

오디오 업믹스를 위한 효율적인 Primary-Ambient 분리 알고리즘

*백용현, *이근상, *전세운, **이석필, *박영철

*연세대학교 MSP 연구실, **전자부품연구원(KETI)

(hyen0330@yonsei.ac.kr)

Efficient Primary-Ambient Decomposition Algorithm for Audio Upmix

*Baek, Yong-Hyun, *Lee, Keun-Sang, *Jeon, Se-Woon, **Lee, Seokpil, *Park, Young-Choel

*Yonsei University, MSP Lab. ** KETI

요약

업믹스(Upmix) 기술은 홈시어터와 같은 다채널 스피커 재생 환경에서 콘텐츠의 대부분을 차지하는 스테레오 음원을 다채널 환경에 재생하기 위한 채널 포맷 변환 기술을 말한다. 업믹스를 위한 전처리 단계로서 특정 방향으로 패닝된 주(primary) 성분과 잔향 및 배경음과 같은 Ambient 성분을 분리하는 과정이 필요하다. Primary와 Ambient를 분리하기 위한 방법으로 채널 간의 상관도, 적응 필터 및 주성분 분석법(principal component analysis, PCA)이 널리 이용되고 있다. 이에 본 논문에서는 비교적 정확하게 Primary와 Ambient를 분리한다고 알려진 주성분 분석법을 이용하여 신호를 분리해 내고 이 때 주성분 분석법이 가지는 문제점을 해결한 향상된 Primary-Ambient 분리 알고리즘을 제안하였다. 제안된 알고리즘은 분리 성능이 Primary 성분이 패닝된 각도에 영향을 받지 않으며 또한 Primary 성분에 섞인 잔여 Ambient를 제거함으로써 기존의 주성분 분석법 보다 더 정확하게 Primary와 Ambient를 분리 할 수 있고 상관성이 없는 Ambient 특성을 좀 더 정확하게 반영한다.

1. 서론

현재 3차원 영상과 더불어 오디오 또한 실감 음원 청취에 대한 사람들의 욕구가 증가하면서 각 가정에 Dolby Surround[1]와 같은 다채널 재생 시스템이 널리 보급 되고 있다. 그러나 현재 대부분의 오디오 콘텐츠들은 여전히 스테레오로 서비스 되는 것이 대부분이며 이것을 다채널 환경에서 재생하기 위하여 채널 포맷 변환 및 분배 기술인 업믹스 기술을 사용한다. 업믹스 기술은 크게 두 단계 과정을 거치는데 첫 번째 단계로 스테레오에서 Primary 성분과 Ambient 성분을 분리하고 두 번째 단계로 분리된 음원을 다채널 스피커 포맷에 맞게 분배하는 과정이다. Primary 성분은 스테레오 음원에서 패닝 기법[2]으로 믹싱되거나 목소리등의 방향성이 있는 주된 성분을 말하고 Ambient 성분은 그 외 잔향이나 배경음 등과 같은 성분을 말한다. 본 논문에서는 업믹스 기술의 첫 번째 단계인 Primary-Ambient를 분리하는 효과적인 방법을 제안한다.

기존의 Primary-Ambient 분리 방법으로 다양한 방법들이 제안되었다 [3~5]. 이들 중 본 논문에서는 주성분 분석법을 이용한 Primary-Ambient 분리 방법에 대하여 고찰하고 주성분 분석법이 가지는 문제점을 완화한 향상된 주성분 분석법을 제안한다. 주성분 분석법의 문제점으로는 크게 두 가지로 첫 번째로 분리해낸 Primary 성분에 잔여 Ambient 성분이 남아 있다는 점, 두 번째로 입력 스테레오 채널에 주성분이 되는 Primary 성분이 존재하지 않는 경우 에너지가 큰 Ambient 성분이 Primary로 분리 된다는 점이다. 이러한 문제를 해결하기 위하여 Modified PCA 알고리즘이 제안되었다[6]. 그러나 Modified PCA 알고리즘은 기존의 PCA 알고리즘이 Primary가 존재하지 않는 경우를 채널간의 정규화된 상관 계수를 이용하여 패닝 신호의 유무를 판단한 뒤 Primary와 Ambient를 분리하는 방법이다. 그러

나 상관 계수가 스테레오 패닝 기법으로 믹싱된 음원의 경우 두 채널에 Ambient만 있을 경우 뿐만 아니라 한쪽 채널로 극단적으로 패닝된 음원의 경우에도 상관 계수가 작아지기 때문에 Ambient로 대부분의 Primary 성분이 같이 빠져 나오는 문제가 있다.

이를 해결하기 위하여 본 논문에서는 상관 계수가 아닌 두 채널간의 고유치값의 비를 이용하여 Primary와 Ambient를 분리한다. 고유치값의 비는 패닝 각도에 영향을 받지 않기 때문에 상호 상관도를 이용할 때의 문제점을 극복할 수 있다. 또한 Ambient를 각 채널별로 상관성이 없는 성분을 두 채널 분리할 수 있으며 primary에 섞인 잔여 Ambient 성분 역시 제거하여 분리할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서 스테레오 입력 신호에 대한 신호 모델링에 대해 설명 하고 3장에서 PCA를 이용하여 패닝 각도를 추정된 뒤 Primary와 Ambient를 분리하는 방법과 Modified PCA에 대해 설명하고 4장에서 제안된 알고리즘을, 5장에서는 모의 실험 결과를 보이고 6장에서 본 논문의 결론을 맺는다.

2. 입력 스테레오 신호 모델링

업믹스 기술에서 입력 스테레오 신호를 분석할 때 단시간 푸리에 변환(short-time Fourier transform, STFT)을 이용하여 분석한다. 이렇게 주파수 도메인으로 변환된 계수들을 사람의 청각 특성에 따라 각각 서브밴드별로 묶어 나누어 처리한다[7]. 이후의 수식들은 주파수 도메인에서의 표현들이며 수식의 간편함을 위해 시간 프레임 인덱스와 주파수 대역 인덱스는 생략한다. 입력 스테레오 신호는 음원이 패닝 기법에 의하여 스테레오 스피커 포맷에 따라 -30° 에서 30° 사이에 레벨 패닝 되어 정위 되었다고 가정하고 여기에 각각의 Ambient가 더해진 형태로 모델링 할 수 있다. 모델링 된 입력 두 채널의 신호는 수식 (1)

과 같이 표현 할 수 있다.

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

여기서 a_L 과 a_R 은 페닝 계인, S 는 페닝 성분이 되고 N_L 과 N_R 은 각 채널에 더해진 Ambient 성분이다. 레벨 페닝 기법에 의해 페닝 계인은 수식 (2)와 같은 관계를 가지며 각 S, N_L, N_R 은 수식 (3)과 같은 관계를 갖는다고 가정한다.

$$a_L^2 + a_R^2 = 1 \quad (2)$$

$$E\{S^H N_L\} = E\{S^H N_R\} = E\{N_L^H N_R\} = 0 \quad (3)$$

따라서 입력 페닝 성분과 Ambient, 또한 각 채널 Ambient간에는 상관성이 없고, 수식 (4)에서 처럼 두 채널에 더해지는 Ambient들의 에너지는 같다고 가정한다.

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

3. PCA 기반의 음원 분리 알고리즘

3-1 PCA를 이용한 음원 분리 방법

PCA를 이용하여 음원을 분리 할 때 먼저 페닝 성분에 곱해진 페닝 계인을 구해야 한다. 페닝 계인은 입력 스테레오 채널의 2x2 공분산 행렬로부터 추정 할 수 있고 공분산 행렬은 아래 식과 같이 계산된다.

$$R = \begin{bmatrix} r_{LL} & r_{LR} \\ r_{RL} & r_{RR} \end{bmatrix}, r_{ij} = E\{X_i^H X_j\} \quad (5)$$

공분산 행렬 R 을 고유치 분해하여 다시 표현 하면 다음과 같다.

$$R = [V_1 \ V_2] \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} V_1^H \\ V_2^H \end{bmatrix}, \lambda_1 \geq \lambda_2 \quad (6)$$

여기서 $\lambda_{1,2}$ 는 고유치 값이고 $V_{1,2}$ 는 각 고유치 값에 대응하는 고유치 벡터가 된다. 이때 두 고유치 값 중 큰 값을 가지는 고유치 값은 Primary에 해당하는 성분의 에너지가 되며 이에 대응하는 고유치 벡터의 방향이 페닝 각도가 되므로 따라서 추정된 페닝 계인은 수식 (7)과 같다[8].

$$\hat{a}_L = v_{1,1}, \hat{a}_R = v_{1,2} \quad (7)$$

추정된 페닝 계인으로부터 Primary 성분과 Ambient 성분은 다음과 같은 식으로 구할 수 있다.

$$\hat{S} = \hat{a}_L X_L + \hat{a}_R X_R \quad (8)$$

$$\hat{N}_L = X_L - \hat{a}_L \hat{S}, \quad (9)$$

$$\hat{N}_R = X_R - \hat{a}_R \hat{S}$$

3-2 Modified PCA를 이용한 음원 분리 방법

3.1 에서 설명한 기존의 PCA 방법으로 음원을 분리할 경우 서론에서 언급했듯이 몇 가지 문제점이 발생한다. 먼저 수식 (8)을 본 논문에서 가정된 입력 신호 모델링 수식 (1)을 적용하여 다시 쓰면 다음과 같다.

$$\begin{aligned} \hat{S} &= \hat{a}_L X_L + \hat{a}_R X_R \\ &= \hat{a}_L (a_L S + N_L) + \hat{a}_R (a_R S + N_R) \\ &= S + \hat{a}_L N_L + \hat{a}_R N_R \end{aligned} \quad (10)$$

따라서 Primary 성분에 잔여 Ambient 성분이 여전히 남아 있는 채로 분리하게 된다. 또한 PCA방법에서 가장 큰 고유치 값을 가지는 방향을 Primary 성분으로 가정하는데 만약 입력 스테레오 채널에서 이러한 페닝 성분이 존재하지 않는 경우 에너지가 큰 Ambient를 Primary 성분으로 분리되어 나오기 때문에 이러한 경우는 두 입력 채널이 그대로 Ambient 성분으로 분리 되어야 된다.

이러한 문제를 해결하기 위하여 Modified PCA 알고리즘이 제안되었다. Modified PCA 알고리즘은 다음과 같은 식을 이용하여 Primary와 Ambient를 분리한다.

$$\begin{aligned} \tilde{N}_L &= |\Phi| \hat{N}_L + (1 - |\Phi|) X_L \\ \tilde{N}_R &= |\Phi| \hat{N}_R + (1 - |\Phi|) X_R \end{aligned} \quad (11)$$

여기서 $|\Phi|$ 은 스테레오 두 채널간의 정규화된 상관 계수를 나타낸다. 따라서 페닝 성분이 존재 할 경우 상관 계수는 1의 가까운 값을 가지게 되어 기존의 PCA 방법을 그대로 사용하게 되고 페닝 성분이 존재 하지 않는 경우 상관 계수가 0에 가까운 값을 가지게 되어 입력 신호를 그대로 Ambient로 내보내게 된다.

페닝 신호 또한 상관 계수를 아래 식을 이용하여 분리한다.

$$\tilde{S} = |\Phi| \hat{S} \quad (12)$$

4. 제안된 음원 분리 알고리즘

4-1 Ambient 신호 분리

Modified PCA 방법에서 상관계수를 이용하여 페닝 성분의 유무를 판단하고 Primary와 Ambient 성분을 분리한다. 그러나 상관 계수가 작아지는 경우가 페닝 성분이 존재하지 않는 경우만이 아니라 페닝 신호가 한쪽 스피커로 극단적으로 페닝 된 경우에도 작아지게 된다. 그림 1은 페닝 각도에 따른 상관 계수와 고유치의 비를 나타낸 것이다.

입력 신호는 각도별로 페닝된 음원에 왼쪽과 오른쪽 채널에 Ambient로 백색 잡음을 더한 신호를 사용하여 측정하였다. 그림 1에서 보듯 상관도가 페닝 신호가 있음에도 불구하고 페닝 각도가 30도인 극단적일 경우에도 0에 가까운 값이 측정됨을 알 수 있다. 이유는 음원이 한쪽으로 극단적으로 페닝 되었을 경우에 한쪽에 0에 가까운 페닝 계인 값이 곱해지게 되어 페닝 음원이 한쪽채널에만 존재하고 다른 채널에는 존재하지 않게 되기 때문이다. 따라서 상관 계수를 이용한 Modified PCA 알고리즘은 페닝 신호의 페닝 각도에 따라 음원 분리 성능이 영향을 받는다.

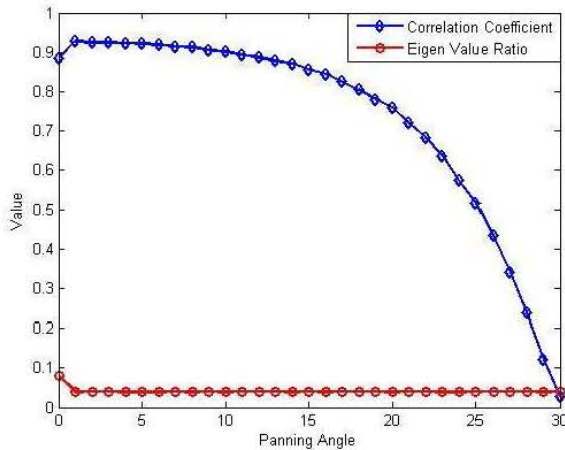


그림 1 페닝 각도에 따른 상관 계수와 고유치의 비

반대로 고유치 값의 비를 이용하면 페닝 각도에 영향을 받지 않기 때문에 상관계수를 이용할 때의 문제점을 해결 할 수 있다. 고유치의 비는 상관계수와는 다르게 페닝 신호가 존재 하지 않는 경우 입력 신호의 공분산 행렬로부터 구한 두 고유치 값이 같아져 비가 1에 가까운 값이 된다. 그림 1에서 보듯 페닝 각도에 대한 영향을 상관계수 보다 적게 받아 각도에 상관없이 비교적 일정한 값을 보이면 알 수 있다.

제안된 알고리즘에서 Ambient를 분리하는 방법은 Modified PCA 알고리즘에서 상관 계수를 고유치의 비로 대체하여 분리하는 것이다. 고유치의 비가 1이 되는 경우는 상관 계수의 비가 1이 되는 경우와 반대이기 때문에 고유치의 비를 이용한 새로운 파라미터 Γ 를 만들고 Γ 는 아래 식과 같이 정의 된다.

$$\Gamma = \left(1 - \sqrt{\frac{\lambda_2}{\lambda_1}} \right) \quad (13)$$

고유치의 비는 Primary 성분 방향의 에너지와 Ambient 성분 방향의 에너지의 비가 되므로 제곱근을 취하고 상관 계수와 동일한 행동을 보일 수 있도록 1에서 뺀 값을 Γ 로 결정한다. 파라미터 Γ 를 이용하여 Ambient를 분리하는 식을 다시 쓰면 다음과 같다.

$$\begin{aligned} \tilde{N}_L &= \Gamma \tilde{N}_L + (1 - \Gamma) X_L \\ \tilde{N}_R &= \Gamma \tilde{N}_R + (1 - \Gamma) X_R \end{aligned} \quad (14)$$

4-2 Panning 신호 분리

페닝 신호는 기존의 PCA 알고리즘에서는 잔여 Ambient가 섞인 채로 분리 되게 된다. PCA로 분리된 페닝 신호에서 추정된 페닝 계인은 이미 알고 있는 파라미터이며 이것과 4-1에서 추정하여 분리해낸 Ambient 신호를 가지고 페닝 신호에 섞인 잔여 Ambient를 제거 할 수 있다. 잔여 Ambient를 제거 하는 방법은 다음 식과 같다.

$$\tilde{S} = \hat{S} - (\hat{a}_L \tilde{N}_L + \hat{a}_R \tilde{N}_R) \quad (15)$$

또한 수식 (14)를 수식 (9)과 (10)를 이용하여 다시 정리하면 다음과 같다.

$$\begin{aligned} \tilde{N}_L &= \Gamma(X_L - \hat{a}_L \hat{S}) + (1 - \Gamma) X_L \\ &= \Gamma \left\{ X_L - (\hat{a}_L^2 X_L + \hat{a}_L \hat{a}_R X_R) \right\} + (1 - \Gamma) X_L \\ &= X_L - \Gamma(\hat{a}_L^2 X_L + \hat{a}_L \hat{a}_R X_R) \\ \tilde{N}_R &= X_R - \Gamma(\hat{a}_R^2 X_R + \hat{a}_L \hat{a}_R X_L) \end{aligned} \quad (16)$$

그리고 수식 (15)를 수식 (16)을 이용하여 다시 정리하면 아래 식과 같이 정리 할 수 있다.

$$\begin{aligned} \tilde{S} &= \hat{S} - \left\{ \hat{S} - \Gamma(\hat{a}_L(\hat{a}_L^2 + \hat{a}_R^2)X_L + \hat{a}_R(\hat{a}_L^2 + \hat{a}_R^2)X_R) \right\} \\ &= \hat{S} - \left\{ \hat{S} - \Gamma(\hat{a}_L X_L + \hat{a}_R X_R) \right\} \\ &= \hat{S} - (\hat{S} - \Gamma \hat{S}) \\ &= \Gamma \hat{S} \end{aligned} \quad (17)$$

따라서 잔여 Ambient가 제거된 페닝 신호 \tilde{S} 는 기존의 PCA를 통해 추정된 페닝 신호 \hat{S} 에 파라미터 Γ 를 곱하여 간단하게 얻을 수 있게 된다.

5. 모의 실험

모의 실험에서는 제안된 알고리즘과 Modified PCA와의 음원 분리 성능을 비교해 보았다. 입력 스테레오 신호는 하나의 모노 페닝 음원을 각도를 바꾸어 가면서 페닝 시키고, 백색 잡음을 이용하여 지수함수적으로 크기가 감소하는 린임펄스 응답을 모델링 한 뒤, 각 스테레오 왼쪽과 오른쪽 채널에 모노 페닝 음원과 린임펄스 응답을 컨볼루션 시켜 얻은 상관성이 없는 Ambient를 더해주었다. 입력 신호의 샘플링 주파수는 44.1Khz이고 PCA 분석은 STFT에서 프레임 길이는 1024샘플 단위로 처리 하였다. 그림 2는 입력 페닝 신호가 30도로 페닝된 신호의 Modified PCA Ambient 신호를, 그림 3는 동일한 신호를 제안된 알고리즘을 이용하여 분리한 Ambient 신호를 나타낸다. 그림 2에서 보듯 극단적으로 페닝된 입력 신호에 대해서도 상관계수가 작아져서 오른쪽 Ambient에 많은 양의 primary 성분이 섞여서 나오게 된다. 반면에 그림 3에서 제안된 알고리즘은 입력 신호가 극단적으로 페닝된 경우에도 Ambient를 페닝각도에 영향을 받지 않고 분리할 수 있다.

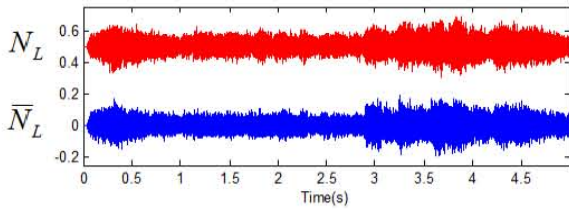


그림 2 추정된 Ambient 신호 (Modified PCA, 30도)

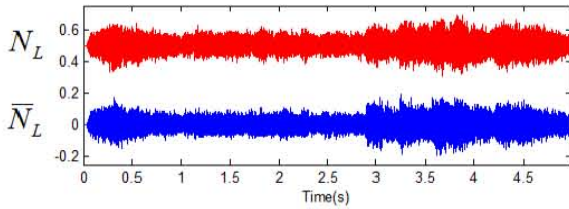


그림 3 추정된 Ambient 신호 (제안된 알고리즘, 30도)

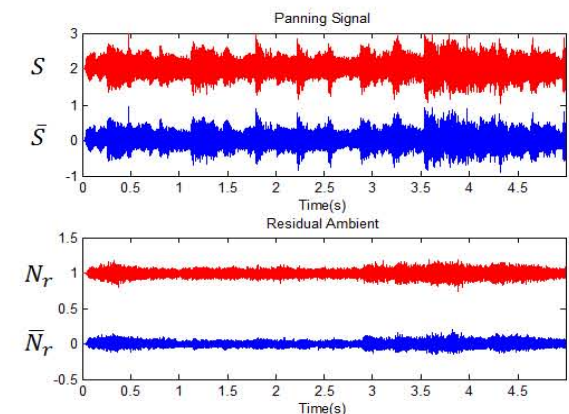


그림 4 추정된 페닝 신호와 잔여 Ambient 신호(제안된 알고리즘)

다음으로 그림 4는 제안된 알고리즘으로 페닝 신호를 분리한 것과 추정된 Ambient로부터 잔여 Ambient를 비교한 것이다. 그림 4에서는 실제 페닝 신호와 제안된 알고리즘으로 분리해낸 페닝 신호와 추정된 Ambient로부터 구한 잔여 Ambient 성분을 비교한 것이다. 여기서

$Nr = \hat{a}_L N_L + \hat{a}_R N_R$, $\bar{N}r = \hat{a}_L \bar{N}_L + \hat{a}_R \bar{N}_R$ 을 의미한다. 제안된 알고리즘은 추정된 Ambient와 페닝 계수로 부터 잔여 Ambient 성분을 제거함으로써 실제 페닝 신호와 배경음을 정확하게 분리하고 있음을 보여준다.

6. 결론

본 논문에서는 PCA를 이용한 향상된 Primary-Ambient 분리 알고리즘을 제안하였다. 기존 Modified PCA 알고리즘이 상관 계수를 사용하여 페닝 각도에 따라 음원 분리 성능이 저하되나 제안한 알고리즘은 두 입력 채널의 공분산 행렬의 고유치 비를 이용함으로써 페닝 성분이 존재하지 않고 Ambient만 있을 경우는 물론 페닝각도에도 상관없이 음원 분리 성능이 강건하게 유지된다. 또한 페닝 신호에서 잔여 Ambient 성분을 제거하여 더 정확한 실제 페닝 신호를 분리해 낼 수 있다.

참고 문헌

- [1]. R. Dressler, "Dolby Surround Prologic II de-coder, principles of operation," Tech. Rep. Dolby Laboratories, 2000
- [2]. Pulkki, Ville. "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", AES convention, vol. 45, Issue 6, pp 456-466, 1997
- [3]. J. Usher and J. Benesty, "Enhancement of spatial sound quality: A new reverberation-extraction audio upmixer," IEEE Trans. Audio, Speech, and Language Processing, vol. 15, no. 7, pp. 2141-2150, 2007
- [4]. C. Faller, "Multiple-loudspeaker playback of stereo signals," J. AES, vol. 54, no.11, pp. 1051-1064, 2006.
- [5]. S.W. Jeon, Y.C. Park, S.P. Lee, D.H. Yoon, "Robust Representation of Spatial Sound in Stereo-to Multichannel Upmix", AES convention, 2010
- [6]. Goodwin. M.M., "Geometric Signal Decompositions for Spatial Audio Enhancement", IEEE ICASSP conf, pp 409-412, 2008
- [7]. E. Zwicker, E. Terhardt, "Analytical Expressions for Critical-Band rate and Critical Bandwidth as a Function of Frequency", ASOA, Vol 68, pp 1523-1525, 1980
- [8] Briand, Manuel, Martin, Nadine, Virette, David, "Parametric Representation of Multichannel Audio Based on Principal Component Analysis", AES convention, 2006