

음원 분리와 음원 위치 추정 통합 알고리즘

*한택진 **박호중

광운대학교

*hanboys87@naver.com

Integrated Algorithm of Sound Source Separation and Localization

*Han, Taek-Jin **Park, Hochong

Kwangwoon University

요약

본 논문에서는 다양한 스테레오 환경에서도 정확한 음원 위치 추정이 가능한 방법을 제안한다. 기존의 음원 위치 추정 방법은 방향성을 가지고 있는 주성분 신호와 방향성이 없는 주변 성분으로 구성된 스테레오 환경에서만 음원의 위치 추정이 가능했다. 그러나 현재 제공되고 있는 스테레오 신호는 방향성을 가지는 다수의 음원으로 구성되어있고, 기존의 음원 위치 추정 방법으로는 정확한 음원 위치 추정이 어렵다. 이와 같은 문제 때문에 다수의 음원을 분리한 뒤, 음원의 위치를 추정하는 방법이 제안되었다. 그러나 음원의 분리 과정에서 생기는 분리 오차가 커서 음원 위치 추정이 정확하지 않다. 이에 본 논문에서는 정확한 음원 위치 추정을 위하여 음원 분리와 음원 위치 추정이 통합된 새로운 알고리즘을 제안한다. 제안한 알고리즘은 음원 위치를 기존의 방법보다 정확하게 추정하는 것을 확인할 수 있었다.

1. 서론

오늘날 멀티미디어의 발전으로 인하여 다채널 오디오 시스템이 개발되어 품질 높은 입체 음향 시스템이 널리 보급되고 있다. 그러나 현재 제공되는 콘텐츠들은 대부분 스테레오 신호이다. 이에 따라 스테레오 신호를 멀티채널 신호로 확장하는 업믹스(Upmix)기술이 활발하게 연구 중이다[1]. 스테레오 신호를 멀티채널 신호로 확장하는 방법은 먼저 방향성을 가지는 음원을 분리하고, 분리한 음원을 정확한 음원 위치로 정위 시켜야 한다. 이에 본 논문에서는 채널 확장에 필요한 음원의 위치를 추정하는 방법을 제안한다.

기존의 음원 위치 추정 방법은 주성분분석(Principal Component Analysis, PCA)를 이용한다. 그러나 PCA만을 이용한 방법은 방향성을 가지는 다수의 음원으로 구성된 스테레오 환경에서 정확한 음원 위치 추정이 어렵다. 이와 같은 문제 때문에 다수의 음원을 분리한 뒤, 음원의 위치를 추정하는 방법이 제안 되었다. 그러나 음원을 분리하는 과정에서 생기는 분리 오차가 크기 때문에 음원 위치 추정 오차가 크다. 이에 본 논문에서는 정확한 음원 위치 추정을 위하여 음원 분리와 음원 위치 추정을 동시에 하는 새로운 통합 알고리즘을 제안한다.

2. 제안하는 음원 위치 추정 방법

2.1 개요

본 논문에서 사용하는 스테레오 신호는 독립적으로 방향성을 가

지는 두 개의 음원이 있다. 각각의 음원은 패닝(Panning) 성분으로 표현하며 각 채널에 포함된 패닝 성분의 양은 패닝 계수로 표현 한다. 이와 같은 패닝 기법이 사용된 스테레오 신호는 식 (1)과 같다[2].

$$\begin{aligned} X_L &= a_{L1}P + a_{L2}N \\ X_R &= a_{R1}P + a_{R2}N \end{aligned} \quad (1)$$

식 (1)은 주파수 영역에서 표현된 값이며 P 와 N 은 각각 에너지가 큰 음원과 에너지가 작은 음원을 나타낸다. 또한, a_{L1} , a_{R1} , a_{L2} , a_{R2} 는 패닝 성분에 곱해지는 패닝 계수를 나타낸다.

$$a_{Li}^2 + a_{Ri}^2 = 1, \quad (i = 1, 2) \quad (2)$$

$$E[P^H P] > E[N^H N] \quad (3)$$

패닝 기법에 의해 패닝 계수는 식 (2)와 같이 정의된다. 식 (3)에서 $E[\cdot]$ 는 에너지를 나타내는 오퍼레이터이고 H 는 허미션(Hermitian) 연산을 나타낸다. 본 논문에서는 식 (3)과 같이 P 의 에너지가 N 의 에너지보다 더 크다고 가정하고, P 의 음원 위치를 추정한다. 음원의 위치는 주파수별로 추정하고 모든 주파수의 추정된 음원 위치를 균일 평균하여 하나의 프레임 음원 위치로 추정한다. 성능평가는 모든 프레임의 평균 음원 위치 오차로 평가한다.

2.2 기존의 음원위치 추정 방법

기존의 음원 위치 추정 방법은 PCA를 이용한다. P 의 음원 위치를 추정하기 위해서는 P 에 곱해진 패닝 계수를 구해야 한다. 패닝 계수는 공분산 행렬을 이용하여 구한다. 공분산 행렬은 식 (4)와 같이 계산된다.

$$R = \begin{bmatrix} r_{LL} & r_{LR} \\ r_{RL} & r_{RR} \end{bmatrix}, \quad r_{ij} = E[X_i^H X_j] \quad (4)$$

공분산 행렬 R 의 고유값을 분해하여 다시 표현하면 식 (5)로 표현할 수 있다.

$$R = [V_1 \ V_2] \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} V_1^H \\ V_2^H \end{bmatrix}, \quad \lambda_1 \geq \lambda_2 \quad (5)$$

식 (5)에서 V_1 과 V_2 는 고유 벡터가 되고 λ_1 과 λ_2 는 각 고유 벡터에 대응하는 고유치가 된다. 두 개의 고유치 중 큰 값을 가지는 고유치가 P 에 해당하는 에너지가 되며 이에 대응하는 고유치 벡터(패닝 계수)의 방향이 패닝 각도가 된다. 고유치를 이용한 추정된 패닝 계수는 식 (6)과 같다[3].

$$\begin{aligned} \widehat{a_{L1}} &= \frac{|r_{LR}|}{\sqrt{(\lambda_1 - r_{LL})^2 + |r_{LR}|^2}} \\ \widehat{a_{R1}} &= \frac{(\lambda_1 - r_{LL})}{\sqrt{(\lambda_1 - r_{LL})^2 + |r_{LR}|^2}} \end{aligned} \quad (6)$$

패닝 각도는 식 (6)에서 구한 패닝 계수를 이용하여 식 (7)로 구할 수 있다.

$$\frac{\tan \Phi}{\tan \Phi_0} = \frac{\widehat{a_{L1}} - \widehat{a_{R1}}}{\widehat{a_{L1}} + \widehat{a_{R1}}} \quad (7)$$

여기서 Φ 는 추정된 패닝 각도를 나타내며 Φ_0 은 스피커 구성에 따른 각도이다. 본 논문에서는 Φ_0 을 30° 로 가정한다.

2.3 제안하는 음원 위치 추정 방법

제안하는 음원 위치 추정 방법은 음원을 분리하기 위하여 비음수 행렬 인수분해(Non-negative Matrix Factorization, NMF) 방법을 사용한다[4, 5]. NMF는 음수가 없는 벡터 V 를 기저(Basis) 벡터 B 와 이득(Gain) 벡터 G 로 분리하는 알고리즘이며 식 (8)로 나타낼 수 있다.

$$V \approx BG \quad (8)$$

여기서 V 는 $(K \times T)$ 크기를, B 는 $(K \times J)$, G 는 $(J \times T)$ 크기의 행렬을 가지고 K 는 주파수 축의 개수, J 는 기저의 개수, T 는 시간 프레임의 개수를 의미하며, B 와 G 의 모든 원소의 값은 음수가 없다. 식 (8)과 같이 V 를 기저 행렬과 이득 행렬의 조합으로 근사적으로 표현

하기 위하여 식 (9)와 (10)으로 표시되는 갱신 방법을 사용한다.

$$B \leftarrow B \times \frac{V}{BG} G^T \quad (9)$$

$$G \leftarrow G \times \frac{B^T V}{B^T 1} \quad (10)$$

제안하는 음원 위치 추정 방법은 P 와 N 을 PCA와 NMF를 이용하여 그림 1과 같이 분리한 뒤 다시 PCA를 이용하여 P 의 음원 위치를 추정한다. 그림 1에서 B_P 와 B_N 은 각각 P 와 N 의 기저를 의미하며 G_P 와 G_N 은 P 와 N 의 이득을 의미한다. 이때 X_L 과 X_R 의 기저는 같으므로 기저는 공유하고 이득행렬은 독립적으로 사용한다.

그림 2는 제안하는 음원 위치 추정의 순서도이다. 먼저 PCA를 이용하여 N 을 추정하고, NMF를 이용하여 B_N 을 구한다. N 은 식 (1)과 (2)를 이용하여 식 (11)로 추정한다.

$$\begin{aligned} \widehat{P} &= P + a_{L1} a_{L2} N + a_{R1} a_{R2} N \\ \widehat{N} &= 0.5(X_L - a_{L1} \widehat{P}) / a_{R1} + 0.5(X_R - a_{R1} \widehat{P}) / a_{L1} \\ &= 0.5N(a_{R1} a_{L2} - a_{L1} a_{R2}) + 0.5N(a_{L1} a_{R2} - a_{R1} a_{L2}) \end{aligned} \quad (11)$$

그 다음 X_L 과 X_R 의 기저와 이득 행렬을 구한다. 이때 기저는 그림 1과 같이 B_P 와 B_N 으로 구성되며 B_P 만을 갱신하고 B_N 은 앞에서 구한 B_N 으로 고정한다. 이렇게 갱신 과정을 거치면 식 (12)와 같이 P 와 N 을 분리 할 수 있다. 마지막으로, P 로만 구성된 스테레오 신호를 식 (13)으로 구하고, 이를 입력신호로 하는 PCA를 다시 이용하여 P 의 음원 위치를 추정한다.

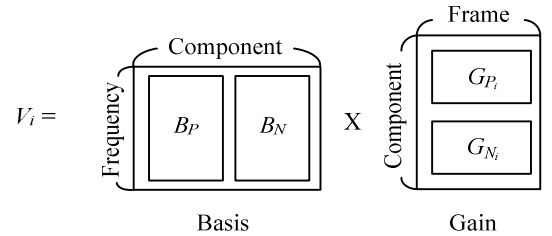


그림 1. 기저와 이득 행렬

Fig. 1. Basis and gain matrix

$$V_i = [B_P \ B_N] \begin{bmatrix} G_P \\ G_N \end{bmatrix}, \quad i = L, R \quad (12)$$

$$\widehat{X}_L = a_{L1} P = B_P G_P$$

$$\widehat{X}_R = a_{R1}P = B_P G_{P_R}$$

(13)

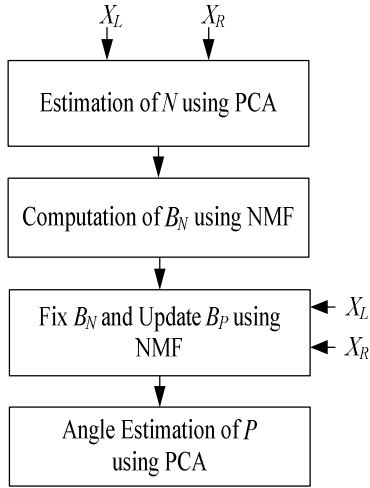


그림 2. 제안하는 음원 위치 추정의 순서도
Fig. 2. Flow chart of proposed sound source localization

3. 성능평가

표 1은 성능평가에 사용한 실험 환경이며, 본 장에서는 표 1과 같은 실험 환경에서 제안한 방법과 기존의 음원 위치 추정 방법의 성능을 평가한다. 또한, 제안한 방법으로 P 와 N 의 기저 개수에 따른 오차를 비교한다. 식 (14)는 성능평가에 사용한 P 와 N 의 패닝 각도들을 나타내고 본 논문에서는 36가지의 P 와 N 의 패닝각도 조합으로 성능을 평가한다.

표 1. 실험 환경
Table 1. Test condition

Sampling rate	48 kHz
Frame length	2048 sample
FFT size	2048 points
Window	Sine window
Overlap	50 %
V size	1024 × 25

$$P = [0^\circ, 5^\circ, 10^\circ, 15^\circ, 20^\circ, 25^\circ]$$

$$N = [0^\circ, 5^\circ, 10^\circ, 15^\circ, 20^\circ, 25^\circ] \quad (14)$$

그림 3은 N 의 패닝 각도를 20도로 고정하고, P 의 패닝 각도에 따른 오차를 비교한다. 그림 3에서 Method1은 PCA로만 P 의 음원 위치를 추정한다. Method2는 P 와 N 을 분리한 후, PCA를 이용하여 P 의 음원 위치를 추정한다. P 와 N 을 분리 하는 과정은 다음과 같다. 먼저 스테레오 신호를 그림 2와 같이 기저와 이득 행렬로 분리한다. 다음으로 에너지가 큰 기저는 B_P 로, 에너지가 작은 기저는 B_N 으로 분리한다. 마지막으로 식 (13)과 같이 P 로만 구성된 스테레오 신호를 만든다.

이 때 사용한 기저의 총 개수는 20개이고 B_P 와 B_N 은 각각 10개이다. Method3는 Method2와 같은 방법으로 P 와 N 을 분리한 후, 식 (13)과

같이 \widehat{X}_L 과 \widehat{X}_R 을 추정 한 뒤, 제안한 방식의 입력 신호를 \widehat{X}_L 과 \widehat{X}_R 으로 사용하여 P 의 음원 위치를 추정한다. Method4는 제안한 방식으로 음원 위치를 추정한다. 음원 위치 추정 오차는 P 와 N 의 패닝 각도 차이가 크게 날수록 큰 오차를 가지게 된다. 여러 가지 방법 중 제안한 방법인 Method4가 가장 작은 오차를 가지며 음원 분리를 한 번 더 포함한 Method3보다 작은 오차를 가진다. Method4가 Method3보다 작은 오차를 가지는 이유는 Method4가 PCA를 이용하여 더 정확한 N 을 추정하기 때문이다.

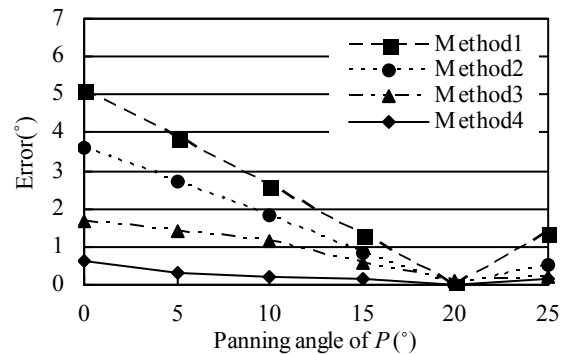


그림 3. P 의 패닝 각도에 따른 음원 위치 추정 오차
Fig. 3. According to panning angle of P , Sound source localization estimation error

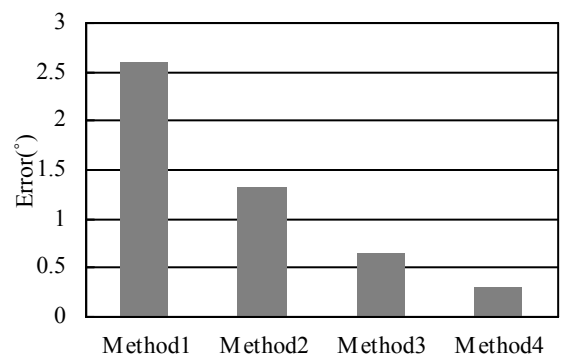


그림 4. 모든 패닝 각도에 대한 평균 음원 위치 추정 오차
Fig. 4. Average sound source localization estimation error about all panning angle

그림 4는 36가지 패닝 각도 조합에 대하여 평균 음원 위치 추정 오차를 나타낸다. PCA만을 이용하여 음원 위치를 찾는 Method1보다 음원을 분리하여 음원 위치를 찾는 Method2, 3, 4 방법이 정확한 음원 위치를 찾는다. 또한, P 의 음원 위치를 추정하는데 오차를 가지게 하는 원인인 N 을 Method4가 더 정확하게 추정하기 때문에 Method3 보

다 P 의 음원 위치 추정이 더 정확하다.

켜 향상된 입체 음향을 제공할 수 있다.

감사의 글

본 연구는 미래창조과학부의 글로벌전문기술개발사업(10047031)의 연구결과로 수행되었음.

참고문헌

[1] Y. H. Baek, S. W. Jeon, S. P. Lee and Y. C. Park, "Efficient Primary-Ambient Decomposition Algorithm for Audio Upmix", *J. of Broadcast Engineering*, vol. 17, No. 6, November 2012.

[2] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", *Journal of the Audio Engineering Society*, vol. 45, Issue 6, p 456-466, 1977.

[3] Y. H. Baek, D. I. Hyun and Y. C. Park, "Robust Primary-ambient Signal Decomposition Method using Principal Component Analysis with Phase Alignment", *Journal of Broadcast Engineering*, vol. 19, No. 1, January 2014.

[4] T. Virtanen, "Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria", *IEEE Transaction on Audio, Speech, Language Processing*, vol. 15, No. 3, March 2007.

[5] M. N. Schmidt, J. Larsen and F. T. Hsiao, "Wind noise reduction using non-negative sparse coding", *IEEE workshop on machine learning for signal processing*, pp. 431-436, August 2007.

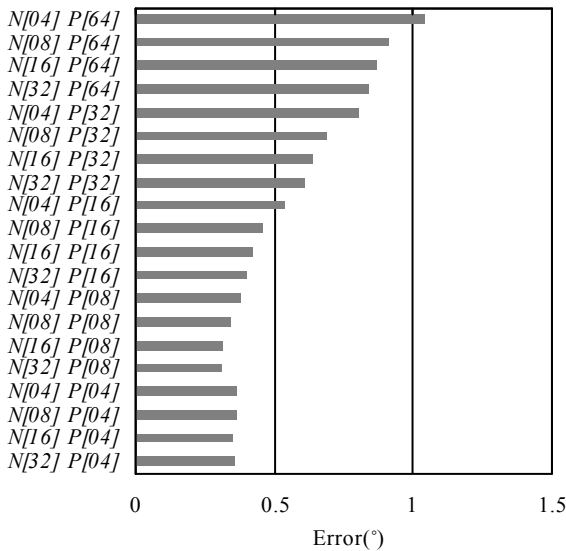


그림 5. P 와 N 의 기저 개수에 따른 평균 음원 위치 추정 오차
 Fig. 5. According to basis number of P and N , Average sound source localization estimation error

그림 5는 P 와 N 의 기저 개수에 따른 평균 음원 위치 추정 오차를 나타내고 음원 위치 추정 방법은 제안한 방법으로 한다. 오차는 36가지 패닝 각도 조합에 대한 평균 음원 위치 추정 오차를 나타낸다. P 와 N 의 기저 개수는 식 (15)와 같이 표시한다. P 의 기저 개수가 고정이고 N 의 기저 개수를 늘릴 경우, P 의 음원 위치 추정 오차를 발생시키는 N 의 기저를 정확하게 모델링(Modeling)하여 음원 위치 추정 오차가 줄어든다. N 의 기저 개수를 고정하고 P 의 기저 개수를 늘릴 경우, P 의 기저 개수와 비례하여 N 의 기저 개수가 줄어들어 오차가 커진다. 그러나 P 의 기저 개수를 너무 작게 사용하면 오히려 오차가 커지는 문제가 발생한다. 그러므로 적절한 기저 개수의 선택이 중요하다.

$$N[N \text{의 기저개수}]P[P \text{의 기저개수}] \quad (15)$$

4. 결론

본 논문에서는 PCA와 NMF를 이용하여, 음원 분리와 음원 위치 추정이 통합된 새로운 알고리즘을 제안하였다. 제안한 알고리즘은 PCA와 NMF를 이용하여 다수의 음원을 분리하고, 에너지가 큰 음원의 위치를 다시 PCA를 이용하여 추정하였다. 본 논문에서 제안한 방법은 다양한 스테레오 환경에서도 기존의 음원 위치 추정 방법보다 정확한 음원 위치를 추정할 수 있다. 또한 음원을 정확한 위치에 정위시