

# 빔공간-영역 다채널 비음수 행렬 분해 알고리즘을 이용한 음원 분리 기법 Part II: 빔공간-변환 기법에 대한 고찰

## Audio Source Separation Method based on Beamspace-domain Multichannel Non-negative Matrix Factorization, Part II: A Study on the Beamspace Transform Algorithms

이석진 · 박상하 · 성평모

(Seokjin Lee, Sang Ha Park, and Koeng-Mo Sung)

서울대학교 전기컴퓨터공학부 뉴미디어통신공동연구소 음향공학연구소

(접수일자: 2012년 2월 27일; 수정일자: 2012년 4월 19일; 채택일자: 2012년 5월 16일)

**초 록:** 빔공간 변환(beamspace transform) 기법은 공간 영역의 신호를 입사각 혹은 그 사인함수의 영역으로 변환하는 기법으로, MUSIC과 같은 음원 정위 및 추적(source localization and tracking) 문제나 적응 빔형성(adaptive beamforming)과 같은 문제에서 많이 사용되는 기법이다. 다채널 음원 분리 기법에 사용될 때에는, 음원의 정보 뿐만 아니라 해당 음원의 이미지(image)를 재구성하여야 하므로 역변환 기법 또한 중요하다. 본 논문에서는 멀티 채널 음원 분리 기법을 위한 빔공간 변환 기법과 그 역변환 기법에 대하여 고찰하였으며, 특히 빔공간-영역 다채널 비음수 행렬 분해 기법에 적용되었을 때 그 성능에 미치는 영향을 중점적으로 살펴보았다.

**핵심용어:** 다채널 음원 분리, 비음수 행렬 분해, 빔공간 변환

**투고분야:** 음향 신호처리 분야(1)

**ABSTRACT:** Beamspace transform algorithm transforms spatial-domain data - such as x, y, z dimension - into incidence-angle-domain data, which is called beamspace-domain data. The beamspace transform method is generally used in source localization and tracking, and adaptive beamforming problem. When the beamspace transform method is used in multichannel audio source separation, the inverse beamspace transform is also important because the source image have to be reconstructed. This paper studies the beamspace transform and inverse transform algorithms for multichannel audio source separation system, especially for the beamspace-domain multichannel NMF algorithm.

**Key words:** Multichannel audio source separation, Non-negative matrix factorization (NMF), Beamspace-transform.

**ASK subject classification:** Acoustic Signal Processing (1)

## 1. 서 론

빔공간 변환(beamspace transform) 기법은 공간 영역의 신호를 입사각의 함수로 나타내는 기법으로, 센서 배열을 이용한 신호처리 시스템에서 여러 목적으로 사용되고 있다. 특히, MUSIC과 같은 신호원 정위(source localization) 알고리즘에 사용되어 그 성능

을 향상시키거나<sup>[1-2]</sup>, 적응 빔형성 문제에 사용되어 그 성능을 높이는 목적으로 널리 사용되어 왔다<sup>[3]</sup>.

다채널 음원 분리 기법들은 센서 간의 공간적인 정보를 이용하는 방향으로 발전하고 있으며<sup>[4]</sup>, 이는 결과적으로 빔형성기에서 간섭 신호를 제거하는 문제와 매우 비슷한 성질을 가진다. 최근에는 빔형성기를 이용하여 다채널 음원 분리 문제를 접근하는 방법들이 시도되고 있는데, 그 중 하나는 음원 분리 기법과 빔형성기를 조합하여 문제를 해결하는 것이고<sup>[5]</sup>, 다른 하나는 빔공간으로 변환하여 음원 분리

\*Corresponding author: 이석진 (panind83@gmail.com)  
151-742 서울시 관악구 대학동 서울대학교 132동 302호  
(전화: 02-880-8427; 팩스: 02-886-0791)

기법을 사용하는 것이다<sup>[3,6]</sup>.

빔공간으로 변환하여 음원 분리 기법을 사용하는 경우, 주의해야 할 문제가 몇 가지 존재한다. 그 중 하나는 빔공간으로 변환할 때 사용되는 빔형성기의 응답이 주파수에 의존적이라는 것이며, 다른 하나는 빔공간 변환 기법의 역변환이 잘 이루어지지 않을 수 있다는 점이다. 본 논문에서는 위와 같은 관점에서 빔공간 변환 기법에 대해 연구하였다.

## II. 배경 지식 및 문제 기술

### 2.1 다채널 비음수 행렬 분해 기법

다채널 음향 신호를 비음수 행렬 분해(Nonnegative Matrix Factorization: NMF)로 음원 분리 하기 위해서 A. Ozerov와 C. Fevotte에 의해 다음과 같은 다채널 비음수 행렬 분해(Multi-channel NMF: MC-NMF) 기법이 개발되었다<sup>[7]</sup>.

다채널 비음수 행렬 기법에서는 입력 신호를 다음과 같이 모델링 한다.

$$x_{i,fn} = \sum_{j=1}^J a_{ij,f} s_{j,fn} + b_{i,fn} \tag{1}$$

여기서  $x_{i,fn}$ ,  $s_{j,fn}$ ,  $b_{i,fn}$ 는 각각 국소 푸리에 변환된 입력 신호, 음원 신호, 잡음 신호를 나타내며,  $a_{ij,f}$ 는 각 주파수 영역의 음원-채널 별 전달 함수를 나타낸다. 여기에 비음수 행렬 분해 기법을 적용하기 위하여 입력 신호의 제곱 값을 취하면 다음과 같은 모델링이 가능하다.

$$\hat{v}_{i,fn} = \sum_j q_{ij,f} \underbrace{\sum_{k \in K_j} w_{fk} h_{kn}}_{P_{j,fn}} \tag{2}$$

여기서  $q_{ij,f} = |a_{ij,f}|^2$ 이며,  $\hat{v}_{i,fn}$ 는 제공된 입력 신호의 추정치이고,  $w_{fk}$ ,  $h_{kn}$ 는 각각 주파수 영역의 기저와 시간 영역의 기저를 나타낸다. 위의 모델링을 이용, 각 파라미터를 추정하여 음원 분리를 수행하는 알고리즘이<sup>[7]</sup>에 자세히 기술되어 있다.

### 2.2 문제 기술

최근에는 위의 알고리즘을 빔공간 영역에서 수행함으로써 성능을 개선시키는 알고리즘이 개발되었으며<sup>[6]</sup>, 이에 대한 시스템 구성도는 그림 1과 같다. 이와 같이 빔공간 영역에서 MC-NMF 알고리즘을 적용하는 이유는 기존의 MC-NMF 알고리즘이 채널 간 위상차이를 이용하지 못하는 문제점을 개선하기 위한 것이며, 빔공간 영역 MC-NMF 알고리즘에서 채널 간 위상차이를 이용할 수 있도록 만들어 주는 것은 바로 빔공간 변환 알고리즘이다<sup>[6]</sup>. 추가적으로, NMF 알고리즘은 데이터의 희박성을 이용하는 알고리즘이므로<sup>[11]</sup>, 빔공간 변환 기법이 데이터의 희박성을 확보하여 준다면 NMF 알고리즘의 성능 또한 개선될 것을 기대할 수 있다. 따라서, 빔공간-영역 MC-NMF의 성능 개선 정도를 좌우하는 것은 빔공간 변환의 정확성과 분해능에 달려있다고 볼 수 있다.

또한,<sup>[6]</sup>의 빔공간-영역 다채널 NMF 알고리즘의 기술 과정에서 빔공간-영역 전달함수  $g_{mj}$ 가 주파수에 무관하다는 가정하에서 알고리즘이 도출되었다. 단순히 각 신호의 조향벡터를 이용하여 빔공간 변환을 수행하는 경우, 각 조향벡터의 빔패턴이 주파수에 의존적일 뿐만 아니라 고주파에서는 앨리어싱과 같은 문제가 발생하게 된다<sup>[3]</sup>. 따라서, 빔공간 변환 기법이 개선된다면(예를 들어, 빔공간 변환 기법이 주파수-불변(frequency-invariant)과 같은 특성을 가진다면), 알고리즘의 가정을 잘 만족시키게 되어 성능이

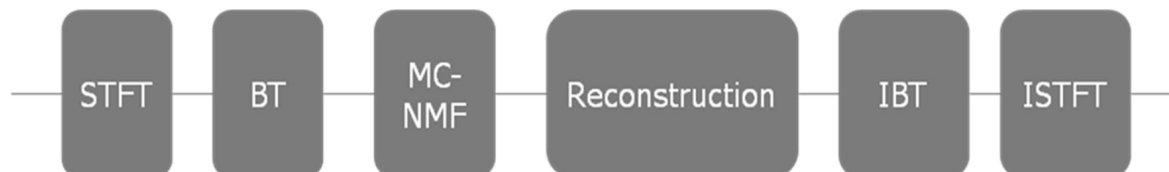


그림 1. 빔공간 영역에서의 다채널 비음수 행렬 분해(MC-NMF) 시스템  
Fig. 1. Beamspace-domain multichannel NMF system.

개선될 것을 기대할 수 있다.

또한, beamspace-MUSIC과 같은 기존의 빙공간-영역 알고리즘들은 빙공간에서 특정 요소를 분석하는 알고리즘이었기 때문에, 빙공간 변환 알고리즘만 필요할 뿐 그 역변환은 고려하지 않았다<sup>[2]</sup>. 그러나 빙공간-영역 MC-NMF 알고리즘과 같은 경우 분리된 음원의 추정 및 재구성을 그 목표로 하기 때문에<sup>[6]</sup>, 역변환 알고리즘이 왜곡을 만들지 않고 음원 이미지(sound source image)를 재구성하는 것이 중요하다.

이와 같이 빙공간 영역에서 MC-NMF 알고리즘을 적용하는 경우, 빙공간 변환이 성능에 큰 영향을 주게 되며, 해당 공간 변환의 역변환 또한 성능에 영향을 줄 수 있다. 본 논문에서는, 이와 같은 문제에서의 빙공간 변환 기법과 그 역변환에 대하여 고찰하고, 실험을 통하여 성능을 평가하였다.

### III. 빙공간 변환 기법 및 역변환 기법

#### 3.1 조향 벡터를 이용한 빙공간 변환 기법

##### 3.1.1 빙공간 변환 기법

일반적으로 협대역 신호의 빙공간 변환은 각 신호의 조향 벡터(steering vector)를 이용하여 수행할 수 있다<sup>[2]</sup>. 이는 조향벡터가 각 센서 간의 위상 차이를 반영하고 있다는 점에 착안한 것이며, 빙공간 변환은 각 조향 벡터로의 정사영을 구함으로써 수행된다. 이는 마치 시간-주파수 영역에서의 이산 푸리에 변환(Discrete Fourier Transform: DFT)과 같이 작동한다.

구체적인 빙공간 변환 기법을 살펴보면 다음과 같다. 먼저, 협대역 신호 혹은 국소 푸리에 변환된 하나의 주파수 구간(frequency bin)에 대하여, 다음과 같은 빙공간 원형 행렬(beam space prototype matrix)를 정의한다.

$$\mathbf{W}_{proto,f} = [\mathbf{a}_f(\theta_0) \quad \mathbf{a}_f(\theta_1) \quad \cdots \quad \mathbf{a}_f(\theta_M)] \quad (3)$$

위의 빙공간 원형 행렬은  $I \times M$  크기의 행렬이며, 각 원소를 이루고 있는  $I \times 1$  크기의 조향 벡터는 다음과 같이 정의된다.

$$\mathbf{a}_f(\theta) = \left[ e^{-j\omega_f(i-1)[d_x \sin(\theta)/c]} \right]_i \quad i=1, \dots, I \quad (4)$$

여기서  $I$ 는 채널의 개수를 나타내고,  $d_x$ 는 마이크 로폰 간의 거리를 나타낸다. 위의 원형 행렬(prototype matrix)를 이용하여 빙공간 변환을 수행하되, 변환이 sub-unitary property를 갖도록 다음과 같이 보정한다.

$$\mathbf{W}_{BT,f} = \mathbf{W}_{proto,f} \left( \mathbf{W}_{proto,f}^H \mathbf{W}_{proto,f} \right)^{-\frac{1}{2}} \quad (5)$$

위와 같이 보정된 빙공간 변환 행렬을 이용하여 다음과 같은 빙공간 변환 작업을 수행한다.

$$\tilde{\mathbf{x}}_{fn} = \mathbf{W}_{BT,f}^H \mathbf{x}_{fn} \quad (6)$$

여기서,

$$\tilde{\mathbf{x}}_{fn} = [\tilde{x}_{1,fn} \quad \cdots \quad \tilde{x}_{M,fn}]^T \quad (7)$$

$$\mathbf{x}_{fn} = [x_{1,fn} \quad \cdots \quad x_{I,fn}]^T \quad (8)$$

이다.

##### 3.1.2 역변환 기법

위에서 설계된 빙공간 변환에 대한 역변환은 다음과 같은 역할을 한다.

$$\mathbf{x}_{fn} = \mathbf{W}_{BT,f}^H \tilde{\mathbf{x}}_{fn} = \mathbf{W}_{BT,f}^H \mathbf{W}_{BT,f}^H \mathbf{x}_{fn} \quad (9)$$

따라서, 빙공간 변환 행렬  $\mathbf{W}_{BT,f}$ 와 역변환 행렬  $\mathbf{W}_{BT,f}^H$ 의 관계는 다음과 같다.

$$\mathbf{W}_{BT,f} \mathbf{W}_{BT,f}^H = \mathbf{I} \quad (10)$$

여기서  $I$ 는 단위 행렬을 나타낸다. 위와 같은 문제의 최소-자승법 해답은 다음과 같은 pseudo-inverse 형태로 얻어진다<sup>[8]</sup>.

$$\mathbf{W}_{BT,f} = (\mathbf{W}_{BT,f}^H \mathbf{W}_{BT,f})^{-1} \mathbf{W}_{BT,f}^H = \mathbf{W}_{BT,f}^\dagger \quad (11)$$

또한, 위에서 빔공간 변환 행렬  $\mathbf{W}_{BT,f}$ 가 sub-unitary 행렬로 설계되었으므로, 역변환 행렬  $\mathbf{W}_{BT,f}$ 는 다음과 같이 간단하게 기술될 수 있다.

$$\mathbf{W}_{BT,f} = \mathbf{W}_{BT,f}^H \quad (12)$$

### 3.2 Tapped delay-line을 이용한 주파수-불변 빔공간 변환

#### 3.2.1 빔공간 변환 기법

선 배열 센서에서 tapped delay-line의 필터 계수  $d(m, n)$ 의 주파수-방위 영역 응답은 다음과 같이 나타난다<sup>[3]</sup>.

$$P(\Omega, \theta) = \sum_{m,n=-\infty}^{+\infty} d(m, n) e^{-jm\Omega \sin \theta} e^{-jn\Omega} \quad (13)$$

여기서  $\mu = d_x / (cT_s)$ 이고,  $T_s$ 는 샘플링 주기,  $\Omega = \omega T_s$ 는 정규화된 각주파수(normalized angular frequency)를 나타낸다. 문제를 단순화하기 위해 다음과 같이 정의하자.

$$\Omega_1 = \mu\Omega \sin \theta \quad (14)$$

$$\Omega_2 = \Omega \quad (15)$$

그러면 위의 주파수-방위 영역 응답을 다음과 같이 나타낼 수 있다.

$$P(\Omega_1, \Omega_2) = \sum_{m,n=-\infty}^{+\infty} d(m, n) e^{-jm\Omega_1} e^{-jn\Omega_2} \quad (16)$$

위의 주파수-방위 영역 응답은 tapped delay-line의 필터 계수의 2차원 DFT(Discrete Fourier Transform)를 취한 꼴로 나타나있다. 이를 역으로 이용하면, 목적하는(혹은 설계된) 주파수-방위 영역의 응답에 대해 2차원 IDFT(Inverse DFT)를 취하면 필터 계수를 얻어 낼 수 있다<sup>[3]</sup>.

이를 이용하여 광대역 빔공간 변환 행렬을 구성하면 다음과 같다.

$$\tilde{\mathbf{x}}_{n,FIB1} = \mathbf{W}_{FIB1-BT}^H \mathbf{x}_n \quad (17)$$

여기서,  $\mathbf{x}_n$ 은 stacked input vector를 나타내며, 다음과 같이 구성된다.

$$\mathbf{x}_n = [\mathbf{x}_n^T \ \mathbf{x}_{n-1}^T \ \cdots \ \mathbf{x}_{n-M+1}^T]^T \quad (18)$$

여기서,

$$\mathbf{x}_n = [x_{1,n} \ x_{2,n} \ \cdots \ x_{I,n}]^T \quad (19)$$

이다. 또한 빔공간 변환 행렬은 다음과 같이 구성된다.

$$\mathbf{W}_{FIB1-BT} = [\mathbf{d}_{\theta_0} \ \mathbf{d}_{\theta_1} \ \cdots \ \mathbf{d}_{\theta_{M-1}}] \quad (20)$$

여기서,

$$\mathbf{d}_\theta = [d_\theta(0,0) \ d_\theta(1,0) \ \cdots \ d_\theta(I-1,1) \ d_\theta(0,1) \ \cdots \ d_\theta(I-1,M-1)]^T \quad (21)$$

이다.

#### 3.2.2 역변환 기법

3.1절의 역변환 기법과 마찬가지로 최소-자승법의 풀이에 의해 역변환 행렬은 다음과 같이 정의될 수 있다.

$$\mathbf{W}_{FIB1-BT} = \mathbf{W}_{FIB1-BT}^H \quad (22)$$

여기서 주목하여야 할 점은, 3.2절의 빔공간 변환-역변환 기법은 시간-영역에서 이루어진다는 점이다. 즉, 변환 행렬이 주파수 별로 다르지 않고 하나의 변환 행렬로 주어지며, 행렬 곱 연산이 시간-영역의 신호에서 이루어지게 된다.

### 3.3 주파수 영역에서의 주파수-불변 빔공간 변환

#### 3.3.1 빔공간 변환 기법

앞서 살펴본 주파수-불변 빔공간 변환을 ULA(Uniform Line Array)상에서 각주파수 개념을 이용하여 다시 나타내면 다음과 같다.

$$P(\omega, \omega) = \sum_{m=-\infty}^{\infty} D(md_x, \omega) e^{-jm\omega} \quad (23)$$

여기서  $\omega_i = (\omega \sin \theta_{d_x})/c$ 이다. 위의 수식을 이용하면, 각 협대역 주파수 구간(frequency bin)마다 1-D DFT를 이용하여 필터 계수를 구할 수 있다<sup>[3]</sup>. 이와 같이 설계된 필터 계수를 이용하여, 3.1절과 같이 각 협대역 주파수 구간마다 빔공간 변환을 수행할 수 있다. 이를 수식으로 나타내면 다음과 같다.

$$\tilde{\mathbf{x}}_{fn, FIB2} = \mathbf{W}_{FIB2-BT, f}^H \mathbf{x}_{fn} \quad (24)$$

여기서 빔공간 변환 행렬  $\mathbf{W}_{FIB2-BT, f}$ 는 다음과 같다.

$$\mathbf{W}_{FIB2-BT, f} = [\mathbf{D}(\theta_0) \mathbf{D}(\theta_1) \cdots \mathbf{D}(\theta_{M-1})] \quad (25)$$

여기서,

$$\mathbf{D}(\theta) = [D_\theta(0, \omega_f) D_\theta(1, \omega_f) \cdots D_\theta(I-1, \omega_f)]^T \quad (26)$$

이다.

#### 3.3.2 역변환 기법

3.1절과 마찬가지로, 역변환 행렬은 다음과 같이 얻어질 수 있다.

$$\mathbf{W}_{FIB2-IBT} = \mathbf{W}_{FIB2-BT}^H \quad (27)$$

## IV. 실험 및 토의

### 4.1 실험 설계

본 논문에서는 앞서 언급하였던 여러 빔공간 변환

기법들이<sup>[7]</sup>의 MC-NMF와 결합되었을 때 그 성능을 살펴보기 위하여 실험을 진행하였다. 실험에 사용된 데이터로는 SiSEC 2010(Signal Separation Evaluation Campaign) 중 “Underdetermined speech and music mixture” task의 dataset을 사용하였으며, 그 중 마이크로폰 사이의 거리가 5 cm인 실험음원 8개를 사용하였다<sup>[9]</sup>. 사용된 실험음원은 각각 3개 혹은 4개의 음성음원이 존재하는 환경에서 2개의 마이크로폰을 사용하여 녹음된 음원들이며, 각 실험음원은 서로 다른 음원 신호를 사용하여 녹음된 음원이다. 사용된 음원데이터 8개의 특성은 표 1에서 살펴볼 수 있다. 즉, 1번 실험음원의 경우 여성음성 3개가 존재하는 환경에서 2개의 마이크로폰을 사용하여 녹음된 음원이다. 또한, 표에 나타난 특성이 같은 음원의 경우(예를 들어, 1번과 2번 신호) 서로 다른 음성음원을 사용하여 녹음되었기 때문에 서로 다른 특성을 가진다.

MC-NMF를 위해 모든 입력 음원들은 국소 푸리에 변환 처리가 되었으며, 이 때 사용된 윈도우는 1024 길이의 Hamming 윈도우이고, 50 % 중첩(overlap)되면서 2048 point로 변환되었다.

앞서 3장에서 언급된 3가지의 빔공간 변환 기법들이<sup>[7]</sup>의 MC-NMF 알고리즘과 결합되어 다채널 음원 분리를 수행하였으며, 조향 벡터를 이용한 빔공간 변환 기법(3.1절)과 주파수 영역에서의 주파수-불변 빔공간 변환 기법(3.3절)은 국소 푸리에 변환 모듈 뒤 단에, Tapped delay-line을 이용한 주파수-불변 빔공간 변환 기법(3.2절)은 국소 푸리에 변환 모듈 앞 단에 각각 사용되었다.

MC-NMF의 기저 개수는 16개로 설정되었고, 알고

표 1. 실험에 사용된 데이터의 특성  
Table 1. Properties of the test data.

Mixed signal	Gender	Sources
1	Female	3 sources
2	Female	3 sources
3	Female	4 sources
4	Female	4 sources
5	Male	3 sources
6	Male	3 sources
7	Male	4 sources
8	Male	4 sources

리즘의 반복횟수는 200으로 설정되었다. 빔공간-영역 MC-NMF의 성능에 대한 고찰은 앞선 논문에서 논의되었고, 본 논문에서는 빔공간-영역 MC-NMF 알고리즘 중 빔공간 변환 기법의 영향에 대해 집중적으로 고찰하고 있으므로, 기존의 빔공간-영역 MC-NMF 알고리즘(3.1 절의 조향벡터 기법을 이용한 MC-NMF 알고리즘)에서 빔공간 변환 기법만 변화시켜 비교 및 도시하였다.

성능평가 지수로는 E. Vincent 등에 의해 개발된 SDR (Signal-to-Distortion Ratio), SIR(Source-to- Interference Ratio), SAR(Source-to-Artifacts Ratio), ISR(source-Image-to-Spatial-distortion Ratio)를 사용하였다<sup>[10]</sup>. SDR은 전체적인 에러를 나타내는 지수이고, SIR은 분리된 음원과 섞여 있는 다른 간섭신호의 에너지 비율, SAR은 분리된 음원과 신호 자체의 결합(artifact)의 에너지 비율 각각 나타낸다. ISR은 공간적인 에러에 대한 성능 지표이다.

위의 성능 평가 지수는 다음과 같이 정의된다. 먼저 추정된 음원 이미지 신호를 식(28)와 같이 나타낸다<sup>[10]</sup>.

$$\hat{s}_{ij}^{img}(t) = s_{ij}^{img}(t) + e_{ij}^{spat}(t) + e_{ij}^{interf}(t) + e_{ij}^{artif}(t) \quad (28)$$

여기서  $s_{ij}^{img}(t)$ 는  $i$ 번째 채널,  $j$ 번째 음원의 실제 이미지 신호를,  $e_{ij}^{spat}(t)$ ,  $e_{ij}^{interf}(t)$ ,  $e_{ij}^{artif}(t)$  신호는 각각 공간적 에러 요소, 간섭 신호 에러 요소, 결합 에러 요소를 나타낸다. 이 때, 각각의 성능 지표는 다음과 같이 정의된다.

$$ISR_j = 10 \log_{10} \frac{\sum_{i=1}^I \sum_t s_{ij}^{img}(t)^2}{\sum_{i=1}^I \sum_t e_{ij}^{spat}(t)^2} \quad (29)$$

$$SIR_j = 10 \log_{10} \frac{\sum_{i=1}^I \sum_t (s_{ij}^{img}(t) + e_{ij}^{spat}(t))^2}{\sum_{i=1}^I \sum_t e_{ij}^{interf}(t)^2} \quad (30)$$

$$SAR_j = 10 \log_{10} \frac{\sum_{i=1}^I \sum_t (s_{ij}^{img}(t) + e_{ij}^{spat}(t) + e_{ij}^{interf}(t))^2}{\sum_{i=1}^I \sum_t e_{ij}^{artif}(t)^2} \quad (31)$$

$$SDR_j = 10 \log_{10} \frac{\sum_{i=1}^I \sum_t s_{ij}^{img}(t)^2}{\sum_{i=1}^I \sum_t (e_{ij}^{spat}(t)^2 + e_{ij}^{interf}(t)^2 + e_{ij}^{artif}(t)^2)} \quad (32)$$

각 에러 요소를 추정하는 과정은<sup>[10]</sup>에 더욱 상세히 기술되어 있다. 위의 모든 성능 지표는 값이 클수록 좋은 성능을 의미한다. 또한 SDR이 종합적인 성능을 의미하기는 하지만, 각 에러 요소들(공간적 에러 요소, 간섭 신호 에러 요소, 결합 에러 요소)이 독립적으로 인지되어 성능에 영향을 주므로<sup>[10]</sup>, SIR, SAR, ISR을 독립적으로 비교하는 것이 유의미하며, SDR의 경우 각 요소에 대한 성능 경향이 다를 경우 종합적 성능에 대한 비교 지표로서 해석될 수 있다.

#### 4.2 실험 결과 및 토의

그림 2에 각 성능 지표에 따른 실험 결과들이 나타나 있다. 실험 결과, 모든 성능 지표에서 조향 벡터 기반의 빔공간 변환 기법에 비해 주파수-불변 빔공간 변환 기법들의 성능이 더 좋게 나타났다. 이는 조향 벡터 기반의 빔공간 변환 기법의 응답이 주파수에 비례하여 크게 변함으로써 음원 분리 기법의 성능에 영향을 준 것으로 해석된다. 주파수-불변 빔공간 변환 기법들의 경우 상대적으로 주파수에 따른 빔공간 변환 응답의 변화가 적은 이점이 있어, 성능 향상에 기여한 것으로 풀이된다.

Tapped delay-line을 이용한 주파수-불변 빔공간 변환 기법의 경우 음원의 종류에 관계없이 상대적으로 안정적인 성능을 보여주는 것을 확인할 수 있다. 다만, ISR적인 측면에서 뚜렷한 성능 저하를 보이고 있어 공간적 에러를 줄이는 것이 목적인 시스템에서는 사용하기가 어렵다는 단점을 가질 수 있다. 주파수 영역에서의 주파수-불변 빔공간 변환 기법의 경우 음원의 종류에 따라 성능 변화를 보이고 있지만, ISR의 성능 저하가 거의 없다는 장점을 가지고 있다.

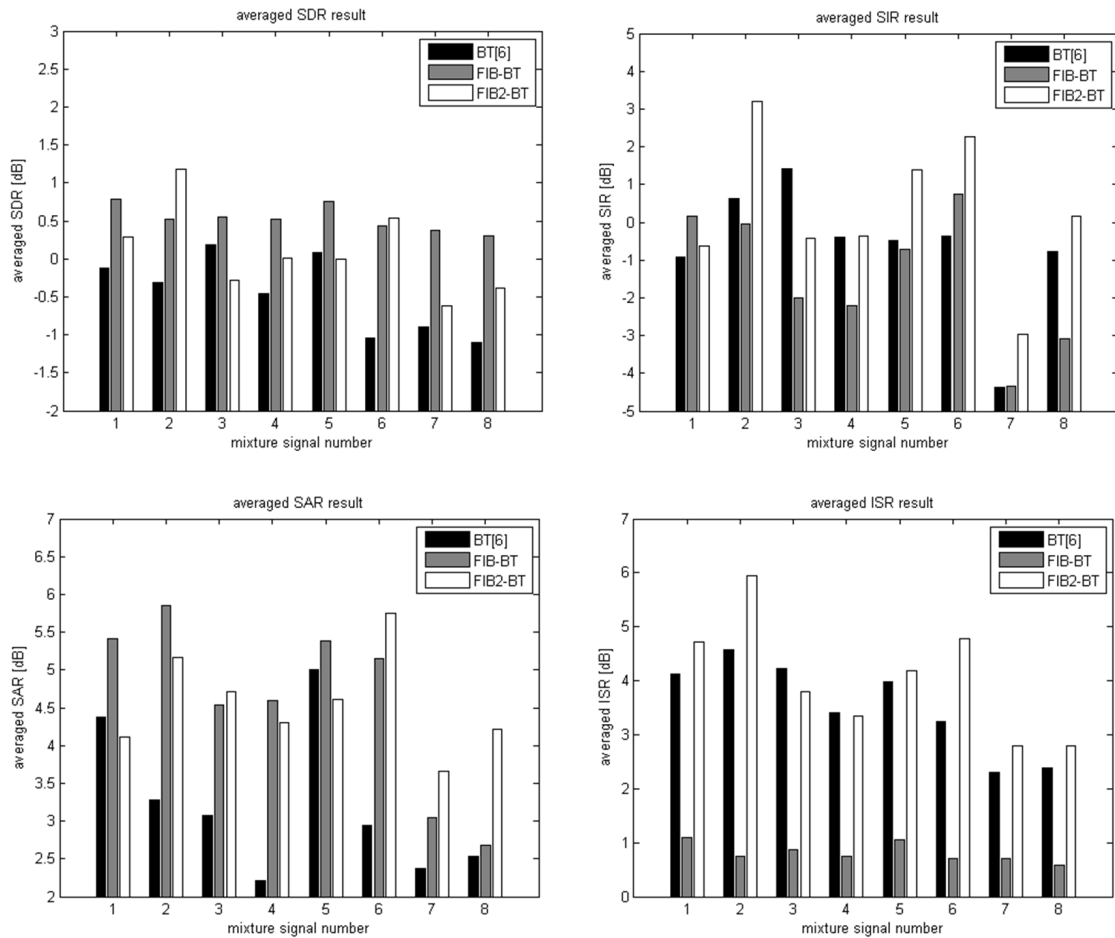


그림 2. 성능 평가 결과  
Fig. 2. Evaluation results.

## V. 결 론

본 논문에서는 다채널 음원 분리 기법을 위한 빔공간 변환 기법에 대해서 연구하였다. 본 논문에서 다른 빔공간 변환 기법들은 다채널 음원 분리에 적용되었을 때 그 성능에 미치는 영향을 염두에 두고 분석되었으며, 특히 변환 기법뿐만 아니라 역변환 기법까지 고려되었다.

본 논문에서 연구된 빔공간 변환 기법은 크게 3가지이다. 첫 번째는 조향 벡터를 사용하는 빔공간 변환 기법이며, 두 번째는 tapped delay-line을 이용한 광대역 주파수 불변 빔공간 변환 기법이고, 세 번째는 주파수 영역에서의 주파수 불변 빔공간 변환 기법이다. 3가지 알고리즘 모두 MC-NMF 알고리즘과 함께 사용되어 다채널 음원 분리를 수행하였다. 실험결과

주파수 불변 빔공간 변환 기법들이 조향벡터 기반의 빔공간 변환 기법에 비해 좋은 성능을 보였으며, tapped delay-line의 광대역 주파수 불변 빔공간 변환 기법이 보다 안정적인 성능을 보인 반면, ISR 측면에서는 주파수 영역의 주파수 불변 빔공간 변환 기법이 더욱 좋은 성능을 보였다.

## 참고문헌

1. Xiao-Liang Xu and Kevin Buckley, "An Analysis of Beam-Space Source Localization," *IEEE Trans. Signal Processing*, vol. 41, no. 1, Jan. 1993.
2. M. D. Zoltowski, G. M. Kautz, S. D. Silverstein, "Beamspace Root-MUSIC," *IEEE Trans. Signal Processing*, vol. 41, no. 1, Jan. 1993.
3. W. Liu and S. Weiss, *Wideband Beamforming Concepts and Techniques*, John Wiley & Sons, 2010.

4. H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, K. Shikano, "Blind Source Separation Based on a Fast -Convergence Algorithm Combining ICA and Beamforming," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 2, 2006.
5. L. -H. Kim, I. Tashev, A. Acero, "Reverberant Speech Signal Separation based on Regularized Subband Feedforward ICA and Instantaneous Direction of Arrival," *Proc. ICASSP 2010*, 2010.
6. S. Lee, S. H. Park, K. -M. Sung, "Beam-space-domain Multichannel Nonnegative Matrix Factorization for Audio Source Separation," Accepted for publications in *IEEE Signal Processing Letters*.
7. A. Ozerov, C. Fevotte, "Multichannel Nonnegative Matrix Factorization in Convolutional Mixtures for Audio Source Separation," *IEEE Trans. Audio, Speech, Language Processing*, vol. 18, no. 3, 2010.
8. T. K. Moon, and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice Hall, New Jersey, 2000.
9. *Signal Separation Evaluation Campaign 2010 (SiSEC 2010)*, <http://www.sisec.wiki.irisa.fr>, 2010.
10. E. Vincent, H. Sawada, P. Bofill, S. Makino, J. P. Rosca, "First stereo audio source separation evaluation campaign: data, algorithms and results," in *Proc. Int. Conf. Ind. Compon. Anal. Blind Source Separation (ICA'07)*, 2007.
11. A. Cichocki, R. Zdunek, A. H. Phan, S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, Wiley, Chichester, 2009.

## 저자 약력

### ▶ 이 석 진(Seokjin Lee)



2006년 8월: 서울대학교 공과대학 전기·컴퓨터공학부(학사)  
 2008년 8월: 서울대학교 공과대학 전기·컴퓨터공학부(석사)  
 2008년 9월 ~ 현재: 서울대학교 공과대학 전기·컴퓨터공학부 박사과정  
 <관심분야> 음원 분리, 음향신호처리, 적응신호처리

### ▶ 박 상 하(Sang Ha Park)



2005년 2월: 서울대학교 음악대학 기악과(학사)  
 2008년 2월: 서울대학교 공과대학 전기·컴퓨터공학부(석사)  
 2008년 3월 ~ 현재 : 서울대학교 공과대학 전기·컴퓨터공학부 박사과정  
 <관심분야> 음원 분리, 악보 전사, 음향신호처리

### ▶ 성 광 모(Koeng-Mo Sung)



1971년: 서울대학교 전자공학과(학사)  
 1973년: 독일 아헨공대 Vordiplom  
 1977년: 독일 아헨공대 전자통신공학 Dipl.-Ing.  
 1982년: 독일 아헨공대 음향공학 Dr.-Ing.  
 1977년 ~ 1983년: 독일 아헨공대 음향공학연구소 연구원  
 1983년 ~ 현재: 서울대학교 공과대학 전기·컴퓨터공학부 교수