객체기반 실감음향 기술 개발

장대영, 이태진
한국전자통신연구원 오디오연구실
dyjang@etri.re.kr, tjlee@etri.re.kr

A Study on Object-based Realistic Audio

Daeyoung Jang, Taejin Lee
ETRI, Audio Lab.

요 약

본 논문에서는 기존의 채널기반의 오디오 기술에 대해 다양한 서비스가 가능하고, 재생환경에 독립적인 객체기반 실감음향 기술에 대해 논하고자 한다. 현재, 극장 사운드를 중심으로 객체기반 오디오 기술이 적용된 사운드가 점차 확산되고 있으며, 미국, 유럽 등 차세대 방송용 오디오에 객체기반 오디오 기술의 도입을 적극적으로 고려하고 있다. 객체기반 오디오 기술은 콘텐츠의 제작단계에서 재생환경을 고려할 필요가 없고, 현장의 음향을 신호와 3 차원 공간 정보로 구분하여 음향 공간의 정보를 그대로 표현함으로써, 재생환경에서는 3 차원 공간 정보를 활용하여 다양한 3 차원 음향 재생 기술을 활용하여 재생할 수 있다. 이러한 객체기반 실감음향 기술 개발을 위해서는 편리한 제작 및 3 차원 공간 정보 표현 기술이 필요하며, 청취환경에서는 객체기반 실감음향 콘텐츠를 제작자의 의도대로 렌더링할 수 있는 재생 및 제어 기술이 필요하다. 이에 객체기반 실감음향 기술의 기술동향과 객체기반 실감음향 서비스를 위한 콘텐츠 표현/제작 및 재생 기술에 대하여 고찰해 보고자 한다.

1. 서론

실감음향 기술은 오랜 역사를 통하여 점차 발전하여 왔으나, 최근 들어, 채널을 늘리는 방법 외에는 실감음향 성능을 획기적으로 개선할 방법이 마땅하지 않았다. 5.1 채널, 7.1 채널 오디오 이후에 9.1, 10.2, 15.1, 22.2, 31.1 채널 등 다양한 채널 방식들이 제안되어 보다 몰입감 있는 실감음향을 재생하기 위해 노력하였지만, 문제는 콘텐츠를 제작하는 일이 복잡해짐에 따라 콘텐츠 확보 및 서비스에 한계가 드러나면서 어느 방식 하나 주도권을 잡지 못하는 실감음향의 춘추전국 시대를 맞이하게 되었다[1].

이러한 국면에 하나의 커다란 파문을 일으키는 실감음향 방식이 객체기반 오디오 방식이다. 돌비의 애트모스, DTS 의 MDA(Multi-Dimensional Audio)에 이어 IOSONO 를 인수한 Barco 의 AuroMax 까지 가세하면서 또다른 오디오 방식 전쟁을 예고하고 있다.

객체기반 오디오 기술은 이전에도, MPEG-4 의 개념에 포함되어 있었고, IOSONO 는 객체기반 오디오 기술과 WFS(Wave Field Synthesis) 기술을 접목하여 실감음향 제작 및 재생 기술을 상업화 하였으나 많은 스피커 개수에 의한 설치비용의 부담으로 크게 확산되지는 못하였다. 한국전자통신연구원과 ㈜오디즌에서도 객체기반 오디오 기술을 음악에 적용하여 대화형 음반인 Music2.0 기술을 상용화하였으나, 콘텐츠 확보에 어려움이 있어 활성화되지는 못하였다.

본 본문에서는 이러한 객체기반 오디오 기술의 활성화에 필수적인 객체기반 오디오 콘텐츠 표현 및 제작 기술과 객체기반 실감음향의 렌더링 및 인터랙션 서비스를 위한 제어 기술 등에 대해 개발 방향 및 전망을 고찰해 보고자 한다.

본 논문의 구성은 다음과 같다. 2 절에서는 객체기반 오디오 기술 동향에 대해 살펴본 후, 3 절 및 4 절에서는 객체기반 오디오 서비스의 활성화를 위한 객체기반 실감음향 콘텐츠의 획득/표현 및 제작 기술과 재생 및 제어 기술의 개발 방향 및 가능성에 대해 고찰한다. 마지막으로 5 절에서는 객체기반 실감음향 기술의 전망 및 차세대 음향 산업으로서의 가능성에 대해 고찰한다.

2. 객체기반 오디오 기술 동향

객체기반 오디오 기술은 20 세기 말 MPEG-4 로부터 본격적으로 출발되었다. MPEG 오디오 기술을 주도하고 있는 독일 FhG 에서 Euro Project CAROUSO(Creating, Assessing and Rendering in Real-time Of high-quality aUdio-viSual envirOnments[2001~2003])[2]의 객체기반 3 차원 음향 결과물을 활용하여 창업한 IOSONO 는 객체기반 오디오 방식과 WFS(Wave Field Synthesis) 기술을 적용하여 수평면을 커버하는 스피커어레이를 통하여 실감음향을 재생하는 기술을 상용화하였다. IOSONO 시스템은 초기에 극장 및 테마파크에 적용되어 호평을 받았지만, 상대적으로 막대한 설치비용을 부담스럽게 생각하는 극장주들에게 효과적으로

어필하지 못하였고, 더 이상 확산시키지는 못하였다.

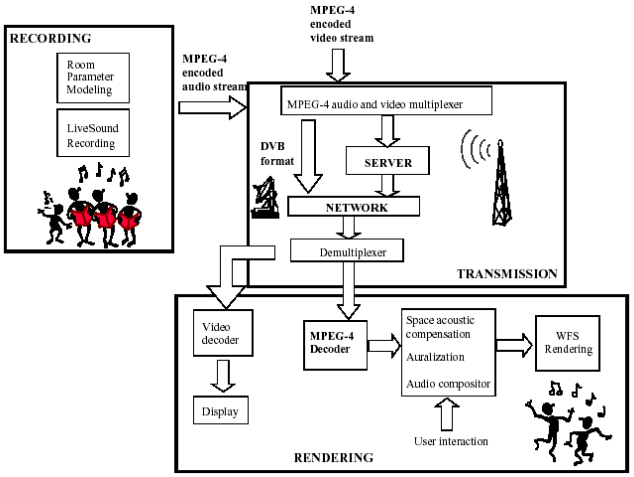

그림 1. CARROUSO 시스템의 개념도

한국전자통신연구원에서는 2002 년부터 객체기반 오디오 기술을 연구하고 있으며[3], 2007 년 ㈜오디즌을 통하여 객체기반 오디오 기술을 음악에 적용한 대화형 음악 서비스, Music2.0 기술을 상용화하였고, 이를 기반으로 MPEG-A IMAF(Interactive Music Application Format) 표준화를 주도하는 한편, MPEG-H 3D Audio 표준화에도 객체+채널 오디오 기술을 제안하며 적극적으로 참여하고 있다. Music2.0 서비스는 일부 매니아 층을 확보하는 성과를 거두기도 했지만, 저작권 협상이 쉽지 않아 콘텐츠를 확보하는데 어려움을 겪으며, 서비스를 더 이상 확산시키지는 못하였다.

그림 2. Music2.0 멀티트랙 음반 서비스 화면

영국 BBC 에서는 윔블던 테니스 실황을 해설 음성과 배경음이 구분된 객체기반 오디오 방식의 방송을 인터넷을 통해 중계하면서, NetMix 라는 툴을 이용하여 사용자가 음성과 배경음을 적절히 조정하여 선호도를 조사한 결과, 각자의 임의대로 조정한 소리를 더 선호하는 경향을 발표한 바 있다.

한편, SRS 를 인수한 DTS 는 객체기반 오디오 포맷인 MDA 를 확산시키기 위한 연합체를 구성하여 2012 년 1 월 CES 에서 공개하였으며, 이후 곧 바로 돌비는 유사한 포맷인 애트모스를 전용 콘텐츠 제작도구, 극장용 오디오 프로세서와 함께 발표하였고, 애트모스 콘텐츠를 제작하여 공급하면서 한발 앞서 상용화에 성공하였다.

애트모스는 특히 그림 3 의 Objects 로 표현된 객체기반 오디오 방식만을 사용하지 않고, 배경음에 대하여 그림 3 의 Beds 로 표현된 채널기반 오디오 방식을 함께 사용함으로써, 객체기반 오디오 방식의 콘텐츠 제작 및 렌더링의 어려움을 다소 해결할 수 있는 절충안을 명시적으로 제시하였는데[4], 이전의 기술들도 완전한 객체기반 오디오 기술의 구현상 어려움을 채널과 객체를 함께 사용하는 방식을 통해 해결하고 있었음을 알 수 있다.

그림 3. 돌비 애트모스 방식의 개념

2015 년 4 월 CinemaCon 전시회에서는 IOSONO 의 3 차원 음향 렌더링 기술을 적용한 Barco 의 객체기반 오디오 방식인 AuroMax 가 공개됨으로써 객체기반 오디오 방식의 주도권을 차지하기 위한 일대 격전이 예고되고 있다. 이러한 가운데, 오디오 콘텐츠 포맷의 주도권을 확보하기 위해서는 콘텐츠 및 상영관의 확보가 무엇보다도 중요하다는 것을 알 수 있으며, 보다 편리한 콘텐츠 제작 기술 및 우수한 렌더링 기술을 통하여 새로운 오디오 방식의 시대에 새로운 기술 경쟁력을 확보할 수 있을 것으로 전망된다. 이에 3 절과 4 절을 통하여 객체기반 실감음향 획득/표현 및 제작, 재생 및 제어 기술에 대한 개발 방향을 고찰해 보고자 한다.

3. 객체기반 실감음향 획득 및 제작 기술

객체기반 실감음향 콘텐츠는 채널기반 오디오 콘텐츠가 채널 수에 따른 채널 신호만을 가진 것에 비해, 객체 수에 따른 객체 신호와 함께 객체 신호가 재생되어야 할 공간 정보를 객체 신호마다 가지는 것이 특징이다. 이러한 공간 정보에는 방향 및 거리를 가장 중요한 요소로 고려할 수 있으며, 그 외에도 음원의 모양 및 지향성 등을 고려할 수 있다.

이러한 실감음향 콘텐츠를 획득하고, 제작하는 방법에 대해서 고려할 수 있는데, 실감음향 콘텐츠를 획득하는 방법은 현장에서 획득하는 방법과 스튜디오에서 획득하는 방법으로 구분할 수 있다. 일반적으로 방송의 경우, 현장에서 획득하는 사례가 많으며, 영화의 경우, 현장에서의 녹음은 일부분이며, 대부분 스튜디오에서 다시 녹음하거나, 음원 라이브러리에 다양한 효과를 적용하여 소리를 만들어 낸다.

현장에서 획득하는 경우, 각 객체 신호를 구분하여 획득할 필요가 있으며, 이와 함께 각 객체 신호의 공간 정보를 함께 획득하여야 한다. 객체 신호를 획득하는 방법은 근접 마이크를 이용한 객체 신호 획득 방법, 초지향성 마이크로폰을 이용한 객체 신호 획득 방법, 그리고, 음원분리 기술에 의한 객체 신호 획득 방법이 있으며, 이들 기술을 적절히 혼용하여 상황에 따라 객체음원을 획득할 수 있다. 객체 신호의 공간 정보 획득은 마이크로폰 어레이에 의한 음원 추적 기술, 영상 정보 분석 기술 등에 의해 획득할 수 있으며, 이러한 공간정보는 적절한 시간으로 샘플링 하여 기록되어야 한다.

스튜디오에서 획득하는 경우에는, 영상의 내부뿐만 아니라 영상 외부 공간의 소리들을 제작자가 영상과 어울리도록 상상하여 제작하게 되는데, 이때 객체 신호와 함께 객체 신호의 패닝 혹은 렌더링 정보를 별도로 저장하여 객체 신호의 공간 정보로 활용할 수 있다. 다만, 하나의 영상 콘텐츠에는 수많은 객체의 소리들이 포함되어 있기 때문에 각 객체 신호를 패닝하고 렌더링하기 위해서는 사용자 입장에서의 편리한 사용자 인터페이스가 요구되며, 기존의 제작 시스템에서의 효과 및 렌더링은 미리 설정된 공간 정보에 따라 자동으로 처리되도록 하는 것이 효과적이다. 즉, 기존의 제작 시스템은 룸의 크기, 공간의 음향특성 등 공간 정보를 시스템에서 알지 못하고, 자동으로 처리될 수 없기 때문에 제작자가 모든 신호의 편집을 직접하였으나, 객체기반 실감음향 콘텐츠 제작에 있어서는 미리 설정된 공간 정보를 통해 자동으로 효과를 적용하는 것도 시도해 볼 수 있다.

4. 객체기반 실감음향 재생 및 제어 기술

3 절에서 기술한 방법으로 제작된 객체기반 실감음향 콘텐츠를 재생하기 위해서는 객체 신호를 함께 전송된 공간 정보에 따라 정확한 위치에 렌더링해 주어야 한다. 렌더링하는 방법은 스피커 개수, 스피커 배치, 청취환경, 사용자 위치, 헤드폰 청취환경 등에 따라 정해져야 하며, 다양한 3 차원 음향 처리 기술이 적용되어야 한다.

먼저 스피커에 의한 재생에 있어 가장 전형적인 방법은 5.1 채널, 7.1 채널 등 멀티채널 스피커에 의한 렌더링 방법을 생각할 수 있다. 그러나, 5.1 채널, 7.1 채널 스피커 배치는 패닝에 의한 렌더링 방법으로는 수평면 상에만 음상을 재현할 수 있기 때문에, 위 혹은 아래에 있는 객체 신호를 렌더링하기 위해서는 다른 방법들을 사용하여야 한다. 이에 바이노럴 렌더링의 수직방향 음상정위 방법을 사용할 수 있을 것이다.

멀티채널 스피커에 의한 렌더링에 있어 한가지 고려할 점은 청취환경에 따라 각 스피커의 배치가 달라 질 수도 있다는 것이다. 이러한 문제는 멀티채널 스피커의 전형적인 배치가 아닌 실제의 스피커 배치 정보를 활용하여 렌더링하면 해결될 수 있으며, 이를 위해서는 스피커 배치 정보 및 청취자의 위치 정보를 포함하여 청취환경의 정보를 시스템에 입력하는 수단이 필요하게 된다.

스피커에 의한 재생에 있어 스피커어레이에 의한 음장합성(WFS) 기술을 이용할 수 있는데, 음장합성은 기본적으로 객체 신호와 객체 신호의 위치 정보를 이용하여 음장을 합성하므로, 객체기반 실감음향의 렌더링에는 가장 적합한 방법이라고 할 수 있다. 그러나 앞에서도 언급했듯이 100 개 이상의 스피커와 이를 구동하기 위한 증폭장치를 필요로 하므로 현실적으로 활용분야는 줄어들 수 밖에 없다.

헤드폰 청취 환경을 고려하면, 일반적인 헤드폰 청취 방법으로는 두 귀 사이 즉, 머리 내부에서만 맴도는 소리만을 재생할 수 있다. 그러나, 인간의 입체음향을 인지하는 청각 시스템의 특성을 활용하는 바이노럴 렌더링을 활용하면, 객체 신호의 공간 정보에 따라 제작자의 의도에 부합되는 실감음향을 재생할 수 있게 된다. 기존의 바이노럴 렌더링에 의한 입체음향 재생의 경우, 채널기반 오디오 신호를 각 스피커의 위치에 렌더링하게 되는데 이때, 각 객체 신호는 여러 개의 스피커에 복제되어 재생되므로 각 스피커 신호 간 상호상관이 높아 지며, 이에 의한 입체음향 효과가 반감되는 단점이 있었다[5]. 그러나, 객체기반 오디오 신호의 경우, 각 객체 신호를 고유의 필터를 사용하여 렌더링함으로써 입체음향 효과를 향상시킬 수 있다.

객체기반 실감음향의 재생에 있어 각 객체 신호는 구분되어 있으므로 재생단에서 선택하여 조정할 수 있는 큰 이점이 있다. 그 중에서도 가장 효과가 크면서도, 간단한 적용 방법이 음성강화 기능이라고 할 수 있는데, 이는 음성과 배경음의 레벨을 상대적으로 가감함으로써, 상황에 따라 음성을 더 강조하여 명료성을 높일 수도 있고, 배경음을 강조하여 현장의 분위기를 더 잘 느낄 수 있도록 조정할 수 있다.

이러한 객체 신호의 제어를 위해서는 객체 신호를 선택하는 과정이 필요한데, 시각적으로 매칭되지 않는 음향 신호를 선택하는 것은 매우 번거롭고 어려운 일이다. 객체기반 실감음향의 충실한 재생도 중요하지만, 객체 신호의 인터랙션을 위한 효과적인 사용자 인터페이스를 고안하는 것도 매우 중요한 일이다.

5. 결론

지금까지 객체기반 실감음향 기술의 동향과 객체기반 실감음향 콘텐츠 서비스의 활성화를 위한 콘텐츠의 획득 및 제작과 재생 및 제어 기술의 개발 방향에 대하여 몇가지 사례를 가정하여 고찰해 보았다.

영화 산업에서는 차세대 오디오 기술로서 돌비의 애트모스, DTS 의 MDA, Barco 의 AuroMax 가 치열한 주도권 경쟁을 벌이고 있으며, 향후 영화용 오디오 포맷으로 자리잡을 것이 확실시 되고 있다. 방송 산업에서도 미국의 ATSC 3.0 표준화에 객체기반 실감음향 기술이 포함되어 있으며, 유럽의 DVB 에서도 차세대 UHDTV 를 위한 오디오 기술에 객체기반 실감음향 기술을 도입하여야 한다는 공감대가 형성되고 있으며, 이를 반영하여 MPEG-H 3D Audio 표준은 객체를 포함하는 실감음향의 압축 부호화뿐만 아니라 이례적으로 오디오 채널 포맷 변환 및 바이노럴 렌더링 기술을 표준화 아이템에 포함시켜 표준화를 진행하고 있다[6].

객체기반 실감음향 기술은 기존의 실감음향 효과를 개선할 수 있는 가능성이 높으며, 그 동안 채널기반의 오디오 콘텐츠 제작 기술과 재생 기술이 서로 종속적 관계를 통해 함께 발전해 왔지만, 산업과 시장의 융통성 측면에서 많은 제한사항을 가질 수 밖에 없었던 오디오 시장이 객체기반 실감음향 기술을 통해 제작 기술과 재생 기술이 종속성을 탈피함으로써 각각 독립적으로 발전할 수 있는 계기가 마련되었다는 측면에서 향후 오디오 시장의 획기적인 변화가 오리라는 추측도 해 볼 수 있다. 객체기반 실감음향 기술은 차후 음향뿐만 아니라 영상 및 오감 기술의 객체기반 서비스의 시발점이 될 것이라 조심스럽게 예측해 본다.

감사의 글

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신.방송 연구개발 사업의 일환으로 수행하였음[R0126-15-1034, 채널/객체 융합형 하이브리드 오디오 콘텐츠 제작 및 재생기술 개발].

[참고문헌]

[1] 장대영, 서정일, 이태진, 강경옥, “초고해상도(UHD) 사운드 기술의 현재와 미래”, 방송공학회지, 2012. 10.

[2] European project CARROUSO Deliverables, http://www.emt.iis.fhg.de/projects/carrouso

[3] Daeyoung Jang, Jeongil Seo, Kyeongok Kang, Hoe-Kyung Jung, “Object-based 3D Audio Scene Representation”, 115th AES Convention Paper, 2003. 10.

[4] Dolby ATMOS White Paper, Next-Generation Audio for Cinema, Dolby Laboratories Inc., 2012.

[5] Daeyoung Jang, Jeong-pyo Hong, Hoe-Kyung Jung, Kyeongok Kang, “Center Channel Separation Based on Spatial Analysis”, 11th Int. Conference on Digital Audio Effects(DAFx-08), 2008. 9.

[6] J. Herre, J. Hilpert, A. Kuntz, J Plogsties, “MPEG-H 3D Audio - The New Standard for Coding of Immersive Spatial Audio”, IEEE Journal of Selected Topics in Signal Processing, 2015