

3DTV 향 3D 영상 정보를 이용한 3D 오디오 원근감 재현 기술

김선민, *이영우, **김승훈, ***이승수

삼성전자, DMC 연구소

{sunmin21.kim, *ywbme.lee, **sh21.kim, ***seungsu.lee}@samsung.com

3D Audio Rendering Method based on 3D Video Information for 3DTV

Sunmin Kim, *Young Woo Lee, **SeungHun Kim, ***SeungSu Lee

DMC R&D Center, Samsung Electronics Co. LTD

요 약

본 논문에서는 3DTV 의 입체감 향상을 위한 3D 음향의 원근감 재현 기술을 제안한다. 먼저, 3D 영상 객체의 깊이를 추출하고 영상 객체의 깊이에 따라 오디오 객체의 거리감을 조절한다. 오디오 거리감 재현을 위해 필요한 오디오 깊이 인자는 3D 영상의 좌/우 이미지의 차이 정보로부터 오디오에 맞도록 비선형 변환을 통해 구해진다. 3D 오디오 재현 알고리즘은 기존의 서라운드 입체음향 기술과 원근감 재현 기술로 구성된다. 원근감 재현 기술은 추정된 오디오 깊이 인자에 따라 신호크기, 초기 반사음, 근거리 머리전달함수, 위상 제어를 통해서 구현된다. 특히, 3D 영상 객체가 화면 앞으로 튀어 나올 때 소리가 튀어나오도록 함으로써 3D 영상 객체와 연동되는 입체 음향을 효과를 통해 3D 방송 시청 시 오디오/비디오 입체감을 향상시켜준다. 상용화된 3DTV 를 활용하여 음질 평가 전문가들의 주관 청취 평가를 통해 제안한 원근감 재현 기술이 3D 방송 시청에 적합함을 검증한다.

1. 서론

3D 영화에 대한 관심이 증가하면서 3D 방송에 대한 표준화가 진행 중이며 주된 논의는 3D 영상에 초점이 맞추어져 있고 3D 음향에 대한 논의는 미비한 실정이다. ITU-R BS.775-2 [1]에서 권장하는 7.1 채널 스피커 시스템은 수평면상의 입체감만을 보장하므로 차세대 스피커 레이아웃에 대한 연구는 고도각 스피커 배치를 통해 수직적 입체음향을 재현하는 연구에 초점을 두고 있다. USC 10.2 채널 [2], NHK 22.2 채널 [3], 삼성 10.2 채널 [4] 시스템 등이 발표되고 있으나 다채널 콘텐츠 부족 및 설치 용이성 측면에서 널리 상용화되고 있지 않은 실정이다.

따라서, 이러한 다채널 스피커 레이아웃의 가정 내 설치 어려움으로 적은 수의 스피커로 가상 입체음향을 생성하는 기술에 대한 필요성이 증대되고 있다. 머리전달함수를 기반으로 2 개의 스피커를 사용하여 수평면 상에서 가상의 스피커를 생성하는 기술 [5-7], 청취 위치의 제약을 극복하기 위해 청취자의 위치를 추정하여 입체음향을 적응적으로 생성해 주는 기술 [8], 수평면 스피커로부터 가상 고도감을 생성하는 기술 [9-10] 등이 연구되어 왔다. 이러한 기술들은 TV, Sound Bar 홈씨어터 등 2 개의 스피커를 사용하는 오디오 시스템의 가상 입체음향 생성 기술로 널리 활용되어 왔다.

최근 들어, 3D 영상 콘텐츠 및 3DTV 에 대한 관심이 급증하면서 3D 음향 기술에 대한 필요성이 증대되고 있다. 3DTV 를 통하여 3D 콘텐츠를 시청할 경우, 3D 영상과

스테레오 스피커를 통해 재생되는 음향 간의 괴리감이 존재한다. 기존의 3D 오디오 기술은 좁은 사운드 스테이지를 넓혀주는 가상 입체음향 기술을 중심으로 발전되어 왔기 때문에 영상 객체와 오디오 객체간의 거리 괴리감을 해결하기 위해서는 3D 영상에 적합한 새로운 방식의 3D 오디오 기술 개발이 필요하다. 영상의 경우 2D 에서 3D 로 바뀌면서 입체감 측면에서 매우 큰 발전을 하고 있는 반면, 오디오의 경우, 기존 기술에서 크게 달라지지 않는 실정이다. 3D 영상에 적합한 오디오 코덱 기술로는 오디오 객체를 각각 분리할 수 있는 Spatial Audio Object Coding [11]을 들 수가 있으나 역호환성 (backward compatibility) 측면에서 2 채널까지만 지원한다. 설령, 오디오 객체가 각각 분리되어 있다고 하더라도 3D 영상에 부합되도록 오디오 객체를 처리하는 원근감 재현 기술 개발이 필요하다.

본 논문에서는 영상 객체가 앞으로 튀어 나오고 뒤로 들어가는 것과 부합되는 새로운 방식의 3D 오디오 재현 기술을 제안한다. 오디오 깊이 정보 추정 알고리즘은 3D 영상의 시차 (Disparity)를 기반으로 오디오 객체가 튀어나와야 하는 타이밍과 튀어나오는 정도를 결정하는 오디오 깊이 인자 (Depth Index)를 추정하고 이 인자를 바탕으로 오디오 객체의 거리를 조절하여 3D 영상의 시청 몰입감을 증대시켜준다. 그림 1 (a)는 기존의 서라운드 입체음향 기술, (b)는 본 논문에서 제안하는 새로운 방식의 3D 오디오 기술의 개요를 보여준다. 그림에서 보여지는 것과 같이 영상 객체의 거리와 연동되는 오디오 객체의 원근감이 재현되는 것이 특징이다.



(a) 기존 입체음향 기술 (b) 제안한 3D 오디오 기술

그림 1. 제안한 입체음향 기술 개요

2. 오디오 깊이 정보 추정 알고리즘

2.1. 3D 영상의 시차 정보 추정 기술

본 절에서는 3D 영상의 시차 영상 (Disparity Map) 추정 기술을 설명한다 [12-13]. 그림 2 에서 보는 것과 같이 왼쪽 영상은 기준으로 오른쪽 영상의 픽셀을 이동하며 두 영상의 차이를 계산하고 두 영상의 차이가 가장 적은 값을 가지는 오른쪽 영상의 위치로부터 시차 값을 구한다. 영상 객체가 시청자에게 가장 가까울 때의 최대 시차 값은 192 로 한정한다. 피로감이 발생하지 않는 Parallax 각도인 1 도에 해당하는 시차 값은 56 이므로 최대 시차 값이 0 에서 56 까지는 쾌적 입체 시청 구간, 56 에서 84 까지는 피로감이 발생하는 구간, 84 이상의 시차 값에서는 인위적으로 3D 효과 부여를 위한 Pop-up 효과 구간이라고 볼 수 있다. 3D 영상의 최대 시차 값은 매 프레임마다 실시간으로 오디오 DSP 에 전달되며 다음 절에서는 오디오 깊이 인자 추정 기술을 제시한다.

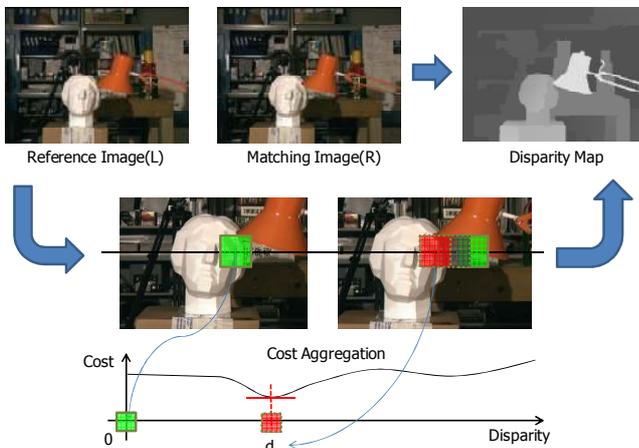


그림 2. 3D 영상의 시차 영상 추정 기술 개요

2.2. 오디오 깊이 인자 추정 기술

3D 영상의 쾌적 입체 구간에서는 자연스러운 음향 원근감을 재현하고 3D 영상의 Pop-up 구간에서는 강한 음향 원근감 효과를 재현하기 위해 그림 3 과 같은 비선형 스케일링을 통해 오디오 깊이 인자를 추정한다. 영상 시차 값이 28 보다 작은 값에서는 오디오 깊이 인자 값을 0 으로 설정하고 124 보다 큰 구간에서는 최대 값인 192 로 설정한다. 그 중간의 시차 값에서는 선형적으로 계산된다. 순간적으로 변하는 영상을 따라가기 위해 오디오 깊이 인자는 영상 시차 값이 커지는 조건에서는 빨리 추종하도록 하고 소리의 자연스러움을 위해서 영상 시차 값이 작아지는 조건에서는 천천히 추종하도록 오디오 깊이 인자를 스무딩한다. 그림 4(a)는 3D

영상의 최대 시차 값의 일 예(Monsters vs Aliens)를 보여주고 그림 4(b)는 최대 값이 1 로 정규화된 오디오 깊이 인자를 보여준다.

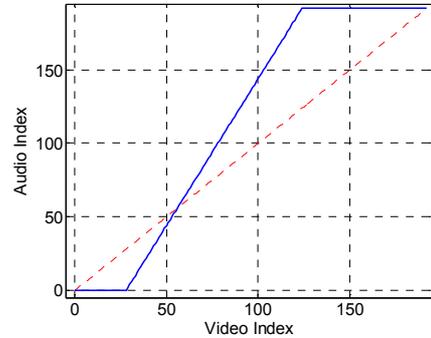
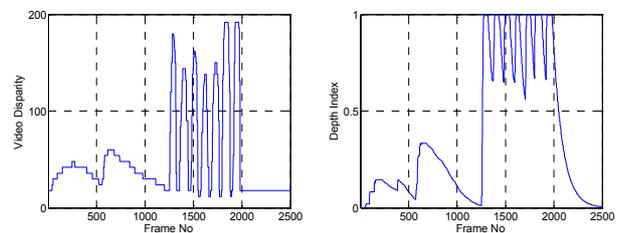


그림 3. 영상 시차 정보의 오디오 깊이 인자 변환



(a) 영상 최대 시차 값 (b) 추정된 오디오 깊이 인자

그림 4. 프레임 별 오디오 깊이 인자 변환

3. 3D 오디오 재현 알고리즘

3.1. 원근감 재현 기술

본 절에서는 2.2 절의 오디오 깊이 인자를 기반으로 오디오의 원근감을 조절하는 기술을 제시한다. 음원의 거리감은 신호 크기, 초기 반사음 특성, 근거리 머리전달함수 (head-related transfer function) 특성 등을 조절함으로써 구현된다 [14-16]. 원거리에 대한 거리감은 초기 반사음 특성 (EDT, C80) 조절 [16]에 의해, 근거리에 대한 거리감은 근거리 머리전달함수의 특성에 의해 나타나는 음향학적 시차 (Auditory Parallax) 모델 [14, 15]에 의해 설명이 된다. 음향학적 시차 모델에 의하면, 음원의 위치 거리에 따라 음원과 두 귀 사이의 음향학적 특성이 바뀌게 된다. 그림 5 와 같이 음원이 정면 0 도에 있더라도 거리가 가까워지면 귀의 위치에서의 실질적인 음원의 각도가 커지게 됨을 알 수 있다.

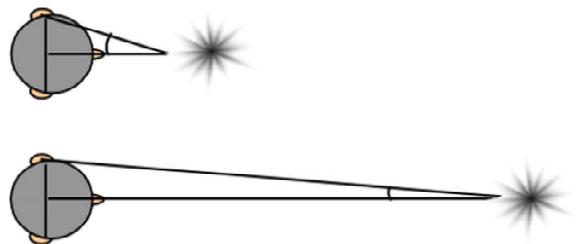


그림 5. 음원 거리에 따른 음향학적 시차

예를 들어, 머리 직경을 18cm 라고 가정하면 음원이 정면 0 도, 20cm 에 위치하게 되면 왼쪽 귀에서 바라본 실질적인

음원의 위치는 25 도가 된다. 이러한 음원 위치의 변화가 음향학적으로 미치는 영향을 보기 위해 CIPIC HRTF [17] 데이터를 사용하여 주파수 특성을 관찰한다. 그림 6 은 음원이 0 도와 25 도에 있을 때 왼쪽 귀의 머리전달함수의 주파수 특성 변화를 나타낸다. 저주파 영역의 크기는 유사하나 고주파 영역의 크기는 0 도에 비해 25 도가 작음을 알 수 있다. 이는 머리에 의해서 고주파 소리가 감쇄되는 두영 효과(head shadowing)에 기인한다. 거리가 가까워질수록 귀에서 바라본 상대적인 음원의 각도가 커지기 때문에 고주파의 감쇄현상은 심화된다.

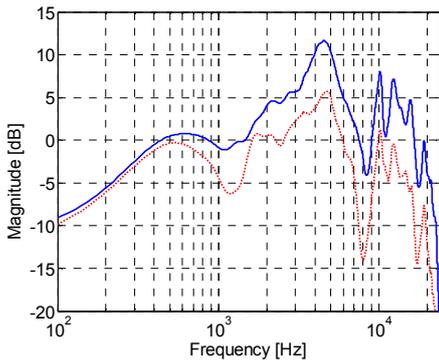


그림 6. 음원 위치에 따른 주파수 특성 (실선: 0도, 점선: 25도)

거리에 반비례하는 음압 특성을 추가로 고려하여 0 도 음원은 100cm, 25 도 음원은 20cm 일때 주파수 특성의 변화량을 그림 7 에 나타낸다. 음원이 가까워지면 저주파 대역이 고주파 대역에 비해 상대적으로 증가가 됨을 알 수 있다.

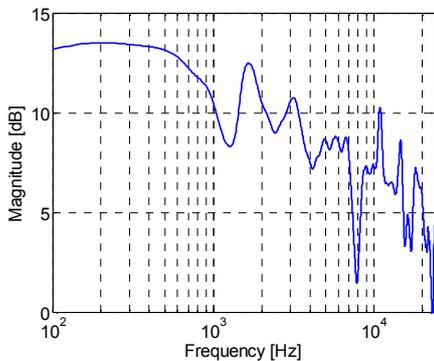


그림 7. 20cm 음원 음압에서 100cm 음원 음압을 뺀 값

위에서 언급한 음원의 거리 인지에 영향을 주는 신호 크기, 초기 반사음, 근거리 머리전달 함수 특성 등을 이용하면 2.2 절의 오디오 깊이 인자에 따라 원하고자 하는 원근감 재현이 가능하다.

3.2. 3D 영상 연동 3D 오디오 재현 기술

3D 영상의 시차 정보로부터 추정된 오디오 깊이 인자를 기반으로 3D 오디오를 재현하는 기술의 블록도를 그림 8 에 제시한다. 제안된 기술은 스테레오 입력을 받아 좌/우 신호의 상관도가 높은 주요 신호를 추출한다 [18,19]. 추출된 주요 신호는 추정된 오디오 깊이 인자를 기반으로 초기 반사음 및 근거리 머리전달함수 특성이 조절된다. 또한, TV 의 좁은 사운드 스테이지를 확장하기 위해 크로스톡 제거[20]를 통해

가상 스피커를 좌/우로 넓게 배치하는 와이드 스테레오 기술을 적용한다 [7]. 와이드 스테레오 블록의 출력 값은 오디오 깊이 인자의 추정치에 따라 신호 크기가 조절 된다. 최종적으로 신호 믹싱을 통해 스테레오 신호가 TV 를 통해 출력된다.

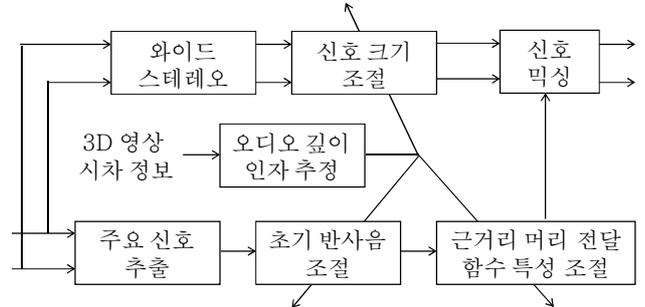


그림 8. 3D 오디오 원근감 재현 기술의 블록도

4. 주관 청취 평가

4.1. 실험 방법

3D 오디오 재현 기술의 효과를 평가 하기 위해 3D 영상 콘텐츠를 이용해 주관평가를 실시하였다. 사용된 콘텐츠는 몬스터 vs 에일리언(드림웍스) 3D 영상의 영상 객체가 튀어나오는 5 개의 클립 구간을 이용하였으며, 3D 오디오 재현 기술이 구현되어 있는 삼성 LED TV 를 이용하여 일반 가정 환경의 음향 평가실에서 진행되었다. 평가항목은 입체원근감 인지효과와 만족도, 서라운드 이미지 향상도, 수평방향 음상 정위(좌/중앙/우)이며 각 평가항목은 MOS (Mean Opinion Score) 최저점 1 에서 최고점 5 점 사이로 평가되었다. 음질 평가 전문가 3 명이 주관 평가에 참여하였으며, 원음을 두 번 청취한 후, 3D 오디오를 두 번 청취하고, 다시 원음을 한 번 청취하였고, 요청에 따라 원음 또는 3D 오디오를 반복해서 청취하였다.

본 논문에서 제안한 3D 오디오 재현 기술은 원근감 효과의 정도에 따라 3 단계(약하게/중간/강하게)로 조절이 가능하도록 개발되었고 각 단계마다 신호크기, 초기 반사음, 근거리 머리전달함수 특성의 최대 조절 정도를 달리 설정하였다.

4.2. 평가 결과

표 1 의 평가 결과에서 보는 것과 같이 세가지 평가 항목 모두 원음 대비 우수함을 보였으며, ‘약하게’ 단계는 입체감 인지 효과는 다소 약하나 자연스러우며, ‘중간’ 단계는 입체감과 공간감이 잘 표현되며, ‘강하게’ 단계는 다소 과할 수 있으나 입체감 효과는 가장 강한 것으로 평가하였다.

표 1. 주관 청취 평가 결과 (3 점: 원음과 유사)

평가항목	효과 단계		
	약하게	중간	강하게
서라운드 이미지 향상도	3.6	4.0	4.1
수평방향 음상정위 (좌/중/우)	3.7	4.0	3.8
입체원근감 인지효과와 만족도	3.5	3.9	4.0

5. 결론

본 논문에서는 3D 영상의 시차 정보를 기반으로 한 3D 오디오 재현 기술을 제안하였다. 3D 오디오 원근감 재현 기술은 추정된 오디오 깊이 인자를 이용하여 오디오 객체의 거리감 및 스테이지감을 조절하였다. 음질 평가 전문가의 주관 청취 평가 결과, 제안한 기술은 3D 영상 청취 시 몰입감을 향상시켜줄 수 있는 3D 오디오 기술임을 입증하였다.

참고문헌

- [1] ITU-R Rec. BS.775-2, "Multichannel stereophonic sound system with and without accompanying picture," International Telecommunications Union, Geneva, Switzerland (1992-1994-2006).
- [2] T. Holman, "5.1 Surround sound, Up and Running," Focal Press, Woburn (2000).
- [3] K. Hamasaki, et al., "The 22.2 Multichannel Sound System and Its Application", presented at 118th AES Convention, Barcelona, Spain (2005).
- [4] Sunmin Kim, Young Woo Lee, and Ville Pulkki, "New 10.2-channel Vertical Surround System (10.2-VSS); Comparison study of perceived audio quality in various multichannel sound systems with height loudspeakers", presented at 129th AES Convention, San Francisco, USA (2010).
- [5] Begault, D. R., 3-D sound for virtual reality and multimedia, pp 158-163, Academic Press, Inc., London (1994).
- [6] Kirkeby, O., et al., The stereo dipole - A virtual source imagining system using two closely spaced loudspeakers, Journal of audio engineering society, Vol. 46, pp 387-395 (1998).
- [7] Sunmin Kim, et al., "Virtual Sound Algorithm for Wide Stereo Sound Stage", presented at 117th AES Convention, San Francisco, USA (2004).
- [8] Sunmin Kim, et al., "Adaptive Virtual Surround Sound Rendering System for an Arbitrary Listening Position", J. Audio Eng. Soc., Vol.56, No.4, 243-254 (Apr., 2008).
- [9] Yusuke Ono, Sungyoung Kim, and Masahiro Ikeda, "Investigating perceptual effects associated with vertically extended sound fields using Virtual Ceiling Speaker", presented at 12^{9th} AES Convention, San Francisco, US (2010).
- [10] Hyun Jo, William L. Martens, Youngjin Park, and Sunmin Kim, "Confirming the Perception of Virtual Source Elevation Effects Created using 5.1 Channel Surround Sound Playback", presented at VRCAI, Seoul, Korea (2010).
- [11] Oliver Hellmuth, et al., "MPEG Spatial Audio Object Coding – The ISO/MPEG Standard for Efficient Coding of Interactive Audio Scenes", presented at 129th AES Convention, San Francisco, USA (2010).
- [12] SCHARSTEIN, D., AND SZELISKI., R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. Journal of Comp. Vision 47, 1/2/3, 7–42.
- [13] K. Yoon et al., "Adaptive Support-Weight Approach for Correspondence Search," IEEE Trans. on PAMI, 2006.
- [14] D. S. Brungart, "Auditory Parallax Effects in the HRTF for Nearby Sources", IEEE Workshop on Application of Signal Processing to Audio and Acoustics, New York, USA (1999).
- [15] H.Y. Kim, et al., "Control of Auditory Distance Perception Based on the Auditory Parallax Model", Applied Acoustics, 62, 245-270 (2000).
- [16] H.G. Moon, et al., "Reverberation Cue as a Control Parameter of Distance in Virtual Audio Environment", IEICE Trans. Fundamentals, vol. E87-A, No.5, 1822-1826, (May, 2004)
- [17] CIPIC Interface Laboratory. CIPIC HRTF database files, released 1.1 August 21, U. C. Davis. <http://interface.cipic.ucdavis.edu/> (2001).
- [18] M.M. Goodwin and J.-M. Jot, "Primary-Ambient Decomposition and Vector-Based Localization for Spatial Audio Coding and Enhancement", Proc. IEEE ICASSP (Apr., 2007)
- [19] Sunmin Kim, "Adaptive Channel Re-mixing Algorithm for Intended Emphasis on the Surround Sound", ICCE, Las Vegas, USA (2010)
- [20] Sunmin Kim and Youngjin Park, "A Direct Design Method of Inverse Filters for Multichannel 3D Sound Rendering," J. sound vib., Vol. 278, Issues 4-5, pp 1196-1204 (2004).