A SPARSE COMPONENT MODEL OF SOURCE SIGNALS AND ITS APPLICATION TO BLIND SOURCE SEPARATION

Yu Kitano, Hirokazu Kameoka†*, Yosuke Izumi, Nobutaka Ono, Shigeki Sagayama*

Graduate School of Information Science and Technology, The University of Tokyo, Japan
{kitano,izumi,onono,sagayama}@hil.t.u-tokyo.ac.jp
†NTT Communication Science Laboratories, NTT Corporation,
kameoka@eye.brl.ntt.co.jp

ABSTRACT

In this paper, we propose a new method of blind source separation (BSS) for music signals. Our method has the following characteristics: 1) the method is a combination of the sparseness-based model of source signals and the factorized basis model in nonnegative matrix factorization (NMF), 2) it is assumed that only one basis which structure source signals is active at each time-frequency bin of the observed signals, in order to degrade the degree of freedom, 3) parameter estimation algorithm is based on the EM algorithm regarding the index of the only one active basis as the hidden variable. We develop the formulation at a different point from NMF and show source separation performance in some simulation experiments.

Index Terms— Blind source separation, nonnegative matrix factorization, EM algorithm, sparseness of source signals, time-frequency masking

1. INTRODUCTION

Blind source separation (BSS) for music signals has been investigated as the basic technique of automatic music transcription, music equalizer and music information retrieval (MIR) [1]. In such case that music performance is recorded using few microphones or remixed to cut a CD, mixing process of sources is unknown. Then the aim of BSS is to separate the mixed music signals into each source.

There are roughly two approaches for BSS. One approach is to use the spatial information of sources. In overdetermined case, where there are more sensors than sources, we can separate observed signals into each source signal using independent component analysis (ICA). However music signals are often stereo recordings and there are more sources than sensors in real environments, therefore we must take into account the source separation in underdetermined case. There are many methods of BSS in underdetermined case based on the sparseness of sources, which have been discussed [2, 3, 9]. The other approach is to use the spectral information of sources. Nonnegative matrix factorization (NMF) is one of the methods to use it, and has been investigated as a separation method of monaural audio signal [6]. In the framework of NMF, it is assumed that acoustic signals in amplitude (or power) domain are composed of just so many components (music signals often satisfy this assumption.), and target signal can be factorized into the spectral patterns which structure the signal and their activations at nonnegative domain. Various extensions of NMF have been researched [5, 7, 8].

In this paper, we propose a new method of BSS for music signal. We have a new assumption that only one basis which structure sources is active at each time-frequency bin, and develop a formulation in combination of the sparseness-based model of source signals and the factorized basis model in NMF to improve separation performance by using the spatial and spectral information of sources.

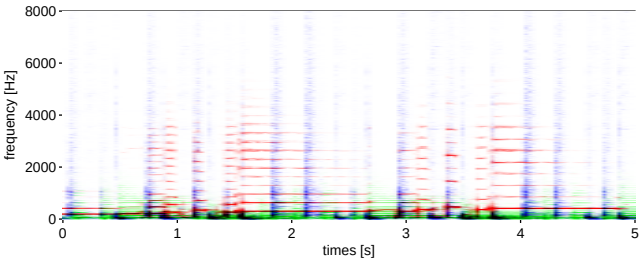

Fig. 1. Spectrogram of mixed signal which includes three sources (red, green, blue).

2. EXISTING APPROACH OF SPARSENESS-BASED BSS

The most typical method of sparseness-based BSS is time-frequency masking. Each source signal has the sparse energy in time-frequency domain as Fig. 1 shows, and it is often assumed only one of them is active on each time-frequency component. Now we consider observed signal $Y_{\omega,t}$ ($\omega = 1,\ldots,\Omega$ is a frequency bin index, and $t = 1,\ldots,T$ is a time frame index) as mixed L source signals $S^{(l)}_{\omega,t}$ ($l = 1,\ldots,L$ is a source index) in time-frequency domain by the Short-Time Fourier Transform (STFT). The mask $\Phi^{(l)}_{\omega,t}$ to extract the source signal l can be written as

$$\Phi^{(l)}_{\omega,t} = \begin{cases} 1 & (i_{\omega,t} = l) \\ 0 & (i_{\omega,t} \neq l) \end{cases}, \tag{1}$$

where $i_{\omega,t}$ is the index of the active source in (ω, t), and estimated source signal l can be written as

$$\hat{S}^{(l)}_{\omega,t} = \Phi^{(l)}_{\omega,t} Y_{\omega,t}. \tag{2}$$

The parameter $i_{\omega,t}$ can be determined by clustering the power ratios and time delays between left and right channel of observed signals at each time-frequency bin [2], or by fitting some distribution using the EM algorithm in the feature domain [3].

There is a little different approach based on the EM algorithm [9]. Let $\mathbf{Y}_{\omega,t} = (Y^{(\mathrm{L})}_{\omega,t}, Y^{(\mathrm{R})}_{\omega,t})^T$ be observed signals recorded by

two microphones in time-frequency domain. $\mathbf{Y}_{\omega,t}$ can be written as

$$\mathbf{Y}_{\omega,t} = \begin{pmatrix} 1 \\ e^{j\delta_{i_{\omega,t}}\omega} \end{pmatrix} S_{\omega,t}^{(i_{\omega,t})} + \mathbf{N}_{\omega,t}, \tag{3}$$

where δ_l is the time delay of the source signal l between two microphones and $\mathbf{N}_{\omega,t} = (N_{\omega,t}^{(\mathrm{L})}, N_{\omega,t}^{(\mathrm{R})})^T$ is the observation error. If $\mathbf{N}_{\omega,t}$ follows some distribution, each parameter can be determined by applying the EM algorithm regarding the active source index $i_{\omega,t}$ as the hidden variable. This algorithm gives the probability that each source is active in each time-frequency bin, and the probability presents the partition function and works as a soft mask for separation.

3. PROPOSED METHOD

3.1. Observation based on a sparse component model

Let $\mathbf{Y}_{\omega,t} = (Y_{\omega,t}^{(\mathrm{L})}, Y_{\omega,t}^{(\mathrm{R})})^T$ be observed signals recorded by two microphones in time-frequency domain. Approximately $\mathbf{Y}_{\omega,t}$ can be written as

$$\mathbf{Y}_{\omega,t} = \sum_{l=1}^{L} \begin{pmatrix} 1 \\ e^{j\delta_l\omega} \end{pmatrix} S_{\omega,t}^{(l)} + \mathbf{N}_{\omega,t}, \tag{4}$$

where $S_{\omega,t}^{(l)}$ is the source signal l, δ_l is the time delay of the source signal l between two microphones, and $\mathbf{N}_{\omega,t} = (N_{\omega,t}^{(\mathrm{L})}, N_{\omega,t}^{(\mathrm{R})})^T$ is the observation error which includes reverberation and background noise. And $\mathbf{N}_{\omega,t}$ is assumed to follow that

$$\mathbf{N}_{\omega,t} \sim \mathcal{N}_{\mathbb{C}}\left(0, \sigma^2\mathbf{I}\right), \tag{5}$$

where $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the complex Gaussian distribution such as

$$\boldsymbol{x} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\pi|\boldsymbol{\Sigma}|} e^{-(\boldsymbol{x}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{6}$$

Now we introduce factorized basis model of NMF to this mixture model in order to utilize the spectral information of source signals. Acoustic signals are assumed to be composed of just so many components which have their own spectral patterns in amplitude domain, and source signal $S_{\omega,t}^{(l)}$ can be factorized as

$$S_{\omega,t}^{(l)} = \sum_{k \in K_l} c_{\omega,t}^{(k)} = \sum_{k \in K_l} H_{\omega,k} U_{t,k} e^{j\phi_{\omega,t,k}}, \tag{7}$$

where $c_{\omega,t}^{(k)}$ is a component which structures acoustic signals, $H_{\omega,k}$ is a static magnitude spectral pattern called a basis, $U_{t,k}$ is a time-varying activation, and $\phi_{\omega,t,k}$ is a phase spectrum. K_l is the class of bases which structure $S_{\omega,t}^{(l)}$ and k $(k = 1, \ldots, K)$ is the index of bases. And we assume

$$\sum_{\omega} H_{\omega,k} = 1 \tag{8}$$

in order to avoid an indeterminacy in the scaling. Some of the authors proposed this factorized model and its parameter estimation algorithm in the framework called complex NMF [7].

Here, it is assumed that the component $c_{\omega,t}^{(k)}$ composed of each basis has sparse energy in time-frequency domain and that only one basis is active in (ω, t) of observed signals. Then $\mathbf{Y}_{\omega,t}$ can be written as

$$\mathbf{Y}_{\omega,t} = \begin{pmatrix} 1 \\ e^{j\delta_{l'}\omega} \end{pmatrix} H_{\omega,k'} U_{t,k'} e^{j\phi_{\omega,t,k'}} + \mathbf{N}_{\omega,t}, \tag{9}$$

where $k'_{\omega,t}$ is the index of the active basis in (ω, t), and $l'_{\omega,t}$ is the index of the source signal which is composed of the basis, $k'_{\omega,t} \in K_{l'_{\omega,t}}$. And for simplicity we abbreviate the frequency and time index of $k'_{\omega,t}, l'_{\omega,t}$. In this paper we call this model a sparse component model, and this framework is a different approach from NMF. Thanks to the assumption, the degree of freedom of observation model can be degraded, and the parameters $H_{\omega,k}$, $U_{t,k}$ can be determined uniquely.

3.2. Applying the EM algorithm to parameter estimation

We can decide these parameters in the framework of Maximum A Posteriori (MAP) estimation. For simplicity, it is defined as $\mathbf{H} \equiv (H_{\omega,k})_{\Omega\times T}$, $\mathbf{U} \equiv (U_{t,k})_{T\times K}$, $\boldsymbol{\phi} \equiv (\phi_{\omega,t,k})_{\Omega\times T\times K}$, $\Psi \equiv \{\mathbf{H}, \mathbf{U}, \boldsymbol{\phi}\}$, $\boldsymbol{\delta} \equiv \{\delta_1, \cdots \delta_L\}$. By Bayes' theorem, the posterior distribution of observed signals $\mathbf{Y}_{\omega,t}$ can be written as

$$p(\Psi, \boldsymbol{\delta}, \sigma^2|\mathbf{Y}) \propto p(\mathbf{Y}|\Psi, \boldsymbol{\delta}, \sigma^2) p(\Psi, \boldsymbol{\delta}, \sigma^2), \tag{10}$$

where $p(\mathbf{Y}|\Psi, \boldsymbol{\delta}, \sigma^2)$ and $p(\Psi, \boldsymbol{\delta}, \sigma^2)$ are the likelihood and prior distribution of the parameters Ψ. For convenience, we assume that the prior distribution of $\mathbf{H}$, $\mathbf{U}$, $\boldsymbol{\phi}$, σ^2 and $\boldsymbol{\delta}$ are independent, that $p(\mathbf{H})$, $p(\boldsymbol{\phi})$, $p(\sigma^2)$ and $p(\boldsymbol{\delta})$ are uniform, and that $\mathbf{U}$ follows a generalized Gaussian distribution given by

$$p(\mathbf{U}) = \prod_{t,k} \frac{1}{2\Gamma(1+p^{-1})b} e^{-\frac{|U_{t,k}|^p}{b^p}}, \tag{11}$$

where p $(0 < p \leq 2)$ and b (> 0) are the parameters that decide the shape of the distribution. The distribution promotes sparsity of $\mathbf{U}$ so that source signals can be composed of as few bases as possible. Then our goal is to obtain the parameter $\theta \equiv \{\Psi, \boldsymbol{\delta}, \sigma^2\}$ that maximizes

$$\Xi(\theta) \equiv \log p(\mathbf{Y}|\Psi, \boldsymbol{\delta}, \sigma^2) + \log p(\mathbf{U}). \tag{12}$$

It is very difficult to maximize $\Xi(\theta)$ directly because the logarithmic posteriori distribution has hidden variable k' in eq. (9), however the parameter θ can be determined applying the EM algorithm as same way as Izumi *et al* [9]. For parameter estimation including missing data, the EM algorithm introduces an auxiliary function called the Q function defined using a tentative parameter. On the assumption that only one basis is dominant at each time-frequency bin, the likelihood $p(\mathbf{Y}_{\omega,t}|\theta)$ can be written as

$$p(\mathbf{Y}_{\omega,t}|\theta) = \frac{p(\mathbf{Y}_{\omega,t}, k'_{\omega,t}|\theta)}{p(k'_{\omega,t}|\mathbf{Y}_{\omega,t}, \theta)} = \frac{p(\mathbf{Y}_{\omega,t}|\Psi_{k'}, \delta_{l'}, \sigma^2)}{p(k'_{\omega,t}|\mathbf{Y}_{\omega,t}, \theta)}, \tag{13}$$

where $k'_{\omega,t}$ represents the index of the dominant basis in (ω, t) and $l'_{\omega,t}$ satisfies $k'_{\omega,t} \in K_{l'_{\omega,t}}$. $p(\mathbf{Y}_{\omega,t}|\Psi_k, \delta_l, \sigma^2)$ represents the likelihood of the observation $\mathbf{Y}_{\omega,t}$ when the basis k is in the direction $\delta_l (k \in K_l)$, and can be written as

$$\log p(\mathbf{Y}_{\omega,t}|\Psi_k, \delta_l, \sigma^2) = -\log(\pi\sigma^4) - \frac{1}{\sigma^2}\left|\mathbf{Y}_{\omega,t} - \boldsymbol{a}_{\omega}^{(l)} H_{\omega,k} U_{t,k} e^{j\phi_{\omega,t,k}}\right|^2, \tag{14}$$

where $\boldsymbol{a}_{\omega}^{(l)} = (1, \exp(j\delta_l\omega))^T$. From eqs. (11) - (14), the Q function can be written as

$$Q(\theta, \theta^{(n)}) = \sum_{k,\omega,t} m_{k,\omega,t}^{(n)} \log p(\mathbf{Y}_{\omega,t}|\Psi_k, \delta_l, \sigma^2) - \sum_{t,k} \frac{U_{t,k}^p}{b^p}, \tag{15}$$

where

$$m_{k,\omega,t}^{(n)} = p(k|\Psi^{(n)}, \delta^{(n)}, \sigma^{2(n)}, \mathbf{Y}_{\omega,t}) \quad (16)$$

$$= \frac{p(\mathbf{Y}_{\omega,t}|\Psi_k^{(n)}, \delta_l^{(n)}, \sigma^{2(n)})}{\sum_{k'} p(\mathbf{Y}_{\omega,t}|\Psi_{k'}^{(n)}, \delta_{l'}^{(n)}, \sigma^{2(n)})}. \quad (17)$$

The parameter θ is estimated by sequential iteration of two steps:

- E-step: calculate $Q(\theta, \theta^{(n)})$ (update $m_{\omega,t,k}^{(n)}$ by eq.(17))
- M-step: update $\theta^{(n)}$ by $\theta^{(n+1)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta, \theta^{(n)})$.

And we are led to get the update rules of θ by differentiating the Q function with respect to each parameter. Update rules of the parameters except $m_{\omega,t,k}^{(n)}$ are summarized as follows:

$$H_{\omega,k}^{(n+1)} = \frac{\sum_t m_{k,\omega,t}^{(n)} U_{t,k}^{(n)} \left|\mathbf{Y}_{\omega,t}^H (\boldsymbol{a}_\omega^{(l)})^{(n)}\right|}{2\sum_t m_{k,\omega,t}^{(n)} (U_{t,k}^{(n)})^2}, \quad (18)$$

$$U_{t,k}^{(n+1)} = \frac{\sum_\omega m_{k,\omega,t}^{(n)} H_{\omega,k}^{(n)} \left|\mathbf{Y}_{\omega,t}^H (\boldsymbol{a}_\omega^{(l)})^{(n)}\right|}{2\sum_\omega m_{k,\omega,t}^{(n)} (H_{\omega,k}^{(n)})^2 + \lambda p \sigma^2 (U_{t,k}^{(n)})^{p-2}}, \quad (19)$$

$$\phi_{\omega,t,k}^{(n+1)} = \arg\left(\mathbf{Y}_{\omega,t}^H (\boldsymbol{a}_\omega^{(l)})^{(n)}\right)^H, \quad (20)$$

$$\delta_l = \underset{\delta_l}{\operatorname{argmax}}\, Q(\theta, \theta^{(n)}), \quad (21)$$

$$\left(\sigma^2\right)^{(n+1)} = \sum_{k,\omega,t} \frac{m_{k,\omega,t}^{(n)}}{2\Omega T} \left|\boldsymbol{\xi}_{\omega,t,k}^{(n)}\right|^2, \quad (22)$$

where $\lambda = b^{-p}$, $\boldsymbol{\xi}_{\omega,t,k}^{(n)} = \mathbf{Y}_{\omega,t} - (\boldsymbol{a}_\omega^{(l)})^{(n)} H_{\omega,k}^{(n)} U_{t,k}^{(n)} e^{j\phi_{\omega,t,k}^{(n)}}$. The update of δ_l is done by calculating the Q function for all the discrete δ_l and selecting the maximum because update rule cannot be solved analytically. In addition, after every update of $\mathbf{H}$, it is standardized as satisfies eq. (8). And from these update rule, the non-negativity of $\mathbf{H}$ and $\mathbf{U}$ is preserved if we start with nonnegative initial conditions for them.

3.3. Mask design based on Wiener filter for separation

There are several methods to extract source signals from observed signals using the estimated parameter θ. Here we design a mask based on Wiener filter for separation in order to utilize the information of power spectra that source signals have. Now the expectation of power spectrum of source signal l can be calculated as

$$E\left[|S_{\omega,t}^{(l)}|^2\right] = \sum_{k \in K_l} p(k)|c_{\omega,t}^{(k)}|^2 + \sum_{k' \notin K_l} p(k') \cdot 0 \quad (23)$$

$$= \sum_{k \in K_l} m_{k,\omega,t} (H_{\omega,k} U_{t,k})^2, \quad (24)$$

so source signal $S_{\omega,t}^{(l)}$ can be written as

$$\hat{S}_{\omega,t}^{(l)} = \frac{E\left[|S_{\omega,t}^{(l)}|^2\right]}{\sum_l E\left[|S_{\omega,t}^{(l)}|^2\right]} \cdot \frac{(\boldsymbol{a}_\omega^{(l)})^H \mathbf{Y}_{\omega,t}}{2}, \quad (25)$$

$$= \frac{\sum_{k \in K_l} m_{k,\omega,t} (H_{\omega,t} U_{t,k})^2}{\sum_k m_{k,\omega,t} (H_{\omega,t} U_{t,k})^2} \cdot \frac{(\boldsymbol{a}_\omega^{(l)})^H \mathbf{Y}_{\omega,t}}{2}. \quad (26)$$

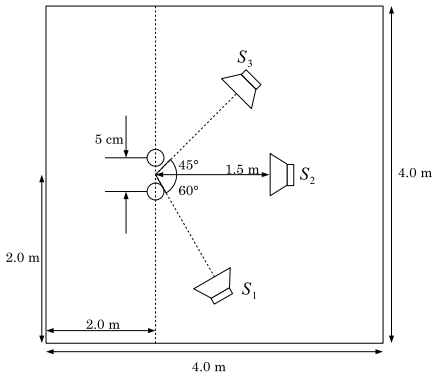

Fig. 2. alignment of sources and sensors.

And our framework can estimate not only source signals but also each component. In the same way as eq. (26), a component $c_{\omega,t}^{(k)}$ of source l can be written as

$$\hat{c}_{\omega,t}^{(k)} = \frac{m_{k,\omega,t} (H_{\omega,t} U_{t,k})^2}{\sum_k m_{k,\omega,t} (H_{\omega,t} U_{t,k})^2} \cdot \frac{(\boldsymbol{a}_\omega^{(l)})^H \mathbf{Y}_{\omega,t}}{2}. \quad (27)$$

4. EXPERIMENT

We evaluated the separation performance of proposed method in reverberant environments through simulations. There were three sources and two microphones in a simulated room illustrated Fig. 2. The observed signals were calculated by mirror method. We synthesized the source signals of RWC database using MIDI (Musical Instrument Digital Interface). The instruments of sources are Tenor Sax, Bass, and Piano. STFT was calculated using Hanning window that was 64ms long with a 32ms overlap at a sampling rate of 16kHz. The algorithm was run for 40 iterations. The parameters were set up at $p = 2$, $\lambda = 0.01$, $K = 30$. How to give the initial values of $\mathbf{H}$ and $\mathbf{U}$ is as following:

1. Initialize $\mathbf{H}$ and $\mathbf{U}$ by applying NMF to the observed signals,
2. Reconstruct components at left and right channels of each factorized basis using the phase spectrum of the observed signals,
3. Classify K components into source signals by comparing signals of left and right channel using cross-correlation.

As a comparison, we used the DUET by Yilmaz *et al.* [2] and the soft masking method using the EM algorithm by Izumi *et al.* [9]. We used the improvement of SNR and SIR as measures of source separation performance. In this paper, SIR and SNR were calculated by $\mathrm{SIR}_i = |< s_i, \hat{s}_i >| / |\sum_{j \neq i}| < s_j, \hat{s}_i > |$ and $\mathrm{SNR}_i = |s_i'|/|\hat{s}_i - s_i'|$, where s_i represents a vector of the true i^{th} source signal in the time domain, $\hat{s}_i$ is the estimated one and $s_i' = s_i \cdot < s_i, \hat{s}_i > /|\hat{s}_i|^2$. Table 1 ,2 and Fig 3 show results. While in case that $T_R = 66.7$ ms Izumi's method and proposed method show better performance for source separation, in case of long reverberation time, proposed method has best performance among these methods.

Fig. 4 shows the spectrogram of source signal $S_{\omega,t}^{(l)}$, estimated source $\hat{S}_{\omega,t}^{(l)}$ and a component of the source $c_{\omega,t}^{(k)}$. We can see a spectral pattern on the spectrogram of $c_{\omega,t}^{(k)}$. Then it is indicated that our method can not only separate the observed signals into sources but also factorize the bases which structure the observed signals.

Table 1. Source separation results (SNR[dB])

Condition	Method	s_1	s_2	s_3	Avg.
$T_R = 66.7$[ms]	DUET	15.0	5.9	10.0	10.3
	Izumi	13.3	9.9	16.1	13.1
	proposed	11.7	4.7	9.4	8.6
$T_R = 273$[ms]	DUET	12.7	-3.1	3.4	4.3
	Izumi	8.1	5.9	13.3	9.1
	proposed	10.1	8.5	14.6	11.0

Table 2. Source separation results (SIR[dB])

Condition	Method	s_1	s_2	s_3	Avg.
$T_R = 66.7$[ms]	DUET	31.4	19.4	20.4	23.7
	Izumi	31.3	19.8	30.2	27.1
	proposed	24.4	25.1	19.6	23.0
$T_R = 273$[ms]	DUET	32.0	12.3	6.1	16.8
	Izumi	32.9	11.1	28.8	24.2
	proposed	28.5	18.8	33.2	26.8

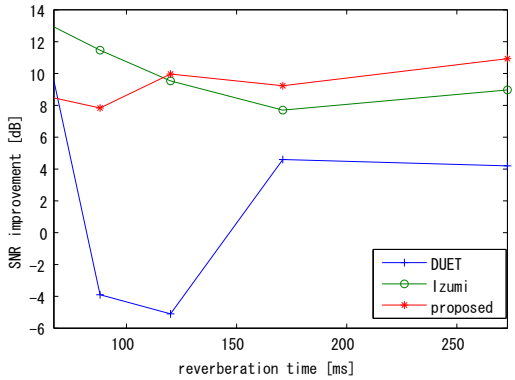

Fig. 3. Average SNR improvement under various conditions.

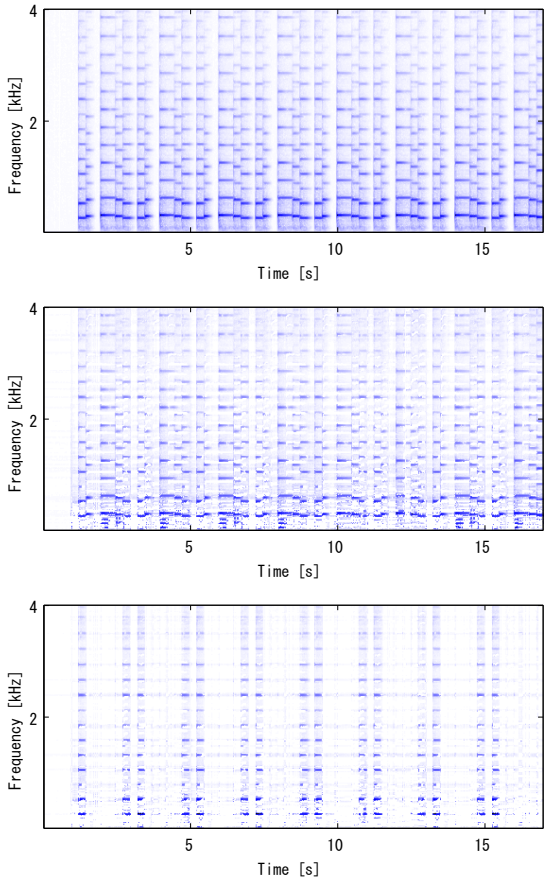

Fig. 4. Spectrogram of source signal (top) and estimated signal (center) and an estimated component of the source (bottom).

5. CONCLUSION

In this paper, we proposed a new method of source separation for music signals based on the assumption that only one basis which structures source signals is active at each time-frequency bin, and developed a parameter estimation method based on the EM algorithm in the same way as [9]. The hidden variable corresponds to the active basis index in our formulation. Our approach is to use the spatial and spectral information, and we confirmed that our method has better performance to separate the observations in long reverberant environments through a simulation experiment, and that bases which structure sources can be learned in our framework. We plan to introduce a noise model appropriate for reverberation and extend the bases to 2-dimensional components with temporal structure like [5].

6. REFERENCES

[1] E. Vincent, H. Sawada, P. Bofill, S. Makino and J. P. Rosca, "First Stereo Audio Source Separation Evaluation Campaign: Data, Algorithms and Results," *Proc. ICA'07*, pp. 552–559, 2007.

[2] O. Yilmaz and S. Rickard, "Blind Separation of Speech Mixtures via Time-Frequency Masking," *IEEE Transaction on Signal Processing*, Vol. 52, No. 7, pp. 1830–1847, 2004.

[3] M. Mandel, D. Ellis and T. Jebara, " An EM algorithm for localizing multiple sound sources in reverberant environments," *Proc, NIPS'06*, 2006.

[4] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," *Proc. NIPS'00*, pp. 556–562, 2000.

[5] P. Smaragdis, "Non-negative Matrix Factor Deconvolution; Extraction of Multiple Sound Sources from Monophonic Inputs," *Proc. ICA'04*, pp. 494–499, 2004.

[6] T. Virtanen, "Monaural Sound Source Separation by Non-Negative Matrix Factorization with Temporal Continuity and Sparseness Criteria," *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 15, No. 3, pp. 1066–1074, 2006.

[7] H. Kameoka, N. Ono, K. Kunio and S. Sagayama, "Complex NMF: A New Sparse Representation for Acoustic Signals," *Proc. ICASSP'09*, pp. 3437–3440, 2009.

[8] A. Ozerov and C. Févotte, " Multichannel Nonnegative Matrix Factorization in Convolutive Mixtures. With Application to Blind Audio Source Separation," *Proc. ICASSP'09*, pp. 3137–3140, 2009.

[9] Y. Izumi, N. Ono and S. Sagayama, "Sparseness-Based 2ch BSS using the EM Algorithm in Reverberant Environment," *Proc. WASPAA'07*, pp. 147–150, 2007.