blind source separation)라고 부른다. 블라인드 신호 분리에 대한 개념을 (그림 8)에 나타낸다[8].

독립된 오디오 객체 신호 s 는 혼합환경 A 에 의해 혼합되어 혼합신호 $x=As$ 를 만들어낸다. 혼합된 신호 x 로부터 s 를 분리해내기 위해서는 $u=Wx=A^{-1}x$ 를 만족하는 분리회로 W 가 필요하다. 여기서 독립된 오디오 객체 신호 s 와 혼합회로 A 는 알려져 있지 않으므로 오직 관찰 가능한 혼합신호 x 만으로부터 분리회로를 예측해야만 한다. 이때 사용되는 가장 중요한 사실은 서로 다른 오디오 객체들이 서로 통계적으로 독립적이라는 점이며 이러한 이유로 신호분리 기술은 독립성분 분석기술의 하나로 분류된다.

오디오 객체 분리 기술은 오디오 객체의 혼합형태에 따라 크게 순시혼합과 지연혼합된 경우로 구분된다. 예를 들어 카테일 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 무향실과 같은 특수한 환경에서의 오디오 혼합은 순시혼합에 해당한다.

3. 오디오 장면 표현 기술

오디오 비주얼 객체들이 시공간상에 배치되어 이

루는 가상공간을 장면(scene)이라고 말한다. 이러한 장면을 구조화된 언어로 기술하는 것을 장면기술(scene description)이라고 하며, MPEG-4와 같이 오디오 비주얼 데이터가 객체별로 부호화되어 단말에 전송될 때는 장면기술 정보에 근거하여 비주얼 데이터를 디스플레이하고 오디오 데이터를 재생하게 된다[9]. 그러므로 객체기반의 멀티미디어 플랫폼에서는 각 객체별 압축/부호화 방식뿐만 아니라 이들을 조합한 인터랙티브한 장면을 어떻게 기술하는가에 대한 내용도 포함하고 있어야 한다.

가. 장면 표현 기술

위와 같은 대화형 3차원 장면을 기술하기 위하여 MPEG-4에서는 BIFS(Binary Format for Scene)라고 하는 오디오 비주얼 장면을 표현하기 위한 언어를 표준화하였다. BIFS는 VRML을 기반으로 하여 개발되었으나 MPEG-4 응용프로그램을 개발하는 데 용이하게 하기 위하여 VRML 노드의 일부를 삭제 또는 수정하였으며, 스트리밍, 2D 장면을 지원하기 위한 노드 등을 추가하였다. 또한 간결한 이진 형식으로 장면을 표현함으로써 장면을 기술하기 위한 메모리 양을 최소화 하였다.

3차원 오디오 장면은 배경음과 3차원 오디오 객체들로 이루어지는 오디오 장면과 이들이 재생되는 공간을 도시하는 3차원 멀티미디어 객체들과 전체 장면을 제어하기 위한 인터페이스 등으로 구성된다. 3차원 오디오 객체들은 3차원 공간상에서 사용자의 제어에 의하여 자유로이 이동(zooming, panning, rotation)할 수 있어야 하고, 새로운 3차원 오디오 객체를 삽입하거나 제거함으로써 새로운 오디오 장면을 구성할 수 있어야 한다. 또한 비주얼 객체와 연동하여 사용자의 시선(view point)이 이동함에 따른 비디오의 이동에도 적절히 대응하여 음원의 위치(sound source position)나 청취자의 위치(listening point)를 이동시켜 주어야 한다.

나. MPEG-4 AudioBIFS

MPEG-4에서는 위와 같은 3차원 오디오 장면을

기술하기 위해 MPEG-4 BIFS 내에 AudioBIFS란 오디오 전용 노드들을 구성하여 사용하도록 하였다. AudioBIFS는 오디오 신호를 오디오 디코더로부터 입력 받아 신호처리 과정처럼 처리하여 최종적인 사운드 노드에서 오디오 신호를 출력하는 구조를 가진다. 따라서 3차원 오디오를 위한 노드들을 풍부하게 제공하고 있는 MPEG-4 AudioBIFS가 3차원 오디오 장면을 기술하는 가장 적절한 방식이라 할 수 있다.

AudioBIFS는 하나의 오디오 장면에 합성 오디오를 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 오디오 도구를 처리하기 위한 AudioFX 노드를 별도로 제공함으로써 신호처리 및 필터링 기능을 유연하게 처리할 수 있도록 하였다. 또한 사용자의 다양한 제어를 받아 들일 수 있도록 AudioBIFS 노드는 오디오 서브트리 또는 다른 BIFS 노드들과의 유연한 상호작용 기능을 제공한다.

초기의 AudioBIFS를 구성하는 오디오 전용 노드들은 AudioBuffer, AudioDelay, AudioFX, Audio-Mix, AudioSwitch, Sound 및 Sound2D로서 주로 VRML 및 합성 오디오 도구의 기능에서 발췌되어 표준화 되었다.

MPEG-4 시스템 규격이 버전 2로 확장함에 따라 AudioBIFS의 기능을 확장한 것이 Advanced AudioBIFS 이다. Advanced AudioBIFS는 3차원 오디오 환경을 보다 적절히 모델링하기 위하여 다음과 같은 4개의 노드들을 새로이 정의하였으며, 이런 노드들을 이용하여 보다 실제적인 3차원 오디오 재생환경을 모델링 해줌으로써 현장감 있는 3차원 오디오 장면을 구성할 수 있게 한다.

- AcousticScene: 음향환경(주로 잔향특성)을 정의
- AcousticMaterial: 음향공간상에서의 반사 및 투과 특성을 정의

- DirectiveSound: 주파수 의존적인 지향특성을 정의
- PerceptualParameters: 지각적인 파라미터에 기반한 음향환경 정의

현재 MPEG에서는 면적이나 부피를 가지는 음원을 표현하는 기능 및 AudioFX의 기능을 단순화하여 상용 AV 단말에서 보다 쉽게 음향 효과를 구현하는 기능 등을 추가하기 위하여 AudioBIFS 버전 3의 표준화가 진행중이다[7]. AudioBIFS 버전 3에서 논의중인 새로운 오디오 노드들은 아래와 같다.

- AdvancedAudioBuffer: AudioBuffer 노드의 기능을 확장하여 오디오 클립을 로드하는 메커니즘을 추가한 노드
- AudioChannelConfig: 멀티채널 또는 엠비소닉 오디오 신호가 오디오 서브트리에서 채널 혼동이 발생하는 것을 막기 위해 각 출력채널의 정확한 재생환경을 기술하는 노드
- AudioFXPROTO: AudioFX 노드의 구현상의 복잡도를 극복하기 위하여 비교적 간단한 음향 효과를 PROTO 메커니즘을 이용하여 표현함으로써 연산능력이 부족한 AV 단말에서도 구현이 가능하도록 하는 노드
- SurroundingSound: 엠비소닉과 같이 제어가능한 멀티채널 오디오 신호를 위한 노드
- Transform3DAudio: 2D 장면에서도 3차원 오디오를 표현할 수 있도록 오디오와 관련한 좌표계를 변환시키는 노드
- WideSound: 점음원이 아닌 선, 면, 체적음원을 표현하기 위한 노드

IV. 대화형 3차원 오디오 기술 개발 동향

대화형 3차원 오디오 처리기술의 최신 동향을 유럽 프로젝트인 CARROUSO와 NHK에서 최근 발표한 VRTV, 그리고 ETRI에서 개발중인 객체기반 3차원 오디오 기술을 중심으로 소개한다.

1. 유럽 CARROUSO 프로젝트

CARROUSO(Creating, Assessing and Rendering in Real time Of high quality aUdio-viSual enviroNments in MPEG-4 context)는 유럽의 IST(Information Society Technologies) 프로젝트의 별칭이다. 그 이름이 말해 주듯이, 비디오 데이터와 연동하여 음장을 녹음 및 재생하되 시청자가 시간적 혹은 공간적 지각특성을 제어할 수 있게 하는 기술을 개발 및 검증하는 것을 목적으로, 유럽 지역의 산업체와 대학 등 10여 개의 기관이 연합하여 추진한 프로젝트이다.

시연 시스템은 (그림 9)와 같이 녹음, 전송, 재생 모듈로 구성된다. 마이크 어레이와 비디오 카메라로 음원의 위치 등 녹음실의 음향특성을 획득하고, MPEG-4 규격에 따라 오디오 및 비디오 데이터를 부호화하여 재생 모듈에 전달하면, 재생 모듈에서 스피커의 개수 및 위치 등 재생환경과 주어진 음향특성에 따라 음장을 재현한다.

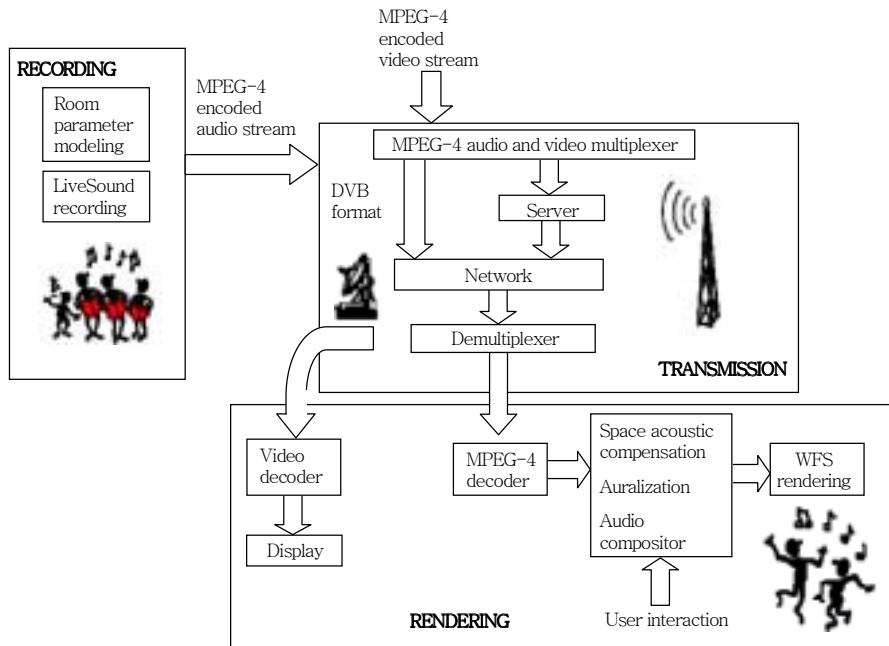
녹음 모듈에서는 음장을 표현하기에 적합한 모델을 미리 정의해 두고 추출한 음향특성을 파라미터로

서 음장을 재생하는 데에 활용한다. 파라미터 기반 음향모델은 오디오 및 비디오 데이터의 부호화 도구와 함께 객체기반 부호화 표준인 MPEG-4 규격 중 일부를 취한 것으로서 새로운 프로파일에 해당한다.

재생 모듈에서는 음향모델을 사용자의 의도에 따라 제어할 수 있는 수단을 제공하고, 오디오 및 비디오 데이터를 음향모델에 따라 재현하는 기능을 수행한다. 음장의 시간적 특성뿐만 아니라, 마이크 어레이를 이용해 획득한 공간적 특성까지 충실히 재현하기 위해 WFS(Wave Field Synthesis) 기술을 이용한다.

CARROUSO 프로젝트에서 WFS를 재생기술로 채택함으로써, 극장 등에서 특정 지점을 제외한 영역에서는 음상이 깨어지는 문제점을 크게 완화하였으며, HRTF(Head-Related Transfer Function)에 기반을 둔 기존의 재생기술을 이용하는 경우보다 음상정위의 정확도를 높였다[10]. 이로 인해 홈 씨어터, 모의비행 시스템, 원격회의의 시스템 등에 응용하여 고품질의 AV 서비스를 제공할 수 있게 되었다.

결국 CARROUSO 프로젝트는 MPEG-4 시스템



(그림 9) CARROUSO 시연 시스템의 구조

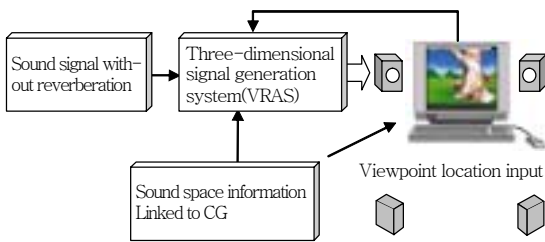
과 WFS 기술을 핵심기술로 활용하여 대화형 3차원 오디오 서비스의 가능성을 보여주었다는 데 의의가 있다.

2. 일본 NHK VRTV

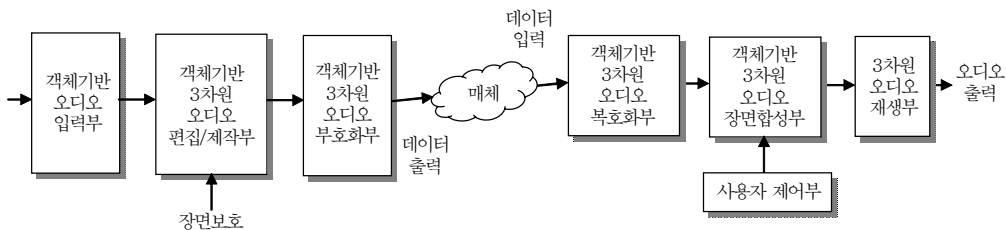
일본 NHK의 STRL(Science & Technical Research Laboratory)은 기존의 수동적 TV에 사용자 제어기능을 첨가함으로써 현장감을 제고한 장치를 VR(Virtual Reality) TV로 명명하고, 이를 구현하기 위한 연구를 수행하였다. 그 결과 (그림 10)과 같이 주어진 음원의 위치와 공간의 잔향특성에 맞추어 원음을 처리 및 재생해 주는 VR 오디오 시스템을 개발하였다.

NHK는 VR 오디오 시스템을 어린이 교육에 응용하여 숨바꼭질 프로그램을 선보였다. 시청자는 숨어 있는 캐릭터가 내는 소리를 듣고 캐릭터를 찾아내는 것이 목표인데, 캐릭터와 시청자 사이의 상대적인 위치와 잔향특성을 고려하여 현실세계에서 시청자가 듣게 될 소리를 충실히 재현하는 데 VR 오디오 시스템을 이용하였다.

다시점(multiview) 비디오 기술과 대화형 3차원 오디오 기술을 방송분야에 응용해 현실감을 제고함



(그림 10) VR 오디오 시스템의 구조



(그림 11) 객체기반 3차원 오디오 방송 시스템의 구조

으로써 TV의 진화 가능성을 보여준 연구이다.

3. 객체기반 3차원 오디오 방송 시스템

2002년부터 지능형정보방송(SmarTV) 프로젝트의 일환으로 ETRI에서 개발중인 객체기반 3차원 오디오 방송 시스템은 (그림 11)에서 나타낸 것과 같이 DMB(Digital Media Broadcasting) 및 디지털 방송에서, MPEG-4 표준을 기반으로 객체기반의 사용자 제어가 가능한 오디오 콘텐츠를 제작하고 부호화하여 전송하며, 수신된 콘텐츠를 3차원 공간상에서 장면을 재구성하고 재생하며, 사용자 제어에 의해 오디오 객체들의 위치 및 특성에 관련된 속성들을 제어하는 기능들을 제공하는 시스템이다. 본 시스템은 3차원 오디오 단말에서의 대화형 콘텐츠, 관련 이미지 혹은 영상 데이터를 부가하여 향상된 멀티미디어 서비스를 제공하기 위한 것이다.

본 시스템의 방송에의 응용 분야로서는 3차원 오디오와 대화형 오디오 서비스를 접목한 서비스라고 할 수 있으며, 3차원 TV 방송, 멀티뷰 방송의 오디오, 가상 전시장의 대화형 객체, AUI(Auditory User Interface) 등 다양한 분야에 응용할 수 있다.

V. 결론

지금까지 방송, 통신 등에 활용될 수 있는 객체기반 3차원 오디오의 개념 소개와 기술 동향을 살펴 보았다. 3차원 오디오 기술은 우리가 일상적으로 경험하고 있는 자연스러운 3차원 오디오의 느낌을 그대로 방송, 통신, 영화 등의 매체를 통해서 재현하고자 하는 것이 그 궁극적인 목적이라고 할 수 있다.

이러한 3차원 오디오 기술은 기본적인 인간의 욕구를 충족시켜 주는 것이라고 할 수 있으며, 사용자들의 요구사항도 보다 자연스럽게 깨끗한 원음을 추구하는 이러한 사항들이 자연스럽게 반영되고 있다.

3차원 오디오의 연구는 오랫동안 발전되어 왔고, 다양한 분야에 응용되고 있다. 그러나 현실적으로 우리가 자주 접하는 매체인 영화, 방송, 음반, 통신 등에 있어서는 아직 3차원 오디오를 적용하기에는 여러 가지 문제점이 따른다. 가장 큰 문제점은 이러한 장비들을 사용하는 형태가 매우 다양하여 실제 3차원 오디오의 효과를 제대로 실현하기가 어려우며, 이러한 문제점을 해결하기 위한 비용이 아직은 매우 높다는 것이다. 현실적으로는 멀티채널에 의한 서라운드 사운드를 3차원 오디오의 대응으로 사용하고 있지만 영화 및 뮤직비디오 외에는 사용자들에게 특별히 호응을 받지 못하고 있는 실정이다.

이러한 3차원 오디오 기술은 결국에는 실감 통신 및 방송 분야에 응용될 것이라는 예측이 가능한데, 이를 위해서는 현장의 3차원 오디오 환경을 그대로 모델링하여 원격지에서 정확하게 재현하는 기술이 필요하며, 사용자와의 인터랙션을 통해 가상현실과 같은 체험이 가능하도록 하여야 한다. 그러나 무엇보다도 중요한 것은 현실적으로 적용 가능한 3차원 오디오 기술들을 적절한 분야에 사용함으로써 3차원 오디오의 효과를 많은 사용자들이 경험할 수 있도록 하여야 하며, 이를 통하여 3차원 오디오의 응용 기술의 발전을 꾀하고 사용자들의 요구사항을 직

접 수렴할 수 있는 기반을 마련하는 것이다.

참 고 문 헌

- [1] Fransis Rumsey et al., *Sound and Recording*, Focal Press, 2003.
- [2] Taejin Lee et al., "3D Audio Acquisition and Reproduction System Using Multiple Microphones on a Rigid Sphere," *116th AES Convention*, May 2004.
- [3] <http://www.schoeps.de/E/kfm6.htm>
- [4] <http://www.theaudio.com/bs-3d.html>
- [5] Hareo Hamada et al., "3D Sound Generation Using Two Loudspeakers - Stereo Dipole System and Its Applications," *AES 15th Int'l Conf.* Oct. 1998.
- [6] Marinus M. Boone, "Acoustic Rendering with Wave Field Synthesis," *Proc. ACM Siggraph and Eurographics Campfire*, Snowbird, Utah, May 26-29, 2001.
- [7] 장대영, 서정일, 이태진, 강경옥, "객체기반 3차원 오디오 방송 시스템," 2002년도 한국음향학회 하계학술발표대회 논문집, 창원, 2002. 7.
- [8] ISO/IEC JTC1/SC29/WG11, *Generic Coding of Moving Pictures and Audio: Systems, Amendment 3: AudioBIFS Extensions*, Dec. 2003.
- [9] Jean-Francois Cardoso, "Blind Signal Separation: Statistical Principles," *Proc. IEEE*, Vol. 86, No. 10, Oct. 1998.
- [10] ISO/IEC 14496-1, *Information Technology-Coding of Audio-Visual Objects: Systems*, International Standard, 1999.