

## 특징행렬 데이터베이스를 이용한 NMF 기반 음악전사

신옥근<sup>†</sup> · 류다현<sup>1</sup>

(원고접수일 : 2012년 11월 5일, 원고수정일 : 2012년 11월 15일, 심사완료일 : 2012년 11월 20일)

### NMF Based Music Transcription Using Feature Vector Database

Ok Keun Shin<sup>†</sup> · Da Hyun Ryu<sup>1</sup>

**요 약 :** NMF를 이용하여 음악을 전사할 때 표적음악에서 특징행렬과 가중치 행렬을 동시에 추출해 내는 방법을 이용하기 위해서는 특징행렬의 크기(특징벡터의 수)를 미리 알아야 할 뿐 아니라, 추출된 각각의 특징벡터들의 음고를 결정하는 어려운 과정을 거쳐야한다. 또 이 방법은 음악에 포함된 음고의 수가 커질수록 특징행렬을 정확하게 추출해 내기 어려워진다는 단점도 있다. 본 연구에서는 이러한 단점을 피하기 위해 특징행렬 데이터베이스를 미리 준비한 다음 이를 실제 음악에 이용하는 방법을 실험한다. 먼저 특징행렬 데이터베이스를 만든 다음, 특징행렬을 추출한 피아노에서 연주된 음악, 그리고 제 3의 피아노에서 연주된 같은 음악을 각각 전사하여 성능을 비교한다. 또 이들 결과와 비교하기 위하여 특징행렬과 가중치행렬을 동시에 추출하는 방법도 실험하여 결과를 비교하였다. 특징행렬 데이터베이스를 이용하는 방법이 특징행렬과 가중치행렬을 동시에 추출하는 방법보다 좋은 성능을 가짐을 확인하였다.

**주제어 :** 음악전사, 비음수행렬분해, 특징행렬, 가중치행렬, 자료구조

**Abstract:** To employ NMF to transcribe music by extracting feature matrix and weight matrix at the same time, it is necessary to know in advance the dimension of the feature matrix, and to determine the pitch of each extracted feature vector. Another drawback of this approach is that it becomes more difficult to accurately extract the feature matrix as the number of pitches included in the target music increases. In this study, we prepare a feature matrix database, and apply the matrix to transcribe real music. Transcription experiments are conducted by applying the feature matrix to the music played on the same piano on which the feature matrix is extracted, as well as on the music played on another piano. These results are also compared to those of another experiment where the feature matrix and weight matrix are extracted simultaneously, without making use of the database. We could observe that the proposed method outperform the method in which the two matrices are extracted at the same time.

**Key words:** Music transcription, NMF, Feature matrix, Weight matrix, Database

### 1. 서 론

자동 음악전사(music transcription)는 연주된 음악 신호로부터 신호처리 기술을 이용하여 음악을 구성하는 음들의 시작시간(onset time), 길이(duration), 음고(음의 높이, pitch)들을 추출해 내는 기술이다. 음악전사의 초창기라 할 수 있는 1990년

대에는 음악적인 지식을 이용한 지식기반 음악전사에 대한 연구가 주류를 이루었는데, 아주 복잡한 일련의 신호처리 과정들로 구성되며, 많은 계산이 필요하였다[1]. 1999년 Lee와 Seung이 비음수행렬 분해(non-negative matrix factorization, NMF) [2]를 제안한 후에는 NMF를 이용한 음악전사에 대한 연

<sup>†</sup> 교신저자(한국해양대학교 IT공학부, E-mail: okshin@hhu.ac.kr, Tel: 051-410-4572

<sup>2</sup> 한국해양대학교 IT공학부, E-mail: yoo3146@hanmail.net, Tel: 051-410-4572

구가 활발하게 진행되어 왔다. NMF를 이용한 음악전사의 기본적인 개념은 진폭 스펙트럼 등과 같이 비음수로 표현된 음악신호를 특징행렬과 가중치행렬의 곱으로 표현한 다음, 이 행렬들로부터 음악적 특성들을 분석해 내는 것이다. 이렇게 분석된 특징행렬의 열벡터(특징벡터)는 음악을 구성하는 기본 구성요소에 대한 정보이며, 가중치행렬은 각 특징벡터들의 가중치를 나타내는 열벡터로 이루어지는데 기본 구성요소 각각에 대한 크기와 시간 정보를 가지고 있다.

NMF를 이용한 음악전사 방법은 특징행렬을 구하는 방법에 따라 두 가지로 나눌 수 있다. 하나는 전사하고자하는 표적음악신호(target music signal)에 NMF를 적용하여 특징행렬과 가중치 행렬을 동시에 계산해 내는 방법이고, 또다른 하나는 off-line으로 특징행렬 데이터베이스를 미리 준비한 다음 표적음악에 적용하는 것이다. 전자의 방법은 표적신호로부터 그 구성성분을 추출해 내기 때문에 표적신호의 특성을 가장 잘 표현하는 성분들로 분해해 낼 수 있는 장점이 있는 반면, 다음과 같은 단점도 동시에 가지고 있다. 우선 NMF를 이용하여 음악을 전사하려면 특징행렬을 구성하는 특징벡터들의 수를 알아야 한다. 각각의 특징벡터들은 악기의 음고에 해당하는데 이들의 수를 정확하게 추정하기가 쉽지 않다. 또 일반적으로 음악에는 많은 수의 음고가 포함되어 있는데, 이 음고들의 수가 커질수록 특징행렬을 정확하게 추출해 내기 어려워 오류가 많아진다. 뿐만아니라, 특징행렬을 추출한 다음에는 특징벡터들 각각의 음고를 결정하는 과정도 필요하게 된다.

특징행렬의 데이터베이스를 이용하는 방법도 많이 연구되어왔는데 이방법의 단점은 특징벡터를 만들 때 이용한 악기와 전사하고자 하는 표적음악의 악기가 다를 경우 악기사이의 미묘한 차이가 전사의 성능에 큰 영향을 미칠 수 있을 것이라는 점이다. 본 연구에서는 미리 준비된 특징행렬(특징행렬 데이터베이스)을 표적음악에 적용하여 가중치행렬을 추출하여 전사하는 방법과 표적음악에서 특징행렬과 가중치 행렬을 동시에 추출하는 방법을 실험을 통해 비교해 본다. 준비된 특징행렬을

이용하려면 특징행렬 데이터베이스를 미리 준비해야 하는 단점은 있으나 표적음악신호를 직접 분해할 때 생기는 어려움은 피할 수 있다. 본 연구에서는 미리 피아노의 특징행렬 데이터베이스를 만든 다음, 베토벤의 월광 소나타를 전사하는 몇 가지 종류의 실험을 수행하고 그 결과를 비교하였다. 다음의 2장에서 NMF에 대해 간단히 설명한 다음 3장에서 NMF를 이용한 음악전사와 특징행렬 데이터베이스를 추출하는 방법을 설명한다. 4장에서는 실험과 결과에 대해 기술하고 마지막의 5장에서 결론을 맺는다.

## 2. 비음수 행렬분해(NMF)

비음수 행렬분해란 식 (1)과 같이 음의 값이 아닌 원소들로 구성된 행렬  $Y$ 를 음의 값이 아닌 원소들로 구성된 두 행렬  $A$ 와  $X$ 의 곱으로 분해하는 방법이다[2].

$$Y = A \cdot X \quad (1)$$

식 (1)에서  $Y$ 의 크기는  $(M \times N)$ ,  $A$ 와  $X$ 의 크기는 각각  $(M \times R)$ ,  $(R \times N)$ 이며, 행렬  $A$ 를 기저행렬 (basis matrix), 또는 특징행렬(feature matrix)이라 하고 이 행렬을 구성하는 열벡터  $a_r$  ( $1 \leq r \leq R$ )는  $Y$ 를 구성하는 의미있는 부분, 또는 기본 특징을 표현하는 특징벡터로 해석할 수 있다. 이 때 행렬  $X$ 의 열벡터  $x_n$  ( $1 \leq n \leq N$ )은  $Y$ 의  $n$ 번째 열  $Y_n$ 의 가중치로써  $A$ 가  $Y$ 를 구성하는 방식으로 설명할 수 있다. 여기서 특히  $R$ 은 행렬  $A$ 의 열벡터(column vector)의 개수로,  $Y$ 를 구성하는 특징벡터들의 수이다. 일반적으로  $R$ 은  $M$ 과  $N$ 보다 적은 값이며  $(M+N) \cdot R < M \cdot N$ 을 만족하여야 한다.

Lee 등은 행렬  $Y$ 를 분해하는 기본적인 NMF 알고리즘 두 가지를 제안하였는데[3], 하나는  $Y$ 와  $AX$ 의 사이의 Frobenius norm을, 그리고 또 다른 하나는 일반화된 KL-Divergence를 최소화시킬 목적함수로 하는 방법이다. 본 연구에서는 식 (2)에 보이는 일반화된 KL-Divergence를 최소화시키는 방법을 택한다.

$$D(Y \parallel AX) = \left\| Y \otimes \ln\left(\frac{Y}{A \cdot X}\right) - A + A \cdot X \right\|_F \quad (2)$$

이식에서  $\|\cdot\|_F$ 는 Frobenius norm이며  $\otimes$ 는 Hadamard 곱으로 행렬의 각 원소별 곱셈을 의미한다. 행렬  $Y$ 가 주어졌을 때 식 (2)의 목적함수를 최소화 하는 ‘multiplicative update rule’ 알고리즘은 먼저  $A$ 와  $X$ 를 임의의 값으로 초기화한 다음, 식 (3-a)을 이용하여  $X$ 를 갱신하고, 이 갱신된  $X$ 를 이용하여 식 (3-b)로 행렬  $A$ 를 다시 갱신하는 과정을 수렴할 때까지 반복하는 것이다.

$$x_{u,n} \leftarrow x_{u,n} \frac{\sum_i a_{i,u} y_{i,n} / (AX)_{i,n}}{\sum_i a_{i,u}} \quad (3-a)$$

$$a_{i,j} \leftarrow a_{i,j} \frac{\sum_n x_{j,n} y_{i,n} / (AX)_{i,n}}{\sum_n x_{j,n}} \quad (3-b)$$

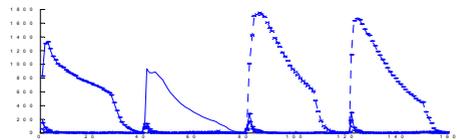
### 3. NMF를 이용한 음악전사

이 장에서는 식 (3-a, 3-b)를 이용하여 표적음악에서 특징행렬과 가중치행렬을 동시에 추출하는 방법의 장단점을 논의한 다음, 악기의 특징행렬  $A$ 를 미리 추출하여 이용하는 방법, 그리고 추출된 가중치의 후처리에 대하여 설명한다.

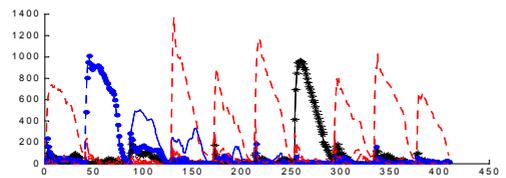
#### 3.1 특징행렬과 가중치행렬을 동시에 추출하는 방법

NMF를 이용한 음악전사에 대한 연구에서 분석하고자 하는 표적음악신호로부터 식 (3-a)와 식 (3-b)를 이용하여 특징행렬  $A$ 와 가중치행렬  $X$ 를 같이 추출해 내는 방법에 대한 연구가 많이 수행되어 왔다[5-8]. 그러나 이러한 접근 방식에는 다음과 같은 몇 가지 어려움이 있다. 먼저, 전사하고자 하는 음악신호에 포함된 모든 음고 각각의 특징벡터를 정확하게 추출해내기 쉽지 않다는 점이다. 예를 들어 단성(monophony)의 피아노음악에서 6~8개 정도까지의 비교적 적은 수의 음고로 구성된 음악 소절에서는 음고 각각에 해당하는 특징 벡터들을 비교적 쉽게 추출할 수 있으나 음고의 수가 이보다 많거나, 또는 다성(polyphony)음악인 경우에는 특징벡터들을 모두 구별해서 추출해내기 어렵다. **Figure 1**에 이러한 현상의 한가지 예를 보인다. 이 그림에 보인 것은 피아노의 연속된 건반 4개

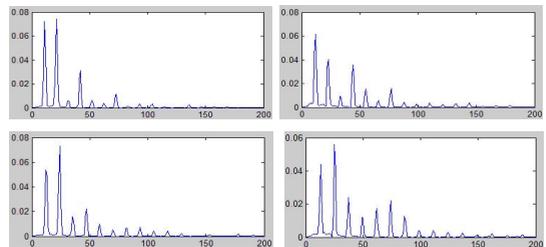
(**Figure 1(a)**)와 10개(**Figure 1(b)**)를 차례로 두드려 음을 녹음한 다음, 식 (3)을 이용하여  $A$ 와  $X$ 로 분해하여 가중치 벡터  $x_n$ 을 그린 것이다. 건반을 차례로 겹치지 않게 두드렸기 때문에  $x_n$ 들은 서로 겹치지 않아야한다. 4개의 음고를 갖는 경우 (**Figure 1(a)**)에는 가중치 벡터  $x_n$ 들이 서로 겹치지 않고 비교적 잘 분해됨을 볼 수 있다. 이 때 얻은 특징벡터들을 **Figure 1(c)**에 보인다. 반면에 10개의 음고가 있는 **Figure 1(b)**에서는 점선으로 표현된 7개의 음고는 비교적 잘 분해되었지만, 3, 4, 5번째의 가중치벡터들은 다른 음고의 일부 벡터와 겹쳐짐을 관찰할 수 있다. 이것은 특징벡터들의 집합인 특징행렬이 정확히 추출되지 않았음을 의미하고, 따라서 정확한 전사가 이루어지기 어려움을 의미한다. 이러한 현상은 표적 음악에 포함된 음고의 수가



(a) Weight vectors of 4 pitches in a source



(b) Weight vectors of 10 pitches in a source



(c) Parts of characteristic vectors  $a_r$  (amplitude spectrum in 0~4.4KHz band) corresponding to  $x_n$  in

(a); Clockwise from top left, A3, A3#, C4, B3

**Figure 1:** Example of weight vectors and characteristic vectors

커질수록 더 심해져서 많은 음고를 포함하는 실질적인 음악을 전사하기에는 많은 어려움이 있다. 이 방법의 두 번째 어려움은 특징벡터의 수를 정확하게 정하기가 어렵다는 것이다. Smaragdis 등[4, 5]은 R의 값을 충분히 크게 추정하여 A와 X를 구한 다음, 특징벡터  $a_r$ 와 이에 해당하는 가중치  $x_n$ 의 모양을 보고 적절한 열벡터를 선택하여 행렬 A를 정하였고, [6]에서는 NMF 분해 이전에 Y에 대해 SVD(singular value decomposition)을 먼저 수행해서 특징벡터의 수를 추정하였으나 항상 정확한 특징벡터의 수를 결정하지는 못하였다.

또 다른 한가지 단점은 특징행렬 A를 추출하고 난 다음 각각의 특징벡터  $a_r$ 의 음고를 결정하는 일이다. 일반적으로 악기의 음고와 음색은 기본음과 이 기본음의 배음의 조합으로 결정되는데, 특징벡터  $a_r$ 의 최대 값에 해당하는 주파수가 반드시 기본음에 해당하는 것은 아니어서 기본음을 찾는 일은 그렇게 간단하지 않다. 박상하 등[7]은 배음통합 방식(Subharmonic Summation)을 이용하여 행렬 A의 각 벡터의 음고를 결정하였으나, 피아노와 같이 비협화음(inharmonic) 악기에서는 배음통합방식도 정확하지 않을 수 있다[8].

### 3.2 악기의 특징행렬 추출

악기의 특징행렬을 추출하기 위해 악기의 파형으로 부터 진폭스펙트럼 (magnitude spectrum)을 추출하여 행렬 Y를 만든 다음, 식 (3)을 이용하여 A와 X로 분해한다. 이과정이 기존의 NMF기반 음악전사와 다른 점은 Y에 포함되는 음고의 파형들이 서로 겹치지 않게 하며, 한번의 NMF 수행에 포함되는 음고의 수를 제한한다는 것이다. 3.1절에서 언급한 것과 같이 한번의 NMF 수행에 포함된 음고의 수가 많아지면 개별 특징벡터들을 명확하게 분리해 내기가 어려워질 뿐 아니라 다른 음고의 영향을 받을 수 있으므로 가능한 한 음고의 수를 제한하는 것이 바람직하다. 따라서 악기의 전체 음역을 잘게 나누어 여러번의 NMF를 수행하고 그 결과로 얻어진 특징벡터들을 정렬하여 특징행렬을 구한다. 이 과정에서 얻어지는 특징행렬의 벡터  $a_r$

들은 각 음고에 해당하는 평균파형이라 할 수 있다. 이 방식의 이점은 파형에 포함된 음고의 수(R)뿐만 아니라 모든 특징벡터의 음고를 미리 알고 있으므로 각 특징 벡터들의 음고를 결정하는 과정이 생략되어 아주 단순한 방법으로 특징행렬 A를 구할 수 있다는 것이다.

### 3.3 가중치 행렬의 후처리

전사하고자 하는 음악신호의 진폭 스펙트럼 Y가 주어지면 3.2절의 방법으로 추출한 특징행렬 A와 Y를 식 (3-a)에 반복적으로 적용하여 X를 구할 수 있다. 이렇게 구해진 가중치  $x_{nr}$ 은 입력 음악행렬 Y의 n번째 열에 포함된 특징벡터  $a_r$ 에 대한 가중치이다. 따라서  $x_n$ 의 크기를 분석하면 음악신호의 n번째 프레임의 음고를 추정할 수 있다. 3.2절에서 기술한 바와 같이 특징행렬의 특징벡터들은 낮은 음고에서 부터 높은 음고순으로 정렬되어 있기 때문에 전사는 간단하게 이루어진다. 그러나 행렬 X에는 여러 가지 요인으로 잡음이 더해져 있으므로 X를 후처리할 필요가 있다. Figure 2에 식 (3-a)를 이용하여 얻은 X로부터 음고를 추정하는 후처리 과정의 순서를 보인다. 후처리의 첫 번째 과정은 잡음의 영향을 최소화하기 위해 식 4를 이용하여 가중치들을 시간축을 따라 평활화하는 것이다. 이 식에서  $n$ 은 시간(프레임) 인덱스,  $u$ 는 특징벡터의 인덱스 ( $2 \leq u \leq n-1$ )이며,  $\alpha$ 는 평활화 계수인데 간단한 실험을 통하여 결정하였다.

$$x_n'(u) = \alpha x_{n-1}(u) + (1-2\alpha)x_n(u) + \alpha x_{n+1}(u) \quad (4)$$

이어서 각각의 프레임이 악보의 효과 등음에 해당하는 휴지부인지를 확인하는데, 프레임 전체의 진폭의 합이 임계치보다 적으면 휴지부로 결정하고, 휴지부가 아닌 프레임 중에 전체적으로는 연속이지만 중간에 한두 개씩 빠지는 공백은 지속되는 것으로 간주하여 연결한다. 이 과정에서 추출된 연속적인 프레임들은 해당 음고의 띠(strip)가 되며 이 띠의 길이가 임계치 이상이면 하나의 음표(note)로 인식된다.

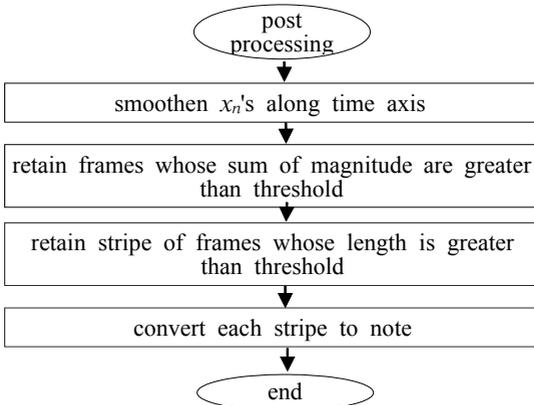


Figure 2: Postprocessing process of the transcription

#### 4. 실험 및 결과

본 연구에서는 Figure 3에 보인 베토벤의 피아노 소나타 월광의 2악장 첫 9마디를 여러 가지 방법으로 전사하는 실험을 하였다. 먼저 피아노의 특징행렬 A(특징행렬데이터베이스, 특징행렬DB)를 구한 다음, 이 행렬을 추출한 피아노에서 피아니스트가 각각 단성, 2성, 다성으로 연주하게 하여 녹음하였다. 여기서 단성이란 악보의 제일 높은음만을 연주한 것이고, 2성은 높은음 순으로 2개, 그리고 다성은 악보의 모든 음을 동시에 연주한 것이다. 또 본 연구에서 준비한 특징행렬DB가 다른 피아노로 연주한 음악에도 적용될 수 있는지를 알아보기 위하여 Wilhelm Kempff의 연주를 녹음한 CD[9]에서 같은 부분을 발췌하였다. 이 음원들에 특징행렬DB를 직접 적용하여 전사하는 실험, 특징행렬DB를 특징행렬의 초깃값으로 이용하여 특징행렬과 가중치행렬을 동시에 추출해 내는 실험, 그리고 특징행렬DB를 사용하지 않고 음원에서 직접 특징행렬과 가중치행렬을 동시에 추출해 내는 실험을 수행하여 그 성능을 비교하였다. 아래에 각각의 실험 결과에 대하여 기술한다.



Figure 3: The score of the music piece used in the experiments (Beethoven, piano sonata 'moonlight, op.27-2, 2nd mov. 1st 9 bars)

#### 4.1 특징행렬DB 추출

특징행렬DB를 구하기 위해 YAMAHA 그랜드 피아노의 건반 88개에 해당하는 파형들을 44.1 kHz, mono로 녹취한 다음, 2048 point 크기로 Hanning-window를 적용하고 FFT를 수행하여 진폭 스펙트럼을 얻었다. 이 때 기술적인 어려움으로 인하여 잡음이나 반향을 제거하기 위한 특별한 처리는 하지 않았다. 특징행렬 A를 구할 때, 88개의 음고 파형을 어떻게 분할하여 NMF를 적용하는 것이 효율적인지를 알아보기 위해 간단한 몇 가지 실험을 수행하였다. 한음씩 차례로 녹음한 88개의 음고 모두를 포함시켜 NMF를 수행했을 때 가중치행렬 X의 벡터들이 서로 겹치는 경우가 많았으며, 특징벡터  $a_i$ 들 중에는 잡음으로 보이는 파형도 많았다. 이는 분해되어야 하는 특징벡터들이 많아지면 따로 추출되어야 할 벡터들이 다른 벡터들의 선형 조합으로 표현되는 등, 개별적인 음고가 잘 분해되지 않았기 때문으로 보인다. 파형들을 소규모로 나눌수록 A와 X의 추출이 쉬워졌으며 4개, 2개의 파형을 대상으로 분해하였을 때 모든 파형이 잘 분해되었다. 본 연구에서는 88개의 파형을 인접한 두 개의 파형씩 묶어 모두 44번의 NMF를 수행하였다. 이렇게 얻어진 특징벡터들을 가장 낮은 음에서부터 높은 음의 순서로 정렬하여 피아노의 특징행렬DB를 완성하였다.

#### 4.2 음악전사

특징행렬DB를 추출한 피아노로 Figure 3의 음악을 연주하여 3.2절에 설명한 과정을 거쳐 전사하였다. 모두 6가지의 실험을 수행하였는데 아래의 표 1에 각각의 실험조건을 요약하였다. 이 표에서 '특징행렬DB 이용'이란 특징행렬DB A와 표적음원 Y로부터 식 (3-a)를 적용하여 가중치행렬 X를 구한 것을 의미한다. 또 '음원에서 A, X 행렬 직접 추출'했다는 것은 특징행렬DB는 이용하지 않고 A와 X를 임의의 값으로 초기화한 다음, 식 (3-a)와 (3-b)를 직접 반복적으로 적용한 실험이다. 한편 행렬 A와 X를 모두 음원에서 추출한 실험들 (D-P, D-K)에서는 음원에 휴지부(검표)를 제외하면 모두 20가지의 음고가 포함되어 있다는 사실을 알고 있으므로 이를 이용하여 행렬 A의 열벡터의 수(R)를

20으로 고정하였다. 이상의 모든 실험들에서 얻어진 각각의 가중치벡터들은 3.3절에 기술한 방법으로 후처리되어 최종적으로 삽입 또는 음표로 인식된다.

**Table 1:** Experiments and their conditions

표기	실험조건
1-line	특징행렬DB 이용, 악보의 가장 높은 음만 연주, 단성, 피아니스트 연주
2-line	특징행렬DB 이용, 가장 높은 2음만 연주, 2성, 피아니스트 연주
A-P	특징행렬DB 이용, 다성, 피아니스트 연주
A-K	특징행렬DB 이용, 다성, W. Kempff 음원
D-P	음원에서 A, X 행렬 직접 추출, 다성, 피아니스트 연주
D-K	음원에서 A, X 행렬 직접 추출, 다성, W. Kempff 음원

**Table 2**에 모든 전사실험의 성능을 나타내었다. 이 표의 Hit는 적중한 음표의 수, Ins는 삽입(insertion) 오류의 수, 그리고 Mis는 삭제(miss)오류의 수, 그리고 Acc는 정확도(Accuracy)로 다음과 같이 정의한다.

$$Acc = \frac{Hit - Ins}{Total} \times 100 (\%) \quad (4)$$

첫 두 실험(1-line, 2-line)은 피아노의 특징행렬을 추출한 다음, 같은 피아노에서 단성(표준음고의 첫줄)과 2성(표준음고 첫 2줄)으로 연주하여 전사한 것으로 각각 사분음표 1박자 음표 23개와 46개를 포함하고 있다. 나머지 실험들은 악보전체를 연주한 것으로 사분음표 77개를 포함한다.

단성(1-line)실험에서는 상당히 높은 hit율을 보였으며 삽입오류가 13%, 탈락오류가 4.3%인데, 이 성능은 녹음된 음악이 단성이라는 정보를 이용하지 않았을 때의 성능이며, 단성이라는 정보를 이용한다면 삽입오류는 전혀 없음을 확인하였다. 2성(2-line)실험에서는 단성의 경우보다 오류가 크게 증가하였다. 특히 탈락 오류가 많아졌고 삽입오류는 파형이 비슷한 옥타브 삽입(C5:C6, Gb6:Gb5, Ab4:Ab3)이 많았다. 다성실험들 중 피아니스트가

**Table 2:** Performances of the transcriptions

	Hit (%)	Ins (%)	Mis (%)	Acc	total beat
1-line	22 (95.7)	3 (13.0)	1 (4.3)	82.6	23
2-line	40 (87.0)	10 (21.7)	6 (13.0)	65.2	46
A-P	66 (85.7)	28 (36.4)	11 (14.3)	49.4	77
A-K	60 (77.9)	31 (40.3)	17 (22.1)	37.7	77
D-P	47 (61.0)	26 (33.7)	30 (39.0)	13.2	77
D-K	44 (57.1)	27 (35.1)	33 (42.8)	8.9	77

연주한 음원에 특징행렬DB를 이용한 실험(A-P)에서 적중률은 85%이지만 삽입, 탈락오류들이 증가하여 정확도는 약 50% 부근이었다. 특징행렬DB를 다른 피아노로 연주한 음원의 가중치행렬 추출에 이용한 실험(A-K)에서는 같은 악기를 이용한 실험들과 비교하면 정확도는 약 10% 가량 떨어진다. 당연한 결과이지만 특징행렬과 전사대상 음악이 서로 다른 악기에서 추출, 연주되었으며 특별한 조율과정을 거치거나 음향효과 등을 감안하지 않았음을 고려한다면 기대 이상의 좋은 전사결과를 얻었다고 할 수 있다.

마지막의 두 가지 실험(D-P, D-K)은 음원으로부터 행렬 A와 X를 직접 추출한 경우로 다른 실험들의 성능과 비교하기 위한 실험이다. 전술한 것처럼 이 실험들에서는 전체 음고의 수가 20개라는 사전 지식을 활용하였다. 또 특징행렬의 각 벡터들의 음고는 수작업으로 정하였다. 두 실험 다 적중률은 60%정도이지만 삽입, 탈락오류가 많아 정확도는 10% 내외에 불과하다. 비교적 난이도가 낮은 음악으로 실험하였지만 이 결과는 음고의 수가 어느 정도 이상인 실제의 다성(polyphony)음악을 단순한 방법으로 직접 전사하여서는 좋은 성능을 얻기 어렵다는 것을 시사한다.

전체적으로 특징행렬DB를 추출한 것과 같은 피아노로 연주한 음악(A-P)의 전사성능이 가장 높았으며, 악기가 달라도 특징행렬DB를 직접 이용하여 전사(A-K)할 수 있음을 알 수 있었다. 삽입오류의 주된 원인은 옥타브 삽입 (F4:F3 Gb5:Gb3,

Ab3:Ab4, G3:G4, C5:C4, 등) 이었으며 인접음 삼입(Ab3:A3, F5, Gb5)등도 있었다.

## 5. 결 론

본 논문에서는 NMF를 이용하여 음악을 전사하는 과정에서 특징벡터를 추출하는 방법의 영향을 실험을 통해 비교하였다. 종래의 NMF를 이용한 음악 전사에서는 표적음악신호에서 특징행렬과 가중치행렬을 동시에 추출하는 방법에 대한 연구가 주를 이루었다. 그러나 이 방법은 많은 음고를 포함하는 실질적인 음악신호를 전사하기에는 해결하기 어려운 문제들이 있는데, 특히 표적음악에 포함된 전체 음고의 수를 결정하거나, 추출된 각각의 특징 벡터들에 대한 음고를 결정하는 일은 아직 많은 연구가 필요한 실정이다. 본 연구에서는 특징행렬을 미리 만들어 놓은 다음, 이를 표적음악에 적용하여 음악에 포함된 가중치행렬을 추출하는 두 가지 방법을 실험하였다. 하나는 특징행렬을 만들 때 이용한 피아노에서 연주한 단성, 2성 및 다성의 피아노 소나타 몇 마디를 전사하였고, 또 다른 하나는 특징행렬과는 관계없는 다른 피아노를 이용한 같은 음악을 CD에서 발췌하여 전사하였다. 다성음악의 전사실험결과 적중률과 정확도가 특징행렬을 추출했던 같은 피아노에서는 각각 85%, 50%가량이었고, 다른 피아노에서는 약 78%와 37% 정도의 성능을 보였다. 이 성능의 의미를 파악하기 위해 음악신호에서 직접 특징행렬과 가중치행렬을 추출하는 실험을 수행하였는데 적중률과 정확도가 각각 60%와 10% 내외의 성능을 보였다. 이 실험에서는 음고의 수가 모두 20개라는 사전 지식을 이용하였으며 추출된 특징벡터들의 음고를 모두 수작업으로 결정하였으므로 실제의 성능은 이보다 훨씬 못할 것으로 예상된다. 따라서 이 성능들을 직접 비교하는 것은 큰 의미가 없을 것이나 악기가 서로 다른 경우에도 특징행렬을 미리 만들어 전사에 이용하는 방법이 더 좋은 결과를 가져 올 수 있을 것으로 보인다. 본 연구에서는 제안하는 방법의 기본적인 가능성을 시험해보았지만, 악기의 특성을 더 세분화한 데이터베이스를 준비하고, 악기를 세심하게 조율하는 한편, 더 발전된

NMF 알고리즘을 이용하면 더 나은 전사가 가능할 것으로 예상된다.

## 참고문헌

- [1] K. D. Martin, Automatic Transcription of Simple Polyphonic Music: Robust Front End Processing, MIT Media Laboratory Perceptual Computing Section Technical Report, no. 399, pp. 1-11, 1996.
- [2] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization", *Nature*, vol. 401, pp. 788-791, 1999.
- [3] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization", *Neural Information Processing Systems*, pp. 556-562, 2000.
- [4] P. Smaragdis, "Redundancy reduction for computational audition, a unifying approach", Thesis, Ph.D, MIT, 2001.
- [5] P. Smaragdis and C. B. Judith, "Non-negative matrix factorization for polyphonic music transcription", *Proceeding IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp.177-180, 2003.
- [6] A. CICHOCKI and R. ZDUNEK, "Regularized alternating least squares algorithms for non-negative matrix/tensor factorization", *Lecture Notes in Computer Science*, Vol. 4493, Springer, pp. 793-802, 2007.
- [7] S. H. Park et al. "Music transcription using non-negative matrix factorization", *The Journal of the Acoustical Society of Korea*, vol. 29, no.2, pp.102-110, 2010. (in Korean).
- [8] L. I. Ortiz-Berenguer et al. "Piano transcription using pattern recognition: aspects on parameter extraction", *Proceeding of 7th international Conference on Digital Audio Effects*, Naples, Italy, pp.212-216, Oct. 5-8, 2004.
- [9] Deutsche Grammophon, Wilhelm Kempff, (CD) Ludwig van Beethoven Klaviersonaten, No. 14 op27, (moonlight), West Germany, 1965.