

## 비음수 행렬의 부분적 공동 분해를 이용한 드럼 음원 분리

유지호<sup>1</sup> 김민제<sup>2</sup> 강경옥<sup>3</sup> 최승진<sup>4</sup>

포항공과대학교 컴퓨터공학과<sup>1,4</sup>, 한국전자통신연구원<sup>2,3</sup>  
{zentasis<sup>1</sup>, seungjin<sup>4</sup>}@postech.ac.kr, {mkim<sup>2</sup>, kokang<sup>3</sup>}@etri.re.kr

### Nonnegative Matrix Partial Co-factorization for Drum Source Separation

Jiho Yoo<sup>1</sup> Minje Kim<sup>2</sup> Kyeongok Kang<sup>3</sup> Seungjin Choi<sup>4</sup>

Department of Computer Science, POSTECH  
Electronics and Telecommunication Research Institute

#### 요 약

본 논문에서는 드럼 음원이 포함된 다성 음악 (polyphonic music) 에서 드럼 신호를 분리해 내는 문제를 다룬다. 기존의 방법은 비음수 행렬 분해 (nonnegative matrix factorization)를 사용하여 찾아낸 주파수 특성들로부터 지지벡터기계 (support vector machine) 을 이용하여 드럼 신호의 주파수 특성들만을 분류해 내는 방식으로 드럼 음원 분리를 수행한다. 하지만 드럼 신호의 주파수 특성과 일반적 악기 신호의 주파수 특성을 분류해 내는 지지벡터기계의 학습에는 드럼이 아닌 악기의 신호가 학습 데이터로 필요하며, 다양한 악기에 대하여 이러한 데이터를 만드는 것은 매우 어렵다. 여기서는 드럼 외의 악기 신호에 대한 추가적인 데이터를 필요로 하지 않고, 드럼 신호에 대한 사전 정보만을 사용하여 드럼 신호를 분리해 내는 방법으로써, 비음수 행렬의 부분적 공동 분해 방법을 제안한다. 대상이 되는 신호의 스펙트로그램 행렬과 드럼 소리로 구성된 사전 정보 신호의 스펙트로그램 행렬을 동시에 분해하되, 분해된 행렬의 일부분을 공유하도록 설정함으로써 자연스럽게 드럼 신호의 특성을 분리해 내는 것이 가능하다. 간단한 곱셈에 의한 업데이트 방식을 통해 이와 같은 작업을 수행할 수 있으며, 실제 음원에 대한 실험을 통해 제안된 방법의 성능을 검증하였다.

#### 1. 서 론

혼합되어 1, 2개의 채널을 통해 제공되는 상용 음악에서 중요한 정보를 추출하는 연구는 다양한 음악 관련 응용분야를 가지고 있다. 이를 통해 추구하고자 하는 것은 텍스트 기반의 검색 이후 획득한 음악을 감상하는 것에 그치던 단순한 음악 콘텐츠의 소비 행태를 대신해 사용자의 다양한 요구에 부합하는 보다 지능화된 응용을 제공하는 것이다. 이와 관련하여, 음악 장르 인식(audio genre classification), 음악 분위기 인식(audio mood classification) 및 음악의 내재된 관계성(similarity)에 기반한 재생 목록의 자동 추천(automatic music recommendation) 등의 응용과, 유관 콘텐츠 검색, 자동 악보 생성(automatic music transcription), 허밍에 의한 질의(query by humming) 등의 응용이 있다. 전술된 음악 콘텐츠 관련 지능 시스템은 음악 신호로부터 많은 특징을 추출 및 활용하는데, 이 때, 대부분의 특징은 음악 콘텐츠에 포함된 악기의 종류에 따라 매우 다르다. 일례로, 음악의 빠르기(tempo)나 분절 정도(beat)와 같은

특징은, 드럼 신호에 대해서 추출되는 것이 혼합 신호로부터 추출되는 것에 비해서 더 좋은 성능을 기대할 수 있다. 반면, 음악에 포함된 다양한 음정 (pitch) 또는 그 음정들로 이루어지는 코드 (chord) 특징은 드럼이 아닌 화성 악기만으로 이루어진 신호에서 추출하는 것이 더 좋은 성능을 낸다. 즉, 객체화된 음악 콘텐츠, 특히 적어도 타악기와 화성악기 정도의 구분 만으로도 상기 음악 콘텐츠 관련 응용의 성능은 비약적으로 높아질 수 있다. 그러나 녹음실 환경에서 미리 혼합되어서 배포되는 상용 음악의 원본 악기 객체를 획득하는 것은 어려운 일이며, 본 논문에서 제시하는 것처럼 음악 신호에 특화된 음원 분리 기술의 필요성이 대두된다.

이와는 별도로, 최근 Music 2.0과 같은 객체기반 오디오 서비스가 음악 시장의 새로운 화두가 되고 있다. 이는 사용자에게 음악에 포함된 다양한 악기를 직접 제어하도록 허용함으로써 보다 다채로운 청취 경험을 제공하는 것을 목표로 한다. 마찬가지로 이들 서비스는 혼합 신호를 단순 제공하는 것이 아닌 객체 단위의 신호를 확보해야 하며, 원본 신호의 확보가 여의치 않은

경우 악기 음원 분리의 활용이 필요하다. 서비스 관점에서 가장 주요한 악기 객체는 가창 (vocal) 신호이지만, 대부분의 가창 신호 추출 방식 역시 타악기 분리와 연동되는 경우 더 나은 성능을 얻는다. 또한 타악기 신호 자체 역시 객체 기반 음악 서비스에서 가창과 더불어 주요한 한 가지 객체이다.

비음수 행렬 분해 (Nonnegative Matrix Factorization) 는 주어진 비음수 입력 행렬을 두 개의 비음수 행렬의 곱의 형태로 표현함으로써 보다 낮은 랭크의 근사를 수행하는 방법이다 [7]. 비음수 행렬 분해는 이미지 처리, 뇌-컴퓨터 인터페이스, 문서 분류, 협업 필터링 등의 다양한 분야에 적용되어 왔으며, 각각의 분야에 최적화된 형태로의 다양한 변형 방법이 제안되어 왔다. 공동 분해 (co-factorization) 는 이와 같은 변형 방법 중의 하나로, 하나의 입력 행렬을 분해하는 기존의 비음수 행렬 분해와는 달리 여러 개의 입력 행렬을 동시에 분해함으로써 그들 사이에 존재하는 다양한 종류의 관계를 파악해 내는 방법이다. 공동 분해는 분류에 있어서 미리 주어진 레이블 정보의 활용 [15], 군집화에 있어서 미리 주어진 연결 정보의 활용 [16], 그리고 뇌파 신호 분석에서 대상별 변화의 활용 [8] 등의 다양한 형태로 사용되어 왔다. 복잡한 형태의 관계형 데이터 (relational data) 를 다루기 위한 집합적 행렬 분해 (collective matrix factorization) [9] 가 제안된 바 있으며, 삼단 분해를 공동 분해의 틀 안에 사용한 비음수 행렬의 공동 삼단 분해 (weighted nonnegative matrix co-tri-factorization) [14] 방법도 제안되어, 협업 필터링 문제에서 추가적인 사용자 정보나 상품 정보를 이용하는 데에 사용되었다. 이처럼 공동 분해 방법은 추가적인 정보의 효과적인 사용을 필요로 하는 문제들에 대한 유용한 해결 방안으로써 활용되어 왔다.

비음수 행렬 분해의 성공적인 활용이 기대되는 분야로 음악 신호 분리 [6,13], 악보 추출 [10,1,12] 등을 포함하는 음악 신호 분석 분야를 들 수 있다. 공동 분해는 원하는 음악 신호의 주파수 특성이나 시간 특성이 사전 정보로 주어진 상황에서, 이와 같은 정보를 활용하여 음악 신호 분석을 수행하는 데에 쓰일 수 있을 것으로 기대된다. 여기서는 음악 신호로부터 주된 리듬 악기 부분을 추출해 내는 드럼 신호 분리 문제 [17,11,5,4] 를 다룬다. 드럼 신호 분리에는 음악 신호의 시간-주파수 영역에서의 스펙트로그램을 각각의 음원 신호들의 합의 형태로 분해한 후, 원하는 음원 신호들을 모아 분리된 신호를 재구성하는 방식이 주로 사용된다. 분리된 신호를 재구성하기 위해서는 어떤 신호들이 드럼에 해당하는지에 대한 사전 지식의 활용이 필요하게 된다. 알려져 있는 드럼 신호를 초기화 과정에서 사용하거나 [3], 분류를 위한 특정 기준을 활용하거나 [2] 하는 방식의 사전 지식 활용 방법이 제안되어 왔으며, 특히 [5] 에서는 미리 학습하여 만든 지지벡터기계 (support vector machine) 을 사용하여 드럼 신호만을 골라 내는 방법이 제안되었다. 하지만 이러한 방법들에서는 사전 정보가 분해 전의 초기화

과정이나 분해 후의 분류 과정에 사용될 뿐, 음악 신호를 분해 하는 과정 자체에서는 활용되지 않으므로, 이러한 사전 정보를 분해 과정 자체에서 활용할 수 있다면 드럼 음원 분리의 성능을 보다 향상시킬 수 있을 것으로 기대된다.

이 논문에서는 음악 신호의 음원 성분으로의 분리 과정에서 사전 정보를 직접적으로 활용할 수 있는 방법으로써, 비음수 행렬의 부분적 공동 분해 (non-negative matrix partial co-factorization) 방법을 제안한다. 이 방법은 음악 신호 행렬을 분해하여 나온 주파수 성분 행렬과 드럼 신호 행렬을 분해하여 얻어지는 주파수 성분 행렬 중의 특정 부분을 공유하는 형태로 구성된다. 공유되도록 설정된 주파수 성분 행렬 부분은 음악 신호와 드럼 신호 양 쪽 모두에 포함된 주파수 특성을 표현하도록 학습되며, 따라서 드럼 신호에 해당하는 주파수 특성만이 공유되는 주파수 성분 행렬에 포함되게 된다. 드럼 신호를 제외한 나머지 화성 악기 신호의 주파수 특성은 자연스럽게 공유되지 않도록 설정된 주파수 성분 행렬에 포함된다. 분리된 드럼 신호는 공유되도록 설정한 주파수 성분 행렬 부분과, 그에 해당하는 시간 성분 행렬의 곱으로 재구성하여 얻을 수 있고, 분리된 나머지 화성 악기 신호는 공유되지 않도록 설정한 주파수 성분 행렬과 그에 해당되는 시간 성분 행렬의 곱을 통하여 얻을 수 있다. 기존의 방법들과는 달리 제안된 방법을 사용하면 별도의 분류기 등을 사용하여 분해된 성분을 분류할 필요 없이, 분해 과정 중에 자연스럽게 음원 분리가 수행된다. 본 논문에서는 이를 위한 간단한 곱셈의 업데이트를 이용한 학습 방법을 제안하고, 실제 음악 신호를 대상으로 실험을 수행하여 성능을 검증하였다.

이 논문의 구성은 다음과 같다. 2장에서 기존에 제안된 비음수 행렬 분해 기반의 드럼 음원 분리 방법을 정리하고, 3장에서 비음수 행렬의 부분적 공동 분해 모델과 개발된 곱셈의 업데이트에 의한 학습 방법을 소개한다. 실제 음악 신호를 대상으로 한 실험 결과가 4장에 정리되었으며, 5장에서 결론을 맺는다.

## 2. 비음수 행렬 분해와 드럼 음원 분리

비음수 행렬 분해를 사용하면 주어진 음악 신호의 스펙트로그램으로부터 음악 신호에 포함된 각각의 음원이 가지는 주파수 영역에서의 특징과 시간 영역에서의 특징을 추출해 낼 수 있다. 주파수-시간 형태로 구성된 세기 스펙트로그램 행렬  $X$  에 비음수 행렬 분해를 다음과 같이 적용하면,

$$X = UV^*$$

주파수 영역에서의 특징을 나타내는 비음수 행렬  $U$  와, 시간 영역에서 해당 주파수 특징이 발생하는 정도를 나타내는 비음수 행렬  $V$  를 얻을 수 있다. 일반적인 음악 신호에 비음수 행렬 분해를 적용하면 주파수 특징 행렬  $U$  의 각각의 열에 드럼 신호의 주파수 특징 벡터와 나머지 화성 악기 신호의 주파수 특징 벡터가 불규칙적으로 섞여서 나타나게 된다. 이 중에서 드럼

소스에 해당하는 주파수 특성을 표현하는 열만을 고르고, 그에 해당하는 시간 영역에서의 발현 정도를 나타내는 열을 행렬  $V$ 에서 골라 드럼 신호의 특징을 나타내는 행렬  $U_D$ 와  $V_D$ 를 구성하면, 분리된 드럼 신호의 스펙트로그램을 다음과 같이 두 행렬의 곱의 형태로 얻을 수 있다.

$$X_D = U_D V_D^*$$

하지만 드럼 음원의 주파수 특성이 행렬  $U$  안에서 특정한 규칙 없이 나타나 있기 때문에, 이 경우 드럼 신호에 해당되는 특성만을 분류해 주는 추가적인 방법이 필요하며, 이를 위해서는 드럼 신호에 대한 사전 지식을 활용할 필요가 있다. 일반적으로 드럼 신호는 화성 악기의 신호와는 달리 주파수 영역에서의 배음 구조가 강하게 나타나지 않고, 시간 영역에서의 지속 시간이 짧은 특성을 가지는 것으로 알려져 있다. 이러한 사전 지식을 이용하여 드럼 신호를 분류해 내는 경험적인 기준을 구성하여 사용하는 것[2]이 가능하다. 또한 드럼 신호의 주파수 특성을 사전 정보 형태로 가지고 있다면, 행렬  $U$ 의 일부분을 알려진 드럼 신호의 주파수 특성으로 초기화한 후 비음수 행렬 분해를 적용하는 방식[3]으로 드럼 신호의 주파수 특징의  $U$  안에서의 위치를 특정 지을 수 있다. 또한, 지지벡터기계를 드럼 신호 특성과 나머지 화성 악기 신호 특성을 분류하기 위해 학습시켜 사용하는 방법[5]도 사용할 수 있다. 이러한 분류기를 학습시키기 위해서는 다양한 드럼 신호와 화성 악기 신호로부터 얻어진 주파수 영역 및 시간 영역에서의 특징들을 이용하여야 한다.

지금까지 제안된 비음수 행렬 분해에 기반한 방법들의 주된 문제점은 사전 지식이 행렬 분해 과정 자체에서는 사용되지 않는다는 데에 있다. 사전 지식은 주로 분해가 된 이후에 드럼 신호를 찾아내는 과정에서 사용되어 왔다. 초기화에 사전에 찾아낸 주파수 특징을 사용하는 방법은 사용할 수 있는 주파수 특징의 개수가 제한적이므로, 다양한 음원에 대해 적용하기에는 무리가 따른다. 비음수 행렬 분해는 일반적으로 드럼 신호의 주파수 특징과 나머지 화성 악기의 주파수 특징을 서로 다른 벡터로 분리해 내는 것으로 알려져 있으나, 이들이 한 벡터에 섞여서 나오지 않으리라는 명확한 보장은 주어져 있지 않다. 게다가 분류기를 사용하는 방법의 경우에는 분류기를 학습하기 위한 추가적인 작업이 필요하며, 분류기에서 발생한 에러에 의해 분리 성능이 떨어질 가능성이 존재한다는 문제점을 가지고 있다.

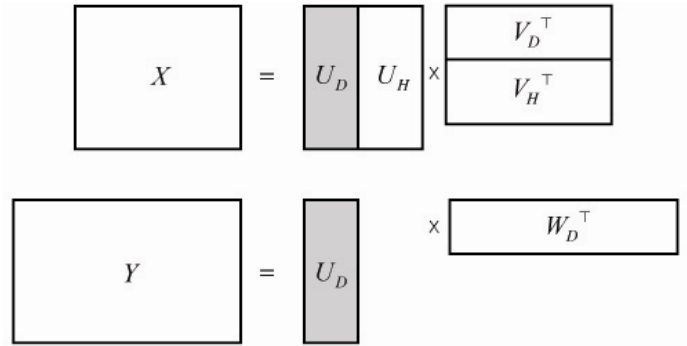


그림 1. 비음수 행렬의 공동 분해 모델. 입력 신호 행렬  $X$ 를 분해하는 데에 사용되는 행렬의 일부분  $U_D$ 가 드럼 신호 행렬  $Y$ 를 분해하는 데에 공동으로 사용되어 드럼 신호의 특징을 표현한다.

### 3. 비음수 행렬의 부분적 공동 분해

여기서는 드럼 음원 분리를 위한 비음수 행렬의 부분적 공동 분해 방법을 소개한다. 기존의 비음수 행렬 분해와는 달리, 비음수 행렬의 공동 분해는 분해에 있어서 다음과 같이 드럼 부분과 나머지 화성 악기 부분을 명시적으로 분리한 모델을 사용한다.

$$X = U_D V_D^* + U_H V_H^* \quad (1)$$

여기서  $U_D$ 와  $V_D$ 는 각각 드럼 음원의 주파수 영역 및 시간 영역에서의 특징을 표현하는 비음수 행렬이며,  $U_H$ 와  $V_H$ 는 각각 나머지 화성 악기 음원의 주파수 영역 및 시간 영역에서의 특징을 표현하는 비음수 행렬이다. 위의 모델 상에서는 드럼 신호와 나머지 화성 악기 신호가 따로 표현되어 있지만, 이를 자동적으로 분해하기 위해서는 추가적인 드럼 신호에 대한 사전 지식이 필요하다. 여기서는 드럼 음원으로만 구성된 신호로부터 얻은 세기 스펙트로그램  $Y$ 를 사전 정보로 활용한다. 사전 정보 행렬은 드럼 음원으로만 구성되어 있으므로 다음과 같이 분해하는 것이 가능하다.

$$Y = U_D W_D^* \quad (2)$$

여기서  $U_D$ 는 (1)에서 쓰인 바와 같이 드럼 신호의 주파수 특징을 나타내는 비음수 행렬이며,  $W_D$ 는  $Y$ 에서 각각의 주파수 특징이 시간대 별로 발현하는 정도를 표현한 비음수 행렬이다. 이러한 형태의 사전 정보 행렬을 이용하는 방법으로  $Y$ 에 비음수 행렬 분해를 적용하여 얻은 주파수 특징 행렬  $U_D$ 를 (1)의 모델의 학습 과정의 초기값으로 주고 따로 학습을 수행하지 않는 것이나, 이러한 방법을 사용할 경우 특히  $Y$ 에 포함된 드럼 신호와  $X$ 에 포함된 드럼 신호가 다른 경우 명확한 분리를 보장하기 힘들다.

여기서는 부분적인 공동 분해를 통해서 행렬  $X$ 와 행렬  $Y$ 를 주파수 특징 행렬  $U_D$ 를 공유한 채 분해하는 방법을 제안한다 (그림 1).  $Y$ 를 분해하기 위하여  $U_D$ 에는 드럼 음원의 주파수 특징들이 자동적으로 모이게 되고,  $U_H$ 에는 주어진 음악 신호의 나머지 부분, 즉 화성 악기의 주파수 특징들이 모이게 된다. 비음수

행렬의 부분적 공동 분해의 목적 함수는 대상 음악 신호 모델 (1) 과 사전 정보 신호 모델 (2) 에서 발생하는 오차를 이용하여 다음과 같이 정의할 수 있다.

$$L = \frac{1}{2} \| X - U_D V_D^* - U_H V_H^* \|_F^2 + \frac{\lambda}{2} \| Y - U_D W_D \|_F^2 \quad (3)$$

여기서  $\lambda$  는 음악 신호와 사전 정보와의 상대적인 중요성을 지정해 주는 인수이다.

위의 목적 함수로부터 분해된 행렬들을 얻기 위해, 각각의 행렬에 대한 편미분을 구한 후 그를 바탕으로 곱셈에 의한 업데이트식을 구성할 수 있다. 만약 편미분

식  $\frac{\partial L}{\partial U}$  를 비음수인 두 부분  $\left[ \frac{\partial L}{\partial U} \right]^+$  와  $\left[ \frac{\partial L}{\partial U} \right]^-$  의 차로

다음과 같이 표현하는 것이 가능하다면,

$$\frac{\partial L}{\partial U} = \left[ \frac{\partial L}{\partial U} \right]^+ - \left[ \frac{\partial L}{\partial U} \right]^-$$

이에 따른 곱셈의 업데이트식을 다음과 같은 형태로 얻을 수 있다.

$$U \leftarrow U e \left( \left[ \frac{\partial L}{\partial U} \right]^+ / \left[ \frac{\partial L}{\partial U} \right]^- \right) \quad (3)$$

위의 식에서  $e$  기호는 행렬의 각 원소끼리 이루어지는 곱셈을,  $/$  기호는 행렬의 각 원소끼리 이루어지는 나눗셈을 의미한다. 위의 곱셈의 업데이트 식은 편미분의 값이 0 이 되는 지역해 (local minimum) 위치에서 수렴하며, 행렬에 주어진 비음수 조건을 벗어나지 않는다.

각각의 행렬에 대한 편미분을 계산하여 위의 (4) 식을 적용하면, 다음과 같은 각 행렬에 대한 곱셈의 업데이트식을 얻을 수 있다.

$$\begin{aligned} U_D &\leftarrow U_D e \frac{X V_D^* + \lambda Y W_D}{U_D V_D^* U_D + U_H V_H^* V_D + \lambda U_D W_D W_D^*} \\ V_D &\leftarrow V_D e \frac{X^* U_D}{V_D U_D^* U_D + V_H U_H^* U_D} \\ U_H &\leftarrow U_H e \frac{X V_H^*}{U_H V_H^* U_H + U_D V_D^* V_H} \\ V_H &\leftarrow V_H e \frac{X^* U_H}{V_H U_H^* U_H + U_D V_D^* V_H} \\ W_D &\leftarrow W_D e \frac{Y^* U_D}{W_D U_D^* U_D} \end{aligned} \quad (5)$$

위의 업데이트식을 반복적으로 수행함으로써 드럼 신호에 해당하는 주파수 특징 행렬  $U_D$  와 그에 해당하는 시간 특징 행렬  $V_D$  를 얻을 수 있다. 분리된 드럼 신호의 세기 스펙트로그램은  $U_D V_D^*$  으로 계산해 낼 수 있으며, 이로부터 분리된 드럼 신호를 얻을 수 있다. 드럼 신호 분리를 위한 비음수 행렬의 부분적 공동 분해 알고리즘을 정리하면 다음과 같다.

#### 알고리즘: 비음수 행렬의 부분적 공동 분해를 이용한 드럼 신호 분리

입력

- 대상 음악 신호의 세기 스펙트로그램  $X$
- 드럼 정보 신호의 세기 스펙트로그램  $Y$
- 중요도 결정 인수  $\lambda$

출력

- 분리된 드럼 신호와 나머지 신호

1. 행렬  $U_D, V_D, U_H, V_H, W_D$  를 임의의 양의 난수값을 사용하여 초기화한다.
2. 일정 횟수, 혹은 수렴할 때까지
  - (5) 의 식을 사용하여 각각의 행렬을 업데이트한다.
3. 분리된 신호를 재구성한다.
  - $X_D = U_D V_D^*$  을 이용하여 드럼 신호를 재구성한다.
  - $X_H = U_H V_H^*$  을 이용하여 나머지 신호를 재구성한다.

#### 4. 실험 결과

제안된 비음수 행렬의 부분적 공동 분해 알고리즘을 드럼 신호 분리 문제에 적용하여 성능을 알아 보았다. 10곡의 실제 가요 음원으로부터 10초 길이의 대상 음악 신호를 각각 10개씩 만들어 실험에 사용하였다. 사전 정보로는 실험 대상 음악과는 다른 13 곡에서 사용된 드럼 신호를 10초씩 모아 총 130초 길이의 사전 정보 신호를 구성하여 사용하였다. 실험에 사용된 음악 신호는 모두 44.1 kHz 의 샘플링 레이트에 16비트로 인코딩 되었으며, 2048 샘플 길이의 윈도우 (약 50 ms) 를 256 샘플씩 이동시켜 스펙트로그램을 구하였다.

분리 실험의 성능을 측정하기 위하여 분리된 신호를 원래의 신호와 비교하였으며, 다음과 같이 계산되는 신호 대 잡음비 (Signal-to-Noise Ratio, SNR) 를 성능 측정의 수단으로 사용하였다.

$$SNR = 10 \log_{10} \frac{\sum s(t)^2}{\left( \sum s(t) - \hat{s}(t) \right)^2}$$

위의 식에서  $s(t)$  는 원래의 신호를 의미하며,  $\hat{s}(t)$  는 분리하여 얻은 신호를 의미한다. 분리 대상으로 사용된 음악의 드럼 트랙과 나머지 트랙을 원래의 신호로 사용하여 분리된 신호의 SNR 을 측정하였다.

비음수 행렬의 부분적 공동 분해를 사용하기 위해서는 인수로 드럼 특징 행렬  $U_D, V_D, W_D$  및 화성 악기 특징 행렬  $U_H, V_H$  에서 몇 개의 특징을 뽑아낼 것인지를 각 행렬의 열의 개수로 설정해 주어야 한다. 다양한 특징 개수에 대해 실험을 수행한 결과 드럼 특징 70개와 화성 악기 특징 100개를 사용한 경우에 좋은 성능이 나타나 이 수치를 실험에 사용하였다.

대상 신호와 사전 정보 신호 사이의 중요도를 결정하는 인수  $\lambda$  값도 두 신호의 길이 비의 역수인  $R=10/130$  값을 기준으로 다양한 인수 값에 대해 실험한 결과  $0.1R=0.0077$  값에서 좋은 성능을 보여 이 값을 전체 실험에서 인수로 사용하였다.

표 1. 비음수 행렬 분해와 지지벡터기계를 사용한 방법 (NMF+SVM) 과 비음수 행렬의 부분적 공동 분해 (NMPCF) 알고리즘의 10 곡의 가요 음악에 대해 측정된 드럼 음원 분리의 SNR 결과. 각 곡에 해당되는 10개의 10초 길이의 신호들에 대한 성능의 평균값으로 5번의 수행에서 얻어진 평균 성능이다. 제안된 방법이 기존의 방법보다 좋은 성능을 보인 경우 굵은 글씨로 표기하였다.

Song	SNR (드럼)		SNR (화성 악기)	
	NMF+SVM	NMPCF	NMF+SVM	NMPCF
1	8.02	<b>8.84</b>	8.43	7.95
2	4.58	<b>5.48</b>	3.49	<b>4.66</b>
3	4.29	<b>5.04</b>	4.69	<b>5.98</b>
4	3.62	3.01	5.14	4.21
5	5.56	5.20	6.17	<b>6.47</b>
6	4.82	<b>6.90</b>	1.35	<b>5.40</b>
7	3.87	<b>3.94</b>	7.08	6.68
8	-0.68	2.76	3.91	<b>6.36</b>
9	4.19	<b>4.32</b>	7.30	7.04
10	7.90	7.81	8.41	8.08
평균	4.62	<b>5.33</b>	5.60	<b>6.28</b>

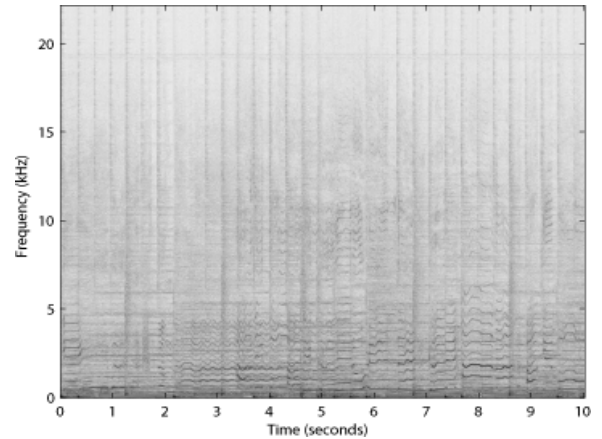
목적 함수의 값은 알고리즘이 반복적으로 수행됨에 따라 상대적으로 천천히 수렴해 가지만, 분리한 신호의 SNR 값은 적은 반복 횟수만으로도 최대치에 도달하는 모습이 관찰되었다. 20회의 반복 횟수를 거치면 충분히 좋은 분리 결과를 얻을 수 있어, 최대 반복 횟수를 20회로 설정하여 실험을 수행하였다.

비음수 행렬의 부분적 공동 분해 알고리즘은 드럼 음원 분리에 대하여 성공적으로 동작함을 확인할 수 있었다 (그림 2). 알고리즘 성능의 비교를 위해 기존에 제안된 비음수 행렬 분해와 지지벡터기계를 사용한 드럼 음원 분리 방법 (NMF+SVM, [5])을 사용하였다. 분해된 드럼 신호의 SNR 과 분해된 나머지 화성 악기 신호의 SNR 을 모두 측정하였다 (표 1). SNR 값은 대체로 곡에 따라 다르게 나타났으며, 같은 곡에 속한 대상 음악 신호의 경우 SNR 값이 유사하게 나타나 곡별로 평균을 내어 SNR 값을 계산하였다. 곡에 따라 차이는 있지만, 제안된 비음수 행렬의 부분적 공동 분해를 사용한 방법이 대체로 기존의 방법보다 우수한 성능을 보임이 확인되었다. 기존의 방법은 몇몇 곡의 경우 심각하게 실패하는 모습을 보였으나, 제안된 방법은 모든 경우에 있어서 기존의 방법보다 낮거나 비슷한 성능을 보였다.

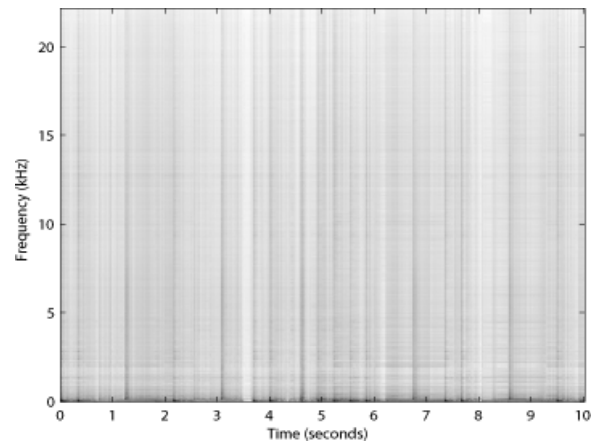
## 5. 결 론

이 논문에서는 드럼 음원 분리를 위하여, 대상 신호 행렬의 분해에 있어서 분해 결과 행렬을 부분적으로 사전 정보 행렬의 분해 결과 행렬과 공유하는 비음수 행렬의 공동 분해 모델을 제안하였다. 간단한 곱셈의 업데이트에 의한 학습 알고리즘이 유도되었으며, 알고리즘의 수행 결과 드럼 음원의 특징을 자동적으로 공유된 행렬에 모아 드럼 음원 분리를 수행할 수

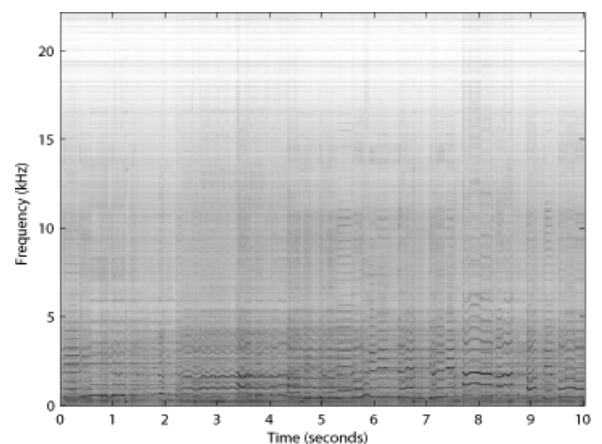
있었다. 실제 음악 신호에 대한 실험에서 제안된 방법이 기존의 비음수 행렬 분해에 지지 행렬 벡터를 사용한 방법에 비해 좋은 분리 성능을 나타냄을 확인하였다.



(a)



(b)



(c)

그림 2. 비음수 행렬의 부분적 공동 분해를 사용한 드럼음원 분리 결과 스펙트로그램 (a) 드럼과 화성 악기가 섞여 있는 원본 신호 (b) 분리되어 나온 드럼 신호 (c) 분리되어 나온 화성 악기 신호

## Acknowledgments

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2009년도 문화콘텐츠산업기술지원사업 및 WCU 프로그램 (Project No. R31-2008-000-10100-0)의 연구결과로 수행되었음.

## 참고 문헌

- [1] N. Bertin, R. Badeau, and G. Richard, "Blind signal decompositions for automatic transcription of polyphonic music: NMF and K-SVD on the benchmark," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, Hawaii, 2007.
- [2] C. Dittmar and C. Uhle, "Further steps towards drum transcription of polyphonic music," in *Proceedings of the Audio Engineering Society Convention*, Berlin, Germany, 2004.
- [3] D. FitzGerald, B. Lawlor, and E. Coyle, "Prior subspace analysis for drum transcription," in *Proceedings of the Audio Engineering Society Convention*, Amsterdam, The Netherlands, 2003.
- [4] O. Gillet and G. Richard, "Transcription and separation of drum signals from polyphonic music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, pp. 529-540, 2008.
- [5] M. Helen and T. Virtanen, "Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine," in *Proceedings of EUSIPCO*, 2005.
- [6] M. Kim and S. Choi, "Monaural music source separation: Nonnegativity, sparseness, and shift-invariance," in *Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation (ICA)*, Charleston, South Carolina, 2006.
- [7] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 13. MIT Press, 2001.
- [8] H. Lee and S. Choi, "Group nonnegative matrix factorization for EEG classification," in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, Clearwater Beach, Florida, 2009.
- [9] A. P. Singh and G. J. Gordon, "Relational learning via collective matrix factorization," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, Las Vegas, Nevada, 2008.
- [10] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 2003, pp. 177-180.
- [11] C. Uhle, C. Dittmar, and T. Sporer, "Extraction of drum tracks from polyphonic music using independent subspace analysis," in *Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation (ICA)*, Nara, Japan, 2003.
- [12] E. Vincent, N. Berlin, and R. Badeau, "Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, NV, 2008.
- [13] T. O. Virtanen, "Monaural sound source separation by perceptually weighted non-negative matrix factorization," Tampere University of Technology, Tech. Rep., 2007.
- [14] J. Yoo and S. Choi, "Weighted nonnegative matrix co-trifactorization for collaborative prediction," in *Proceedings of the 1st Asian Conference on Machine Learning (ACML)*, Nanjing, China, 2009.
- [15] K. Yu, S. Yu, and V. Tresp, "Multi-label informed latent semantic indexing," in *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, Salvador, Brazil, 2005.
- [16] S. Zhu, K. Yu, Y. Chi, and Y. Gong, "Combining content and link for classification using matrix factorization," in *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, Amsterdam, The Netherlands, 2007.
- [17] A. Zils, F. Pachet, O. Delerue, and F. Gouyon, "Automatic extraction of drum tracks from polyphonic music signals," in *Proceedings of the 2nd International Conference on Web Delivering of Music*, Darmstadt, Germany, 2002.