자유로운 스피커 배치환경을 위한 가상음원 표현기술

서정일, 장대영, 강경옥
한국전자통신연구원
seoji@etri.re.kr, dyjang@etri.re.kr, kokang@etri.re.kr

Virtual Sound Source Representation for Flexible Loudspeaker Position

Jeongil Seo, Dae-young Jang, Kyeongok Kang
Electronics and Telecommunications Research Institute (ETRI)

요 약

본 논문에서는 멀티채널 오디오 재생시스템을 구성하는 라우드스피커들이 표준위치와 상이한 위치에 존재할 경우 입력된 멀티채널 오디오 신호가 표현하는 음장감을 최대한 유지하면서 주어진 라우드스피커 배치환경에 최적으로 가상음원과 멀티채널 오디오 신호를 재생하는 방법을 제안한다. 능동잡음제어 기술에 이용되는 LSM 기술을 활용하여 원음장과 스피커를 통해 재생되는 복원음장 사이의 오차를 최소화하는 음향공간필터를 설계하여 적용하였으며, 청취평가를 통하여 음장감이 유지됨을 확인하였다.

1. 서론

5.1 채널로 대표되던 멀티채널 오디오 서비스는 UHDTV 와 같은 새로운 실감방송 서비스 및 Dolby ATMOS 와 같은 3 차원 영화음악 서비스의 도래와 더불어 채널 수가 증가하고 라우드스피커들이 3 차원으로 배치되어 3 차원 공간감을 제공하는 서비스로 확장되고 있다. NHK 22.2 채널이 가장 대표적인 3 차원 라우드스피커 포맷이다.

또한, 다양한 라우드스피커 포맷환경에서 음향 저작자가 의도한 3 차원 공간상에 정확하게 음원을 표현하기 위하여 객체음원을 활용하는 기술도 널리 활용되고 있다.

MPEG-H 3D Audio 는 이러한 환경변화에 대응하기 위하여 22.2 채널과 같은 초다채널 오디오 신호와 객체음원을 함께 부호화하고 다양한 출력 라우드스피커 채널 환경에 최적으로 재생하기 위한 기술을 표준화하고 있다[1].

MPEG-H 3D Audio 가 표준화하고자 하는 기술 중에 하나인 Flexible Rendering 은 사용자가 구비하고 있는 라우드스피커들이 표준 설치위치와 상이한 위치에 설치되어 있다고 하더라도 입력 멀티채널 오디오 신호와 객체음원 신호를 라우드스피커 설치위치에 맞도록 변환하여 재생하는 기술이다.

본 논문은 이러한 Flexible Rendering 기술의 요구사항을 만족시키기 위하여 VBAP[2]와 같이 3 차원 공간상의 패닝기술을 활용하는 방법이 아닌, 라우드스피커와 청취공간으로 구성되는 음향공간필터를 예측하여 음원의 위치에 따라 적응적으로 음향공간필터를 적용함으로써 가상음원 또는 입력 오디오 채널 신호를 렌더링하는 기술을 제안한다.

2. 음향공간필터 설계

LSM(Least Squared Method)은 그림 1 과 같이 마이크로폰 어레이를 통해 녹음된 원음장 신호 $\mathbf{p}$와 라우드스피커 어레이에 의해 재생되는 복원음장 신호 $\hat{\boldsymbol{p}}$ 사이의 식 1 과 같은 평균자승오차를 최소로 하도록 라우드스피커 출력신호를 조정하는 방법이다[3].

$$\mathrm{e} = \|\boldsymbol{p} - \boldsymbol{q}\|^2 \qquad \text{(식 1)}$$

간섭 및 반사가 없는 공간에서 라우드스피커 $\mathrm{q}(y_n)$에 의하여 발생된 음파는 마이크로폰 $\mathrm{p}(x_m)$에서 녹음된다. 이때 라우드스피커와 마이크로폰 사이의 음향학적 전달함수는 $\mathrm{Z}(x_m/y_n)$로 표현된다. 따라서 라우드스피커에서 발생된 신호와 마이크로폰에서 녹음한 신호는 식 2 와 같이 행렬형태로 표현할 수 있다.

$$\boldsymbol{q}^T = [q(y_1), q(y_2), \ldots, q(y_L)],\ \boldsymbol{p}^T = [p(x_1), p(x_2), \ldots, p(x_M)]$$
$$\hat{\boldsymbol{p}} = \mathbf{Z}\mathbf{q} \qquad \text{(식 2)}$$

따라서 원음장 신호와 최대한 유사한 복원음장을 얻기 위해서는 식 1 과 같이 원음장 신호 $\mathbf{p}$와 복원음장 신호 $\hat{\boldsymbol{p}}$ 사이의 평균자승오차를 최소화 시켜야 한다. 복원음장신호 $\hat{\boldsymbol{p}}$ 는 식 2 와 같이 표현할 수 있으므로 두 신호 사이의 평균자승오차로 표현되는 비용함수는 식 3 과 같이 표현된다.

$$\mathbf{J} = \|\boldsymbol{p} - \hat{\boldsymbol{p}}\| = \sqrt{\textstyle\sum_{q=1}^{Q} |p_q - \hat{p}_q|^2} = \|\boldsymbol{p} - \boldsymbol{Z}\boldsymbol{q}\| \qquad \text{(식 3)}$$

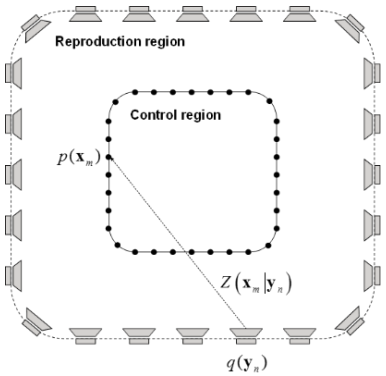

그림 1. 라우드스피커 및 마이크로폰 어레이 환경에서의 음압

음향전달함수 $\mathbf{Z}$ 가 singular value 가 존재하지 않는 행렬이면 라우드스피커에 인가할 신호 $\mathbf{q}$ 는 식 4 와 같이 의사역행렬 연산으로 결정된다.

$$\mathbf{q} = (\boldsymbol{Z}^{\boldsymbol{H}}\,\boldsymbol{Z}\,)^{-1}\,\boldsymbol{Z}^{\boldsymbol{H}}\,\boldsymbol{p} = \boldsymbol{Z}^{+}\,\boldsymbol{p} \qquad \text{(식 4)}$$

3. 성능평가

MPEG-H 3D Audio 실험음원 중 Rain 은 22.2 채널 배경음과 모노채널 발자국 소리와 발자국 소리의 3 차원 공간상의 재현위치를 포함하는 메타데이터로 구성된다. 본 논문에서 제안한 기술의 성능을 평가하기 위하여 MPEG-H 3D Audio 에서 정의한 28.2 채널 라우드스피커들 중 5 개 및 10 개를 임의로 선택하여 Rain 음원을 주어진 라우드스피커 환경에 맞게 재현하였다. 실험에 사용된 5 채널(Rand5) 및 10 채널(Rand10) 라우드스피커 위치는 표 1 과 같다.

음향공간필터는 Rand5 및 Rand10 라우드스피커 배치환경에서 각각 설계되었으며, 22.2 채널 배경음을 구성하는 각 채널 신호들을 고정음원으로 가정하여 각 채널신호의 재생위치에 렌더링 하였으며, 발자국 소리는 1024 샘플단위로 정의된 재생위치를 반영하여 매 프레임마다 재생위치를 변경하여 렌더링 하였다.

4. 청취평가

3 장에서 설명된 과정으로 생성된 5 채널(Rand5) 및 10 채널(Rand10) 오디오 신호를 22.2 채널 배경음과 VBAP 방식으로 발자국 소리를 렌더링하여 생성된 원음(HR)과 이를 3.5kHz 저대역필터를 통과시킨 앵커신호(LP35)로 시험 음원을 구성하여 MUSHRA 방법으로 청취평가를 수행하였다. 숙련된 26 명의 평가자가 참여하였으며, 그 결과는 그림 2 및 그림 3 과 같다. 단 Rand5 와 Rand10 렌더링을 수행하는 입력 오디오 신호는 512kbps 로 부호화된 신호를 활용하였다.

제안된 기술은 Rand5 의 경우 평균 60.6, Rand 10 의 경우 65.8 정도의 MUSHRA 점수가 측정되었다. 비교적 낮은 점수이기는 하지만 512kbps 로 압축된 오디오 신호를 입력신호로 사용하였기 때문에 부호화 과정에서 발생된 열화도 함께 반영되어 있으므로 원음과 상당히 유사한 음장감을 제공한다고 해석할 수 있다.

5. 결론

본 논문에서는 LSM 방법을 이용하여 음향공간필터를 구성하고 이를 임의의 라우드스피커 배치환경에 적응적으로 활용하기 위한 방법을 제안하였다. 청취평가 결과 원음과 유사한 음장감을 제공함을 확인하였고, 이러한 재생 기술은 멀티채널 오디오 재생시스템을 구성하는 라우드스피커들이 어떠한 위치에 설치되어 있다 하더라도 최적의 음장감을 제공할 수 있음을 확인하였다.

감사의 글

본 연구는 미래창조과학부의 산업원천기술개발사업(정진)/방송통신 ETRI 지원사업(방발)의 연구결과로 수행되었음. [초고품질 콘텐츠 지원 UHD 실감방송/디지털 시네마/사이니지 융합서비스 기술 개발]

참고문헌

[1] ISO/IEC JTC1/SC29/WG11, N13411, Call for Proposals for 3D Audio, Jan. 2013.

[2] V. Pulkki, “Compensating displacement of amplitude-panned virtual sources,” AES 22th Int. Conf. on Virtual, Synthetic and Entertainment Audio, pp. 186-195. 2002 Espoo, Finland.

[3] F. M. Fazi, P. A. Nelson, J. E. N. Christensen, and J. Seo, “Surround system based on three-dimensional sound field reconstruction,” the 125th AES Convention, 2008.

표 1. 실험에 사용된 라우드스피커의 위치

Rand5			Rand10					
LS#	수평각	수직각	LS#	수평각	수직각	LS#	수평각	수직각
1	0	−15	1	0	90	6	135	35
2	45	35	2	0	−15	7	−30	0
3	135	35	3	30	35	8	−45	35
4	−30	0	4	45	−15	9	−90	0
5	−110	0	5	110	0	10	−110	35

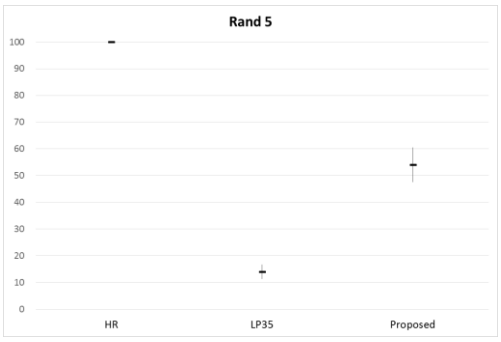

그림 2. Rand5 조건에서의 청취평가 결과

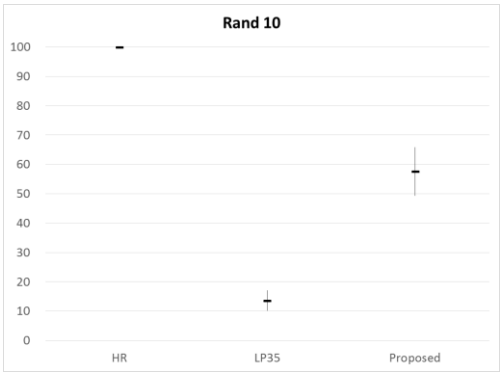

그림 3. Rand10 조건에서의 청취평가 결과