

정규논문 (Regular Paper)

방송공학회논문지 제19권 제1호, 2014년 1월 (JBE Vol. 19, No. 1, January 2014)

<http://dx.doi.org/10.5909/JBE.2014.19.1.64>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 위상 정렬을 이용한 주성분 분석법의 강인한 스테레오 음원 분리 성능 유지 기법

백 용 현<sup>a)</sup>, 현 동 일<sup>b)</sup>, 박 영 철<sup>c)‡</sup>

### Robust Primary-ambient Signal Decomposition Method using Principal Component Analysis with Phase Alignment

Yong-Hyun Baek<sup>a)</sup>, Dong-il Hyun<sup>b)</sup>, and Young-Cheol Park<sup>c)‡</sup>

#### 요 약

스테레오 신호에서 방향성을 갖는 주음원 성분과 배경음 성분을 분리하는 과정은 스테레오 음원을 다채널 포맷으로 변환 하는 업믹스 기술의 전처리 단계로 필요하다. 이를 위하여 PCA(Principal Component Analysis)가 널리 사용되고 있으나, 기존의 방법에서는 스테레오 음원이 진폭 패닝만 되었다고 가정하고 있기 때문에 채널 신호 간 위상 차이가 발생하는 경우 성능이 저하된다. 본 논문에서는 채널 신호 간의 위상차에 의한 주성분 분석 방법의 오차를 분석하고 이를 보완하여 주성분 분석법 성능 저하를 방지하는 방법을 제안한다. 제안된 알고리즘은 주파수 빈(bin) 단위에서 채널 간 위상차를 추정하여 보상하고 임계대역 단위로 주성분과 주변 성분을 분리한다. 실험 결과로부터 위상차가 있는 스테레오 음원에 대하여 제안된 알고리즘이 PCA 분리 성능의 저하를 방지 하며, 패닝 각도, PAR(Primary to Ambient energy ratio)에 영향 받지 않는 강건한 분리 성능을 보임을 확인하였다.

#### Abstract

The primary and ambient signal decomposition of a stereo sound is a key step to the stereo upmix. The principal component analysis (PCA) is one of the most widely used methods of primary-ambient signal decomposition. However, previous PCA-based decomposition algorithms assume that stereo sound sources are only amplitude-panned without any consideration of phase difference. So it occurs some performance degradation in case of live recorded stereo sound. In this paper, we propose a new PCA-based stereo decomposition algorithm that can consider the phase difference between the channel signals. The proposed algorithm overcomes limitation of conventional signal model using PCA with phase alignment. The phase alignment is realized by using inter-channel phase difference (IPD) which is widely used in parametric stereo coding. Moreover, Enhanced Modified PCA(EMPCA) is combined to solve the problem of conventional PCA caused by Primary to Ambient energy Ratio(PAR) and panning angle dependency. The simulation results are presented to show the improvements of the proposed algorithm.

Keyword : PCA, decomposition, primary signal, ambient signal, inter-channel phase difference

a) 연세대학교 전산학과(Computer Science. Dep., Yonsei Univ.)

b) 연세대학교 전기전자공학부(Electrical & Electronic Eng., Yonsei Univ.)

c) 연세대학교 컴퓨터정보통신공학부(Computer & Telecommunications Eng., Yonsei Univ. )

‡ Corresponding Author : 박영철(Young-Cheol Park)

E-mail: young00@yonsei.ac.kr

Tel: +82-33-760-2756

· Manuscript received October 4, 2013 Revised December 13, 2013 Accepted December 13, 2013

## I. 서론

업믹스<sup>[1]</sup>는 스테레오 음원을 홈 시어터 환경과 같은 5.1 채널 서라운드 재생 환경에서 재생하기 위한 다채널 포맷 변환 기법이다. 스테레오 신호의 업믹스 과정은 흔히 두 단계로 나누어진다. 첫 번째 단계는 방향성이 존재하며 두 채널간의 상관도가 높은 주(primary) 성분과 방향성이 없고 두 채널간의 상관도가 없는 주변(ambient) 성분을 분리하는 것이고, 두 번째 단계는 분리된 음원을 재생하고자 하는 다채널 스피커 포맷에 맞게 렌더링 하는 과정이다.

지난 수년간 주성분과 주변성분을 분리하기 위한 다양한 알고리즘이 제안되었다<sup>[2][3][4]</sup>. 적응 필터를 이용하여 서로 반대 채널을 참고하여 상관도가 높은 주성분을 찾아내는 채널 예측 방법<sup>[2]</sup>, 필터 오차 신호에 대한 최소 자승법 (Least Square Method)를 이용하여 주성분과 주변 성분을 분리 하는 방법<sup>[3]</sup>, PCA를 이용하여 음원의 패닝 계수를 추정하여 주성분과 주변 성분을 분리하는 방법<sup>[4]</sup> 등이 제안되었다. [4]에서는 실제로 스테레오 신호에 상관도가 높은 주성분과 두 개의 독립적인 주변성분이 존재하는데 PCA로는 하나의 주변성분만 분리할 수 있기 때문에 추정한 패닝 계수와 신호를 행렬의 형태로 표현한 뒤 의사역 (Pseudo inverse)하여 두 개의 주변성분을 분리할 수 있도록 한 방법이다. 그러나 대부분의 기존 방법들은 스테레오 신호에 포함된 음원들이 진폭 패닝(amplitude panning)만<sup>[5]</sup> 되었다고 가정한다. 따라서 채널 신호에 포함된 음원 성분 간의 위상차는 고려되지 않는다. 그러나 스테레오 음원은 진폭 패닝은 물론 채널 간의 시간지연을 포함하는 매우 다양한 형태로 믹싱될 수 있으며, 특히 마이크로폰을 이용하여 라이브로 녹음된 음원의 경우, 음원의 위치에 따라 두 채널 신호 간 위상차가 존재하게 된다. 이러한 위상차를 갖는 음원에 기존의 음원 분리 알고리즘을 적용할 경우, 상당한 성능 저하가 발생하게 된다<sup>[6]</sup>. PCA의 분리 성능 저하의 요인으로서는 크게 위상차로 인한 패닝 계수 추정의 오차와 다운 믹스시 동기화가 되지 못해 발생하는 오차로 크게 두가지로 분석 할 수 있다. 참고 문헌 [6]에서는 오차의 원인을 위상차로 인하여 주성분의 상관도가 실제보다 떨어지면서 실제로 주성분인 성분이 주변

성분으로 분리되어 나오기 때문에 발생하는 것으로 보고 시간 영역에서 두 채널의 위상을 보상하는 방법을 사용한다.

본 논문에서는 앞서 설명한 분리 방법들 중 PCA를 이용하되 이러한 채널 간 음원의 위상차가 발생하는 경우를 고려한 음원 분리 알고리즘을 제안한다. 제안하는 알고리즘에서는 파라메트릭 스테레오 코딩에서 주로 사용하는 채널 간 위상차를 이용하여 주파수 영역에서 위상을 사전에 정렬한 뒤 PCA를 사용하여 음원을 분리한다. 참고 문헌 [6]에서와 같이 시간영역에서 동기화 하게 되며 주성분이 하나가 아니라 여러 개가 존재하는 경우 동기화 시키기 어려운 문제점이 있다. 이러한 한계를 극복하기 위해 음원분리는 주파수 영역에서 임계 대역으로 나누며 임계 대역 내에서는 하나의 주성분만 존재한다는 가정 하에 분리하게 된다<sup>[4]</sup>. 그러나 실제로 신호의 분석은 일정 프레임 단위로 단 구간 푸리에 변환하여 주파수 영역으로 변환하게 되는데 프레임 단위에서는 위상차를 보상한다 하더라도 패닝 계수 추정의 오차를 보상할 수 없게 된다. 그러나 본 논문에서 분석을 통하여 이러한 패닝 계수오차가 실제로 PCA 성능 저하에 큰 영향을 미치지 않고 다운믹스시 동기화가 주요 인임을 이용하여 주파수 영역에서 동기화를 통해 PCA 성능이 위상차로 인해 크게 저하 되지 않고 강건하게 유지되게 하였다. 또한 PCA가 패닝 각도 및 PAR의 따라 그 성능이 저하되는 문제점을 해결한 EMPCA(Enhanced Modified Principal Component Analysis)<sup>[7]</sup>을 적용하여 단순히 진폭 패닝된 스테레오 음원 뿐만 아니라 어떠한 스테레오 음원에 대해서도 분리 성능이 강건하게 유지되도록 한다.

본 논문의 구성은 다음과 같다. 2장에서 스테레오 신호의 모델링, 3장에서는 위상 정렬 EMPCA를 이용한 주성분과 주변성분 분리방법을 설명한다. 4장에서는 제안된 알고리즘의 성능을 평가하고 5장에서 결론을 맺는다.

## II. 스테레오 신호 모델링

업믹스 기술에서 입력 스테레오 신호를 분석할 때 단시

간 푸리에 변환(Short-Time Fourier Transform, STFT)을 이용하여 분석한다. 이렇게 주파수 도메인으로 변환된 계수들을 사람의 청각 특성에 따라 각각 서브밴드별로 묶어 나누어 처리한다<sup>[8]</sup>. 스테레오 신호에서 음원은 -30°와 30°에 배치된 두 개의 스피커를 가정하고 진폭 패닝(Amplitude Panning, AP)을 통해 정위 되었고, 여기에 각각의 상관도가 없는 주변 성분이 더해진 형태로 모델링 할 수 있다. 모델링된 입력 두 채널의 신호를 주파수 영역에서 표현하면 수식 (1)과 같이 나타낼 수 있다.

$$\begin{aligned} X_L &= P_L + N_L = a_L S e^{j\theta_L} + N_L \\ X_R &= P_R + N_R = a_R S e^{j\theta_R} + N_R \end{aligned} \quad (1)$$

여기서  $P_L, P_R$ 은 주성분,  $N_L, N_R$ 은 각각의 채널에 더해진 주변 성분이다. 또한  $S$ 는 음원신호,  $a_L, a_R$ 은 음원 패닝 계수이며  $\theta_L, \theta_R$ 은 각 채널 음원 성분의 위상이다. 수식의 간편함을 위해 시간, 주파수 인덱스는 생략되었다. 기존의 음원 분리 알고리즘<sup>[2][3][4]</sup>에서는 수식 (1)의 신호의 모델링이 다음 수식 (2)(3)(4)와 같은 관계를 만족한다고 가정한다.

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

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

$$P_L = \frac{a_L}{a_R} P_R e^{j(\pm \Delta\theta)}, \Delta\theta = \theta_L - \theta_R \quad (4)$$

수식(2)는 스테레오 신호에 포함된 주성분의 에너지가 실제 음원의 에너지와 동일함을 뜻하고, 이는 채널 간의 위상 차, 즉 시간 지연이 없을 때 만족한다. 수식(3)은 각각의 주성분과 주변성분간의 상관도가 없음을 나타낸다. 수식 (4)는 패닝 된 신호가 한쪽 채널에 시간 지연이 되어 있음을 나타내며 여기서  $\Delta\theta$ 는 두 채널간의 위상차이다. 추가적으로 주변성분보다 에너지가 우세하며 두 채널 사이의 주변 성분간의 에너지는 같다고 가정한다<sup>[9]</sup>. 또한 각 밴드 별로 하나의 주성분만이 존재한다고 가정한다.

### III. 위상 정렬 주성분 분석 음원 분리 알고리즘

#### 1. PCA를 이용한 음원 분리

PCA를 이용하여 음원을 분리 할 때 스테레오 두 채널의 주성분에 곱해진 패닝 계수  $a_L$ 과  $a_R$ 은 채널의 2x2 공분산 행렬로부터 추정할 수 있고 공분산 행렬은 아래 식과 같이 계산된다.

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

수식 (5)의 공분산 행렬로부터 고유치 분해를 통하여 다시 표현하면 다음과 같다.

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

여기서  $\lambda_{1,2}$ 는 고유치 값이고  $V_{1,2}$ 는 각 고유치 값에 대응하는 고유치 벡터이다. 이때 두 고유치 값 중 큰 값을 가지는 고유치 값은  $\lambda_1 = \sigma_S^2 + \sigma_N^2$ 으로 표현할 수 있고 이는 주성분에 해당하는 성분의 에너지가 된다. 또한 이에 대응하는 고유치 벡터의 방향이 패닝 각도가 되므로 AP 계수를 수식 (7)과 같이 얻을 수 있다.

$$\begin{aligned} v_{1,1} &= \frac{|r_{LR}|}{\sqrt{(\lambda_1 - r_{LL})^2 + |r_{LR}|^2}} = a_L \\ v_{1,2} &= \frac{(\lambda_1 - r_{LL})}{\sqrt{(\lambda_1 - r_{LL})^2 + |r_{LR}|^2}} = a_R \end{aligned} \quad (7)$$

한편 기존의 PCA에서는  $\Delta\theta = 0$ 으로 가정하여, (즉, 수식 (1)에서  $\theta_L = \theta_R = 0$ ) 추정된 패닝 계수로 부터 주성분과 주변성분은 수식 (1), (2)를 이용하여 다음 수식 (8), (9)와 같이 추정된다.

$$\hat{S} = v_{1,1} X_L + v_{1,2} X_R = S + a_L N_L + a_R N_R \quad (8)$$

$$\begin{aligned}\widehat{N}_L &= X_L - v_{1,1}\widehat{S} \\ &= (1 - a_L^2)N_L - a_L a_R N_R \\ &= a_R^2 N_L - a_L a_R N_R \\ &= a_R(a_R N_L - a_L N_R), \\ \widehat{N}_R &= X_R - v_{1,2}\widehat{S} = a_L(a_L N_R - a_R N_L)\end{aligned}\quad (9)$$

따라서 PCA가 주성분과 주변성분을 직교화 하여 상관관계가 없도록 분리하는 장점이 있는 반면에 수식 (8)에서와 같이 PCA를 통해 추정된 주성분에는 실제 주성분뿐만 아니라 주변 성분 또한 포함하고 있음을 알 수 있다. 수식 (9)에서도 추정된 주변 성분은 페닝 계수의 값에 영향을 받는다. 즉, 주성분의 페닝 각도가 어느 한쪽 방향으로 극단적으로 치우쳐 있을 경우 정확하게 주변성분을 분리하지 못하는 단점도 가지고 있다<sup>[7]</sup>.

## 2. 위상차에 의한 분리 성능 저하

3.1절에 설명된 바와 같이 기존의 PCA를 이용한 음원 분리 방법은 두 스테레오 채널간의 위상차가 존재 하지 않는 경우이다. 만약 채널 음원 신호 간 시간 지연이 존재하는 경우, 오른쪽 채널을 기준으로 왼쪽 채널 신호에 시간 지연이 발생했다고 가정하면, 수식(1)의 신호의 모델링은 아래와 같이 다시 쓸 수 있다.

$$\begin{aligned}X_L &= a_L(1 + \delta)S e^{j\Delta\theta} + N_L \\ X_R &= a_R S + N_R\end{aligned}\quad (10)$$

여기서  $\delta$ 는 왼쪽 채널 신호의 시간 지연에 의해 동기화 되지 못해 발생하는 두 채널의 음원 신호간의 스펙트럼 오차이다. 이 스펙트럼 오차  $\delta$ 는 채널이 동기화 되었을 때의 상관도가 1이라면 시간 지연으로 인하여 두 채널에 존재하는 주성분 사이의 상관도가 떨어지게 되므로 나타나는 오차이다. 실제로 상관도가 있지만 시간 지연으로 인해 상관도가 떨어지게 되고 이것이 페닝 오차를 발생시키는 이유 중 하나가 된다<sup>[6]</sup>. 수식 (10)으로부터 공분산 행렬은 다음식과 같이 구할 수 있다.

$$\mathbb{R} = \begin{bmatrix} a_L^2(1 + \delta)^2\sigma_S^2 + \sigma_N^2 & a_L a_R(1 + \delta)\sigma_S^2 e^{j\Delta\theta} \\ a_L a_R(1 + \delta)\sigma_S^2 e^{-j\Delta\theta} & a_R^2\sigma_S^2 + \sigma_N^2 \end{bmatrix}\quad (11)$$

수식 (11)을 고유치 분해하여 구한 주성분에 해당하는 고유치 값은 다음과 같다.

$$\begin{aligned}\widetilde{\lambda}_1 &= a_L^2(1 + \delta)^2\sigma_S^2 + a_R^2\sigma_S^2 + \sigma_N^2 \\ &= \sigma_S^2 + (2\delta + \delta^2)a_L^2\sigma_S^2 + \sigma_N^2\end{aligned}\quad (12)$$

수식 (12)에서 첫번째 항은 실제 주성분의 에너지, 세번째 항은 주변성분의 에너지이다. 이것은 시간 지연이 없는 경우에도 주성분에 주변성분 에너지가 포함되어 있는 일반적인 주성분 분석법의 문제이다. 두번째 항은 시간지연에 의해 발생하는 에너지 오차로 앞서 언급했듯이 두 채널 사이의 주성분간의 상관도가 떨어져서 발생하는 오차 에너지이다. 오차가 발생한 고유치 값에 대응하는 고유치 벡터를 수식 (7)을 이용하여 구하면 다음과 같다.

$$\begin{aligned}\widetilde{a}_L = v_{1,1} &= \frac{a_L(1 + \delta)}{1 + a_L\delta} \\ \widetilde{a}_R = v_{1,1} &= \frac{a_R}{1 + a_L\delta}\end{aligned}\quad (13)$$

수식 (13)에서 보듯 페닝 계수 오차  $\delta$ 가 클수록, 또는 시간 지연이 있는 채널의 페닝 계수가 클수록 페닝 계수 오차가 크게 나타난다. 그러나 일반적인 스테레오 녹음 환경을 생각해보면 두 채널의 위상차는 2ms이내이고 이로 인한 스펙트럼 오차는 무시 할 수 있을 정도로 작다고 예측해 볼 수 있다. 따라서 참고 문헌 [6]에서 언급한 오차  $\delta$ 로 인한 PCA의 성능 저하는 실제로 미미하다.

PCA의 분리 성능이 떨어지는 가장 큰 요인은 위상차로 인한 스펙트럼 오차 (즉, 상관도가 떨어져서 발생하는 오차)  $\delta$ 가 아니라 수식 (8)과 같이 페닝 계수로부터 스테레오 음원을 하나의 모노음원으로 다운믹스 할 때의 두 채널간의 동기화 문제이다. 기본적으로 다운믹스는 두 채널의 시간 지연이 없다는 가정 하에 이루어지지만 시간 지연이 있는 음원을 수식 (8)을 이용하여 주성분을 추정한다면 아래와 같다.

$$\begin{aligned}\widehat{S} &= \widetilde{a}_L X_L + \widetilde{a}_R X_R \\ &= \left( \widetilde{a}_L^2 e^{j\theta_{pvd}} + \widetilde{a}_R^2 \right) S + \widetilde{a}_L N_L + \widetilde{a}_R N_R\end{aligned}\quad (14)$$

수식 (14)에서처럼 시간 지연이 있는 음원은 두 채널간의 동기화가 되지 못하기 때문에 패닝 계수가 정확하게 추정 되었다 하더라도 원래 주성분을 제대로 추정하지 못하게 된다. 정리하자면 위상차이로 인해 주성분간의 상관도가 떨어져서 발생하는 스펙트럼 오차  $\delta$ 는 패닝 계수 추정에 큰 영향을 미치지 않지만, 다운 믹스 단계에서 위상차가 두 채널의 동기화를 이루지 못하게 하여 실제 주성분과 추정 한 주성분의 오차가 크게 나타나게 된다.

### 3. 위상 정렬 주성분 분석법

3.2절에서 살펴 보았듯이 위상차로 인해 발생하는 오차는 PCA를 이용한 음원 분리 성능을 저하시킨다. 이 문제를 해결하기 위해서 두 스테레오 채널간의 위상차를 보상해주고 주성분 분석법을 적용해야 한다. 수식 (10)으로부터 상호 위상차  $\Delta\theta$ 는 파라메트릭 스테레오 코딩에서 널리 사용되는 것으로 전체 위상차(Overall Phase Difference, OPD)를 이용한 각 채널의 위상차(Channel Phase Difference, CPD)를 구하여 두 채널의 위상차를 보상하여 준다. 파라메트릭 스테레오 코딩에서 각 채널의 CPD는 아래와 같은 식으로 추정할 수 있다<sup>[10]</sup>.

$$\begin{aligned} \theta_{CL} &= \tan^{-1}\left(\frac{c_2 \sin(\Delta\theta)}{c_1 + c_2 \cos(\Delta\theta)}\right) \\ \theta_{CR} &= \tan^{-1}\left(\frac{c_1 \sin(\Delta\theta)}{c_2 + c_1 \cos(\Delta\theta)}\right) \end{aligned} \quad (15)$$

여기서  $c_1, c_2$ 는 두 채널의 상호 레벨차(Inter Intensity Difference, IID)로부터 구해지는 레벨 계수로서 아래와 같이 정의 된다.

$$IID = 10 \log_{10} \left( \frac{r_{LL}}{r_{RR}} \right) \quad (16)$$

$$c_1 = \sqrt{\frac{10^{\frac{IID}{10}}}{1 + 10^{\frac{IID}{10}}}}, c_2 = \sqrt{\frac{1}{1 + 10^{\frac{IID}{10}}}} \quad (17)$$

수식 (16)의 CPD를 이용하여 위상을 정렬한 뒤 PCA를

통해 구한 패닝 계수로부터 수식 (8),(9)를 이용하여 주성분과 주변 성분을 분리한다. 그러나 3.1절에서 설명 했듯이 기존의 주성분 분석법은 패닝 각도 및 PAR에 따라 분리 성능이 저하되므로 이러한 문제를 해결한 EMPCA를 이용하여 분리 한다. EMPCA는 다음 수식 (18), (19) 그리고 (20)으로 음원을 분리한다.

$$\hat{S} = \overline{a_L} X_L e^{-j\theta_{CL}} + \overline{a_R} X_R e^{j\theta_{CR}} \quad (18)$$

$$\overline{S} = \sqrt{\frac{\lambda_1 - \lambda_2}{\lambda_1}} \hat{S} \quad (19)$$

$$\begin{aligned} \overline{N_L} &= X_L e^{-j\theta_{CL}} - \left(1 - \sqrt{\frac{\lambda_2}{\lambda_1}}\right) \overline{a_L} \hat{S} \\ \overline{N_R} &= X_R e^{j\theta_{CR}} - \left(1 - \sqrt{\frac{\lambda_2}{\lambda_1}}\right) \overline{a_R} \hat{S} \end{aligned} \quad (20)$$

여기서  $\overline{\bullet}$ 로 표현된 파라미터들은 위상 정렬 이후에 추정 한 패닝 계수와 주성분 주변성분을 나타낸다. 제안된 방법은 PCA를 수행하기 전에 입력 스테레오 신호의 위상을 보상하여 주고 여기에 주성분의 패닝각도와 PAR에 따라 분리 성능이 저하되는 기존에 PCA가 가지고 있는 문제를 해결한 EMPCA 방법을 결합함으로써 기존의 분리방법들이 진폭만 패닝된 스테레오 신호만을 가정했던 한계를 벗어나 입력신호의 위상차, 패닝 각도 그리고 PAR에 상관없이 분리 성능이 강건하게 유지되도록 하였다.

일반적으로 PCA를 적용할 때 사람의 청각 특성을 고려한 임계 대역(Critical Band)으로 나누어 처리하는데 이는 임계 대역 내에서의 대역 잡음만이 음의 마스킹(Masking)에 기여하고 대역 외 잡음은 이에 기여하지 않기 때문이며 보통 오디오 신호처리에서는 이러한 음의 라우드니스와 마스킹 역치가 급격히 변하는 구간을 기준으로 나누어 처리하게 된다.

그러나 수식 (8)을 이용하여 주성분을 다운 믹스 할 때 약간의 위상차만 존재하더라도 원래 신호를 정확하게 복원하지 못하게 된다. 따라서 위상 정렬을 할 경우에는 임계 대역 별로 처리하기 보다는 각 Bin별로 정확 하게 구해주어 보상

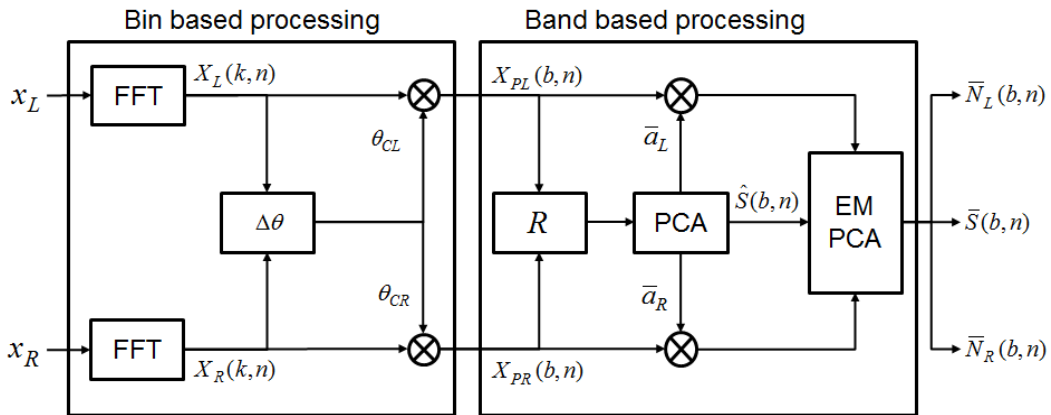


그림 1. 위상 정렬 주성분 분석법 블록도  
 Fig 1. Phase aligned PCA algorithm block diagram

하는 것이 위상차 추정치의 오차를 줄일 수 있다. 아래 그림 1은 전체 음원 분리 시스템의 블록 다이어그램을 나타낸다.

#### IV. 모의 실험

모의 실험을 통하여 제안된 알고리즘의 성능을 평가하였

다. 평가 방법은 위상 정렬을 한 경우와 하지 않은 경우의 주성분과 주변성분의 분리 성능을 비교하였다. 실험에 사용된 스테레오 음원은 하나의 모노 신호를 레벨 페닝을 통해 음상을 정위시키고 이상적인 비교를 위해 상관도가 없는 백색 잡음을 주변성분으로 하여 두 채널에 각각 더해주었다. 입력 신호의 샘플링 주파수는 44.1KHz이고 주성분 분석은 STFT에서 프레임 길이는 1024샘플 단위로 처리 하

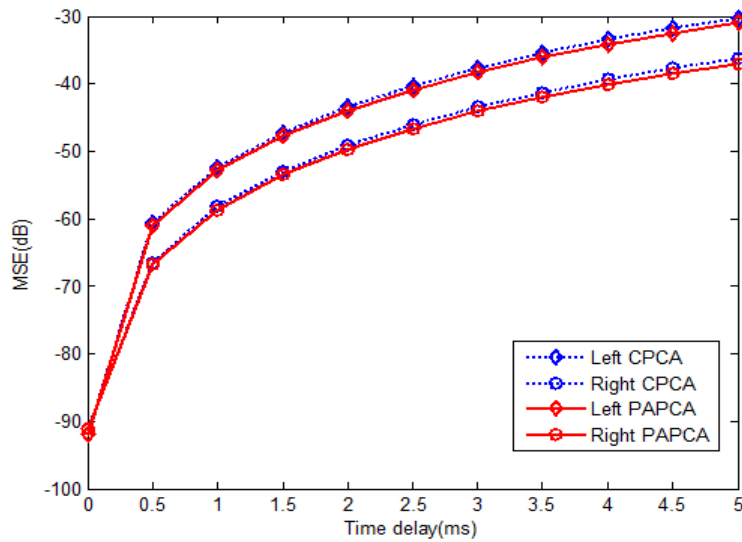


그림 2. 페닝 계수 추정 결과 (주성분 페닝 각도 : 10°)  
 Fig 2. Panning Gain Estimation results (Primary panning angle : 10°, Conventional PCA(CPCA), Phase Aligned PCA(PAPCA))

였다. 그림 2는 입력 스테레오 신호의 주성분의 패닝 각도가 10도, 이 때 패닝 계수는  $a_L = 0.4697, a_R = 0.8828$ 이다. 여기에 왼쪽 채널의 시간지연이 0~5ms까지 0.5ms단위로 변할 때 패닝 계수의 MSE를 측정한 결과이고 MSE는 아래 식과 같이 구해진다.

$$E_{MSE} = 10 \log_{10} \left( E \{ a_i - \tilde{a}_i \}^2 \right), i = L, R \quad (21)$$

그림 2에서 보면 위상차를 보상한다 하더라도 시간차가 길어질수록 패닝 계수오차는 증가한다. 제안된 PAPCA로 추정된 패닝 계수의 오차가 CPCA보다는 조금 더 적게 남을 확인 할 수 있지만 그 향상된 정도는 미미하다. 이런 경우 시간 영역에서 위상차를 보상하지 않고 주파수 영역에서 위상차를 보상해 준다는 것은 일정 프레임 단위로 나누어 처리할 때 한 프레임 내에서는 위상차 보상이 불가능하다. 즉, 스펙트럼 오차  $\delta$ 가 STFT로 분석 할 때부터 필연적으로 발생하는 주어진 오차이며 이를 줄이기 위해서는 시간 영역에서 두 채널간의 위상차를 보정한 뒤 STFT 분석이 이루어져야 한다. 그러나 앞서 언급하였듯이 스테레오

신호에서 패닝 된 주성분이 하나가 아니라 여러 개가 존재할 수 있다. 이런 경우 시간 영역에서 채널 간 동기화가 어렵고 주파수 임계 대역 내에 하나의 주성분만 존재한다는 가정하에 대역 통과 필터를 이용하여 신호를 분리하여 시간 영역에서 처리 한다 하더라도 연산량이 매우 많아지게 된다.

그러나 그림 2에서 보듯 시간 지연이 5ms로 제법 긴 시간차가 나더라도 패닝 계수의 MSE는 -30dB 정도로 그 차이가 작다. 일반적인 스테레오 녹음 기법에서 두 마이크 간의 시간차는 2ms이내이고 그림 2에서 시간차 2ms일때의 MSE는 -50에서 -40dB사이이므로 그 오차가 더욱 작다고 할 수 있다. 따라서 위상차가 나더라도 일반적인 스테레오 녹음 환경을 고려할 때 CPCA와 PAPCA모두 패닝 계수는 거의 정확히 추정한다.

그림 3은 특정 시간 지연 시간 2ms에서 추정된 패닝 계수를 시간 축으로 관찰한 것이다. 그림 2에서 위상 정렬을 통해 스펙트럼 오차를 줄이지 못한다 하더라도 그림 3의 결과와 같이 위상차로 인한 스펙트럼의 오차가 무시 할 수 있을 만큼 작기 때문에 패닝 계수를 실제와 거의 동일하게

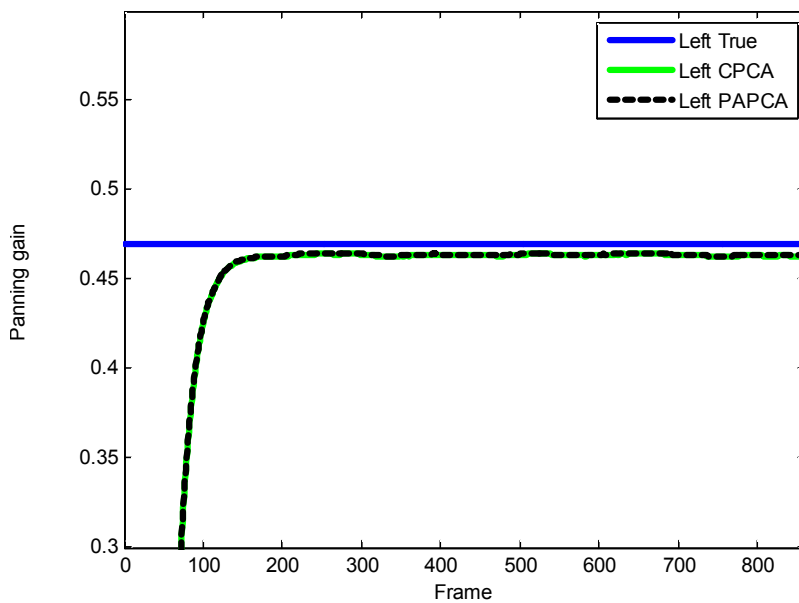


그림 3. 패닝 계수 추정 결과 (주성분 패닝 각도 : 10°, 시간 지연 : 2ms)  
 Fig 3. Panning Gain Estimation results (Primary panning angle : 10°, Time delay : 2ms)

추정함을 확인 할 수 있다. 즉, 참고 문헌 [6]에서와 같이 시간영역에서 위상차를 보상하지 않고 주파수 영역에서 하더라도 실험 결과와 같이 오차  $\delta$ 로 발생하는 PCA 분리 성

능 저하는 미미하기 때문에 위상 정렬은 주파수 영역에서 이루어지는 것이 알고리즘의 복잡도 측면에서 이득을 볼 수 있다.

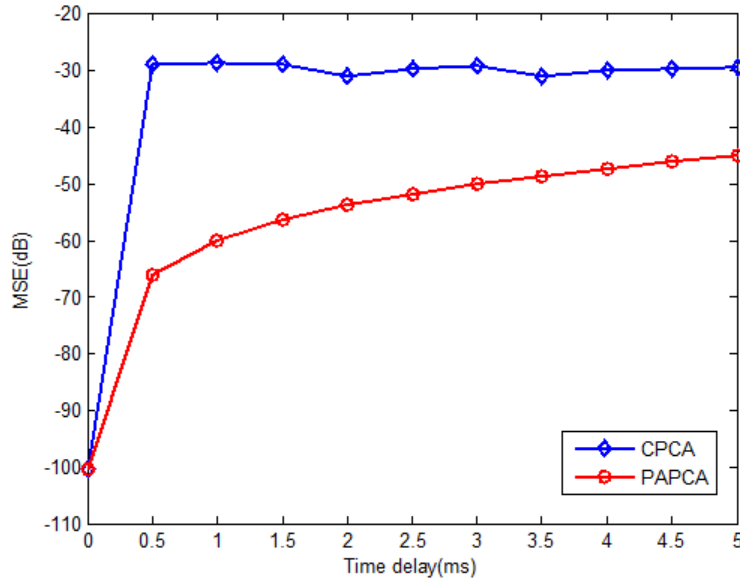


그림 4. 주성분 신호 추정 결과 (주성분 패닝 각도 : 10°)  
 Fig 4. Primary signal estimation results (Primary panning angle : 10°)

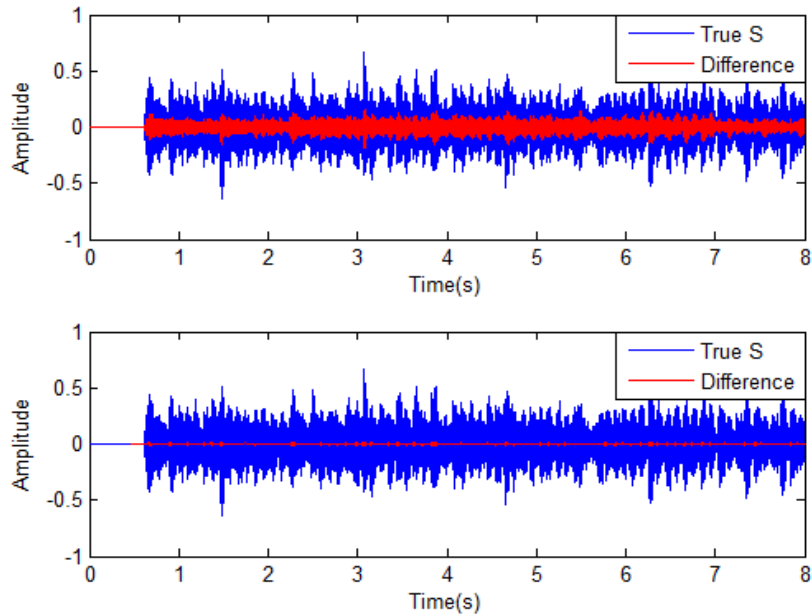


그림 5. 주성분 신호 추정 결과 (b) 기존의 PCA (c) 제안된 위상 정렬 PCA (주성분 패닝 각도 : 10°)  
 Fig 5. Primary signal estimation results (a) CPCA (b) PAPCA (Primary panning angle : 10°)



그림 4는 그림 2와 동일한 환경에서 기존에 PCA방법과 제안된 위상 정렬 PCA방법으로 추정된 주성분과 실제 주성분과의 MSE를 추정한 결과이며 수식 (21)의 식을 이용해  $S$ 와  $\bar{S}$ 의 차이를 구한 것이다. 페닝 계수 오차와는 다르게 그림 4에서 보면 위상차가 조금만 발생하더라도 주성분의 MSE가 매우 크게 나타난다. 따라서 앞서 설명했듯이 페닝 계수의 오차보다는 위상차로 인하여 동기화 되지 못하여 생기는 오차가 크며 그림 4는 제안된 알고리즘이 두 채널을 동기화 시킴으로써 정확히 주성분을 추정함을 보여준다.

그림 5 또한 특정 시간 지연 시간 2ms에서 추정된 주성분을 시간 축으로 관찰한 것으로 빨간 선은 실제 주성분과 추정된 주성분과의 오차를 나타낸다. 그림 5(a)처럼 비동기화가 주성분 신호의 큰 오차를 유발하지만 5(b)의 제안된 알고리즘의 경우 실제와 오차가 매우 작게 나타남을 확인할 수 있다.

마지막으로 그림 6은 주변성분 분리 성능을 비교한 것으로 입력 신호는 그림 2에서 사용한 신호와 동일하며 이상적인 실험을 위해 주변성분으로 각 채널간 상관도가 없는 백

색 잡음을 사용하였다. (b)와 같이 위상 정렬을 하지 않는 경우 이 때 제대로 추정하지 못한 주성분의 오차 신호가 주변성분 신호로 분리 되어 나오는 반면 위상 정렬을 할 경우 주성분의 추정 오차가 적어지고 그로 인하여 실제 주변 성분과 유사하게 분리해 내며 동기화가 되고 난 이후로는 실제 위상차가 없는 신호의 모델을 사용할 때와 동일한 환경이 되기 때문에 여기에 EMPCA를 적용하여 주성분의 페닝 각도, PAR에 무관하게 정확히 주변 성분을 추정할 수 있다.

### V. 결론

본 논문에서는 기존의 주성분-주변성분 분리 알고리즘이 단순히 진폭 페닝된 신호만을 다룬 한계점을 극복하여 신호의 위상차까지 고려한 음원 분리 알고리즘을 제안하였다. 제안된 알고리즘은 파라메트릭 스테레오 코딩에서 널리 사용되는 상호 위상차를 이용하여 시간지연이 있는 스테레오 두 채널을 동기화 시킴으로써 위상차로 인하여 발생하는

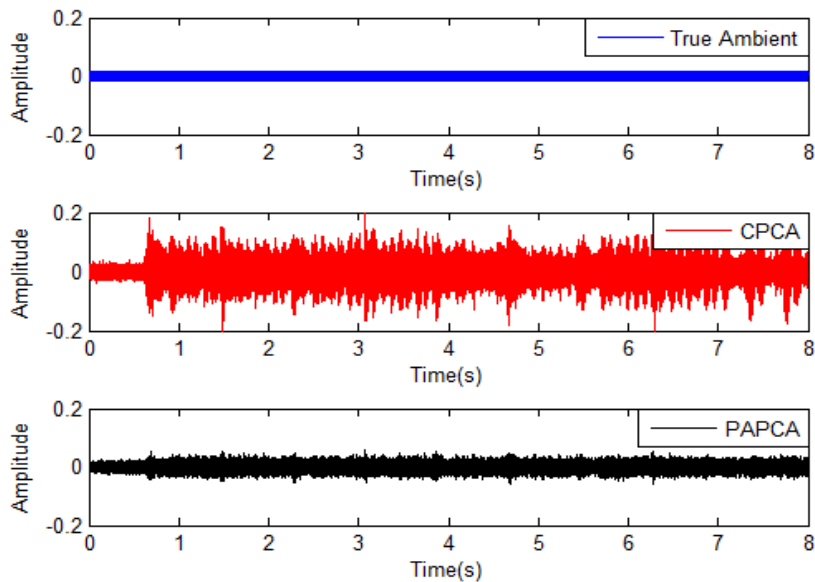


그림 6. 주변 성분 추정 결과 (a) 실제 주변성분 (b) 기존의 PCA (c) 제안된 위상 정렬 PCA  
 Fig 6. Ambient estimation results (a) : True ambient (b) : CPCA (c) : PAPCA

주성분 분리 성능을 향상 시켰고 더불어 기존의 주성분 분석법이 가지는 문제를 해결한 변형된 주성분 분석법과 결합함으로써 입력 신호의 위상차, 패닝 각도 및 PAR에 상관 없이 그 분리 성능이 강건한 알고리즘을 제안하였다. 모의 실험을 통하여 위상차를 보상함으로써 PCA 성능 향상을 확인 하였으며 단순히 진폭 패닝만 가정한 신호에 국한 되지 않고 위상차가 존재하는 실제 라이브로 녹음된 스테레오 신호에 대해서도 주성분과 주변 성분을 분리 하는 성능이 저하되지 않고 정확히 추정할 수 있다고 기대된다.

### 참 고 문 헌 (References)

- [1] R. Dressler, "Dolby Surround Prologic II de-coder principles of operation," Tech. Rep. Dolby Laboratories, 2000
- [2] J. Usher and J. Benesty, "Enhancement of spatial sound quality: A new reverberation-extraction audio upmixer," Audio, Speech, and Language Processing, IEEE Trans, vol. 15, no. 7, pp. 2141-2150, 2007
- [3] C. Faller, "Multiple-loudspeaker playback of stereo signals," J. AES, vol. 54, no.11, pp. 1051-1064, 2006.
- [4] 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
- [5] Pulkki, Ville. "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", AES convention. vol 45, Issue 6, pp 456-466, 1997
- [6] Jianjun He, Ee-Leng Tan, Woon-Seng Gan, "Time-Shifted Principal Component Analysis Based Cue Extraction for Stereo Audio Signal", ICASSP, 2013
- [7] Yong-Hyun Baek, Se-Woon Jeon, Seok-pil Lee, and Young-Cheol Park, "Efficient Primary-Ambient Decomposition Algorithm for Audio Upmix" JBE Vol. 17, No. 6, 924~932, November 2012
- [8] Petersen, T, and S. Boll., "Critical band analysis-synthesis." Acoustics, Speech and Signal Processing, IEEE Trans. Vol 31, no. 3, pp. 656-663, 1983.
- [9] MERIMAA, Juha, GOODWIN, Michael M., JOT Jean-Marc. "Correlation-based ambience extraction from stereo recordings", In AES convention123. Oct. 2007.
- [10] Hyun, Dong-il, Young-cheol Park, and Dae Hee Youn. "Estimation and quantization of ICC-dependent phase parameters for parametric stereo audio coding." EURASIP Journal on ASMP, pp. 1-12, 2012

---

### 저 자 소 개



백 용 현

- 2009년 : 연세대학교 정보기술학부 학사 졸업
- 2009년 ~ 현재 : 연세대학교 전산학과 석박사 통합과정
- 주관심분야 : 음성 신호처리, 오디오 신호처리, 3D 오디오



현 동 일

- 2005년 : 연세대학교 기계 전자공학부 학사 졸업
- 2007년 : 연세대학교 전기전자공학과 석사 졸업
- 2013년 : 연세대학교 전기전자공학과 박사 졸업
- 주관심분야 : 오디오 신호처리, 3D 오디오, 오디오 부호화

---

저 자 소 개

---



박 영 철

- 1986년 : 연세대학교 전자공학과 학사 졸업
- 1988년 : 연세대학교 전자공학과 석사 졸업
- 1993년 : 연세대학교 전자공학과 박사 졸업
- 2002년 ~ 현재 : 연세대학교 컴퓨터 정보통신공학부 교수
- 주관심분야 : 음성/오디오 신호처리, 3D 오디오, 적음 신호처리