STRUCTURED NON-NEGATIVE MATRIX FACTORIZATION WITH SPARSITY PATTERNS

Hans Laurberg[1]*, Mikkel N. Schmidt*[2]*, Mads Græsbøll Christensen*[1] *and Søren Holdt Jensen*[1]

[1] Department of of Electronic Systems, Aalborg University,
Niels Jernes Vej 12, DK-9220 Aalborg, Denmark,
[2] Informatics and Mathematical Modelling, Technical University of Denmark
Richard Pedersens Plads, Bldg. 321, DK-4000 Kgs Lyngby, Denmark,
emails: {hla}@kom.aau.dk,{mns}@imm.dtu.dk

ABSTRACT

In this paper, we propose a novel algorithm for monaural blind source separation based on non-negative matrix factorization (NMF). A shortcoming of most source separation methods is the need for training data for each individual source. The algorithm proposed in this paper is able separate sources even when there is no training data for the individual sources. The algorithm makes use of models trained on mixed signals and uses training data where more than one source is active at the time. This makes the algorithm applicable to situations where recordings of the individual sources are unavailable. The key idea is to construct a structure matrix that indicates where each source is active, and we prove that this structure matrix, combined with a uniqueness assumption, is sufficient to ensure that results are equivalent to training on isolated sources. Our theoretical findings is backed up by simulations on music data that show that the proposed algorithm trained on mixed recordings performs as well as existing NMF source separation methods trained on solo recordings.

1. INTRODUCTION

Separation of a single source in a monaural recording, such as a single instrument in polyphonic music or the cocktail party problem[1] is a difficult task. An unsupervised approach is to decompose the signal into basic "atoms", and then group these to form auditory objects—se e.g. [2, 3, 4, 5, 6]. Another unsupervised approach is to form flexible source models, and fit these to the mixture—se e.g. [7, 8, 9]. A supervised approach is to learn source models from isolated recordings of each source, and use these to separate the mixture subsequently. These source models can be based on, e.g., neural networks [10, 11], factorial hidden Markov models [12, 13], vector quantization [14, 15], independent component analysis [16, 17], or non-negative matrix factorization [1, 18].

When a reasonable amount of training data with isolated sources is available, supervised, model based methods generally yield very good results; however, there are many applications where suitable training data cannot be obtained—for example in instrument separation where many instruments and singers never occur alone. Thus, to use model based methods to separate sources, it is desirable to learn source models directly from the available mixture.

In this paper, we propose a method for learning models of individual sources directly from mixture, in a single-channel source separation framework [18] based on non-negative matrix factorization (NMF). We show that, under certain conditions, training on mixtures works equally well as training on isolated sources. There has been proposed algorithms to learn source models directly from mixtures, by locating areas in the training data, where only one source is active [19]. Our approach does not require this; however, we do require areas, in which each source is *inactive*. The proposed algorithm is successfully tested on music data.

The paper is structured as follows. In Section 2, we introduce NMF and discuss its computation. Next, in Section 3, we describe a general framework for single-channel source separation based on NMF. Our proposed method for learning source models directly from mixed recordings is described in Section 4 and experimentally evaluated on music recordings in Section 5. Finally, we conclude with our conclusions in Section 6 and a detailed description of the simulations in Appendix A.

This research was supported by the Intelligent Sound project, Danish Technical Research Council grant no. 26–02–0092.

2. NON-NEGATIVE MATRIX FACTORIZATION

Non-negative matrix factorization[1] (NMF) is the approximate factorization of a non-negative matrix, $\mathbf{V} \in \mathbb{R}_+^{n\times m}$, into the product of two non-negative matrices, $\mathbf{W} \in \mathbb{R}_+^{n\times r}$ and $\mathbf{H} \in \mathbb{R}_+^{r\times m}$

$$\mathbf{V} \approx \mathbf{WH}. \tag{1}$$

In [20] a simple iterative NMF algorithm has been proposed, that minimizes

$$E(\mathbf{W},\mathbf{H}) = \|\mathbf{V} - \mathbf{WH}\|_F^2, \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Further, they have proven [21], that each iteration reduces the objective function. In addition to the Frobenius norm, numerous NMF cost functions have been suggested [22, 23], and many different algorithms for computing the NMF have been proposed—for an overview, see [24]. Much effort has been put into finding solutions that are sparse, starting with the sparse NMF method proposed by [25][2]. Many papers from different areas report that sparse NMF algorithms outperform traditional NMF algorithms, which indicates that data in those papers are sparse—se e.g. [1, 25, 26, 27, 28]. We believe that there are two reasons for the success of sparse NMF. Firstly, the NMF research has started in areas where it is known that there are understandable underlying data (which often means sparse underlying data). Secondly, if the underlying data is not sparse at all (no elements are close to zero) the NMF is not unique [29].

[1] In some literature NMF is also called non-negative matrix approximation and positive matrix factorization.

[2] In the work of Hoyer, the method is called non-negative sparse coding.

Algorithm 1 NMF source separation

1: For each source, n, compute NMF of isolated training data,

$$\mathbf{V}'_n \approx \mathbf{W}'_n\mathbf{H}'_n.$$

Store $\mathbf{W}'_n$ and discard $\mathbf{H}'_n$.

2: Compute $\mathbf{H}_1, \cdots, \mathbf{H}_N$

$$\mathbf{V} \approx \sum_{n=1}^{N} \mathbf{V}_n = \mathbf{W}'\mathbf{H} = [\,\mathbf{W}'_1 \cdots \mathbf{W}'_N\,] \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_N \end{bmatrix}$$

3: Return $\mathbf{V}_n = \mathbf{W}'_n\mathbf{H}_n$ as an estimate of the n'th source.

In this paper, we will use the sparse NMF formulation of [27] that is based on the following cost function

$$C(\mathbf{W},\mathbf{H}) = \frac{1}{2}\|V - \overline{\mathbf{W}}\mathbf{H}\|_F^2 + \lambda \sum_{i,j} \mathbf{H}_{i,j} \tag{3}$$

$$\overline{\mathbf{W}}_n = \frac{\mathbf{W}_n}{\|\mathbf{W}_n\|} \quad , n \in \{1, \ldots, r\}, \tag{4}$$

where $\mathbf{W}_n$ is the n'th column vector in $\mathbf{W}$, and the parameter λ controls the trade-off between sparsity of $\mathbf{H}$ and approximation error, $E(\mathbf{W},\mathbf{H})$.

3. SOURCE SEPARATION USING NMF

A supervised approach [18] to source separation is described in Algorithm 1. In the first step of the algorithm, training data, consisting of isolated recordings of each source, are used to build a model of each source. Step 1 in the algorithm has only to be calculated once for each source, and the computational complexity of this step is thus not crucial. For the cost function in Equation 3, Step 2 in the algorithm can be computed efficiently using quadratic programming. To ensure that the $\mathbf{W}'$ in Algorithm 1 Step 1 can be used for separation, it is desirable that the estimated $\mathbf{W}'$ is unique up to a permutation and a scaling—for further analysis of uniqueness of NMF see[29]. In [29] a NMF is called unique if all factorisations are on the form

$$\mathbf{V} = \underbrace{\mathbf{W}'}_{=\mathbf{W}\mathbf{D}^{-1}\mathbf{P}^{-1}} \underbrace{\mathbf{H}'}_{=\mathbf{P}\mathbf{D}\mathbf{H}}, \tag{5}$$

where $\mathbf{P}$ and $\mathbf{D}$ is a permutation and a scaling, respectively. So using this terminology, Algorithm 1 will produce reproducible results if all $\mathbf{V}'_n$ are unique.

4. LEARNING SOURCE MODELS FROM MIXED SOURCES

To explain the new algorithm, we start by reformulating the first step in Algorithm 1. If all training data are gathered in one matrix, say $\mathbf{V}' = [\,\mathbf{V}'_1 \cdots \mathbf{V}'_N\,]$, Step 1 can be computed for all instruments by solving

$$\mathbf{V}' \approx \mathbf{W}'\mathbf{H}' = [\,\mathbf{W}'_1 \cdots \mathbf{W}'_N\,] \begin{bmatrix} \mathbf{H}'_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{H}'_N \end{bmatrix}. \tag{6}$$

Implementing step one of Algorithm 1 in this manner is computationally inefficient, but it makes it clear, that prior knowledge of zeros in $\mathbf{H}'$ makes it possible to find $\mathbf{W}'_n$ for each source. In the following, we call a matrix $\mathbf{H}'$ with zeros in patterns a *structured* $\mathbf{H}'$ matrix, and we refer to NMF, with *structured* $\mathbf{H}$, as *structured NMF*. The following theorem shows, that most matrices $\mathbf{H}$ with structure can be used to indentify the model for each source.

Algorithm 2 Structured NMF source separation

1: Gather all training data in a data matrix

$$\mathbf{V}_{train} = [\mathbf{V}_1 \cdots \mathbf{V}_M].$$

Let $\mathbf{H}'$ be a structured matrix, and solve

$$\mathbf{V}_{train} \approx [\mathbf{W}'_1 \cdots, \mathbf{W}'_N]\mathbf{H}',$$

maintaining the structure in $\mathbf{H}$. Store $[\mathbf{W}'_1 \cdots \mathbf{W}'_N]$ and discard $\mathbf{H}'$.

2: Compute $\mathbf{H}_1, \cdots, \mathbf{H}_N$

$$\mathbf{V} \approx \sum_{n=1}^{N} \mathbf{V}_n = \mathbf{W}'\mathbf{H} = [\,\mathbf{W}'_1 \cdots \mathbf{W}'_N\,] \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_N \end{bmatrix}$$

3: Return $\mathbf{V}_n = \mathbf{W}'_n\mathbf{H}_n$ as an estimate of the n'th source.

Theorem 1 *Let*

$$\mathbf{V} = [\,\mathbf{V}_1 \cdots \mathbf{V}_N\,] = \mathbf{W}\mathbf{H} = [\,\mathbf{W}_1 \cdots \mathbf{W}_N\,] \begin{bmatrix} \mathbf{H}_1^1 & & \mathbf{H}_N^1 \\ & \ddots & \\ \mathbf{H}_1^N & & \mathbf{H}_N^N \end{bmatrix}$$

be a unique NMF, where $\mathbf{H}_n^n = \mathbf{0}$ *for all* n*, and let*

$$\mathbf{V} = \hat{\mathbf{W}}\hat{\mathbf{H}} = [\,\hat{\mathbf{W}}_1 \cdots \hat{\mathbf{W}}_N\,] \begin{bmatrix} \hat{\mathbf{H}}_1^1 & & \hat{\mathbf{H}}_N^1 \\ & \ddots & \\ \hat{\mathbf{H}}_1^N & & \hat{\mathbf{H}}_N^N \end{bmatrix}$$

be any NMF of $\mathbf{V}$*, where* $\hat{\mathbf{H}}_n^n = \mathbf{0}$ *for all* n*. If there are no* $n \neq m$ *such that* $\mathbf{H}_n^m$ *has a row of zeros then*

(a) $\mathbf{W}_n\mathbf{H}_m^n = \hat{\mathbf{W}}_n\hat{\mathbf{H}}_m^n$*, for all* n *and* m*.*

(b) For all n*, there is a permutation matrix,* $\mathbf{P}_n$*, and a diagonal scaling matrix,* $\mathbf{D}_n$*, such that* $\hat{\mathbf{W}}_n = \mathbf{W}_n\mathbf{P}_n\mathbf{D}_n$*.*

Proof outline. The NMF of $\mathbf{WH}$ is unique, and therefore $\hat{\mathbf{W}} = \mathbf{W}\mathbf{D}^{-1}\mathbf{P}^{-1}$ and $\hat{\mathbf{H}} = \mathbf{PDH}$. The proof is concluded by realizing that the permutation $\mathbf{P}$ must be block diagonal,

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{P}_N \end{bmatrix}, \tag{7}$$

in order for $\hat{\mathbf{H}}_n^n = \mathbf{0}$ for all n and therefore

$$\mathbf{PD} = \begin{bmatrix} \mathbf{P}_1\mathbf{D}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{P}_N\mathbf{D}_N \end{bmatrix}. \tag{8}$$

□

In terms of modelling sources, the theorem states, that if one wants to estimate N source models, and has N training recordings, each with one source missing, then there is a unique solution, if all source components are active in all training files where it is not missing. Theorem 1 leads naturally to Algorithm 2. The training data used in step 1 of Algorithm 2 does not need to be isolated recordings of each and, and Theorem 1 shows, that if the assumptions are fulfilled, the result is the same as for Algorithm 1. Note that step 2 and 3 of Algorithm 2 is the same as in Algorithm 1.

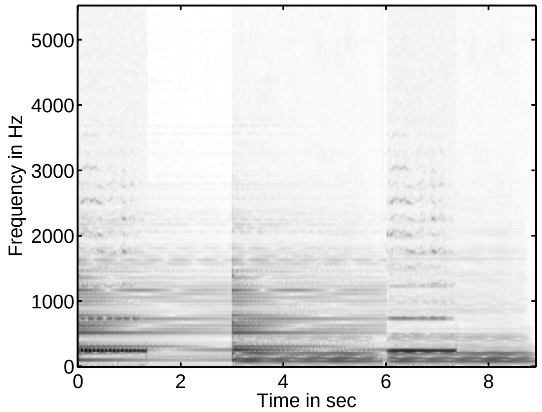

Fig. 2. The amplitude spectrogram $\mathbf{V}$ of a bas, a flute and a piano that plays together two and two.

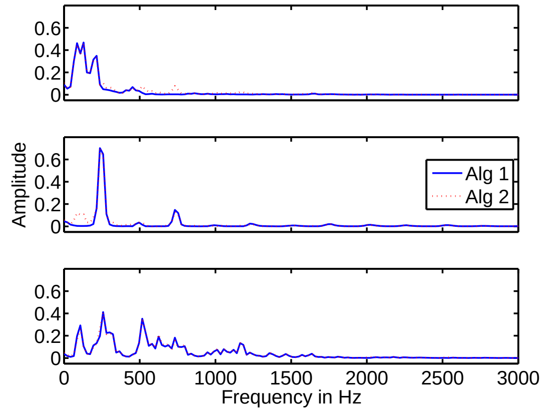

Fig. 3. The estimated basis spectras of the bas (top), flute (middle) and piano (bottom) from Figur 2.

5. RESULTS

We have constructed three tests, in which we compare Algorithm 1 and 2. Throughout the test, Algorithm 1 always has solo recordings for the training where as Algorithm 2 always use mixture recordings. The first is a simulation shows that both algorithms can separate three and four artificial sources. The second test is an example of instrument tone separation in a single channel recording of musical notes and the third test is an example of instrument separation in a single channel recording of mixed polyphonic music.

In the first test, artificial sources are separated. The data, V', is a square matrix, and each source has $\{2, 4, 6\}$ components. In Figure 1, the estimation error is shown for Algorithm 1 (trained on individual training data for each source) and Algorithm 2 (trained on mixed training data). For a detailed description of the experiment, see Appendix A. From the simulation it can be seen, that when the amount of data is sufficient, the two algorithms perform equally well. In the second test, Algorithm 2 is used on amplitude spectra of three instruments form the Iowa Music Database [30]. Each $\mathbf{V}'_n$ consist of two instruments that both plays one note. In this test the averaged cosine of the angles between the basis vectors calculated using Algorithm 2 and the basis vectors calculated using solo recordings above 0.99, which in practise means that they are equal. Figure 2 shows an example of $\mathbf{V}_{train} = [\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3]$ and there corresponding three basis spectres are shown in figure 3. It can be seen that the basis vectors are estimated almost correct even though the spectres are heavily overlapping. It can also be seen that the small errors occur at a frequency in a basis vector when the there are a lot of energy in both the other basis vectors at that frequency. A reason for this is that the NMF problem might not be unique and the non-uniqueness is that it is possible to raise the energy of one basis vector by decreasing the other basis vectors when the tones starts and stops at the same time.

In the third test, Algorithm 2 is used on amplitude spectra of midi music. The instrument models were trained on three 10-second training files, each with two instruments playing. These models were used to separate the three instruments from a 10-second evaluation file, as shown in Figure 4. In this test, the mixing of the instruments is performed in the time domain, which makes the amplitude spectra non-additive, due to phase differences, when there is overlap between the spectra. In this simple experiment, it is possible to separate the three instruments with minor artefacts. In the estimated piano the artefacts do not sound like an instrument but in the estimated drum signal, it is possible to hear the bas in the background and in the estimated bas signal there is the piano in the background. It is possible to download the sound files from our website (http://kom.aau.dk/ hla/structuredNMF).

6. CONCLUSION

An algorithm for source separation based on training source models on mixed audio recordings was presented. In contrast to existing algorithms, the proposed algorithm uses training data where more than one source is active, which makes the algorithm applicable to situations, where individual recordings of sources are unavailable.

The proposed algorithm is based on the non-negative matrix factorization (NMF), and can be used with most NMF algorithms. The novel idea in this paper is to construct a structure matrix, that indicates where each source is active, and the proof that this structure matrix, together with a uniqueness assumption, is enough to ensure, that results are equivalent to training on isolated sources. The theoretical results are backed up by simulations that show that the proposed algorithm performs as well as existing NMF source separation methods, when sufficient training data is available.

A. SIMULATION DETAILS

In the first test $\mathbf{W}$, $\mathbf{H}_{train}$ and $\mathbf{H}_{test}$ are generated as uniform IID values raised to the power of 8. All NMF calculation in this simulation use the sparse NMF algorithm [27] with $\lambda = 0.001$, 200 iterations and 20 different starting points. The error plotted in Figure 1 is a Monte Carlo simulation of the mean square error of between the test sources and the estimated test sources. There are used 20 Monte Carlo runs in the simulation. In order to make the plot more dense, the error is divided by the number of basis vectors r to compensate for different amplitudes of the matrices.

In the second test, notes with the length of one to three seconds were used and the data was downsampled to 11.025 kHz. In the third test was the sampling frequency of the sound files is 44.1 kHz. The algorithm setup for both music tests is the sparse NMF with $\lambda = 0.1$, 500 iterations, one starting point and the amplitude spectrogram of the music is calculated using a (46.4 ms) Hanning window and 50% window overlap. To estimate the instrument time signal the phase of the mixed spectrogram is used directly.

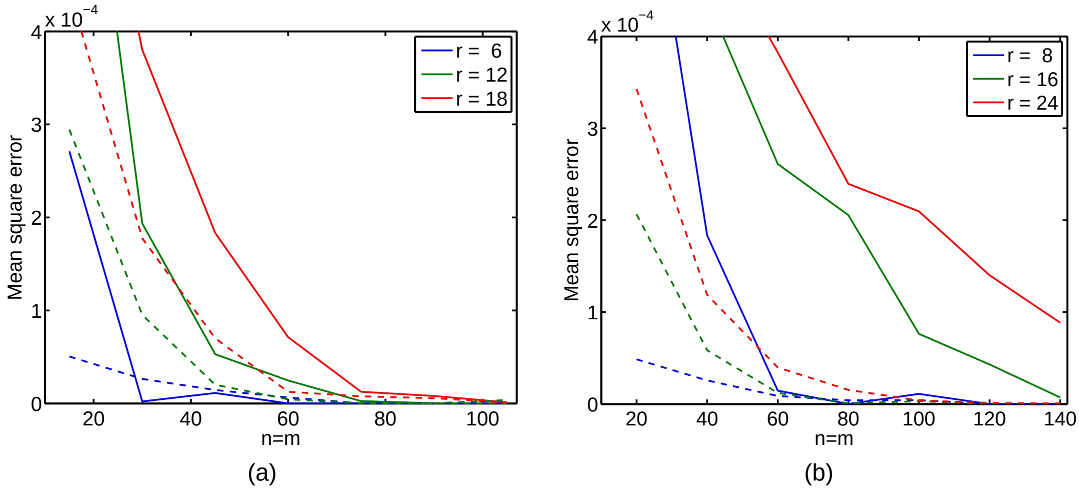

Fig. 1. The mean error of separation of (a) three sources and (b) four sources, using Algorithm 1 (dashed lines) and Algorithm 2 (solid lines). The simulation is computed with different model orders r and size of training data $m = n$.

B. REFERENCES

[1] Mikkel N. Schmidt and Rasmus K. Olsson, "Single-channel speech separation using sparse non-negative matrix factorization," in *Spoken Language Proceeing, ISCA International Conference on (INTERSPEECH)*, 2006.

[2] Tuomas Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 3, pp. 1066–1074, Mar 2007.

[3] Francis R. Bach and Michael I. Jordan, "Blind one-microphone speech separation: A spectral learning approach," in *Neural Information Processing Systems, Advances in (NIPS)*, 2005, pp. 65–72.

[4] Guoning Hu and DeLiang Wang, "On amplitude modulation for monaural speech segregation," in *Neural Networks (IJCNN), Proceeding of the International Joint Conference on*, 2002, pp. 69–74.

[5] Tero Tolonen, "Methods for separation of harmonic sound sources using sinusoidal modeling," in *Audio Engineering Society (AES) Convention*, May 1999.

[6] T. Virtanen and A. Klapuri, "Separation of harmonic sound sources using sinusoidal modeling," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on (ICASSP)*, 2000, vol. 2, p. 765.

[7] Jonathan Le Roux, Hirokazu Kameoka, Nobutaka Ono, Alain de Cheveigné, and Shigeki Sagayama, "Single channel speech and background segregation through harmonic-temporal clustering," in *Applications of Signal Processing to Audio and Acoustics, IEEE Workshop on (WASPAA)*, Oct 2007, pp. 279–282.

[8] Ron J. Weiss and Daniel P. W. Ellis, "Monaural speech separation using source-adapted models," in *Applications of Signal Processing to Audio and Acoustics, IEEE Workshop on (WASPAA)*, 2007, pp. 114–117.

[9] Mikkel N. Schmidt and Morten Mørup, "Nonnegative matrix factor 2-d deconvolution for blind single channel source separation," in *Independent Component Analysis and Blind Signal Separation, International Conference on (ICA)*, vol. 3889 of *Lecture Notes in Computer Science (LNCS)*, pp. 700–707. Springer, Apr 2006.

[10] J. Klingseisen and M. Plumbley, "Experiments on musical instrument separation using multiple cause models," in *In Cambridge Music Processing Colloquium, Cambridge, England, 30 Sept*, 1999.

[11] J. Klingseisen and M. Plumbley, "Towards musical instrument separation using multiple-cause neural networks," in *Proceedings of the International Workshop on Independent Component Analysis And Blind Signal Separation*, June 2000.

[12] Sam T. Roweis, "One microphone source separation," in *Neural Information Processing Systems, Advances in (NIPS)*, 2000, pp. 793–799.

[13] Trausti Kristjansson, John Hershey, Peder Olsen, Steven Rennie, and Ramesh Gopinath, "Super-human multi-talker speech recognition: The IBM 2006 speech separation challenge system," in *Spoken Language Proceeing, ISCA International Conference on (INTERSPEECH)*, 2006, pp. 97–100.

[14] Sam T. Roweis, "Factorial models and refiltering for speech separation and denoising," in *Speech Communication and Technology, European Conference on (EUROSPEECH)*, 2003, pp. 1009–12.

[15] Daniel P. W. Ellis and Ron J. Weiss, "Model-based monaural source separation using a vector-quantized phase-vocoder representation," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on (ICASSP)*, May 2006, pp. 957–960.

[16] Gil-Jin Jang and Te-Won Lee, "A probabilistic approach to single channel source separation," in *Neural Information Processing Systems, Advances in (NIPS)*, 2003.

[17] Gil-Jin Jang and Te-Won Lee, "A maximum likelihood approach to single-channel source separation," *Journal of Machine Learning Research*, vol. 4, pp. 1365–1392, Dec 2003.

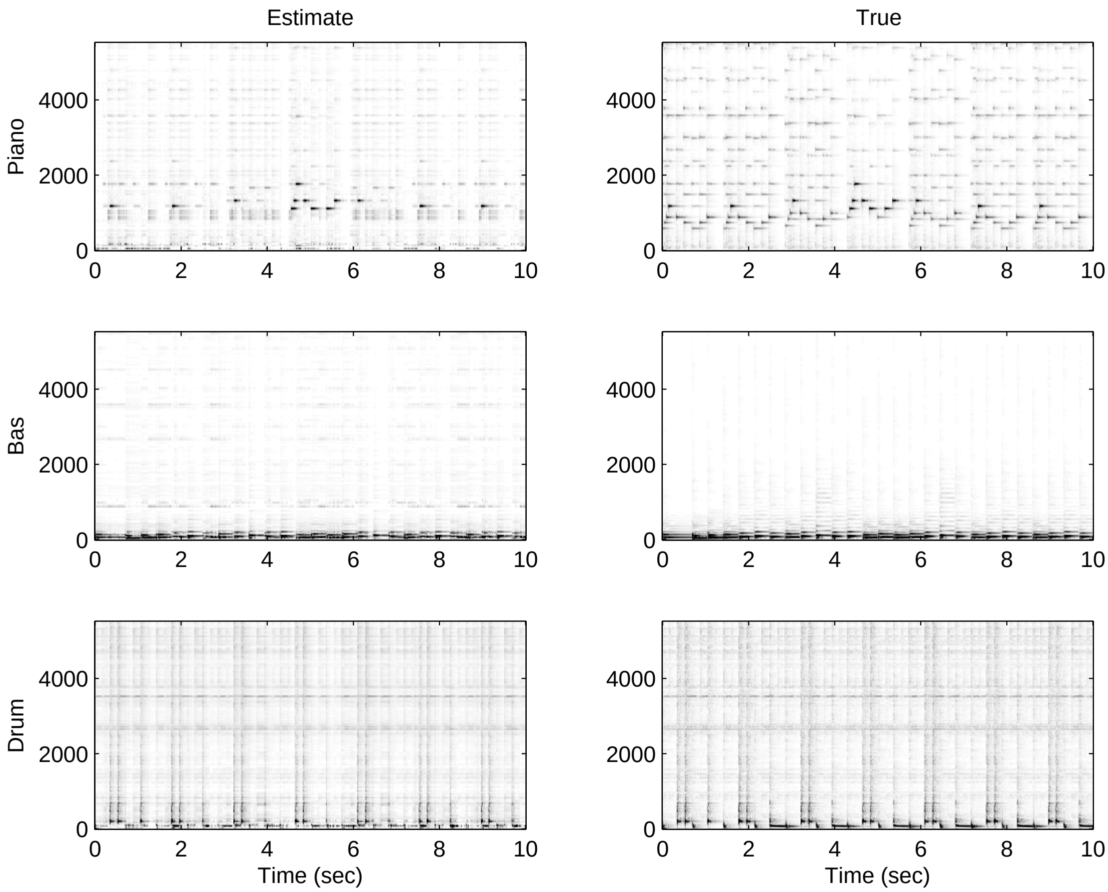

Fig. 4. The figure shows the spectrograms of the separation of a MIDI music piece with a piano (top), a bas (middle) and a drum (bottom). The left column shows the estimate and the right column shows the spectrogram of each instruments.

[18] Beiming Wang and Mark D. Plumbley, "Investigating single-channel audio source separation methods based on non-negative matrix factorization," in *ICA Research Network International Workshop, Proceedings of*, Sep 2006, pp. 17–20.

[19] Minje Kim and Seungjin Choi, "On spectral basis selection for single channel polyphonic music separation," in *Artificial Neural Networks, International Conference on (ICANN)*. Sep 2005, vol. 3697 of *Lecture Notes in Computer Science (LNCS)*, pp. 157–162, Springer.

[20] Daniel D. Lee and H. Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization.," *Nature*, vol. 401, no. 6755, pp. 788–791, October 1999.

[21] Daniel D. Lee and H. Sebastian Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*, 2000, pp. 556–562.

[22] Raul Kompass, "A generalized divergence measure for non-negative matrix factorization," *Neural Computation*, vol. 19, no. 3, pp. 780–791, 2007.

[23] Andrzej Cichocki, Rafal Zdunek, and Shun ichi Amari, "Csiszar's divergences for non-negative matrix factorization: Family of new algorithms," in *Lecture Notes in Computer Science (LNCS)*. 2006, vol. 3889, pp. 32–39, Springer.

[24] Michael W. Berry, Murray Browne, Amy N. Langville, Paul V. Pauca, and Robert J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics and Data Analysis*, vol. 52, pp. 155–173, September 2007.

[25] P. O. Hoyer, "Non-negative sparse coding," in *IEEE Proc. Neural Networks for Signal Processing*, 2002, pp. 557–565.

[26] Bruno A. Olshausen and David J. Field, "Sparse coding of sensory inputs," *Current Opinion in Neurobiology*, vol. 14, no. 4, pp. 481–487, Aug 2004.

[27] J. Eggert and E. Körner, "Sparse coding and NMF," in *Proc. Int. Joint Conf. on Neural Networks*, 2004, pp. 2529–2533.

[28] Arshia Cont, "Realtime multiple pitch observation using sparse non-negative constraints," in *Music Information Retrieval, International Conference on (ISMIR)*, October 2006.

[29] H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, and S. H. Jensen, "Theorems on positive data: On the uniqueness of nmf," *Computational Intelligence and Neuroscience*, 2008.

[30] "The university of iowa musical instrument samples database," [Online]. Available: http://theremin.music.uiowa.edu.