

## 스테레오-멀티채널 업믹스 시스템에서의 초기 반사음 생성 기법

이태규\*    백용현\*\*    박영철\*\*    윤대희\*

\*연세대학교 전기전자공학과

\*\*연세대학교 컴퓨터정보통신공학과

tglee@dsp.yonsei.ac.kr

## Early Reflection Generation Technique in Stereo-to-Multichannel Upmix System

\*Taegy Lee    \*\*Yong-Hyun Baek    \*\*Young-cheol Park    \*Dae Hee Youn

\*Dept. of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea

\*\*Computer and Telecommunication Engineering Division, Yonsei University, Wonju, Korea

## 요약

스테레오-멀티채널 업믹스(Stereo-to-Multichannel Upmix)기술은 시장의 대부분을 차지하는 스테레오 음원을 홈시어터와 같은 멀티채널 스피커 재생 환경에서 효과적으로 재생하기 위한 채널 포맷 변환 기술을 말한다. 수평적인 공간감뿐만 아니라 수직적인 공간감을 재현하기 위하여 10.2, 22.2채널 환경의 경우 고도 채널이 추가되었다. 본 논문에서는 이러한 채널 환경에서 수직, 수평적인 공간감을 재현하기 위하여 새로운 초기 반사음을 생성하는 알고리즘을 제안하였다. 제안된 알고리즘은 진폭 페닝방법을 이용하여 전방의 가상 음원들에 의하여 발생하는 리시버 위치에서의 음압을 일치시킴으로써 기존 알고리즘이 가지지 못했던 이론적인 배경을 갖는다. 청취 평가 실험을 통하여 제안된 알고리즘은 음장의 수직, 수평적인 폭을 증가시키며, 좀 더 안정적인 음상 정위 성능을 가지는 것을 확인하였다.

## 1. 서론

현재 HDTV(High-Definition TV)에 비하여 4배 이상의 스크린 크기와 시야각 100도에 이르는 UHDTV(Ultra High-Definition TV) 기술이 경쟁적으로 개발이 마무리 되고 있다. 사용자에게 이에 걸맞은 높은 몰입감을 부여하기 위하여 5.1, 7.1, 10.2, 22.2 채널과 같은 멀티채널 환경에서의 오디오 신호 처리 기술이 경쟁적으로 개발되고 있다.

그러나 멀티미디어 시장에서 영화 콘텐츠를 제외한 대부분의 오디오 콘텐츠는 2채널 스테레오 신호이다. 따라서 멀티채널 시스템을 갖추고 있더라도, 해당 시스템에 적합한 멀티미디어 콘텐츠가 부족하기 때문에 제대로 활용되기 어려운 상황이다. 이러한 시장 특성을 극복하기 위하여 2채널 스테레오 신호를 멀티채널 신호로 사전정보 없이 업믹스하는 시스템이 개발되어 왔으며, 이는 스테레오-멀티채널 블라인드 업믹스 시스템이라 불린다[1].

스테레오-멀티채널 블라인드 업믹스 방법은 크게 음장 분석과 음장 합성인 두 가지 과정으로 이루어져 있다[2]. 음장 분석 단계에서는 국소 푸리에 변환(short-time Fourier transform)을 이용하여 입력 신호를 주파수 축으로 변환한 후 좌-우 채널간의 공분산과 같은 정보를 이용하여 공간적 단서를 추출하고 입력 스테레오 신호를 방향성 신호(primary signal)와 배경음 신호(ambient signal)로 분리한다[3, 4, 5, 6, 7]. 음장 합성 단계에서는 분리된 방향성 신호는 추출된 공간적 단서를 바탕으로 전방 채널에 페닝 되어 분배되고, 분리된 배경음 신호는 후방 채널에 분배된다. 이 때 후방 채널 간 상관도를 떨어뜨려 사용자에게 포위감을 제공하기 위하여 각 후방 채널 신호는 비상관

(decorrelation)된다. 또한 Haas 효과에 의하여 전방 채널의 음상을 흘뜨리지 않도록 1.5ms이상의 딜레이를 준다[8].

5.1, 7.1 채널과 같은 적은 스피커 개수를 가지는 멀티채널 환경에서의 블라인드 업믹스 알고리즘은 수평적인 공간감을 향상시키는 것에 초점을 두었다. 그러나 10.2, 22.2채널과 같이 보다 많은 수의 스피커를 가지는 멀티채널 환경에서는 수평적인 공간감뿐만 아니라 수직적인 공간감을 부여하기 위하여 스피커들이 수평면 밖에 배치되었다. 따라서 이러한 스피커 환경에 맞게 업믹스 된 멀티채널 신호에는 수직적인 공간감도 부여되어야 한다. 수직적인 공간감을 부여하기 위하여 기존의 알고리즘은 공간적으로 초기 반사음을 분리하여 측면 채널과 고도 채널에 분배하였다. 이 방법의 경우 수직, 수평적인 공간감을 향상시키고 음상은 흘뜨리지 않았다는 것에 의의가 있으나, 이론적인 배경이 없다는 것에 한계를 지닌다[9].

본 논문에서는 수평면 밖에 존재하는 스피커를 포함하는 스테레오-멀티채널 업믹스 시스템에서의 새로운 초기 반사음 합성 알고리즘을 제안한다. 제안하는 알고리즘은 청자 위치에서의 초기 반사음의 음압을 유지시켜준다는 점에서 물리적인 타당성을 가진다. 더불어 제안하는 알고리즘은 기존 알고리즘보다 음장의 수직, 수평적인 폭을 증가시키며, 좀 더 안정적으로 음상을 정위시키는 것을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서 스테레오 입력 신호에 대한 신호 모델링을 통한 음장 분석 방법을 설명하고, 3장에서 기존의 초기 반사음 합성 알고리즘을 설명한다. 4장에서는 제안된 알고리즘을, 5장에서는 청취 실험 결과를 보이고 6장에서 본 논문의 결론을 맺는다.

## 2. 음장 분석

음장을 분석하기 위하여 국소 푸리에 변환을 이용하여 입력 스테레오 신호를 주파수 축으로 변환한다. 변환된 푸리에 계수들을 사람의 청각 특성을 반영하기 위하여 임계 대역(critical band)으로 나눈 서브밴드별로 처리한다. 주파수 축으로 변환된 신호 모델은 식 (1)과 같이 표현될 수 있다.

$$\begin{aligned} X_L &= g_L S + N_L \\ X_R &= g_R S + N_R \end{aligned} \quad (1)$$

시간과 주파수 인덱스는 편의상 생략한다. 방향성 신호는 레벨 패닝을 이용하여 정위되고 방향성 신호  $S$ 와 좌, 우 배경음  $N_L, N_R$ 의 경우 서로 상관도가 없다고 가정하자. 이는 수식적으로 다음과 같이 표현 가능하다.

$$E\{S^H N_L\} = E\{S^H N_R\} = E\{N_L^H N_R\} = 0 \quad (2)$$

레벨 패닝 기법에 의하여 좌, 우 패닝 계인  $g_L, g_R$ 은 다음과 같은 수식을 만족시킨다.

$$g_L^2 + g_R^2 = 1 \quad (3)$$

마지막으로 수식 (4)와 같이 배경음은 방향성 신호의 특징과 방향에 관계없이 좌, 우 채널의 에너지가 같다고 가정할 수 있다.

$$E\{N_L^H N_L\} = E\{N_R^H N_R\} \quad (4)$$

위의 신호 모델을 이용하여 다양한 제약 조건을 바탕으로 방향성 성분과 두 배경음 성분을 분리할 수 있다. 추정된 방향성 성분과 두 배경음 성분 각각의 자승 오차를 최소화 하는 최소 자승법[3], 좌, 우 채널간의 자승 오차를 최소화 하는 크로스채널 예측법[4], 의사역행렬을 이용한 방법[6], 주성분 분해를 이용하는 방법[5], 주성분 분해 방법의 변형된 방법[5, 7]들이 존재한다. 본 논문에서는 주성분 분해방법의 변형된 방법[7]을 이용하였다. 이 방법은 방향성 성분의 에너지가 작을 때에도 음원 분리 성능이 저하되지 않으며, 패닝 각도의 영향에 강건한 음원 분리 성능을 가진다. 또한 방향성음원에서 잔여 배경음 성분을 제거하여 더 정확한 실제 패닝 신호를 분리해 낼 수 있다.

## 3. 기존의 초기 반사음 합성 알고리즘

기존의 초기 반사음 합성 알고리즘은 추정된 측면 채널과 고도 채널들 각각의 초기 반사음의 충격 응답에 방향성 신호를 컨볼루션하여 음장의 폭과 높이를 향상시켰다[9]. 각 측면, 고도채널의 초기 반사음 충격 응답을 추정하는 방법은 Image source method를 바탕으로 한다[10]. 자세한 방법은 다음과 같다.

먼저 Image source method를 이용하여 80ms 이내의 시간에 도달

하는 가상 음원의 위치를 구한다. 얻어진 가상 음원들은 각 측면, 고도 채널마다 다른 공간적인 필터링을 통하여 걸러진다. 고도 채널의 경우 고도 채널 위치 위쪽에 해당하는 초기 반사음들을 재생하고, 측면 채널의 경우 측면 채널보다 측면에 존재하는 초기 반사음들을 재생한다. 이를 수식적으로 나타내기 위하여 전방 벽면의 위쪽 두 꼭짓점과 리시버의 위치를 지나는 평면의 방정식을  $a_h x + b_h y + c_h z + d_h = 0$ , 전방 벽면의 좌측 두 꼭짓점과 리시버의 위치를 지나는 평면의 방정식을  $a_l x + b_l y + c_l z + d_l = 0$ , 전방 벽면의 우측 두 꼭짓점과 리시버의 위치를 지나는 평면의 방정식을  $a_r x + b_r y + c_r z + d_r = 0$  이라 가정하자.  $HL, HR, WL, WR$  채널 각각의 공간적인 필터인  $\delta_{HL}, \delta_{HR}, \delta_{WL}, \delta_{WR}$ 은 다음과 같이 정의될 수 있다.

$$\begin{aligned} \delta_{HL} &= \begin{cases} 1 & a_h x + b_h y + c_h z + d_h \geq 0, x \geq 0, \text{ and } y < 0 \\ 0 & \text{otherwise} \end{cases} \\ \delta_{HR} &= \begin{cases} 1 & a_h x + b_h y + c_h z + d_h \geq 0, x \geq 0, \text{ and } y > 0 \\ 0 & \text{otherwise} \end{cases} \\ \delta_{WL} &= \begin{cases} 1 & a_l x + b_l y + c_l z + d_l \geq 0, z \geq 0, \text{ and } y < 0 \\ 0 & \text{otherwise} \end{cases} \\ \delta_{WR} &= \begin{cases} 1 & a_r x + b_r y + c_r z + d_r \geq 0, z \geq 0, \text{ and } y > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (5)$$

가상 음원의 개수를  $N$ ,  $p$ 번째 가상 음원의 위치벡터를  $\vec{R}_p$ , 리시버의 위치 벡터를  $\vec{R}_r$ , 각 벽면의 반사계수를  $\beta_{x1}, \beta_{x2}, \beta_{y1}, \beta_{y2}, \beta_{z1}, \beta_{z2}$ , 각 벽면을 통과한 횟수를  $i, j, k, l, m, n$ 이라고 할 때, 공간적인 필터를 적용한 초기 반사음의 충격 응답(impulse response)을 수식적으로 나타내면 수식 (6)과 같다.

$$\begin{aligned} R_s(t) &= \sum_{p=0}^N \beta_{x1}^i \beta_{x2}^j \beta_{y1}^k \beta_{y2}^l \beta_{z1}^m \beta_{z2}^n \frac{1}{4\pi |R_r + R_p|} \\ &\quad \times \delta\left(t - \frac{|R_r + R_p|}{c}\right) \delta_s \\ s &\in \{HL, HR, WL, WR\} \end{aligned} \quad (6)$$

## 4. 제안된 초기 반사음 합성 알고리즘

제안하는 초기 반사음 합성 알고리즘은 리시버 위치에서 초기 반사음들에 의하여 발생하는 음압을 전방에 존재하는 세 개의 스피커들의 출력을 이용하여 일치시킨다. 초기 반사음의 음압 벡터를  $p_I$ , 세 개의 스피커 출력의 음압 벡터를 각각  $p_A, p_B, p_C$ , 세 개의 스피커 출력 이득 값을  $w_A, w_B, w_C$ 라 하자. 초기 반사음의 음압 벡터를 세 개의 스피커 출력의 선형 결합으로 표현하면 식 (7)과 같다.

$$p_I = [w_A \ w_B \ w_C][p_A \ p_B \ p_C]^T = WP \quad (7)$$

이득 값을 얻기 위하여 양변에  $P$ 의 역행렬을 곱하면 식 (8)과 같

다.

$$W = p_I P^{-1} \quad (8)$$

위의 알고리즘은 벡터 기반 진폭 패닝(VBAP, vector base amplitude panning)과 동일하다[11]. 이를 미루어 보았을 때 제안하는 알고리즘은 초기 반사음을 진폭 패닝을 이용하여 렌더링하는 것과 같다.

제안하는 알고리즘은 다양한 멀티채널 환경에서 적용 가능하나 본 논문에서는 USC 10.2채널 환경을 예로 들어 설명하겠다[12]. 10.2 채널 시스템에서 가능한 삼각형 망은 표 1과 같다.

	스피커 구성
1	FC, HL, HR
2	FC, FL, HL
3	FC, FR, HR
4	FL, HL, WL
5	FR, HR, WR

표 1 가능한 삼각형 망

식 (8)을 이용하여 임의의 초기 반사음의 음압 벡터  $p_I$ 가 특정 삼각형 망을 지나는 것을 판별할 수 있다. 식 (8)을 이용하여 얻어진 이득 값이 적어도 하나 이상이 음수인 경우 초기 반사음의 음압 벡터는 해당 삼각형 망 밖을 지나게 된다. 따라서 이득 값은 다음과 같이 얻어진다. 해당 이득 값은 방향만을 반영하므로 정규화 되었다.

$$w_s = \begin{cases} w_s / \|W\| & \forall w_s \geq 0 \text{ where } s \in \{A, B, C\} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

해당 삼각형 망을 지나는 경우 반사와 거리에 따른 감쇄를 반영하면 식 (10)과 같다.

$$R_s(t) = \sum_{p=0}^N \beta_{x1}^i \beta_{x2}^j \beta_{y1}^k \beta_{y2}^l \beta_{z1}^m \beta_{z2}^n \frac{1}{4\pi |R_r + R_p|} \times \delta \left( t - \frac{|R_r + R_p|}{c} \right) w_s \quad (10)$$

$s \in \{FC, FL, FR, WL, WR, HL, HR\}$

그림 1은 방 크기가 4x4x4m일 때 제안하는 알고리즘을 이용하여 얻어진 FC, FL, FR, WL, WR, HL, HR 채널의 초기 반사음 충격 응답이다.

### 5. 청취 평가 실험

제안된 초기 반사음 합성 알고리즘을 기존의 초기 반사음 합성 알고리즘과 비교하였다[9]. 스피커는 USC 10.2채널 구성[12]에 따라 배치되었다. 4명의 실험자가 참가하였으며, 음장의 수평적인 폭

(horizontal width)과 수직적인 폭(vertical width)의 증가량과 음상 정위의 안정도(stability of sound image localization)를 스테레오 신호

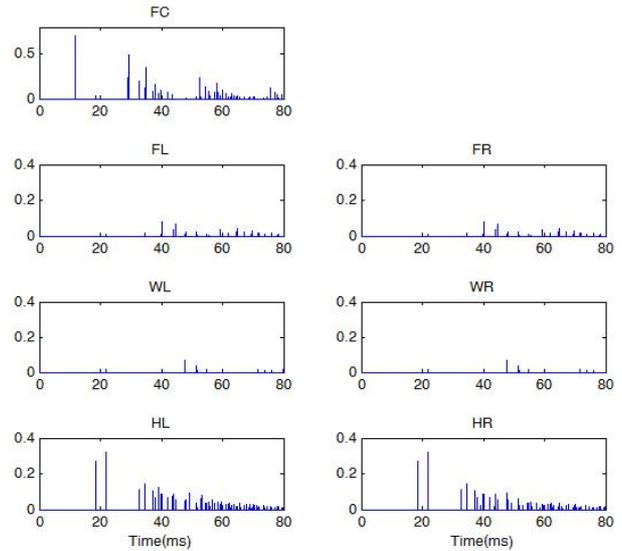


그림 1 제안된 알고리즘을 이용하여 얻어진 초기 반사음 충격 응답

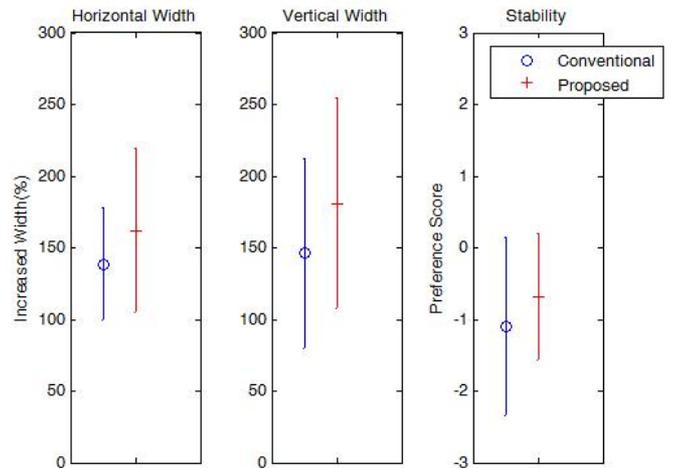


그림 2 청취 평가 실험 결과

와 비교하여 측정하였다. 음장의 수평, 수직적인 폭은 스테레오 신호 대비 증가한 정도의 비율을 측정하였으며, 음상 정위의 정확도는 -3.0에서 3.0점의 스케일로 상대적인 선호도 점수(preference grade)를 주도록 하였다. 그림 2는 청취 실험의 결과를 나타낸다. 제안된 초기 반사음 합성 알고리즘이 이전 방법에 비하여 음장의 수직, 수평적인 폭은 소폭 상승하였으며, 안정적인 음상 정위 성능을 가짐을 보여준다.

### 6. 결론

본 논문에서는 진폭 패닝방법을 이용하여 전방의 가상 음원들에 의하여 발생하는 리시버 위치에서의 음압을 일치시키는 초기 반사음 생성 알고리즘을 제안하였다. 제안된 알고리즘은 기존 알고리즘과는 달리 물리적인 타당성을 가진다. 제안한 알고리즘은 청취 실험을 통하여 기존 알고리즘보다 음장의 수직, 수평적인 폭을 상승시키고 안정적

인 음성 정위 성능을 가진다.

## 7. Reference

- [1] Avendano, Carlos, and Jean-Marc Jot. "A frequency-domain approach to multichannel upmix." *Journal of the Audio Engineering Society* 52.7/8 (2004): 740-749.
- [2] Breebaart, Jeroen, and Erik Schuijers. "Phantom materialization: A novel method to enhance stereo audio reproduction on headphones." *Audio, Speech, and Language Processing, IEEE Transactions on* 16.8 (2008): 1503-1511.
- [3] Faller, Christof. "Multiple-loudspeaker playback of stereo signals." *Journal of the Audio Engineering Society* 54.11 (2006): 1051-1064.
- [4] Usher, John, and Jacob Benesty. "Enhancement of spatial sound quality: A new reverberation-extraction audio upmixer." *Audio, Speech, and Language Processing, IEEE Transactions on* 15.7 (2007): 2141-2150.
- [5] Goodwin, Michael M. "Geometric signal decompositions for spatial audio enhancement." *Acoustics, Speech and Signal Processing, ICASSP IEEE International Conference on*. IEEE, 2008.
- [6] 전세운, 박영철, 이석필, 윤대희, "다채널 포맷 변환과 공간적인 입체 음향 정보의 효과적인 유지에 대한 연구," *전자공학회 2010년도 하계종합학술발표회 논문집*, 2010
- [7] Baek, Yong-Hyun, et al. "Efficient Primary-Ambient Decomposition Algorithm for Audio Upmix." *Audio Engineering Society Convention* 133. 2012.
- [8] Litovsky, Ruth Y., H. Steven Colburn, William A. Yost, and Sandra J. Guzman. "The precedence effect." *The Journal of the Acoustical Society of America* 106 (1999): 1633.
- [9] Choi, Sunwoong, Dong-il Hyun, Young-cheol Park, Seokpil Lee, and Dae Hee Youn. "Blind Upmixing for Height and Wide Channels Based on an Image Source Method." *Audio Engineering Society Convention* 133. 2012.
- [10] Allen, Jont B., and David A. Berkley. "Image method for efficiently simulating small room acoustics." *The Journal of the Acoustical Society of America* 65 (1979): 943.
- [11] Pulkki, Ville. "Virtual sound source positioning using vector base amplitude panning." *Journal of the Audio Engineering Society* 45.6 (1997): 456-466.
- [12] Holman, Tomlinson. "The history and future of DSPs in consumer audio equipment-part I: history and current conditions." *Consumer Electronics, ICCE, Digest of Technical Papers. International Conference on*, vol., no., pp.1,2, 9-13 Jan. 2008