한국음향학회지 제29권 제2호 pp. 102~110 (2010)

비음수 행렬 분해 (NMF)를 이용한 악보 전사

Music Transcription Using Non-Negative Matrix Factorization

박 상 하*, 이 석 진*, 성 굉 모*
(Sang Ha Park*, Seokjin Lee*, Koeng-Mo Sung*)

*서울대학교 전기 · 컴퓨터공학부 뉴미디어통신 공동연구소 음향공학 연구실
(접수일자: 2009년 12월 8일; 채택일자: 2010년 2월 11일)

악보 전사란, 오디오 파일로부터 음고 (음표의 높낮이)와 리듬 (음표의 길이) 정보를 추출하여 악보를 만드는 것이다. 본 논문에서는 음원 분리 및 데이터 분류에 자주 사용되는 Non-Negative Matrix Factorization (NMF)와 Non-Negative Sparse Coding (NNSC) 방식을 사용하여 오디오 파일을 주파수와 리듬 성분으로 분류하였다. 또한 배음 통합 (subharmonic summation) 방법으로 분류된 주파수들로부터 기본 진동 주파수를 계산하였고, 이로써 악보를 이루는 음표의 높낮이를 정확히 얻을 수 있었다. 제안한 방식으로 악보 전사가 성공적으로 이루어졌고, NMF 혹은 NNSC만 사용하여 악보 전사를 하였던 기존의 논문들에 비해 향상된 결과를 얻을 수 있었다.
핵심용어: 비음수 행렬 분해, 악보 전사, 배음 통합 방법
투고분야: 음악음향 및 음향심리 분야 (8)

Music transcription is extracting pitch (the height of a musical note) and rhythm (the length of a musical note) information from audio file and making a music score. In this paper, we decomposed a waveform into frequency and rhythm components using Non-Negative Matrix Factorization (NMF) and Non-Negative Sparse Coding (NNSC) which are often used for source separation and data clustering. And using the subharmonic summation method, fundamental frequency is calculated from the decomposed frequency components. Therefore, the accurate pitch of each score can be estimated. The proposed method successfully performed music transcription with its results superior to those of the conventional methods which used either NMF or NNSC.
Keywords: Non Negative Factorization, Music Transcription, Subharmonic Summation
ASK subject classification: Musical Acoustics and Psychoacoustics (8)

I. 서 론

악보 전사란 음악 파일로부터 음고와 해당 음고의 길이를 추출하는 것을 말하며, 연주자가 음악을 연주하기 위해 필요한 악보를 만들어 내는 작업이라고 할 수 있다. 악보 전사는 음악 분석 및 음원 분리 등의 응용 분야를 위한 기본적이면서도 핵심적인 작업이지만, 이는 오랫동안 풀리지 않는 문제였다. 하지만 Goto가 2000년에 멜로디 부분을 성공적으로 전사한 이후 [1], 많은 연구자들이 악보 전사 분야에 관심을 가지고 본격적으로 연구하기 시작했다. 이 문제를 해결하기 위하여 현재까지도 많은 방식들이 제안되고 있으며, 그 중 대부분은 음악 신호를 분석하여 많은 정보들을 추출함으로써 음표의 높이와 길이를 예측, 계산해 내는 방식이다 [2]. 하지만 음악 신호를 분석 및 훈련의 과정을 거쳐 악보를 전사하는 방법은 매우 복잡할 뿐 아니라 많은 시간과 계산량을 필요로 한다. 이에 따라 음악 구조적인 분석을 배제하고, 보다 간단하면서도 효과적인 방식인 비음수 행렬 분해 (NMF)를 이용한 악보 전사 방식이 제안되었다 [3].

비음수 행렬 분해 (Non-negative Matrix Factorization, NMF)는 뇌 과학 분야에서 주로 사용되고 있는 방식으로, '사람의 뇌는 사물의 의미 있는 부분적 특징들을 결합함으로써 전체 사물을 인지한다'라는 사람의 인지 과정에 기본 개념을 두고 있다 [4]. 즉, 하나의 사물은 그 사물을 구성하는 기본적인 특징들 (부분적 특징들, basis vector)과 그에 부합하는 가중치 (weight vector)로 나타낼 수 있다는 것이다. 양수의 값들만으로 이루어진

책임저자: 박 상 하 (bosetom@acoustics.snu.ac.kr)
151-742 서울특별시 관악구 신림동 서울대학교 공과대학
뉴미디어통신공동연구소 132동 302호
(전화: 02-880-8427; 팩스:02-880-8207)

행렬을 양수의 값을 갖는 두 개의 기저 행렬 (basis and weight)로 나누는 비음수 행렬 분해 (NMF)는 Lee와 Seung에 의해 제안된 후, 뇌 과학 및 이미지 분야에서 이미지 분리와 데이터 분류 등에서 많이 사용되었고 음향 분야에서도 음원 분리 및 악보 전사에 사용되기 시작했다.

비음수 행렬 분해 (NMF)를 이용한 악보 전사 방식은 기존의 방식보다 간편하고 계산이 빠르다는 장점이 있다. 하지만 기존의 논문들은 음악 신호를 나타내는 행렬을 주파수 벡터들과 리듬 벡터로 분리하는 데에 그쳤고, 주파수 벡터들로부터 음고를 정확히 추출하지는 못하였다. 이에 따라 본 논문에서는 비음수 행렬 분해 (NMF)로 음악 신호를 분리한 후, 배음 통합 (subharmonic summation) 방식을 이용하여 분리된 주파수 벡터들로부터 기본 진동 주파수를 계산하였다. 비음수 행렬 분해 (NMF)와 배음 통합 방식을 결합한 새로운 방식으로 악보 전사를 수행한 결과, 기존의 비음수 행렬분해 (NMF)만을 사용하였을 때보다 음고 탐색 부분이 많이 보완되어 악보 전사의 성능이 향상되었다.

II. 악보 전사 방법

2.1. Non-Negative Matrix Factorization

비음수 행렬 분해 (Non-Negative Matrix Factorization, NMF)는 양수의 값들로 이루어진 n×m 행렬 V를 두 개의 저차 행렬의 곱으로 분해하는 방식이다. 이를 수식으로 나타내면 다음과 같다.

$$V_{i\mu} \approx (WH)_{i\mu} = \sum_{a=1}^{r} W_{ia} H_{a\mu} \qquad (1)$$

식 (1)에서 W는 기저 행렬이며, H는 가중 행렬이다. 행렬 W와 H의 크기는 각각 n×r, r×m이며, 행렬 W와 H 모두 양수의 값만을 가지는 행렬이다. 차수 r의 크기는 항상 (n+m)r < nm을 만족해야 하며, 행렬 W와 H의 곱으로 이루어진 행렬 WH는 V의 축소된 형태라고 볼 수 있다 [4].

식 (1)을 만족시키는 W와 H값을 찾기 위한 목적 함수로 행렬 W와 H의 Euclidean distance의 제곱을 최소화 하는 방식과 Kullback-Leibler divergence값을 최소화 하는 방식의 두 가지 방식이 제안되었다 [5]. 본 논문에서는 식 (2)와 같이, 간단하면서도 계산량이 더 적은 방식인 Euclidean distance를 최소화 하는 방식을 사용하였다.

$$\| V - WH \|^2 = \sum_{ij} (V_{ij} - (WH)_{ij})^2 \qquad (2)$$

식 (2)의 목적함수가 최소값으로 수렴할 때까지 행렬 W와 H는 계속 갱신되어야 한다. 이러한 W와 H의 update rule은 식 (3)과 같다.

$$H_{a\mu} \leftarrow H_{a\mu} \frac{(W^T V)_{a\mu}}{(W^T WH)_{a\mu}}$$
$$W_{ia} \leftarrow W_{ia} \frac{(VH^T)_{ia}}{(WHH^T)_{ia}} \qquad (3)$$

식 (3)의 update rule을 사용하면 W와 H값이 갱신되면서도 Euclidean distance $\| V - WH \|$ 값은 증가되지 않고, Euclidean distance가 최소값에 도달할 때까지 W와 H값이 갱신된다. 반복된 갱신 (iteration update)을 통해 행렬 W와 H이 계산되고, 행렬 V는 두 개의 행렬로 분리된다.

2.2. Non-Negative Sparse Coding

앞의 NMF을 이용하여 행렬 V를 두 개의 저차 행렬의 곱으로 나타내기 위해서는 W와 H의 차수 r을 최적의 값으로 지정해 주어야 한다. 하지만 H가 희박한 분포 (sparse distribution)를 갖는 경우에는 적은 개수의 가중 행렬로 입력 행렬 V를 나타낼 수 있다. 이러한 방식의 행렬 분해를 Non-Negative Sparse Coding (NNSC)이라 하며, NNSC의 목적함수는 식 (4)와 같이 NMF의 목적함수 (식 (2))에 희박 정도를 나타내는 부분이 추가된 형태이다 [6].

$$V_{i\mu} = \frac{1}{2} \| V - WH \|^2 + \lambda \sum_{ij} f(H_{ij}) \qquad (4)$$

식 (4)에서 λ가 포함된 항이 희박 정도 (sparseness)와 정확도 (accuracy) 간의 정도를 조절하는 부분이다. 이와 같은 NNSC의 수식을 사용하면, W와 H의 차수 r이 실제보다 높게 지정되었을 경우, 필요한 차수 이상의 부분들은 낮은 에너지의 잡음 (noise)신호로 분리되어 차수 r의 범위를 알기 힘든 경우에 유용하다는 장점이 있다. 하지만 희박 정도 (sparseness) 변수 λ의 적당한 값을 미리 지정해야 한다는 단점이 있다.

식 (4)의 목적 함수가 최소값으로 수렴하도록 하는 iteration update rule은 식 (5)와 같다.

$$H \leftarrow H \frac{(W^T V)}{(W^T WH + \lambda)}$$
$$W \leftarrow W - \mu (WH - V)(H)^t \tag{5}$$

식 (5)에서 μ는 기울기 하강 (gradient descent)의 정도를 나타내는 단계 크기 (step size)이며, 작은 양수의 값인 μ가 $\mu > 0$을 만족한다면 반복 갱신 (iteration update)을 통해 목적함수는 점점 감소하게 된다. 이와 같은 iteration update rule을 통해 행렬 V는 NNSC의 방식으로 행렬 H와 W로 나눠지게 된다.

2.3. 비음수 행렬 분해 (NMF)를 이용한 악보 전사

음악 신호 $x(t)$는 식 (6)과 같이 여러 개의 정현파 (sinusoidal wave)들로 이루어져 있다.

$$x(t) = \sum_{r=1}^{R} A_r(t) cos(\omega_r t + \theta_r(t)) \tag{6}$$

식 (6)에서 $A_r(t)$는 순간진폭 (instantaneous amplitude)이고, ω_r는 각속도 (angular frequency)이며, $\theta_r(t)$는 순간위상 (instantaneous phase)이다. $x(t)$를 L의 시간 구간만큼 나눈 후, Short-time Fourier Transform (STFT)를 수행하면 진폭 스펙트럼 (amplitude spectrum)과 위상 스펙트럼 (phase spectrum)으로 나눠진다. 이 때 진폭 스펙트럼은 시간에 따른 주파수와 그 크기들로 이루어지게 된다. 비음수 행렬 분해 (NMF)나 NNSC을 이용하여 진폭 스펙트럼을 두 개의 행렬로 분해하면 주파수 행렬과 해당 주파수의 길이 (리듬) 행렬로 나누어진다. 이후 주파수 행렬로부터 보다 정확한 음고 (음표의 높낮이)를 계산하고자 배음 통합 (subharmonic summation) 방식 [7]을 적용시키면 음고 벡터와 리듬 벡터가 추출되게 된다. 본 논문에서 제안한 NMF와 배음 통합 방식을 이용한 악보 전사 방법을 블록 다이어그램으로 나타낸 것이 다음의 그림 1이다.

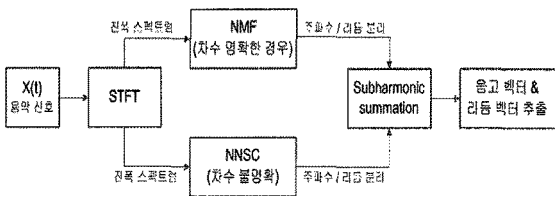

그림 1. 제안하는 방식의 블록 다이어그램
Fig. 1. Block diagram of the proposed method.

다음의 그림은 소리 신호 (sound signal)에 STFT을 취한 후의 진폭 스펙트럼이다.

그림 2에서 보는 바와 같이 총 1.8초의 시간의 음악 파일의 0~0.5초와 1~1.5초에 300 Hz의 주파수가 존재하고, 0.5~1.2초와 1.5~1.8초에 440 Hz의 주파수가 존재한다. 위와 같은 소리 신호의 진폭 스펙트럼을 NMF를 이용해 두 개의 행렬로 분류하면(차수 r=2), 시간 성분과 주파수 성분으로 분류되게 되고 이는 그림 3, 4과 같다.

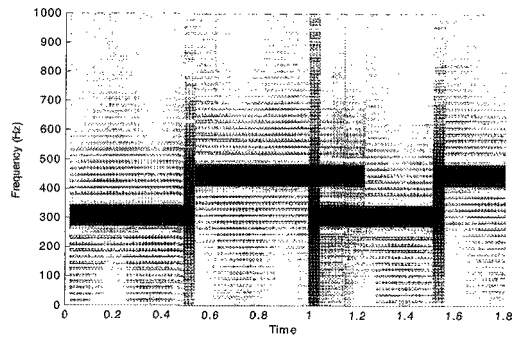

그림 2. 소리 신호의 진폭 스펙트럼 1
Fig. 2. Magnitude spectrum of sound signal 1.

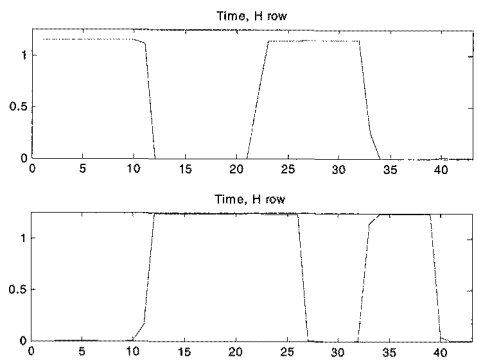

그림 3. 행렬 H의 열(row), 시간 성분
Fig. 3. The rows of matrix H, time component.

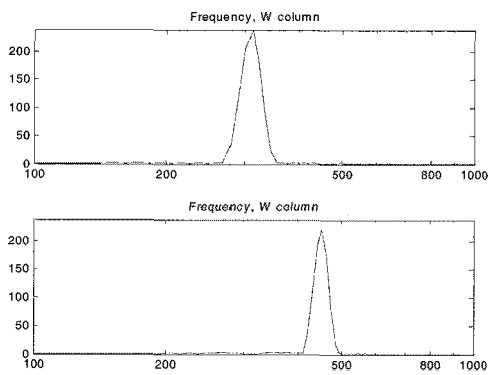

그림 4. 행렬 W의 행(column), 주파수 성분
Fig. 4. The columns of matrix W, frequency component.

위의 그림에서 볼 수 있듯이 H행렬의 열 (row)은 기저 (basis)들의 시간에 따른 위치를 나타내고, W행렬의 행 (column)은 주파수 도메인에서의 특징, 즉 음표의 높이를 나타낸다.

이와 같은 방법으로 비음수 행렬 분해 (NMF)를 이용하여 음악 파일을 주파수 성분들과 시간 성분들로 분류할 수 있다. 하지만 실제 음악은 앞에서 예로 든 바와 같이 단순히 두 개의 주파수만으로 이루어 있지 않으며, 수많은 배음들이 추가로 존재한다. 이 경우 비음수 행렬 분해 (NMF)를 이용하여 행렬을 분해할 때 존재하는 음표의 개수만큼의 r 값을 지정해 주지 못한다면, 행렬 분해가 정확히 이루어지지 않게 된다. 그 예로 다음의 그림 5에서 보는 바와 같이, 그림 2의 신호에 0.5~1초에 880 Hz가 추가로 존재하는 경우를 들 수 있다.

위의 소리 신호를 비음수 행렬 분해 (NMF)를 이용하여 차수 r을 2와 3으로 놓고 분리한 결과는 각각 다음의 그림 6, 7과 같다.

그림 5의 소리 신호를 비음수 행렬 분해 (NMF)를 이용한 행렬 분해 결과, r=2인 경우 880 Hz의 신호를 440 Hz의 신호의 배음으로 분류하였고, r=3인 경우 300 Hz, 440 Hz, 880 Hz의 신호를 각각 다른 성분으로 분류하였다. 따라서 비음수 행렬 분해 (NMF)를 이용하여 악보 전사를 할 때, 음표의 개수가 몇 개인지 지정하는 일은 매우 중요하다고 할 수 있다. 하지만 길이가 긴 음악의 경우 정확한 r의 값을 찾아내기란 쉽지 않으며, 이러한 단점을 보완하기 위해 NNSC를 사용하는 것이 더 유용하다. NNSC를 이용하면 r의 값을 실제 필요한 값보다 약간 크게 지정하여도 목적 함수에 포함되어 있는 희박 정도 (sparseness)를 조절하는 부분에 의해 적당한 r개의 음표로 분류되는 결과를 얻을 수 있다. 실제 음악에서의 NMF와 NNSC를 이용한 음악 분류의 결과는 다음 장에서 비교하도록 하겠다.

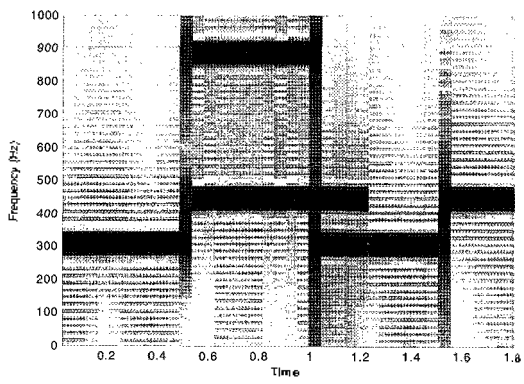

그림 5. 소리 신호의 진폭 스펙트럼 2
Fig. 5. Magnitude spectrum of sound signal 2.

이로써 정확한 r값을 알 수 없는 경우에도 NNSC를 이용함으로써 음악 분류가 가능하게 되었다. 하지만 분류된 결과를 이용하여 음고 (음표의 높낮이)와 해당 음고의 길이 (리듬)를 추출하는 데에 한 가지 문제가 더 존재한다. 이는 그림 6의 오른쪽 하단의 그림과 그림 7의 오른쪽 상단에서 보는 바와 같이 W의 하나의 column에 여러 주파수 값들이 존재하는 경우이다. 위의 경우에는 두 개의 피크값 중 하나의 값이 우월하게 크기 때문에 그 값을 음고로 간주할 수 있다. 하지만 실제 음악에는 배음들이 많이 존재하기 때문에 행렬 W의 하나의 column에 수많은 피크값들이 나타나게 된다. 이 때 다른 음고의 배음과의 섞임 현상으로 인해 해당 음고의 기본 진동 주파수의

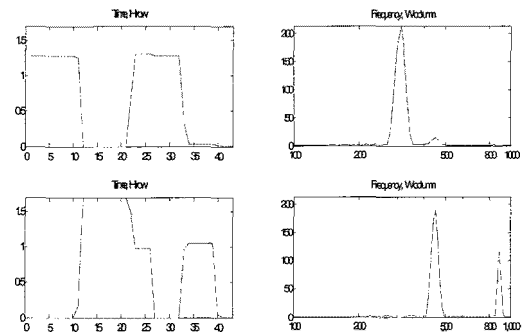

그림 6. NMF를 이용한 그림 5의 행렬 분해
(좌: 행렬 H의 열, 우: 행렬 W의 행, r=2)
Fig. 6. Matrix decomposition of pictured in fig.5 using NMF. (left: rows of matrix H, right: columns of matrix W, r=2)

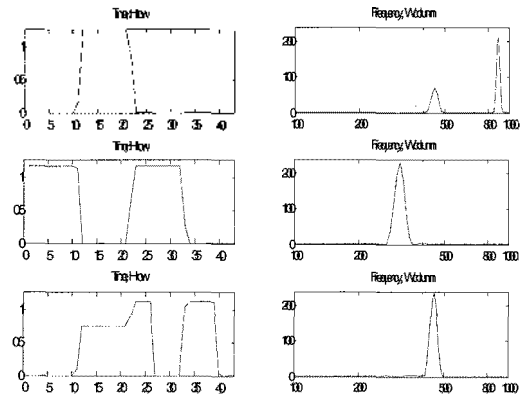

그림 7. NMF를 이용한 그림 5의 행렬 분해
(좌: 행렬 H의 열, 우: 행렬 W의 행, r=3)
Fig. 7. Matrix decomposition of pictured in fig.5 using NMF. (left: rows of matrix H, right: columns of matrix W, r=3)

크기 값이 그 배음의 크기 값보다 작은 경우가 발생하게 되고, 음고값이 정확히 추출되지 않을 수 있다는 문제가 존재한다. 이를 막기 위하여 본 논문에서는 subharmonic summation을 이용하여 해당 음고의 기본 진동 주파수와 배음값들을 묶어 하나의 음고값으로 나타내도록 함으로써 음고 추출에서의 정확도를 높이고자 하였다.

2.4. 배음 통합 (Subharmonic Summation) 방식

앞 절에서 언급하였듯이 주파수를 나타내는 행렬 W의 column 값들 중에서 음고 값을 찾아야 하지만 행렬 W의 column의 최대값들이 기본 진동 주파수 값을 보장하지는 않는다. 따라서 본 논문에서는 주파수 값들 중에서 기본 진동주파수를 찾고자 배음 통합 (subharmonic summation) 방식을 이용하였다 [7].

배음 통합 방식은 여러 개의 배음 성분들을 통합하여 음고 성분을 계산하는 방법이다. 짧은 구간의 소리 신호를 FFT한 후, 음고 결정에 사용하는 주파수 범위를 1250 Hz 이하의 값들로 제한한다. 1250 Hz 이하의 주파수들 중 피크값들을 찾은 후, 피크값으로부터 대략 20 Hz 범위 내에 위치한 주파수 값들은 0으로 놓는다. 이후 해닝 필터 (Hanning filter)를 이용하여 다음의 수식 (7)과 같이 스펙트럼을 스무딩 (smoothing)한다.

$$B_k(n) = \frac{1}{4}A_{k-1}(n) + \frac{1}{2}A_k(n) + \frac{1}{4}A_{k+1}(n) \quad (7)$$

식 (7)에서 $A_k(n)$와 $B_k(n)$는 각각 스무딩 전, 후의 스펙트럼이고, n은 시간 인덱스이며, k는 주파수 번호를 의미한다. 피크값들이 강조된 $B_k(n)$에 식 (8)과 같이 arc tangent 함수 $W_k(n)$를 곱해 줌으로써 사람의 귀가 주파수에 따라 다르게 인지하는 정도를 보완하였다.

$$P_k(n) = W_k(n)B_k(n) \quad (8)$$

또한, 높은 배음들은 음고 결정에 적은 영향을 주기 때문에 저차 배음들에 해당하는 가중치인 h_i을 식 (8)에서 계산된 값에 곱한다. 이 때 i는 배음의 개수이며, 실험적으로 h_i은 0.84^{i-1}가 적당한 값임이 밝혀져 있다.

모든 주파수에 대해 각각의 가중치 h_i가 곱해진 후, 모든 주파수 값들은 더해져서 식 (9)의 형태를 이루게 된다.

$$H_k(n) = \sum_{i=1}^{N} h_i P_i(n) \quad (9)$$

식 (9)의 함수가 subharmonic sum spectra이며, 모든 배음들이 더해진 값을 나타낸다. 식 (9)의 H의 최대값을 찾음으로써 음고값을 계산해 내는 것이 배음 통합 방식의 핵심이다.

다음 절에서는 NMF 혹은 NNSC로 행렬 분해 후, 배음 통합 방식의 적용 전과 후의 계산된 음고 값을 비교해 보았다. 그 결과, 특히 첼로와 같은 현악기에서 배음 통합 방식을 적용하였을 때 음고 탐색 성능이 월등히 좋아짐을 확인할 수 있었다.

III. 실험 결과

3.1. 피아노 단성 악보 전사

우선 피아노 곡의 한 마디를 NMF와 NNSC를 이용하여 분류해 보았다. 다음의 악보는 '바흐 평균율 클라비어곡집 1권 제6번 d단조'의 앞부분 한 마디이다.

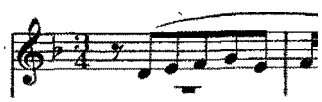

그림 8. 피아노 악보 - 단선율 악보
Fig. 8. Piano score - monophony.

다음의 그림 9는 위의 음악 파일을 STFT한 후의 진폭 스펙트로그램이다. STFT시 샘플링 주파수는 44100 Hz 이고, 2048 샘플의 해닝 윈도우를 사용하였다.

위의 음원을 NMF를 이용해 주파수와 시간의 행렬로 분리한 결과는 다음의 그림 10, 11과 같다. 이 때 그림

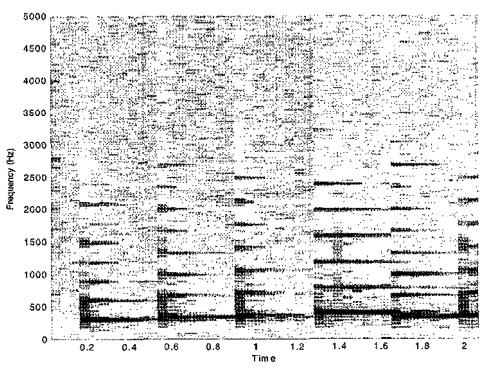

그림 9. 단선율 피아노의 진폭 스펙트럼
Fig. 9. Magnitude spectrum of monophony piano sound.

8에서 보는 바와 같이, 본 음원에서 사용된 음표는 'D4', 'E4', 'F4', 'G4'의 4개의 음이기에 r은 4로 놓았다.

그림 10의 아래 부분 (행렬 W부분)의 동그라미 부분은 행렬 W의 행들 중에서 최고값을 나타낸 것으로, 음고를 나타낸다고 볼 수 있다. 행렬 H의 4개의 열들은 각 음고들이 나타난 시간을 의미하기에 각각 위로부터 'D4', 'G4', 'E4', 'F4'음을 나타내는 것을 확인할 수 있으며, 이는 행렬 W로부터도 확인할 수 있다. 행렬 W로부터 찾아낸 음고값은 301 Hz, 398 Hz, 334 Hz, 355 Hz이며, 행렬 W의 각 행의 값들 중 최대값을 찾은 후 배음 통합 방식을 사용하지 않더라도 음고 값을 찾는데 무리가 없었다. 또한 차수 r의 크기를 음원에서 사용된 음표의 개수와 같도록 하였기에 NMF와 NNSC의 결과에 큰 차이가 없었다.

하지만 r을 5로 놓았을 경우, NMF를 사용하여 음원을 행렬 분해 후 행렬 W로부터 찾아낸 음고값은 301 Hz, 355 Hz, 398 Hz, 334 Hz, 398 Hz이다. 이 경우 398 Hz의 음고를 두 번 찾아낸 것을 확인할 수 있고, 이 때의 행렬 H의 열과 행렬 W의 행을 그림으로 나타내면 다음의 그림 11과 같다.

위의 그림 11와 그림 10을 비교하면, 그림 11에서 H의 마지막 열과 W의 마지막 행이 r을 잘못 지정함에 따라 추가로 더 나온 것이라는 것을 확인할 수 있다. 하지만 음원 파일이 긴 경우에는 음원 파일에 해당하는 음표를 세어 r을 지정한다는 것은 무리라고 볼 수 있다. 이 경우에는 NNSC를 사용하는 것이 편리한데, 그림 12는 그림 9의 음원 파일을 r=5로 놓고 NNSC의 방법으로 행렬 분해한 결과이다.

NMF에서의 결과와는 다르게 행렬 H와 W의 두 번째 결과에서 noise처럼 분리되었음을 금방 확인할 수 있다. 이로부터 정확한 음표의 개수를 알 수 없는 경우는 NMF보다 NNSC를 이용하는 것이 보다 정확한 결과를 도출하는데 유용하다고 볼 수 있다.

또한 NNSC를 이용하여 행렬을 분리하였을 때, 행렬 W의 피크값은 301 Hz, 215 Hz, 366 Hz, 398 Hz, 334

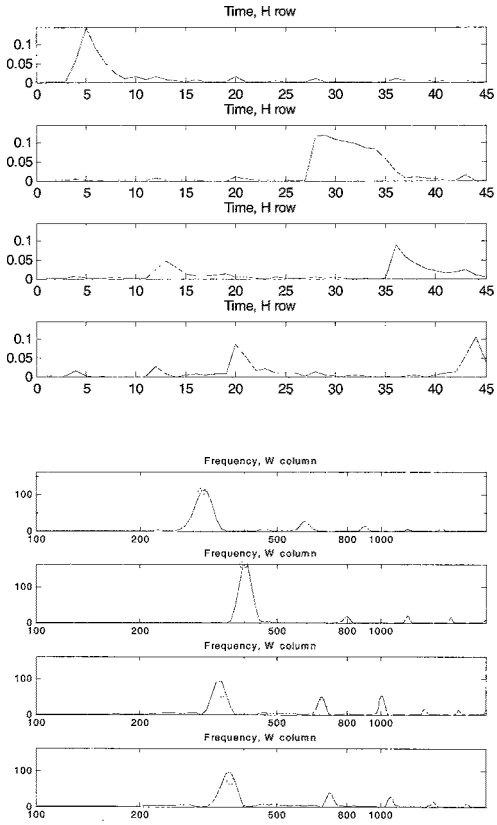

그림 10. NMF를 이용한 피아노 음원의 행렬 분해
(위: 행렬 H의 열, 아래: 행렬 W의 행, r=4)

Fig. 10. Matrix decomposition of piano sound using NMF. (upper: rows of matrix H, lower: columns of matrix W, r=4)

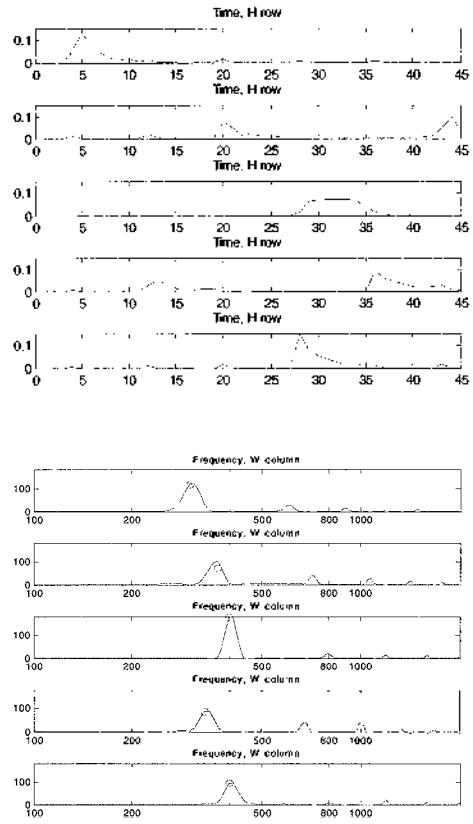

그림 11. NMF를 이용한 피아노 음원의 행렬 분해
(위: 행렬 H의 열, 아래: 행렬 W의 행, r=5)

Fig. 11. Matrix decomposition of piano sound using NMF. (upper: rows of matrix H, lower: columns of matrix W, r=5)

Hz이다. 하지만 이 경우 배음 통합 방식으로 찾아낸 결과는 301 Hz, 96 Hz, 355 Hz, 398 Hz, 334 Hz이다. 특히 행렬 W의 두 번째 값은 배음 통합 방식을 사용하지 않았을 때 (215 Hz)와 사용하였을 때 (96 Hz) 큰 차이를 나타냈으며, 배음 통합 방식을 사용하였을 때 보다 유의미한 결과가 나왔다. 이 경우 배음 통합 방식을 사용하여 찾아낸 음고의 결과 값만 보고도 제외되어야 하는 음고를 명확히 알 수 있기에 악보 전사에 매우 유용하다고 볼 수 있다.

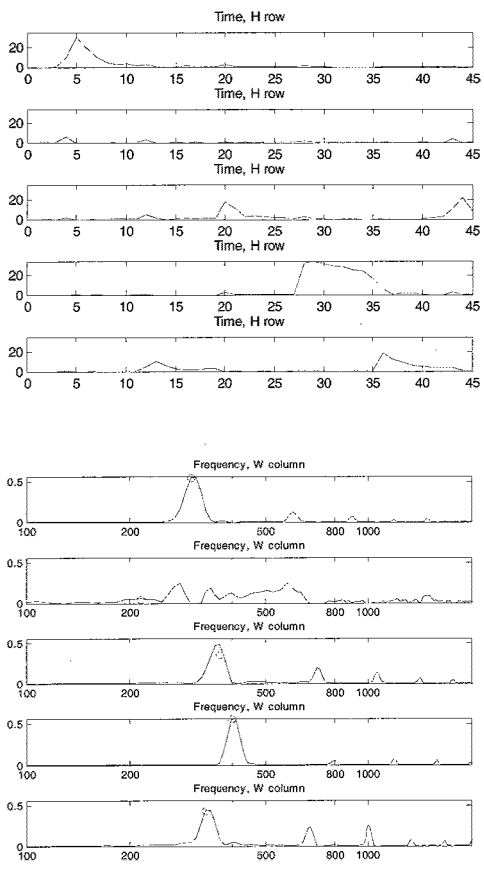

그림 12. NNSC를 이용한 피아노 음원의 행렬 분해
(위: 행렬 H의 열, 아래: 행렬 W의 행, r=5)
Fig. 12. Matrix decomposition of piano sound using NNSC.
(upper: rows of matrix H, lower: columns of matrix W, r=5)

3.2. 첼로 단성 악보 전사

다음은 첼로의 단선율 음원으로 NMF와 NNSC, 그리고 배음 통합 방식을 이용하여 음고와 리듬을 찾아보았다. 분석에 사용된 음원은 '바흐의 무반주 첼로 모음곡 1번 전주곡'의 앞부분 두 마디이다. 다음의 그림 13은 이 음원의 진폭 스펙트로그램이다.

위의 그림 13에서 보는 바와 같이, 피아노의 경우와 달리 첼로의 경우 저주파에 에너지 많은 편이다. 이는 첼로의 음역이 피아노의 음역에 비해 상대적으로 낮기 때문이다. 그림 14는 단선율 첼로 음원을 r=4로 놓고 NMF의 방법으로 행렬 분해한 결과이다.

첼로 음원에서 사용된 음은 'G2', 'D3', 'A3', 'B3'이며, 행렬 W로부터 찾아낸 최대값, 즉 NMF만 사용하고 배음

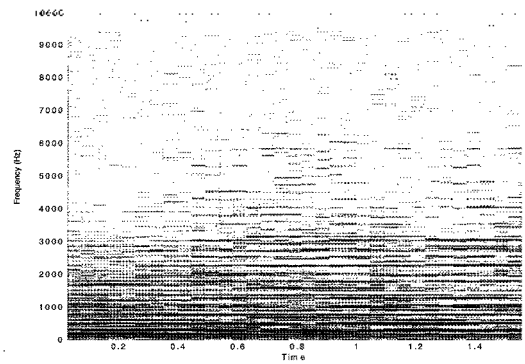

그림 13. 단선율 첼로의 진폭 스펙트럼
Fig. 13. Magnitude spectrum of monophony cello sound.

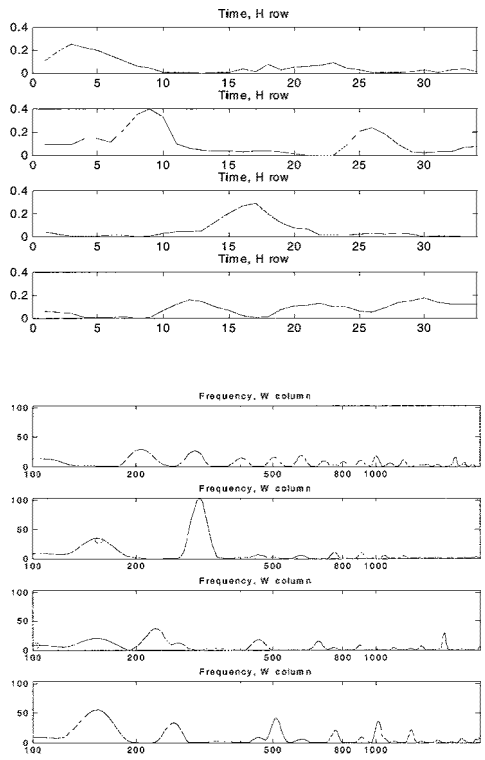

그림 14. NMF를 이용한 첼로 음원의 행렬 분해
(위: 행렬 H의 열, 아래: 행렬 W의 행, r=4)
Fig. 14. Matrix decomposition of cello sound using NNSC.
(upper: rows of matrix H, lower: columns of matrix W, r=4)

통합 방식을 적용시키기 전의 결과값은 65 Hz, 156 Hz, 102 Hz, 59 Hz이다. 65 Hz, 59 Hz는 실제값보다 낮은 값들로, 첼로 음원의 경우에서는 더더욱 NMF만을 이용해서는 정확한 음고값을 찾지 못하였다.

이 값들에 배음 통합 방식을 적용시키면 102 Hz, 156 Hz, 226 Hz, 253 Hz로 각각 'G2', 'D3', 'A3', 'B3'에 대응되는 음고를 정확히 찾았다고 볼 수 있다. 첼로 음원의 경우 피아노의 경우에서와는 달리 배음 통합 방식의 적용 전과 후에 차이가 많이 나타났다. 이는 음원이 저주파에 많은 에너지를 갖고 있기 때문에, 배음 통합 방식을 적용하지 않으면 사람 귀의 인지 정도나 배음에 대한 가중치 함수가 적용되지 않아 실제보다 낮은 주파수를 음고로 찾아내는 경향이 있었다. 또한 첼로의 경우 피아노와는 달리 순정율의 음계를 가지고 있어서 'G2'음과 'D3'음, 'D3'음과 'A3'음의 배음 성분이 겹치게 된다. 따라서 행렬 분해시 다른 음의 배음 성분을 자신의 배음 성분으로 가져오는 경향이 있어 NMF만을 이용해서는 음고 추출에 어려움이 존재하며 이 경우 배음 통합 방식을 적용해야 올바른 음고를 찾아낼 수 있다.

첼로 단성 악보 전사시, 앞의 피아노를 이용한 경우에서와 마찬가지로 차수 r이 사용된 음표의 개수와 같은 경우에는 NMF와 NNSC사이의 성능 차이가 그다지 존재하지 않았으며, 차수 r을 정확히 알 수 없는 경우에는 NNSC가 더 좋은 성능을 나타냈다.

3.3. 피아노 다성 악보 전사

우선 피아노 곡의 한 마디를 NMF를 이용하여 분류해 보았다. 다음의 악보는 '바흐 평균율 클라비어곡집 1권 제6번 d단조'의 앞부분 네 마디이다.

그림 15. 피아노 악보 - 다성 악보
Fig. 15. Piano score - polyphony.

다음의 표 1에서는 위의 음악 파일에서 실제 사용된 음표와 위의 음악 파일을 SIFT한 이후 기존의 방식 (NMF만 사용)과 제안된 방식 (NMF에 배음 통합 방식 결합)으로부터 계산된 시간 행렬, 주파수 행렬 결과를 비교해서 결과값을 대응시켜 보았다.

악보에서 총 14개 음표가 반복 사용되었고, 기존의 방식으로는 8음에 대해서만 정확한 값을 찾아낸 반면, 제안한 방식으로는 12음에 대해서 시간과 주파수 상에서 정확한 값을 찾아내었다. 다만 제안한 방식으로도 표의 2번째 값인 G#3음에 대해서는 한 옥타브 위의 주파수 값을 찾아냈고, 표의 11번째 값인 A3음에 대해서는 두 옥타브 위의 주파수 값을 찾아내었다. 이 두 음 모두 시간 벡터에서도 최대값이 작아 다소 모호한 값이 출력되었다.

표 1. 피아노 다성 악보 전사 결과
Table. 1. Polyphonic piano music transcription result.

번호	사용된 음표	기존(NMF)		제안한 방식	
		주파수 결과	정확도	주파수 결과	정확도
1	A4	447Hz	정확	447Hz	정확
2	(G#3)	393Hz	부정확	393Hz	부정확
3	C5	221Hz	부정확	533Hz	정확
4	F4	361Hz	정확	355Hz	정확
5	B♭4	312Hz	부정확	474Hz	정확
6	G4	398Hz	정확	398Hz	정확
7	G#4	291Hz	부정확	420Hz	정확
8	B♮4	501Hz	정확	501Hz	정확
9	C♯4	280Hz	정확	285Hz	정확
10	B3	248Hz	정확	248Hz	정확
11	(A3)	404Hz	부정확	888Hz	부정확
12	C4	269Hz	정확	269Hz	정확
13	E4	334Hz	정확	334Hz	정확
14	D4	269Hz	부정확	301Hz	정확

총 14개 음표가 반복 사용되며 멜로디에 35개의 음표, 베이스에 13개의 음표가 사용된 피아노 다성 악보를 NMF와 배음 통합 방식을 이용하여 악보 전사를 한 결과, 12개의 음표는 시간과 주파수 모두 정확하게 찾아내었고, 2개의 음표에 대해서는 시간에서 모호한 값을, 그리고 주파수에서는 옥타브 위의 값을 찾았다. 이 때 잘못 찾은 두 개의 주파수는 곡에서 사용된 음들 중 가장 저음에 해당하는 음들일 뿐만 아니라, 곡 전체에 걸쳐 각각 한 번과 두 번 정도만 사용된 음들이었다.

이로써 NMF와 배음 통합 방식을 이용해 다성 악보를 전사해 보았고, 총 48개의 사용된 음표 중에서 3개의 음표를 제외한 45개의 음표를 음고와 리듬 면에서 정확히 분리하였다.

V. 결론

본 논문에서는 NMF와 NNSC의 방법을 이용한 악보 전사 방법을 제안하였다. 기존의 논문과는 달리 배음 통합 방식을 추가하여 음고 추출에서의 정확도를 높였다. 제안한 방법으로 단성 피아노, 단성 첼로, 다성 피아노 음원

로 실험해 보았고, 단성 피아노와 단성 첼로 음원에서 음고와 리듬을 정확하게 추출하였다.

단성 피아노 음원의 경우 배음 통합 방식의 추가 유무가 결과에 큰 영향이 미치지 않았으나 단성 첼로의 경우 배음 통합 방식 추가 후 음고 추출에서 정확도가 많이 향상되었다. 이는 음역이 낮은 악기의 경우, 다른 음의 배음 성분을 자신의 배음 성분으로 가져와 분류가 잘 이루어지지 않거나 배음에 해당하는 음들을 음고로 인식하는 등 악보 전사가 상대적으로 잘 되지 않았다. 이 경우 배음 통합 방식을 통해 저음부에 가중치를 준 후 배음 성분을 통합하여 음고를 계산하면 정확도가 향상된 결과를 도출하였다.

다성 피아노 음원의 경우 저음과 사용 빈도가 적은 음을 제외한 음들에서 음고와 리듬을 비교적 정확하게 추출하였다. 다성 음원의 경우 배음 성분이 겹치는 경우가 단성 음악에 비해 많이 존재하기에 악보 전사의 정확도는 상대적으로 떨어지는 경향이 있었다.

사용된 음표의 개수를 정확히 알고 있는 경우 NMF와 NNSC 방법 간에 큰 차이가 존재하지 않았으나, 사용된 음표의 개수가 모호한 경우에는 NNSC가 음원 종류에 상관없이 보다 나은 결과를 도출하였다.

앞으로 STFT이 아닌 wavelet을 이용하여 낮은음 부분의 해상도를 높여 에러를 줄이는 것과 다성 음악 전사에서 추가적인 단서들을 사용하여 악보 전사의 정확도를 높이는 방향으로 추가 연구를 진행할 것이다.

참 고 문 헌

1. M. Goto, "A robust predominant-f0 estimation method for real-time detection of melody and bass lines in CD recordings", *ICASSP*, pp. 757-706, 2000.
2. M.P. Ryynänen and A.P. Klapuri, "Automatic transcription of melody, bass line, and chords in polyphonic music", *Computer Music Journal*, vol.32, no.3, pp. 72-86, 2008.
3. P. Smaragdis and J.C. Brown, "Non-negative matrix factorization for polyphonic music transcription", *in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 177-180, 2003.
4. D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization", *Nature*, vol. 401, no. 6755, pp.788-791, 1999.
5. D.D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization", *in Advances in Neural Information Processing systems*, pp. 556-562, MIT Press, 2001.
6. P.O. Hoyer, "Non-negative sparse coding", in Neural Networks for Signal Processing, *IEEE Workshop*, pp. 557-565, 2002.
7. D. Hermes, "Measurement of pitch by subharmonic summation", *J. Acoust. Soc. Am.*, vol. 83, no. 1, pp. 257-264, 1988.

저자 약력

•박 상 하 (Sang Ha Park)

2005년 2월 : 서울대학교 음악대학 기악과 졸업 (학사)
2008년 2월 : 서울대학교 공과대학 전기·컴퓨터공학부 졸업 (공학석사)
2008년 3월~현재: 서울대학교 공과 대학 전기·컴퓨터공학부 박사 과정

•이 석 진 (Seokjin Lee)

2006년 8월 : 서울대학교 공과대학 전기·컴퓨터공학부 졸업 (학사)
2008년 8월 : 서울대학교 공과대학 전기·컴퓨터공학부 졸업 (공학석사)
2008년 9월~현재: 서울대학교 공과대학 전기·컴퓨터공학부 박사 과정

•성 굉 모 (Koeng-Mo Sung)

1965년~1971년 : 서울대학교 전자 공학과
1971년~1973년 : 독일 아헨공대 Vordiplom
1973년~1977년 : 독일 아헨공대 전자통신공학 Dipl.-Ing.
1977년~1982년 : 독일 아헨공대 음향공학 Dr.-Ing. (공학박사)
1977년~1983년 : 독일 아헨공대 음향공학연구소 연구원
1983년~현재 : 서울대학교 공과대학 전기·컴퓨터공학부 교수