스테레오 패닝 음원을 위한 음원 분리 알고리즘

백용현* · 박영철**

A Source Separation Algorithm for Stereo Panning Sources

Yong-Hyun Baek*, Young-Cheol Park**

요 약

본 논문에서는 패닝 기법을 이용하여 믹싱된 스테레오 음원에서 음원을 분리하는 방법에 대하여 고찰한다. 음원 분리 알고리즘은 다채널 포맷 변환을 위한 업믹스나 음질 개선, 고품질 음원 분리 등 다양한 응용분야에 사용될 수 있다. 본 논문에서 사용하는 음원 분리 알고리즘은 믹싱된 스테레오 채널을 시간-주파수 별로 PCA(Principal Component Analysis) 분석 방법을 이용하여 각각의 음원들이 패닝된 방향을 추정하며, 추정된 방향의 성분만을 추출하는 방향 필터링 과정을 거쳐 음원들을 독립적으로 분리 해 낸다. 실험을 통해 각 음원 분리 알고리즘의 성능을 평가하였다.

Key Words : source separation, panning gain, stereo, PCA, eigen vector ratio.

ABSTRACT

In this paper, we investigate source separation algorithms for stereo audio mixed using amplitude panning method. This source separation algorithms can be used in various applications such as up-mixing, speech enhancement, and high quality sound source separation. The methods in this paper estimate the panning angles of individual signals using the principal component analysis being applied in time-frequency tiles of the input signal and independently extract each signal through directional filtering. Performances of the methods were evaluated through computer simulations.

Ⅰ. 서 론

현재 상용화 되고 있는 대부분의 오디오 컨텐츠들이 스테레오로 녹음, 혹은 합성 되어 있다. 다채널 포맷 변환 방법(upmix) 혹은 오디오 코딩과 같은 응용 분야에서 이러한 스테레오 음원에서 추출한 공간 정보를 기반으로 다양한 연구가 진행되었다. 업믹스에서는 음원이 패닝된 방향과 같은 공간 정보를 추출해 내어 다채널 포맷 재생 환경에 맞는 신호를 합성한다[1]. 코딩에서는 공간 정보와 함께 다운 믹스 된 신호를 전송하여 전송률을 줄이고 다시 합성할 때 함께 보낸 공간정보를 이용하여 원래 전송된 스테레오 음원의 장면(scene)을 유지하면서 합성하게 된다[2]. 본 논문에서는 여러 개의 음원이 패닝 기법을 이용하여 믹싱 된 스테레오 신호에서 추출한 공간 정보를

* 연세대학교 전산학과
** 교신저자 연세대학교 컴퓨터정보통신 공학부 교수 (young00@yonsei.ac.kr)
접수일자 : 2011년 4월 1일, 수정일자 : 2011년 5월 2일, 심사완료일자 : 2011년 6월 2일

이용하여 각각의 음원을 분리하는 알고리즘에 대하여 고찰하고자 한다. 스테레오 음원은 여러 개의 모노 신호가 패닝 되어 합성되었다고 가정하고 주성분 분석(PCA: principal component analysis)을 통하여 패닝 게인(panning gain)을 추정한 뒤 추정된 패닝 게인으로부터 각 음원의 방향 각도를 계산한다. 마지막으로 스테레오 채널에서 패닝된 모노 음원들을 각도에 따라 분리해 낸다. 본 논문의 구성은 다음과 같다. 2장에서 입력신호의 모델에 대해 설명하고 3장에서 패닝 게인을 추정하여 음원을 분리하는 방법을 설명한다. 4장에서 모의실험 결과를 분석하고 5장에서 결론을 맺는다.

Ⅱ. 음원 분리 알고리즘

논문에서 입력신호는 여러 개의 모노 음원들이 서로 다른 방향으로 패닝 된 스테레오 신호로 가정하고 수식 (1)과 같이 모델링 할 수 있다.

$$x_L = \sum_{i=1}^{k} a_{Li} s_i + n_L \qquad (1)$$
$$x_R = \sum_{i=1}^{k} a_{Ri} s_i + n_R$$

k는 패닝 되어 섞여 있는 입력 신호의 개수이며 $a_{L1,\cdots,Lk}$와 $a_{R1,\cdots,Rk}$는 패닝 게인을 의미하고 $s_{1,\cdots,k}$는 패닝 게인에 의해 방향을 가지는 모노의 패닝 음원이며 n_L과 n_R은 각 채널에 섞여 있는 잔향 성분이다. 입력 스테레오 신호를 분석할 때 단 구간 푸리에 변환(Short Time Fourier Transform)을 이용하여 주파수 도메인에서 이루어지게 되는데 이 때 패닝된 신호들이 각 주파수 도메인에서 상당히 크게 중첩 되지 않고 주파수 별로 하나의 패닝 음원만 존재한다고 가정하면 입력 스테레오 신호를 주파수 도메인에서 다음과 같이 표현할 수 있다.

$$X_L = a_L S + N_L \qquad (2)$$
$$X_R = a_R S + N_R$$

각 주파수 별로 좌우 두 채널에서 a_L과 a_R은 패닝 게인이 되고 S, A_L과 A_R은 수식 (1)에서 정의 했던 패닝 음원과 잔향 성분들의 주파수 도메인 표현이며 S, A_L과 A_R은 각각 통계적으로 독립적이라고 가정한다. 패닝 게인은 패닝 이전의 모노 음원의 에너지가 스테레오로 패닝 된 이후 에너지 정규화(energy normalizarion)를 위한 관계를 갖는다.

$$a_L^2 + a_R^2 = 1 \qquad (3)$$

이 때 섞여 있는 두 채널의 잔향들의 에너지는 같고 패닝된 음원의 에너지보다는 작다고 가정한다.

그림 1은 위와 같이 가정된 스테레오 입력 신호에서 각각의 방향으로 패닝된 음원들을 분리하는 알고리즘의 블록도이다.

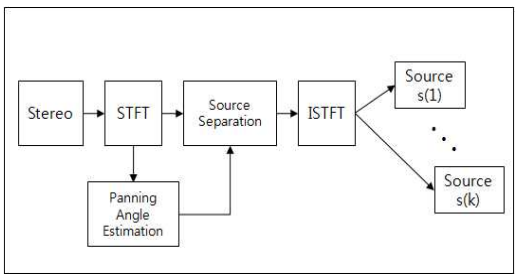

그림 1 . 음원 분리 알고리즘 블록도
Fig. 1. Block diagram of source separation algorithm

입력 스테레오를 STFT를 통하여 시간-주파수 축에서 주성분 분석을 이용하여 패닝 각도를 추정하고 추정된 패닝 각도로부터 각각의 음원들을 분리해 낸다.

Ⅲ. 음원 분리 알고리즘

3.1 패닝 게인 추정

모노 음원들은 패닝 게인에 의해 두 채널에서 레벨 차이에 의해 방향감을 가지게 된다. 따라서 음원의 평면상에서의 각도를 패닝 게인을 추정함

으로써 구할 수 있다. 이를 위해 두 채널 신호로부터 패닝 인덱스(panning index)를 구하여 음원의 공간 정보를 추정하는 방법이 제안되었다[3]. 본 논문에서는 주성분 분석 방법을 통하여 패닝 게인을 추정한다. 단 구간 푸리에 변환을 이용하여 주파수 도메인으로 변환한 뒤 스테레오 입력 신호의 좌,우 채널의 공분산 행렬을 구하면 수식 (4)와 같다.

$$R_{cov} = \begin{bmatrix} r_{LL} & r_{LR} \\ r_{RL} & r_{RR} \end{bmatrix} = \begin{bmatrix} \sigma_{X_L}^2 & \sigma_{X_L X_R}^2 \\ \sigma_{X_R X_L}^2 & \sigma_{X_R}^2 \end{bmatrix} \quad (4)$$

위 식에서 $\sigma_x^2 = E\{|xx^*|\}$ 로 정의 되고 수식 (2)를 이용하여 다시 표현 하면 다음과 같다.

$$R_{cov} = \begin{bmatrix} a_L^2\sigma_S^2 + \sigma_{N_L}^2 & a_L^2\sigma_S^2 \\ a_R a_L \sigma_S^2 & a_R^2\sigma_S^2 + \sigma_{N_R}^2 \end{bmatrix} \quad (5)$$

이 때 두 채널에 섞인 잔향 성분들의 에너지는 같다고 가정하였으므로 $\sigma_N^2 = \sigma_{N_L}^2 = \sigma_{N_R}^2$로 두고 고유 벡터(eigenvector)의 비를 구하면 수식 (6)과 같은 식을 구할 수 있다.

$$\begin{bmatrix} r_{LL} - \lambda_0 & r_{LR} \\ r_{RL} & r_{RR} - \lambda_0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (6)$$

$$\frac{v_1}{v_2} = \frac{r_{LR}}{\lambda_0 - r_{LL}}$$

여기서 λ_0는 공분산 행렬의 가장 큰 고유치(eigenvalue)이고 이것은 패닝 음원 에너지와 잔향 성분 에너지의 합과 같다.[4] 수식 (5)를 이용하여 고유벡터의 비를 정리하면 아래와 같다.

$$\begin{aligned} \frac{v_1}{v_2} &= \frac{a_L a_R \sigma_S^2}{\sigma_S^2 + \sigma_N^2 - \left(a_L^2\sigma_S^2 + \sigma_N^2\right)} \\ &= \frac{a_L a_R \sigma_S^2}{\sigma_S^2 - a_L^2\sigma_S^2} \\ &= \frac{a_L a_R}{a_R^2} \\ &= \frac{a_L}{a_R} \end{aligned} \quad (7)$$

수식 (7)로부터 고유벡터의 비는 패닝 게인의 비와 같음을 알 수 있다. 패닝 게인과 고유벡터는 놈 (norm)값이 1이 되므로 정규화 하여 각각의 패닝 게인을 구하면 최종적으로 추정된 패닝 게인은 수식 (8)과 같다.

$$\begin{aligned} a_L = v_1 &= \frac{r_{LR}}{\sqrt{(\lambda_0 - r_{LL})^2 + r_{LR}^2}} \\ a_R = v_2 &= \frac{\lambda_0 - r_{LL}}{\sqrt{(\lambda_0 - r_{LL})^2 + r_{LR}^2}} \end{aligned} \quad (8)$$

그리고 추정된 패닝 게인으로부터 패닝 기법을 통하여 가정된 패닝 음원들의 방향은 탄젠트 법칙에 의해 평면 공간상의 각도(θ)를 계산할 수 있다.

$$\frac{a_L - a_R}{a_L + a_R} = \frac{\tan\theta}{\tan 30^o} \quad (9)$$

스테레오 채널에서 패닝 각도의 범위는 좌30°~우30°에 해당하고 이렇게 추정된 각도를 기반으로 분리해 내고자 하는 방향에 해당 하는 신호만을 가우시안 윈도우 필터를 통과시키면 원래의 모노 음원 신호를 얻을 수 있다.

3.2. 가우시안 윈도우를 이용한 음원 분리

추정된 패닝 게인으로부터 얻은 패닝 각도를 기반으로 가우시안 윈도우(Gaussian window) 함수를 적용하여 음원을 분리 해 낸다. 가우시안 윈도

우 함수는 다음과 같이 주어진다.

$$G_{\theta_i}(m,l) = \nu + (1-\nu)e^{-\frac{1}{2\epsilon}(\theta(m,l)-\theta_i)^2} \quad i = 1,\ldots,k \tag{10}$$

위 식에서 m은 시간축으로 프레임 인덱스가 되고 l은 주파수축에서 각 주파수 빈(bin) 인덱스가 된다. θ_i는 가우시안 윈도우의 중앙에 해당 하는 각도가 됨과 동시에 분리해 내고자 하는 음원의 각도가 되며 $\theta(m,l)$은 패닝 게인으로부터 얻은 각 주파수 빈에 해당하는 각도이다. ϵ은 가우시안 윈도우의 너비를 결정하는 파라미터가 되며 너비가 좁을수록 분리성능은 좋아지나 분리된 음원의 왜곡이 심해진다. 반대로 너비가 넓으면 왜곡은 줄어들지만 다른 음원 성분이 많이 섞여 있는 상태로 분리 되게 된다. ν는 가우시안 윈도우의 최소값를 설정하는 파라미터로 STFT 성분이 0이 되는 것을 방지하여 뮤지컬 잡음이 생기는 것을 최소화 한다. 그림 2는 θ_0가 20°일 때 가우시안 윈도우를 그린 것이다. 가우시안 윈도우를 통한 음원 분리는 다음과 같이 이루어진다.

$$S_{\theta_i}(m,l) = G_{\theta_i}(m,l)\{a_L X_L(m,l) + a_R X_R(m,l)\} \tag{11}$$

분리된 음원은 좌우 주파수 빈에 패닝 게인을 곱하여 이전의 원래 신호의 에너지를 유지하고 여기에 가우시안 윈도우 함수를 곱하여 얻을 수 있다. 따라서 가우시안 윈도우 함수는 분리해 내고자 하는 방향에 가중치를 주는 함수라고 할 수 있다. 그림 2에서 보듯 θ_0가 20°일 때 20°에 해당하는 주파수 성분에 대하여 가중치를 1로 최대로 주고 점점 20°에서 멀어질수록 다른 방향에 대해서는 가중치를 점점 낮추어 분리해 내고자 하는 방향의 성분$S_{\theta_i}(m,l)$를 계산한 뒤에 ISTFT(Inverse Short Time Fourier Transform) 취하여 합성한다.

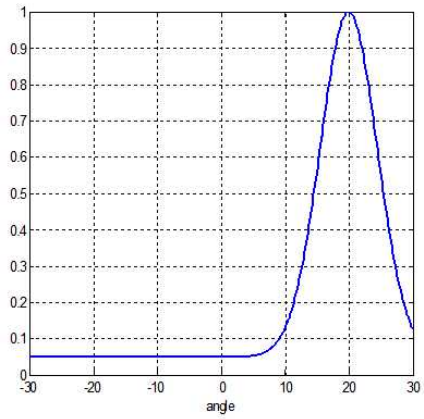

그림 2. 가우시안 윈도우 함수 (Θ_0=20°, ν=0.03, ε=20)

Fig. 2. Gaussian window function

Ⅳ. 모의 실험 및 평가

모의실험은 모노 음원들은 패닝 기법을 통하여 스테레오 신호로 합성하여 사용하며 사용된 음원은 샘플링 주파수 44.1khz의 20초짜리 모노 신호로서 드럼, 음성, 기타를 사용하였다.

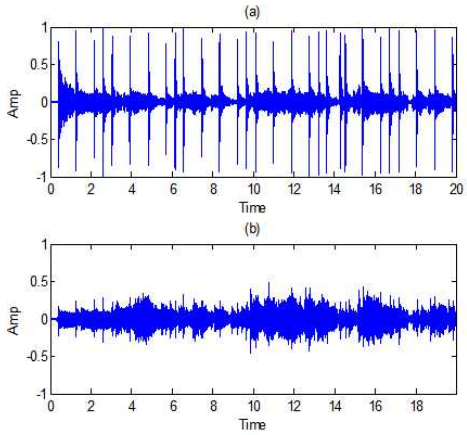

그림 3. 입력 스테레오 신호 - (a)좌 (b)우

Fig. 3. Input stereo signal - (a) Left channel (b) Right channel

드럼은 −20°, 음성은 0° 그리고 기타는 20°로 패닝하여 만든 스테레오 입력 신호는 그림 3과 같

다. 이 때 잔향 성분에 대해서는 고려하지 않았다.

STFT를 수행할 때 프레임의 크기는 4096이며 50% 오버랩을 시키고 해닝(Hanning) 윈도우를 사용하였으며 가우시안 윈도우의 파라미터는 ν =0.03, ϵ =10을 사용하였다. 프레임 사이즈가 크면 클수록 주파수 해상도가 좋아지기 때문에 믹싱된 신호와 혹은 잔향 성분이 있을 경우 신호사이에 중첩되는 주파수 성분을 줄일 수 있어 음원 분리 성능이 좋아진다. 그림 4는 스테레오로 믹싱하기 전의 모노의 원신호이며 그림 5는 음원 분리 알고리즘을 통하여 분리해낸 출력신호이다.

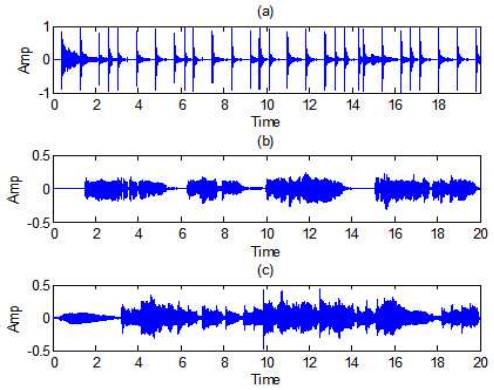

그림 4. 원신호 (a) 드럼 (b) 음성 (c) 기타
Fig. 4. Original signal - (a) Drum (b) Voice (c) Guitar

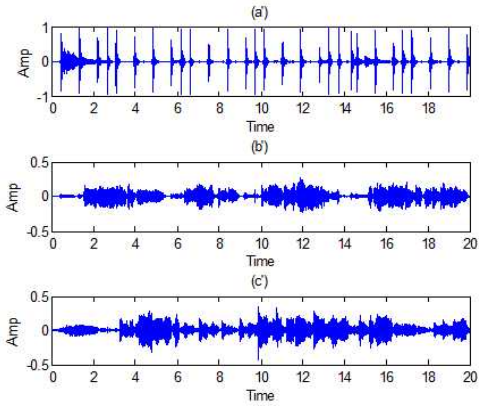

그림 5. 분리된 음원 (a') 드럼 (b') 음성 (c') 기타
Fig. 5. Separated signal - (a') Drum (b') Voice (c') Guitar

패닝 각도를 통하여 스테레오 신호에 믹싱된 원래 신호를 유사하게 분리 할 수 있음을 확인 할 수 있다.

분리된 음원의 성능 평가는 SDR (Source to Distortion Ratio), SIR (Source to Interferences Ratio), SAR (Source to Artifact Ratio)를 측정하여 평가한다.[5] 잔향 혹은 노이즈에 대한 성분은 고려하지 않았기 때문에 SNR (Source to Noise Ratio)는 평가에서 제외하였다. 먼저 스테레오 신호에서 분리해낸 음원을 $\hat{s_i}$라고 하면 수식 (12)와 같이 나눌 수 있다.

$$\hat{s_i} = s_{target} + e_{intf} + e_{noise} + e_{artif} \tag{12}$$

여기서 s_{target}은 분리해 내고자 하는 음원이 되고 e_{intf}, e_{noise} 그리고 e_{artif}는 각각 간섭신호, 노이즈, artifact에 대한 오차에 해당한다. SDR, SIR, SAR은 수학적으로 다음과 같이 정의한다.

$$SDR = 10\log\frac{\| s_{target} \|^2}{\| e_{intf} + e_{artif} \|^2} \tag{13}$$

$$SIR = 10\log\frac{\| s_{target} \|^2}{\| e_{intf} \|^2} \tag{14}$$

$$SAR = 10\log\frac{\| s_{target} + e_{intf} \|^2}{\| e_{artif} \|^2} \tag{15}$$

표 1은 avendano 방법과 제안한 방법과의 SDR, SIR, SAR을 측정하여 성능을 비교한 것이다.

표 1. 성능 평가
Table 1. Performance evaluation

알고리즘	음원	SDR	SIR	SAR
Avendano (Panning Index)	드럼	13.5897	40.3635	13.5992
	음성	9.5550	17.3868	10.4153
	기타	8.9239	28.7348	8.9753
Angle Estimation	드럼	14.5756	37.4906	14.5986
	음성	9.1629	20.8603	9.5025
	기타	9.3073	27.6641	9.3786

가우시안 윈도우는 동일한 너비와 최소값을 가지도록 설계하여 동일한 환경에서 획득한 분리된 음원을 사용하였다. 주성분 분석법을 통한 음원 분리 알고리즘이 기존의 방법(참고문헌 [3])과 유사한 성능을 보임을 확인 할 수 있다.

V. 결 론

본 논문에서 주성분 분석법을 이용하여 추정된 각도를 기반으로 음원을 분리하는 알고리즘에 대해 고찰 하였다. 주성분 분석법을 통해 두 채널간의 공분산 행렬의 가장 큰 아이겐 벨류에 해당하는 아이겐 벡터의 비가 패닝 게인의 비와 같음을 이용하여 각도를 추정하고 가우시안 윈도우 함수를 이용하여 분리해내고자 하는 각도별로 가중치를 주어 음원을 분리 해 낸다.

본 논문에서는 잔향과 노이즈와 같은 엠비언트(Ambient) 성분에 대해서는 고려하지 않았다. 잔향과 같은 성분이 섞여 있는 경우 이는 음원을 분리함에 있어 성능을 저하시키는 요인이 된다. 따라서 음원 분리 성능을 저하시키는 엠비언트 성분을 추출하는 방법과 같은 향후 이를 해결 할 연구가 필요할 것으로 보인다.

참 고 문 헌

[1] R. IRWAN, RONALD M. AARTS "Two-to-Five Channel Sound Processing", AES. vol 50, No 11, pp. 914-925, 2002

[2] SW Jeon, DG Hyun, JI Seo, YC Park, DH Yoon, "Enhancement of Principal to Ambient Energy ratio for PCA-bsed Parametric audio coding" , IEEE ICASSP, pp. 14-19, 2010

[3] C. Avendarno, "Frequency-Domain Source Identification and Manipulation Stereo Mixes for Enhancement, Suppression and Re-Panning Applications", IEEE Workshop, pp. 55-58, 2003

[4] M Briand, D. Virette, N. Martin, "Parametric Representation of Multichannel Audio Based on Principal Component Analysis", AES 120th Convention, 2006

[5] C. Fevotte, R. Gribonval, E. Vincent, "BSS_EVAL toolbox User Guide Revision 2.0" , IRISA, 2005

저자약력

백 용 현(Yong-Hyun Baek) **학생 회원**

2009년 연세대학교 컴퓨터 정보통신학부 학사
2009년-현재 연세대학교 전산학과 이학석박 통합 과정

<관심분야> 디지털 신호처리, 음성/오디오 신호처리, 3D 오디오

박 영 철(Young-Cheol Park) **정회원**

1986년 연세대학교 전자전기공학과 학사
1988년 연세대학교 전자전기공학과 석사
1993년 연세대학교 전자전기공학과 박사
2002년-현재 연세대학교 컴퓨터 정보통신공학부 교수

<관심분야> 디지털 신호처리, 음성/오디오 신호처리, 적응 필터, 3D 오디오