석 사 학 위 논 문

비음성, 희소성 및 이동일치성을 이용한 단일 채널 음악 음원 분리

김민제 (金珉弟)

전자컴퓨터공학부

(컴퓨터공학 지능형 멀티미디어)

포항공과대학교 대학원

2006

비음성, 희소성 및 이동일치성을 이용한
단일 채널 음악 음원 분리

Monaural Music Source Separation:

Nonnegativity, Sparseness, and

Shift-Invariance

Monaural Music Source Separation: Nonnegativity, Sparseness, and Shift-Invariance

Advisor : Seungjin Choi

by

Minje Kim
Division of Electrical and Computer Engineering
(Computer Science and Engineering)
Pohang University of Science and Technology

A thesis submitted to the faculty of Pohang University of Science and Technology in partial fulfillment of the requirements for the degree of Master of Science in the Division of Electrical and Computer Engineering (Computer Science and Engineering)

Pohang, Korea
12. 19. 2005
Approved by

Major Advisor

비음성, 희소성 및 이동일치성을 이용한 단일 채널 음악 음원 분리

김민제

위 논문은 포항공과대학교 대학원 석사 학위논문으로 학위논문 심사위원회를 통과하였음을 인정합니다.

2005년 12월 19일

학위논문 심사위원회 위원장 최승진 (인)

위 원 방승양 (인)

위 원 김대진 (인)

MECE 20042104

김민제 Minje Kim, Monaural Music Source Separation: Nonnegativity, Sparseness, and Shift-Invariance. 비음성, 희소성 및 이동일치성을 이용한 단일 채널 음악 음원 분리, Division of Electrical and Computer Engineering, 2006, P, Advisor : Seungjin Choi. Text in English.

ABSTRACT

In this paper we present a method for polyphonic music source separation from their monaural mixture, where the underlying assumption is that the harmonic structure of a musical instrument remains roughly the same even if it is played at various pitches and is recorded in various mixing environments. We incorporate with *nonnegativity, shift-invariance*, and *sparseness* to select representative spectral basis vectors that are used to restore music sources from their monaural mixture. Experimental results with monaural instantaneous mixture of voice/cello, trumpet/viola and monaural convolutive mixture of saxophone/viola, are shown to confirm the validity of our proposed method.

Contents

List of Figures

1 Introduction

Blind Source Separation(BSS) can be grouped into two categories by the number of observed and source signals: over-determined is the case that we have observed signals more than the signals to estimate and under-determined is the opposite case. Over-determined BSS problems can be attacked by assuming the statistical independence of the sources [1]. In case of under-determined ones, however, we need to have more special information about the mixing process. Extreme, but most general applications of BSS in music separation are single channel source separation ironically, when we have only one mixture of signals played by several musical instruments.

The nonnegative matrix factorization (NMF) [2] or its extension such as nonnegative matrix deconvolution (NMD) [3] and sparse coding [4], was shown to be useful in polyphonic music description [5, 6], in the extraction of multiple music sound sources [3, 7], and in general sound classification [8]. Some of these methods regard each note as a source, which might be appropriate for music transcription and work for source separation in a very limited case. It is true that music source separation is a subarea of blind source separation (BSS), but it has some peculiar characters. So we want to refine the term, music source separation, by adding a constraint that we have to solve this problem regarding the special property of musical sound not merely applying current BSS algorithms. In this paper, our goal to estimate is the whole melody generated by each instrument, not the fragmented notes. The music source separation can be a useful preprocessing of the automatic music transcription task which will be devoted for each instrument.

Furthermore, our proposed method can also be automatic music transcription itself since it is a combination of the previous transcription method [5] and the music source separation defined before.

Another issue about music source separation which has to be pointed out is that usually there is only one observation of musical mixture in real world. General methods for BSS require multiple observations recorded by enough number of sensors. In the case of monaural mixture, we have to introduce intrinsic characters of target signal (it was the main requirement of music source separation, too). There was a previous work which exploits harmonic structure of speech for single channel dereverberation [9], but music source separation has difficulty in that the interference signals also have their own harmonic structure.

In this paper we present a method for monaural polyphonic music separation, the goal of which is to restore the whole melody generated by each musical instrument from a single channel mixture of several polyphonic musical sounds. We assume that the harmonic structure of a musical instrument approximately remains the same, even if it is played at different pitches and is recorded in different environments. Different musical instruments are assumed to have different spectral characteristics (harmonic structure).

The main idea is to select a few representative spectral basis vectors in the auditory spectrogram of measurement data, assuming that there are some sections in the auditory spectrogram where only a single note from a single source appears. Rather than learning basis vectors, we select a few appropriate nonnegative basis vectors using the sparseness of spectral coefficients. These shift-invariant nonnegative basis vectors are fixed and associated encoding variables are learned by the overlapping NMF [10] which incorporates with the shift-invariant representation, in order to restore music sources. The method is related to our earlier work [11] and the generalized prior subspace analysis [12]. However, the key distinction lies in a way of selecting shift-invariant basis vectors. Promising results with monaural instantaneous mixture of voice/cello, trumpet/viola and convolutive mixture of saxophone/viola, are presented to confirm the validity of our proposed method.

2 Overlapping NMF: Nonnegativity and Shift-Invariance

Nonnegative matrix factorization (NMF) is a simple but efficient factorization method for decomposing multivariate data into a linear combination of basis vectors with nonnegativity constraints for both basis and encoding matrix [2].

Given a nonnegative data matrix $\boldsymbol{V} \in \mathbb{R}^{m \times N}$ (where $V_{ij} \geq 0$), NMF seeks a factorization

$$\boldsymbol{V} \approx \boldsymbol{W}\boldsymbol{H}, \tag{1}$$

where $\boldsymbol{W} \in \mathbb{R}^{m \times n}$ $(n \leq m)$ contains nonnegative basis vectors in its columns and $\boldsymbol{H} \in \mathbb{R}^{n \times N}$ represents the nonnegative encoding variable matrix. Appropriate objective functions and associated multiplicative updating algorithms for NMF can be found in [2].

The overlapping NMF is an interesting extension of the original NMF, where transform-invariant representation and a sparseness constraint are incorporated with NMF [10]. Some of basis vectors computed by NMF could correspond to the transformed versions of a single representative basis vector. The basic idea of the overlapping NMF is to find transformation-invariant basis vectors such that fewer number of basis vectors could reconstruct observed data. Given a set of transformation matrices, $\mathcal{T} = \left\{\boldsymbol{T}^{(1)}, \boldsymbol{T}^{(2)}, \ldots, \boldsymbol{T}^{(K)}\right\}$, the overlapping NMF finds a nonnegative basis matrix $\boldsymbol{W}$ and a set of nonnegative encoding matrix $\left\{\boldsymbol{H}^{(k)}\right\}$ (for $k = 1, \ldots, K$) which minimizes

$$\mathcal{J}(\boldsymbol{W}, \boldsymbol{H}) = \frac{1}{2}\left\|\boldsymbol{V} - \sum_{k=1}^{K} \boldsymbol{T}^{(k)} \boldsymbol{W} \boldsymbol{H}^{(k)}\right\|_F^2, \tag{2}$$

where $\|\cdot\|_F$ represents Frobenious norm. The multiplicative updating rules for the overlapping NMF were derived in [10], which are summarized below.

Algorithm Outline: Overlapping NMF [10]

Step 1 Calculate the reconstruction: $\boldsymbol{R} = \sum_{k=1}^{K} \boldsymbol{T}^{(k)} \boldsymbol{W} \boldsymbol{H}^{(k)}$.

Step 2 Update the encoding matrix by

$$\boldsymbol{H}^{(k)} \leftarrow \boldsymbol{H}^{(k)} \odot \frac{\boldsymbol{W}^{\top} \left[\boldsymbol{T}^{(k)}\right]^{\top} \boldsymbol{V}}{\boldsymbol{W}^{\top} \left[\boldsymbol{T}^{(k)}\right]^{\top} \boldsymbol{R}}, \quad k = 1, \ldots, K, \tag{3}$$

where $\odot$ denotes the Hadamard product and the division is carried out in an element-wise fashion.

Step 3 Calculate the reconstruction $\boldsymbol{R}$ again using the encoding matrix $\boldsymbol{H}^{(k)}$ updated in Step 2, as in Step 1.

Step 4 Update the basis matrix by

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \odot \frac{\sum_{k=1}^{K} \left[\boldsymbol{T}^{(k)}\right]^{\top} \boldsymbol{V} \left[\boldsymbol{H}^{(k)}\right]^{\top}}{\sum_{k=1}^{K} \left[\boldsymbol{T}^{(k)}\right]^{\top} \boldsymbol{R} \left[\boldsymbol{H}^{(k)}\right]^{\top}}. \tag{4}$$

3 Spectral Basis Selection: Sparseness

Overlapping NMF was devised to learn both the basis vectors and encoding matrices set, but we select and fix the basis vectors preliminarily to the main overlapping NMF update routine, so that the equation (4) is not used in our framework. We can avoid trivial solutions and make the system faster with introducing the spectral basis selection method.

The goal of spectral basis selection is to choose R representative vectors $\boldsymbol{V}_r = [\boldsymbol{v}_{r_1} \cdots \boldsymbol{v}_{r_R}]$ (R is the number of music sources) from $\boldsymbol{V} = [\boldsymbol{v}_1 \cdots \boldsymbol{v}_N]$ where $\boldsymbol{V}$ is the data matrix associated with the spectrogram of mixed sound. Each column vector $\boldsymbol{v}_t$ corresponds to the power spectrum of the mixed sound at time $t = 1, \ldots, N$. Selected representative vectors are fixed as basis vectors that are used to learn an associated encoding matrix set through the overlapping NMF with sparseness constraint, in order to restore unmixed musical sound.

Our spectral basis selection method is based on the assumption that there are some sections where only a single note from a single source appears. In the spectrogram of mixed sound, solo sections are searched partly through the sparseness value of $\boldsymbol{v}_t$ over time. Our earlier work can be found in [11].

Fig. 1 shows the schematic diagram of the spectral basis selection method, consisting of two parts. The first part is to select several candidate vectors $\boldsymbol{V}_c = [\boldsymbol{v}_{c_1} \, \boldsymbol{v}_{c_2} \cdots \boldsymbol{v}_{c_K}]$ from $\boldsymbol{V}$ using a sparseness measure and a clustering-elimination method. The second part involves determining representative basis vectors from candidate vectors, through the overlapping NMF. More detailed description is summarized below.

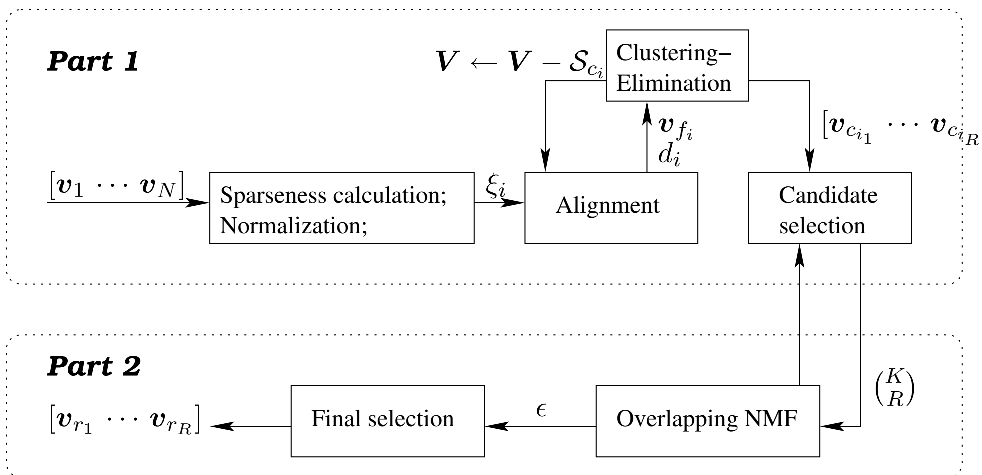


Figure 1: Schematic diagram of our spectral basis selection method, is shown, where 'Part 1' involves the selection of candidate vectors and 'Part 2' determines a few representative spectral basis vectors from candidate vectors found in 'Part 1'.

Part 1

1. **Sparseness calculation:** We calculate the sparseness value for input vectors $\boldsymbol{v}_t$ for $t = 1, \ldots, N$, using the measure in [13],

$$\xi_t = \text{sparseness}(\boldsymbol{v}_t) = \frac{\sqrt{m} - (\sum_i |v_{it}|)/\sqrt{\sum_i v_{it}^2}}{\sqrt{m} - 1}, \tag{5}$$

where v_{it} is the ith element of the m-dimensional vector $\boldsymbol{v}_t$.

2. **Normalization:** We normalize input vectors $\boldsymbol{v}_t$ for $t = 1, \ldots, N$ such that each vector has unit Euclidean norm,

$$\boldsymbol{v}_t \leftarrow \frac{\boldsymbol{v}_t}{\|\boldsymbol{v}_t\|}. \tag{6}$$

3. **Alignment:** We calculate the index $f_i = t^*$ which involves the largest sparseness value among $\{\xi_t\}_{t=1}^N$, i.e,

$$t^* = \arg\max_{1 \le t \le N} \xi_t. \tag{7}$$

The vector $\boldsymbol{v}_{f_i}$ associated with the index $f_i = t^*$, is referred to as a *foundation vector* that has the largest sparseness value among $\{\boldsymbol{v}_t\}$. Then we align each vector $\boldsymbol{v}_j$ in L remaining input vectors (initially $L = N$ but L represents the number of remaining vectors after the clustering-elimination procedure in step 4) with respect to the current foundation vector $\boldsymbol{v}_{f_i}$ such that the Euclidean distance between $\boldsymbol{v}_{f_i}$ and vertically shift-up or -down version of $\boldsymbol{v}_j$, is minimized. In other words, vectors $\boldsymbol{v}_j$ are vertically shifted-up or -down such that their shifted version provides the minimal Euclidean distance from the foundation vector $\boldsymbol{v}_{f_i}$.

4. **Clustering-Elimination:** The goal of the clustering-elimination step is to eliminate vectors belonging to the cluster where the foundation vector is contained, since those vectors are regarded as redundant vectors. To this end, we first apply the k-means clustering method to dichotomize the aligned vectors (including the foundation vector), leading to two groups $\mathcal{S}_{c_i}$ and $\bar{\mathcal{S}}_{c_i}$. The cluster containing the foundation vectors, $\mathcal{S}_{c_i}$, is further grouped into R sub-clusters, producing $\{\boldsymbol{v}_{c_{i_1}}, \ldots, \boldsymbol{v}_{c_{i_R}}\}$ that is a collection of mean vectors of R sub-clusters.

5. **Candidate selection:** Add the mean vector of the cluster $\mathcal{S}_{c_i}$ to the candidate set.

6. **Repeat:** Repeat steps 3-5 with data excluding vectors in $\mathcal{S}_{c_i}$, i.e, $\boldsymbol{V} - \mathcal{S}_{c_i}$, until we choose a pre-specified number of candidate vectors or there is no remaining input vector.

Part 2

This second part involves determining the final representative spectral basis vectors $\{\boldsymbol{v}_{r_1}, \ldots, \boldsymbol{v}_{r_R}\}$ from $K \geq R$ candidate vectors $\{\boldsymbol{v}_{c_1}, \ldots, \boldsymbol{v}_{c_K}\}$ (where K is the integral multiples of R, depending on the number of loops in the clustering-elimination) found in the first part.

1. **Overlapping NMF** Repeat the following step for all possible $\binom{K}{R}$ combination. Construct a small set of input vectors $\widetilde{\boldsymbol{V}}$ by random sampling and treat them at input vectors for the overlapping NMF. Choose R candidate vectors from $\{\boldsymbol{v}_{c_1}, \ldots, \boldsymbol{v}_{c_K}\}$ and fix them (denoted by $\widetilde{\boldsymbol{W}}$)

as basis vectors. Run the overlapping NMF with these $\widetilde{\boldsymbol{V}}$ and $\widetilde{\boldsymbol{W}}$ to calculate the reconstruction error.

2. **Final selection** Choose spectral basis vectors that give the lowest reconstruction error.

4 Numerical Experiments

We present two simulation results for monaural instantaneous mixtures of voice/cello and trumpet/viola and monaural convolutive mixtures of saxophone and viola. We apply our spectral basis selection method with the overlapping NMF to these two data sets transformed to auditory spectrograms using the NSL toolbox [14]. Experimental results are shown in Fig. 2, 3 and 4 where figure captions describe detailed results. Note that the mixture in Fig. 4 (c) is a convolutive mixture and we can apply our framework even in that case without any modification if the reverberation time is not too long.

Fig. 5 shows the reusability of our obtained spectral basis vectors. The mixture in Fig. 5 (c) is another part of the same song used in Fig. 4. In this example, we do not have to find out the spectral basis vectors of saxophone and viola again, but can simply reuse the previous results of Fig. 4. Note that if some input data do not satisfy the horizontal sparseness, which means that there is no section occupied by only one instrument, our spectral basis selection method will fail in this case. However we can attack this problem by reusing the previously obtained spectral basis vectors of the same source instruments.

As an additional experiment, we selected only a single shifted version of each basis vector at a given time which has the highest corresponding encoding value of all the possible vertical shifting and then unmixed with them. On the assumption that an instrument can not play multiple notes at a given time, that simple post-processing step could wipe out unwanted interference

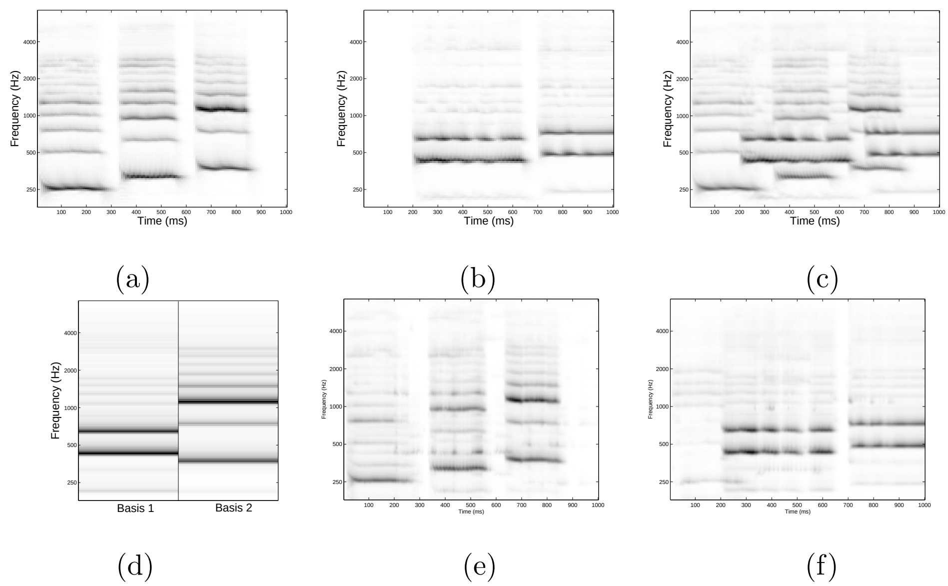


Figure 2: Auditory spectrograms of original sound of /*ah*/ voice and a single string of a cello are shown in (a) and (b), respectively. Horizontal bars reflect the harmonic structure. One can see that every note is the vertically-shifted version of each other if their musical instrument sources are the same. Monaural mixture of voice and cello is shown in (c) and final two representative spectral basis vectors in (d) which give the smallest reconstruction error in the overlapping NMF are selected by our algorithm in Fig. 1. Each of these two basis vectors is a representative one for voice and a string of cello. Unmixed sound is shown in (e) and (f) for voice and cello, respectively.

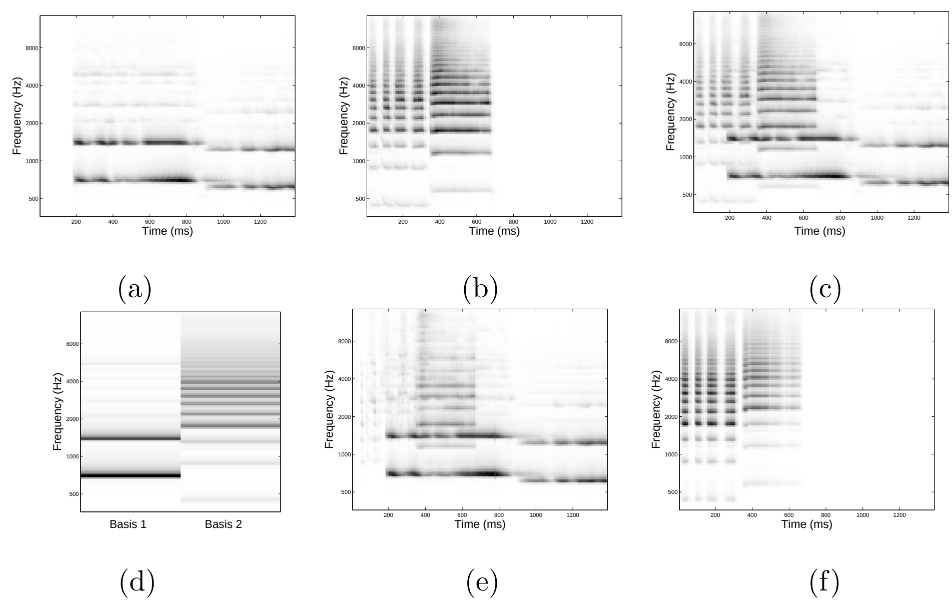


Figure 3: Auditory spectrograms of original sound of trumpet and a single string of viola are shown in (a) and (b), respectively. Monaural mixture of trumpet and viola is shown in (c) and final two representative spectral basis vectors in (d). Unmixed sound is shown in (e) and (f) for trumpet and viola, respectively.

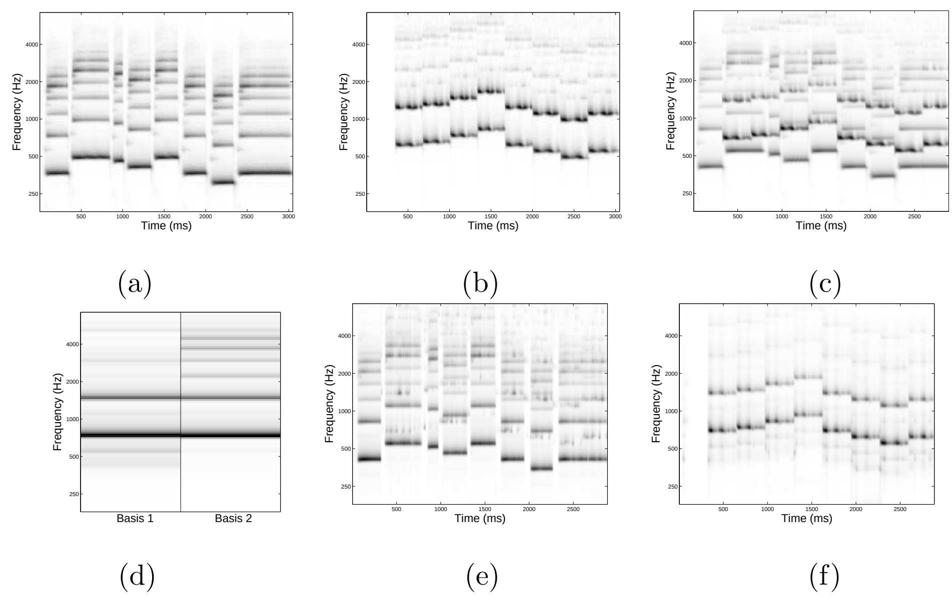


Figure 4: Auditory spectrograms of original sound of saxophone and viola are shown in (a) and (b), respectively. Every note is artificially generated by changing the frequency of a real sample sound, so that the spectral character of each instrument is constant in all the variations of notes. We mixed these two signals by convolving them with two impulse response signals measured in a studio environment (reverberation time is about 150ms and the frequency response makes a peak at around 27Hz). The monaural convolutive mixture is shown in (c) and finally selected two representative spectral basis vectors are in (d). Unmixed sound is shown in (e) and (f) for saxophone and viola, respectively.

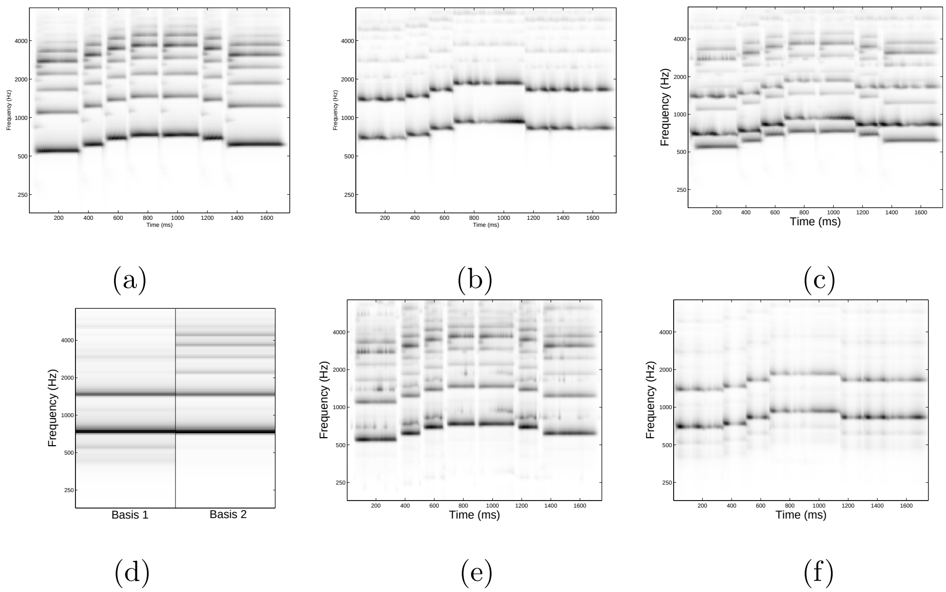
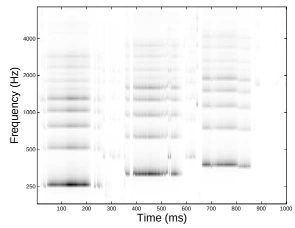
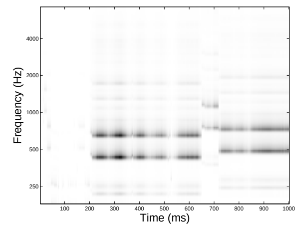


Figure 5: These figures show the reusability of spectral basis vectors. Auditory spectrograms of original sound of saxophone and viola are shown in (a) and (b), respectively. Every note is generated in the same manner of Fig. 4 but the melody is totally different from it since this is another part of the same song. The mixing process is also the same with the previous experiment. The monaural convolutive mixture is shown in (c). Instead of finding out representative basis vectors, we reused the basis vectors (d) found in previous example. Unmixed sound is shown in (e) and (f) for saxophone and viola, respectively.

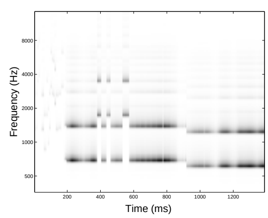
sounds generated from other sources with some decrease of robustness. Additional experimental results are shown in Fig. 6. Audio demo can be found in `http://home.postech.ac.kr/~minjekim/demo.php`.



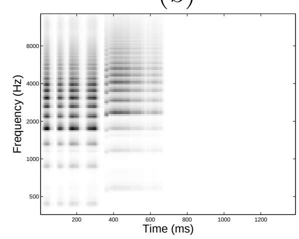
(a)



(b)



(c)



(d)

Figure 6: Unmixed sounds using the highest encoding value, (a), (b), (c) and (d) for voice, cello, viola and trumpet respectively. Compare these results with Fig. 2 (e), Fig. 2 (f), Fig. 3 (e) and Fig. 3 (f) respectively.

The set of transformation matrices, $\mathcal{T}$, that we used, is

$$\mathcal{T} = \left\{ \boldsymbol{T}^{(k)} \,\middle|\, \boldsymbol{T}^{(k)} = \overset{k-m}{\overset{\longmapsto}{\boldsymbol{I}}}, \quad 1 \leq k \leq 2m-1 \right\}, \tag{8}$$

where $\boldsymbol{I} \in \mathbb{R}^{m \times m}$ is the identity matrix and $\overset{j}{\overset{\longmapsto}{\boldsymbol{I}}}$ leads to the shift-up or shift-down of row vectors of $\boldsymbol{I}$ by j, if j is positive or negative, respectively. After shift-up or -down, empty elements are zero-padded.

For the case where $m = 3$, $\boldsymbol{T}^{(2)}$ and $\boldsymbol{T}^{(5)}$ (they means that $k = 2$ and

$k = 5$) are defined as

$$\boldsymbol{T}^{(2)} = \overset{2-3}{\mapsto}{\boldsymbol{I}} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{T}^{(5)} = \overset{5-3}{\mapsto}{\boldsymbol{I}} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{9}$$

Multiplying a vector by these transformation matrices, leads to a set of vertically-shifted vectors.

5 Discussions

We have presented a method of spectral basis selection for monaural music source separation, where we incorporated with the harmonics, sparseness, clustering, and the overlapping NMF. Rather than learning spectral basis vectors from the data, our approach is to select a few representative spectral vectors among given data and fix them as basis vectors to learn associated encoding variables through the overlapping NMF, in order to restore unmixed sound. The success of our approach lies in the two assumptions. The one is that the distinguished timbre of a given musical instrument can be expressed by a transform-invariant time-frequency representation, even though their pitches are varying. The other is that there is solo sections in a musical sound where the contribution of each source instrument appears. Our experimental results showed that the proposed methods are reasonable in convolutive mixture case if the frequency response is somewhat flat and the reverberation time is short enough.

Music source separation is quite different from speech separation. Unfortunately, the assumption that every notes from same instrument hold the harmonics structure in same cannot be applied to speech sounds. Some consonants are, for instance, don't have any harmonics structure because human pronounces them without vibrating his or her vocal cord. Moreover, even though vibrations are generated from same vocal cord, the structure of harmonics are changed when they pass through the mouth and then, finally, sounds a bunch of different vowels. So we can't distinguish the source(vocal cord) of speech with just looking into the harmonics structure of it.

There was another point at issue in progress of our work. Stringed instruments, such as violin, cello and viola, have multiple harmonics structures because the instruments are made of several distinguished strings and they should be viewed as independent instruments. So we have to suppose five basis vectors actually, when the input data are consist of trumpet and cello: one for trumpet and the others for four kinds of cello strings. In our experiments, the two notes were played on the same string of viola or cello. This fact tells us that we can get the separated string sounds of a stringed instrument. We expect that this results in sophisticated score which guide not only the notes and length of the sound but which string to use. A note can be played on different strings, so if the score exactly tells which one to use, the results of automatic music transcription will be finer than before.

요 약 문

비음성, 희소성 및 이동일치성을 이용한 단일 채널 음악 음원 분리

서로 다른 종류의 악기는 기본적으로 다른 배음 구조를 가지고 있다는 것과, 이러한 배음의 구조적 특성은 하나의 악기가 여러가지의 음역에서 연주되거나 실세계의 여러 다른 환경에서 녹음된다고 하더라도 대략적으로 비슷하게 유지된다는 가정을 바탕으로, 이 논문에서는 단일 채널에서 획득한 하나의 혼합 신호로부터 다성음악의 여러 음악 음원을 분리하는 방법론을 제시하고 있다. 우리는 희소성(sparseness)을 이용하여 대표 주파수 기본 벡터들을 선택하는 알고리즘을 제시하였고, 이렇게 선택된 기본 벡터들과 비음성(nonnegativity), 이동일치성(shift-invariance)을 이용하여 단일 혼합 신호로부터 음악 음원 들을 분리 및 획득하였다. 음성/첼로의 순간적인 혼합 신호와 트럼펫/비올라의 순간적인 혼합신호, 반향 있는 환경에서 혼합된 색소폰/비올라의 콘볼루션된 혼합 신호를 이용하여 제시된 방법론의 타당성을 검증하였다.

References

[1] Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons, Inc. (2001)

[2] Lee, D.D. Lee and H.S. Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401 (1999) 788.791

[3] Smaragdis, P.: Non-negative matrix factor deconvolution: Extraction of multiple sound sources from monophonic inputs. In: Proc. Int'l Conf. Independent Component Analysis and Blind Signal Separation, Granada, Spain (2004) 494.499

[4] Plumbley, M.D., Abdallah, S.A., Bello, J.P., Davies, M.E., Monti, G., Sandler, M.B.: Automatic transcription and audio source separation. Cybernetics and Systems (2002) 603.627

[5] Smaragdis, P., Brown, J.C.: Non-negative matrix factorization for polyphonic music transcription. In: Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY (2003) 177.180

[6] Abdallah, S.A., Plumbley, M.D.: Polyphonic music transcription by non-negative sparse coding of power spectra. In: Proc. Int'l Conf. Music Information Retrieval, Barcelona, Spain (2004) 318.325

[7] Helén, M., Virtanin, T.: Separation of drums from polyphonic music using nonnegative matrix factorization and support vector machine. In: Proc. European Signal Processing Conference, Antalaya, Turkey (2005)

[8] Cho, Y.C., Choi, S.: Nonnegative features of spectro-temporal sounds for classfication. Pattern Recognition Letters 26 (2005) 1327.1336

[9] Nakatani, T., Miyoshi, M.: Blind dereverberation of single channel speech signal based on harmonic structure. In: Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing. (2003)

[10] Eggert, J., Wersing, H., K¨orner, E.: Transformation-invariant representation and NMF. In: Proc. Int'l Joint Conf. Neural Networks. (2004)

[11] Kim, M., Choi, S.: On spectral basis selection for single channel polyphonic music separation. In: Proc. Int'l Conf. Artificial Neural Networks. Volume 2., Warsaw, Poland, Springer (2005) 157.162

[12] FitzGerald, D., Cranitch, M., Coyle, E.: Generalised prior subspace analysis for polyphonic pitch transcription. In: Proc. Int'l Conf. Digital Audio Effects. (2005)

[13] Hoyer, P.O.: Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research 5 (2004) 1457.1469

[14] Ru, P., Chi, T., Shamma, S.: NSL Toolbox (1997)

감 사 의 글

지난 2년간의 대학원 생활에서 얻게 된 가장 큰 가치관의 변화는 바로 자신감에 관한 것이었습니다. 이전에 생각하던 자신감이란, 단지 남에게 자신을 보여줄 때의 태도와 관련된 것이고, 그렇게 자신감을 표현함으로써 타인에게서 좋은 평가를 이끌어 내는 것은 어쩐지 편법 같다는 막연한 생각만을 하고 있던 제가, '자신감'에 대해 다르게 생각한 계기는 바로 지난 2년간의 대학원 생활이었습니다.

다른 사람들 앞에서 당당하기 위해서, 스스로가 학문적으로 뼈를 깎는 노력과 인생을 걸 수 있을 정도의 목표 의식을 가져야 한다는 사실을 알고부터, 자신감이란 단지 남에게 자신을 어떻게 표현하느냐의 문제가 아닌, 자기 자신과의 치열한 싸움으로부터 얻은 고귀한 가치라는 것을 새롭게 알게 되었습니다. 학문적으로든 인간적으로든 이러한 치열함과 과학자로서의 완벽주의, 연구에 대한 열정을 몸소 보여 주시고 저를 채찍질해 주셨던 최승진 지도 교수님께 깊은 감사와 존경의 마음을 이 자리를 빌어서 전해드립니다. 그리고 IM 랩이라는 작지 않은 조직에 미래 지향적이고 긍정적인 사명감을 불러일으켜 주신 방승양 교수님께 학자로서 뿐만이 아니라, 교육자의 귀감으로서 존경의 말씀을 전합니다. 또한 항상 열정적으로 연구하시는 김대진 교수님의 모습을 뵈면서, 게으른 제 자신을 돌아볼 수 있었습니다. 두 분 교수님들께도 진심으로 감사드립니다.

처음 랩에 왔을 때 제게 멋진 별명을 만들어주셨던 홍모형, 앞으로도 힘든 일이 있을 때는 형이 하셨던 좋은 말씀 기억하면서 살겠습니다. 재원이형, 결혼하시면 꼭 축가를 불러드리려고 했는데 이렇게 그냥 나가서 미안합니다. 앞으로도 힘든 일이 있을 때는 형의 성실하시던 모습을 기억하면서 저를 추스르겠습니다. 봉진이형, 후배가 되어서 붙임성있게 굴지 못했던 것 죄송합니다. 앞으로도 외로운 일이 있을 때는, 제게 술 한 잔 하자고 권하시던 형의 따뜻한 마음을 기억하겠습니다. 숙정누나, 친누나처럼 저를 돌봐주시어 감사합니다. 누나가 해 주셨던 많은 기대, 저버리지 않고 열심히 살겠습니다. 평생 좋은 친구로 기억하겠습니다. 재모형, 인사도 드리지 못하고 떠나서 죄송합니다. 연구에 좋은 성과 있길 빌겠습니다. 다소 건방지고 어린 후배의 말을 잘 들어주시고, 언제나 위로와 질책을 해주시던 희열이형, 건투를 빕니다. 제 앞가림 하기 급급한 힘든 상황이 닥치더라도 주변을 살피는 따뜻한 형의 마음을 기억하고 본받겠습니다. 재환이형, 함께 했던 좋은 추억들 만들어 주셔서 감사합니다. 좀이 쑤실 때면, 책상에 누구보다 오래 앉아 계시던 형의 진중함을 기억하겠습니다. 지혜누나, 버릇없이 굴어서 죄송합니다. 기분이 안 좋을 때면, 공학 2동 전체에 울려 퍼지던 누나의 웃

음소리를 기억하겠습니다. 귀찮은 물음들에도 성심성의껏 도와주시던 현철이형 감사드립니다. 열심히 연구하여 형에게서 받은 도움을 환원하며 살겠습니다. 저의 형처럼 무뚝뚝하지만, 저의 형만큼 저를 아껴주셨다고 생각되는 형수형, 많은 추억 주셔서 감사합니다. 대환이형, 운동 하자고 권하실 때 같이 하지 않아서 늘 죄송했습니다. 의젓하게 중심을 잃지 않던 형의 모습을 잘 본받겠습니다. 상기형과 혜경누나, 제일 힘들 때, 가장 곁에서 지켜봐 주셔서 감사드립니다.

함께 입학하여 힘들 때나 즐거울 때나 그것을 나누고 위로해왔던 저의 입학 동기 일곱 분들께도 진심으로 감사를 드립니다. 프로젝트와 논문으로 고생할 때도 언제나 웃음을 잃지 않고 우리의 정신적인 안식처가 되어 주신 진영이형, 감사합니다. 방돌이 동수형, 지난 일년간 형과 같이 살 수 있어서 행복했습니다. 아무리 힘들고 어려운 일이 있어도, 돌아가 쉴 수 있는 따뜻한 방을 만들어주신 것, 감사드립니다. 크고 작은 일이 있을 때 언제나 앞장서서 도맡아 하시던 근호형, 제가 시덥잖은 얘기를 할 때도 크게 웃어 주셔서 감사합니다. 종경이형, 형과 같이 부지런한 동료와 같이 공부할 수 있어서 행복했습니다. 어리석은 의문이 많을 때, 많이 도와 주셔서 감사합니다. 상재형, 형은 똑똑하신 분이니 어려운 일이 있어도 잘 헤쳐나갈 것이라고 믿습니다. 힘내세요. 박선호에게는 재미없는 농담만 한다고 면박을 많이 줘서 미안하게 생각합니다. 본심은 그렇지 않았다는 걸 알아줬으면 좋겠습니다. 똑똑하고 성실하면서도 겸손함을 잃지 않았던 우한이의 모습에서 많은 것을 배울 수 있었습니다. 어떤 곳에서나 자신이 갖고 있는 장점을 놓치지 않길 바랍니다.

후배님들께는 죄송한 마음 뿐입니다. 학문적으로든 인간적으로든 도움을 드리고 싶었는데, 제 앞가림하기 급급해서 좋은 선배가 되지 못한 것 같습니다. 필원이형, 이선호, 용덕이, 계현이는 지금 하고 있는 그대로 열심히 달리시기 바랍니다. 든든한 랩의 기둥들이 될 수 있을 것으로 믿어 의심치 않습니다. 우주와 상호, 석원이와 인호는 아마도 바쁘고 힘든 일이 많을 내년에도 지금처럼 서로 잘 의지하며 즐겁게 랩을 꾸려 주시기 부탁드립니다. 특별히 도움을 주지 못한 선배이지만, 서로 잘 지내는 모습들에 위안을 삼고 떠나겠습니다.

제가 학부에 있을 때부터 지금까지, 언제나 제 공부를 걱정해 주셨던 은사님이신 위규범 교수님께도 이 자리를 빌어 감사의 말씀을 전합니다. 교수님께서 몸소 보여주셨던 학문을 대하는 태도는 앞으로도 계속 제게 귀감이 될 것입니다. 이미 졸업하셔서 함께 랩생활을 하지는 않았지만, 지난 일년간 함께 프로젝트를 하며 여러 가지로 돌봐주신 인선누나와 제 취업에 대해 많은 조언을 해주신 용춘이형, 용진이형, 혜진누나에게도 감사드립니다. 또한 랩의 살림살이를 도맡아 해주시고, 제가 잘 모르는 여러 업무들을 도와주신 미화누나께도 감

사드립 니다. 외로운 포항 생활에서 힘들 때면 불러낼 수 있었던 좋은 벗이 되어 준 국필이 형에게도 감사합니다. 내가 없어도 외로워하지 않기 바랍니다.

대학원 생활이 힘들 때면 생각나는 따뜻한 고향 같은 곳이 있었습니다. 불쑥 찾아가도 언제나 귀 기울여 제 얘기를 들어주며, 제게 제 2의 고향을 만들어 준 수원에 있는 친구들에게 감사의 말씀을 전합니다. 이 친구들과 함께 했던 대학 4년은 앞으로도 제 인생을 풍요롭게 만들어 줄 따뜻한 경험이 될 것 같습니다. T.H.i.S, 들꽃회 여러분들, 밴드 진실 멤버 및 전국에 계신 팬 여러분들, 김민의 뮤직 살롱 회원 여러분들, 이 같이 많은 이름 으로 저와 함께해 주신 친구들에게 감사드립니다. 특히 지난 5년이 넘는 시간 제 곁에서 누구보다 저의 공부를 응원해 주고, 건강을 걱정해 주었으며, 제 마음을 따뜻하게 해 주었던 최가에게 감사의 마음을 전합니다.

마지막으로, 제가 이 자리에서 마음놓고 공부할 수 있게 된 것에는 오로지 제 가족의 사랑 덕분이라고 해도 과언이 아닙니다. 제가 성실함, 긍정적인 사고방식, 독립심, 창의력과 같은 좋은 덕목들을 조금이라도 갖고 있다면, 그것은 모두 저를 아끼고 사랑해주시는 가족들 덕분입니다. 제가 힘들어할 때 흔들리지 않게 저를 붙들어 주고, 넘어져도 손털고 다시 일어날 수 있게 해준, 제 가슴 속 가장 깊은 곳에 있는 에너지는 모두 우리 가족, 저의 부모님과 형이 제게 준 제 인생 최고의 선물입니다. 감사드립니다.