RELAXATION OF RANK-1 SPATIAL CONSTRAINT IN OVERDETERMINED BLIND SOURCE SEPARATION

Daichi Kitamura[*], *Nobutaka Ono*[†*], *Hiroshi Sawada*[‡], *Hirokazu Kameoka*[‡§], *Hiroshi Saruwatari*[§]

[*] SOKENDAI (The Graduate University for Advanced Studies), Kanagawa, Japan
[†] National Institute of Informatics, Tokyo, Japan
[‡] Nippon Telegraph and Telephone Corporation, Tokyo, Japan
[§] The University of Tokyo, Tokyo, Japan

ABSTRACT

In this paper, we propose a new algorithm for overdetermined blind source separation (BSS), which enables us to achieve good separation performance even for signals recorded in a reverberant environment. The proposed algorithm utilizes extra observations (channels) in overdetermined BSS to estimate both direct and reverberant components of each source. This approach can relax the rank-1 spatial constraint, which corresponds to the assumption of a linear time-invariant mixing system. To confirm the efficacy of the proposed algorithm, we apply the relaxation of the rank-1 spatial constraint to conventional BSS techniques. The experimental results show that the proposed algorithm can avoid the degradation of separation performance for reverberant signals in some cases.

Index Terms— Blind source separation, overdetermined, nonnegative matrix factorization, rank-1 spatial constraint

1. INTRODUCTION

Blind source separation (BSS) is a technique for separating specific sources from a recorded sound without any information. In a determined or overdetermined case (number of microphones ≥ number of sources), independent component analysis (ICA) [1] is the method most commonly used, and many ICA-based techniques have been proposed [2, 3]. For an underdetermined case (number of microphones < number of sources), nonnegative matrix factorization (NMF) [4] has received much attention. BSS is generally used to solve speech separation problems [5], but recently the use of BSS for music signals has also become an active research area [6].

To solve the BSS problem even in an underdetermined case, multichannel NMF (MNMF) has been proposed [7, 8]. MNMF estimates a mixing system for the sources so that the decomposed bases (spectral patterns) can be clustered into specific sources. However, MNMF has a high computational cost and sometimes lacks robustness because of its dependence on the initial values.

This work was partially supported by Grant-in-Aid for JSPS Fellows Grant Number 26·10796.

For an overdetermined case, independent vector analysis (IVA) [9], which is an extension of frequency domain ICA (FDICA), has been proposed. We have also proposed an efficient algorithm of MNMF with the rank-1 spatial constraint (Rank-1 MNMF) [10]. These methods estimate a demixing matrix while assuming linear time-invariant mixing in the time-frequency domain. This assumption corresponds to the rank-1 spatial constraint. However, for reverberant signals, the separation performance of these methods markedly degrades because the rank-1 spatial assumption is not valid.

In this paper, we propose a new algorithm for overdetermined BSS, which enables us to achieve good separation performance even for reverberant signals. The algorithm utilizes extra observations (channels) to estimate the reverberant components of each source. The efficacy of the proposed algorithm is experimentally confirmed using music signals.

2. CONVENTIONAL METHODS

2.1. Linear time-invariant assumption

Let the numbers of sources and observations (channels) be N and M, respectively. The multichannel sources, observed signal, and estimated (separated) sources in each time-frequency slot are described as

$$\boldsymbol{s}_{ij} = (s_{ij,1} \cdots s_{ij,N})^{\mathrm{T}}, \tag{1}$$

$$\boldsymbol{x}_{ij} = (x_{ij,1} \cdots x_{ij,M})^{\mathrm{T}}, \tag{2}$$

$$\boldsymbol{y}_{ij} = (y_{ij,1} \cdots y_{ij,N})^{\mathrm{T}}, \tag{3}$$

where $i = 1, \ldots, I$; $j = 1, \ldots, J$; $n = 1, \ldots, N$; and $m = 1, \ldots, M$ are the integral indexes of the frequency bins, time frames, sources, and channels, respectively, $^{\mathrm{T}}$ denotes a vector transpose, and all the entries of the vectors are complex values. If we assume that the mixing system is linear time-invariant, we can define an $M \times N$ mixing matrix $\boldsymbol{A}_i = (\boldsymbol{a}_{i,1} \cdots \boldsymbol{a}_{i,N})$ ($\boldsymbol{a}_{i,n}$ denotes a steering vector) at each frequency. The observed signal $\boldsymbol{x}_{ij}$ is represented as

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij}. \tag{4}$$

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

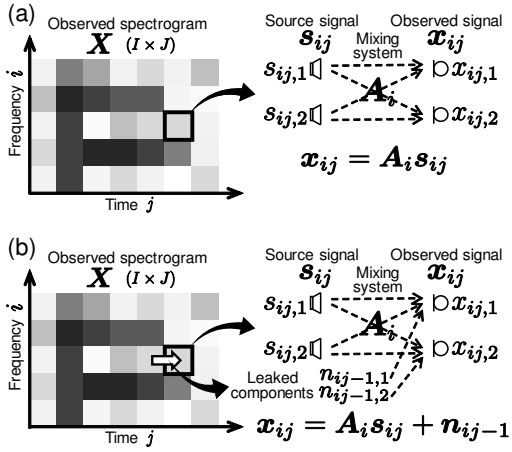

Fig. 1. Mixing system of each spectrogram slot when $N = M = 2$; (a) has a linear time-invariant mixing system and there is no reverberation; (b) has some leaked components from the previous frame because of reverberation.

Figure 1 (a) shows the mixing system corresponding to (4). In linear time-invariant mixing, all the time frames are independent of other time frames, meaning that they do not affect each other. However, for the case of reverberant recording, reverberant components can leak from the previous frame as shown in Fig. 1 (b), and the mixed signal $\boldsymbol{x}_{ij}$ cannot be represented using only $\boldsymbol{A}_i$. Therefore, the assumption of linear time-invariant mixing holds only when the lengths of all impulse responses between the sources and microphones are sufficiently shorter than the length of the window function in the short-time Fourier transform (STFT).

When the assumption is valid and $M = N$, the estimated source $\boldsymbol{y}_{ij}$ can be represented by a demixing matrix $\boldsymbol{W}_i = (\boldsymbol{w}_{i,1} \cdots \boldsymbol{w}_{i,N})^{\mathrm{H}}$ ($\boldsymbol{w}_{i,n}$ denote demixing filters) as

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij}, \tag{5}$$

where $^{\mathrm{H}}$ denotes a Hermitian transpose.

2.2. Principal component analysis for overdetermined BSS

When $M > N$, in a typical separation method using FDICA or IVA, principal component analysis (PCA) is applied in advance and the dimension of $\boldsymbol{x}_{ij}$ is reduced so that $M = N$. This preprocessing is performed with the expectation that the reverberant components in the observed signal are eliminated by the dimensionality reduction. Therefore, PCA is applied to make the assumption of linear time-invariant mixing valid even in a reverberant environment. However, if the purpose of source separation is to obtain each source image including the reverberation, PCA degrades the separation performance by removing the reverberation components. Moreover, if the source powers in mixtures are unbalanced (e.g., music signals), PCA can even remove direct components of weak sources, which leads to a greater risk of poor separation.

The assumption of linear time-invariant mixing is made valid by using a sufficiently long window function in the STFT. However, if we use a too long window function for FDICA or IVA, the independence assumption collapses in each frequency band [11]. Therefore, the separation performance has a trade-off based on the length of the window function in terms of the assumptions of linear time-invariant mixing and the independence of sources.

2.3. MNMF with rank-1 spatial constraint

In MNMF [8], the decomposition model of an observed signal $\mathsf{X}_{ij} = \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}}$ is represented as

$$\mathsf{X}_{ij} \simeq \hat{\mathsf{X}}_{ij} = \textstyle\sum_k \left(\sum_n \mathsf{H}_{i,n} z_{nk}\right) t_{ik} v_{kj}, \tag{6}$$

where $k = 1, \ldots, K$ is the integral index of the NMF bases (spectral patterns), $\mathsf{H}_{i,n}$ is an $M \times M$ spatial covariance matrix for frequency i and source n, z_{nk} ($\in \mathbb{R}_{[0,1]}$) is a latent variable that clusters K bases into N sources and satisfies $\sum_n z_{nk} = 1$, and t_{ik} ($\in \mathbb{R}_{\geq 0}$) and v_{kj} ($\in \mathbb{R}_{\geq 0}$) are the elements of the basis matrix $\boldsymbol{T}$ ($\in \mathbb{R}_{\geq 0}^{I \times K}$) and activation matrix $\boldsymbol{V}$ ($\in \mathbb{R}_{\geq 0}^{K \times J}$), respectively. MNMF estimates the spatial covariance matrix H corresponding to each source and the source components $\boldsymbol{TV}$. The estimated source $\boldsymbol{y}$ is obtained by clustering $\boldsymbol{TV}$ into H using cluster indicator $\boldsymbol{Z}$ ($\in \mathbb{R}_{[0,1]}^{N \times K}$). The variables H, $\boldsymbol{Z}$, $\boldsymbol{T}$, and $\boldsymbol{V}$ are estimated by minimizing the divergence between X_{ij} and $\hat{\mathsf{X}}_{ij}$ [8]. However, this optimization has a high computational cost, and the separation results strongly depend on the initial values.

As an efficient optimization method for MNMF, Rank-1 MNMF has been proposed [10]. In this method, we assume an overdetermined case, $M \geq N$, and the spatial covariance $\mathsf{H}_{i,n}$ is approximated by a rank-1 matrix. This approximation corresponds to the assumption of linear time-invariant mixing. Rank-1 MNMF can estimate the demixing matrix $\boldsymbol{W}_i$ using fast IVA update rules [12] and the NMF variables $\boldsymbol{T}$ and $\boldsymbol{V}$ using simple NMF update rules. When $N = M$, the fast update rules of IVA are obtained as [12]

$$V_{i,n} = J^{-1} \textstyle\sum_j \left(\sum_l t_{il,n} v_{lj,n}\right)^{-1} \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}}, \tag{7}$$

$$\boldsymbol{w}_{i,n} \leftarrow (\boldsymbol{W}_i V_{i,n})^{-1} \boldsymbol{e}_n, \tag{8}$$

$$\boldsymbol{w}_{i,n} \leftarrow \boldsymbol{w}_{i,n} \left(\boldsymbol{w}_{i,n}^{\mathrm{H}} V_{i,n} \boldsymbol{w}_{i,n}\right)^{-\frac{1}{2}}, \tag{9}$$

$$y_{ij,n} = \boldsymbol{w}_{i,n}^{\mathrm{H}} \boldsymbol{x}_{ij}, \tag{10}$$

where $\boldsymbol{e}_n$ denotes the unit vector with the nth element equal to unity. The update rules of NMF are obtained as

$$t_{il,n} \leftarrow t_{il,n} \sqrt{\frac{\sum_j |y_{ij,n}|^2 v_{lj,n} \left(\sum_{l'} t_{il',n} v_{l'j,n}\right)^{-2}}{\sum_j v_{lj,n} \left(\sum_{l'} t_{il',n} v_{l'j,n}\right)^{-1}}}, \tag{11}$$

$$v_{lj,n} \leftarrow v_{lj,n} \sqrt{\frac{\sum_i |y_{ij,n}|^2 t_{il,n} \left(\sum_{l'} t_{il',n} v_{l'j,n}\right)^{-2}}{\sum_i t_{il,n} \left(\sum_{l'} t_{il',n} v_{l'j,n}\right)^{-1}}}, \tag{12}$$

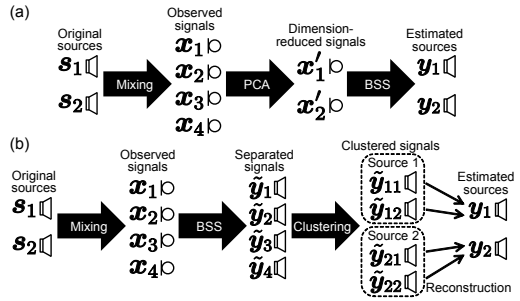

Fig. 2. Algorithms of (a) conventional and (b) proposed methods ($N=2, M=4, P=2$).

where $l = 1, \ldots, L$ is the integral index of the NMF bases for each source, and $t_{il,n}$ and $v_{lj,n}$ are the basis and its activation that represent only source n, respectively. Similarly to (6), we can extend Rank-1 MNMF so that the bases for each source are adaptively determined by the latent variable $\boldsymbol{Z}$ [10].

In Rank-1 MNMF, we can optimize all the variables $\boldsymbol{W}_i$, $\boldsymbol{T}$, and $\boldsymbol{V}$ faster and more robustly than in conventional MNMF. However, if the reverberation time of the recorded environment increases, the separation performance markedly degrades because the approximation of the rank-1 spatial model collapses. Conventional MNMF can achieve a certain level of separation even for reverberant signals because this method can estimate full-rank spatial covariance matrix $\mathsf{H}_{i,n}$.

3. PROPOSED METHOD

3.1. Relaxation of rank-1 spatial constraint utilizing extra observations

To relax the constraint of the rank-1 spatial model in Rank-1 MNMF, we propose the utilization of extra observations for modeling the reverberant components. In this method, we consider that the number of observations M is P times the number of sources N, namely, $M = PN$. In conventional overdetermined BSS, PCA is applied before the separation so that M equals N as shown in Fig. 2 (a). In the proposed algorithm, we estimate M separated signals $\tilde{\boldsymbol{y}}$ as shown in Fig. 2 (b). In this approach, the leaked component from previous frames ($\boldsymbol{n}_{ij-1}$ in Fig. 1 (b)) of each source is modeled as an additional new source, namely, each original source is represented with rank-P spatial model. To obtain an estimate of the source including both direct and reverberant components, the separated signals must be clustered using some criteria, which is a kind of permutation problem. The clustered separated signal $\tilde{\boldsymbol{y}}$ is represented as follows:

$$\tilde{\boldsymbol{y}}_{ij} = \left(\tilde{y}_{ij,11} \cdots \tilde{y}_{ij,1P}\ \tilde{y}_{ij,21} \cdots \tilde{y}_{ij,2P} \cdots \tilde{y}_{ij,NP}\right)^{\mathrm{T}}, \quad (13)$$

$$y_{ij,n} = \textstyle\sum_p \tilde{y}_{ij,np}, \quad (14)$$

where $\tilde{y}_{ij,n1}, \ldots, \tilde{y}_{ij,nP}$ correspond to the direct and reverberant components of one source n. Finally, each estimated source $y_{ij,n}$ is reconstructed by summing of the clustered components as represented by (14).

3.2. Clustering with spectral correlations

In Sect 3.1, the complex-valued spectrograms of the sources are estimated by assuming the independence between them. However, we can expect that the power spectrograms of the direct and the reverberant components for the same source have a correlation. Based on this assumption, we propose to use cross-correlation between the power spectrograms $\tilde{Y}_{ij,np} = |\tilde{y}_{ij,np}|^2$ to determine which separated signal $\tilde{y}_{ij,np}$ corresponds to the direct or reverberant component of which source:

$$C(\boldsymbol{A}\|\boldsymbol{B}) = \max\left(\left\{\textstyle\sum_{i,j} a_{ij} b_{ij+\tau} \mid \tau = 0, 1, \ldots, \tau_{\max}\right\}\right), \quad (15)$$

where $\boldsymbol{A}$ ($\in \mathbb{R}_{\geq 0}^{I \times J}$) and $\boldsymbol{B}$ ($\in \mathbb{R}_{\geq 0}^{I \times J}$) are the power spectrograms, a_{ij} and b_{ij} denote the elements of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, and τ is an index of the delay in the time frame. For clustering, we first calculate (15) between all separated signals $\tilde{y}_{ij,np}$. Then, the signals are merged in descending order of C until the number of clusters becomes N, with all the clusters (signal sets) required to have the same number of signals (see Fig. 3).

3.3. Auto-clustering with basis-shared Rank-1 MNMF

For Rank-1 MNMF, we can consider another approach for clustering the signals $\tilde{y}_{ij,np}$. Since the reverberation consists of a sum of time-delayed versions of the direct component, it is represented by the convolution. Even in the power spectrogram domain, this model is approximately valid [13]. If we assume that the impulse response in the power spectrogram domain is identical over all frequency bins, the direct and reverberant components of the same source can be modeled by the same bases $\boldsymbol{T}_n$ (spectral patterns) and different activations $\boldsymbol{V}_{np}$ (time-varying gains) as follows:

$$\tilde{\boldsymbol{Y}}_{n1} \simeq \boldsymbol{T}_n \boldsymbol{V}_{n1}, \ \ \tilde{\boldsymbol{Y}}_{n2} \simeq \boldsymbol{T}_n \boldsymbol{V}_{n2}, \ \ \ldots, \ \ \tilde{\boldsymbol{Y}}_{nP} \simeq \boldsymbol{T}_n \boldsymbol{V}_{nP}, \quad (16)$$

where $\tilde{\boldsymbol{Y}}_{np} (\in \mathbb{R}_{\geq 0}^{I \times J})$ is the power spectrogram of signal $\tilde{y}_{ij,np}$, $\boldsymbol{T}_n$ ($\in \mathbb{R}_{\geq 0}^{I \times L}$) is a shared basis matrix whose elements are $t_{i1,n}, \ldots, t_{iL,n}$, and $\boldsymbol{V}_{np}$ ($\in \mathbb{R}_{\geq 0}^{L \times J}$) is an activation matrix whose elements are $v_{1j,np}, \ldots, v_{Lj,np}$. This basis sharing leads to the separated signals $\tilde{y}_{ij,n1}, \ldots, \tilde{y}_{ij,nP}$ representing the direct and reverberant components of one source n. The cost function of basis-shared Rank-1 MNMF can be defined as

$$Q = \textstyle\sum_{i,j} \left[\sum_{n,p} \frac{|\tilde{y}_{ij,np}|^2}{\sum_l t_{il,n} v_{lj,np}} - 2 \log |\det \boldsymbol{W}_i| + \sum_{n,p} \log \sum_l t_{il,n} v_{lj,np} \right]. \quad (17)$$

The update rules of $\boldsymbol{W}_i$ for minimizing (17) are the same as (7)–(10) if we consider $N \leftarrow M = NP$, and the update rules of the NMF variables are obtained as follows:

$$t_{il,n} \leftarrow t_{il,n} \sqrt{\frac{\sum_{j,p} |y_{ij,np}|^2 v_{lj,np} \left(\sum_{l'} t_{il',n} v_{l'j,np}\right)^{-2}}{\sum_{j,p} v_{lj,np} \left(\sum_{l'} t_{il',n} v_{l'j,np}\right)^{-1}}}, \quad (18)$$

$$v_{lj,np} \leftarrow v_{lj,np} \sqrt{\frac{\sum_i |y_{ij,np}|^2 t_{il,n} \left(\sum_{l'} t_{il',n} v_{l'j,np}\right)^{-2}}{\sum_i t_{il,n} \left(\sum_{l'} t_{il',n} v_{l'j,np}\right)^{-1}}}. \quad (19)$$

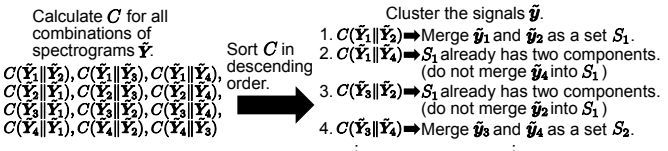

Fig. 3. Hierarchical clustering using correlation C ($N=2$, $M=4$, $P=2$), where all sets must have the same number of signals.

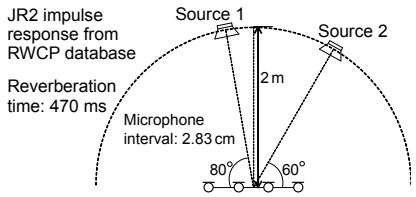

Fig. 4. Recording conditions of room impulse response.

However, the clustering result fluctuates depending on the initial values of the variables. To avoid this problem, we used IVA and the clustering method described in Sect 3.2 to obtain initial value of demixing matrix $\boldsymbol{W}_i$.

4. EXPERIMENT

4.1. Conditions

To confirm the efficacy of the proposed algorithm, we conducted an evaluation experiment using professional music signals. In this experiment, we produced observed signals with $M = 4$ channels and $N = 2$ sources by convoluting the impulse response JR2 (see Fig. 4) from the RWCP database [14] with each source. Table 1 shows the songs and sources used, which were obtained from SiSEC [15]. We compared IVA with PCA (PCA+IVA) and Rank-1 MNMF with PCA (PCA+Rank-1 MNMF), which both assume the rank-1 spatial constraint. In addition, two types of conventional MNMF [8] were also evaluated: MNMF w/o MWF and MNMF+MWF. In MNMF w/o MWF, the maximum SNR beamformer [16], which is calculated from the estimated spatial covariance $\mathsf{H}_{i,n}$, was used for separation. MNMF+MWF utilizes multichannel Wiener filtering (MWF) to enhance the estimated sources. As the proposed methods, constraint-relaxed IVA with the clustering method in Sect. 3.2 (Proposed IVA) and constraint-relaxed Rank-1 MNMF with basis sharing (Proposed Rank-1 MNMF) were evaluated, where the pretrained and clustered demixing matrix was used for the initial value in Proposed Rank-1 MNMF. Moreover, we evaluated the limit separation performance of linear filtering (Ideal linear filter), which is the maximum SNR beamformer calculated using the ideal spatial covariances of each source. It is necessary to apply a back-projection technique [17], except for in MNMF+MWF, to the estimated sources. The characteristics of each method are shown in Table 2 and the other conditions are described in Table 3. Note that we used a 128-ms-long window in the STFT for the signals that have 470-ms-long reverberation,

Table 1. Music sources

ID	Song	Source (1/2)
1	bearlin-roads_snip_85_99	acoustic_guit_main/piano
2	fort_minor-remember_the_name_snip_54_78	drums/vocals
3	ultimate_nz_tour_snip_43_61	guitar/vocals

Table 2. Characteristics of each method

Method	# of filters per source	Postfilter
PCA+IVA	1	None
PCA+Rank-1 MNMF	1	None
MNMF w/o MWF	1	None
MNMF+MWF	1	MWF
Ideal linear filter	1	None
Proposed IVA	2	None
Proposed Rank-1 MNMF	2	None

which means that the rank-1 spatial model collapses. As the evaluation scores, we used the signal-to-distortion ratio (SDR) [18], which indicates the total separation performance.

4.2. Results

Figure 5 shows the average scores and their deviations in 10 trials with various initializations. The methods using PCA cannot achieve good separation because they require the rank-1 spatial approximation. The scores of MNMF w/o MWF indicate poor separation accuracy and strong dependence on the initial values because it is difficult to estimate the full-rank spatial covariance H. However, MWF with NMF variables can greatly enhance the estimated sources. Proposed Rank-1 MNMF separates the sources with high accuracy. In particular, this method outperforms the limit performance of linear filtering (Ideal linear filter) as shown in Figs. 5 (b) and (c). This is because ground truth sources include reverberations, which can span more than two dimensional space, and the proposed algorithm can effectively relax the rank-1 spatial constraint. Table 4 shows actual computational times for the separation of song ID3, where the calculations were performed using MATLAB 8.3 (64-bit) with an Intel Core i7-4790 (3.60 GHz) CPU. The computational time of Proposed Rank-1 MNMF includes the initialization time for $\boldsymbol{W}_i$, which is the same as that of Proposed IVA. We confirm that Proposed Rank-1 MNMF can maintain efficient optimization and achieve good separation performance.

5. CONCLUSION

In this paper, we proposed a new relaxation method for the rank-1 spatial constraint. This method utilizes extra observations to estimate reverberant components while maintaining the rank-1 model. The efficacy of the proposed method was confirmed with IVA and Rank-1 MNMF, and they achieved a good separation performance even for reverberant signals.

REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol.36, no.3, pp.287–314, 1994.

Table 3. Experimental conditions

Sampling frequency	Downsampled from 44.1 kHz to 16 kHz
FFT length	128 ms
Window shift	64 ms
Number of bases	$L=15$ ($K=30$)
Maximum delay in time frame	$\tau_{\max}=2$
Number of iterations	200

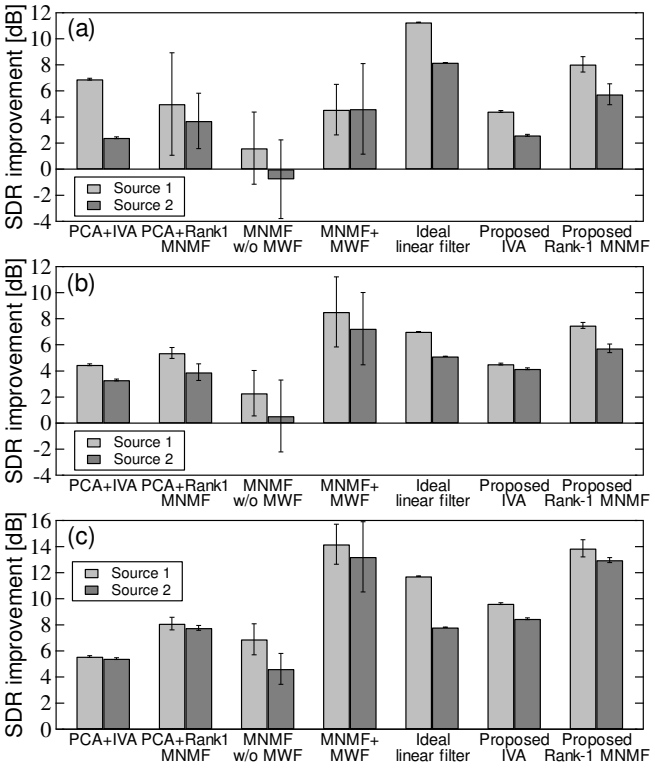

Fig. 5. Average SDR improvements for (a) song ID1, (b) song ID2, and (c) song ID3.

Table 4. Computational times for separation of song ID3 (s)

PCA+IVA	PCA+Rank-1 MNMF	MNMF+MWF	Proposed IVA	Proposed Rank-1 MNMF
23.4	29.4	3611.8	60.1	143.9

[2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol.22, pp.21–34, 1998.

[3] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. ASLP*, vol.14, no.2, pp.666–678, 2006.

[4] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Proc. NIPS*, vol.13, pp.556–562, 2001.

[5] S. Makino, T.-W. Lee and H. Sawada, "Blind Speech Separation," *Springer*, 2007.

[6] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino and S. Sagayama, "Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints," *Proc. ICASSP*, pp.5365–5368, 2012.

[7] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. ASLP*, vol.18, no.3, pp.550–563, 2010.

[8] H. Sawada, H. Kameoka, S. Araki and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complex-valued data," *IEEE Trans. ASLP*, vol.21, no.5, pp.971–982, 2013.

[9] T. Kim, H. T. Attias, S.-Y. Lee and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. ASLP*, vol.15, no.1, pp.70–79, 2007.

[10] D. Kitamura, N. Ono, H. Sawada, K. Kameoka and H. Saruwatari, "Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model," *Proc. ICASSP*, pp.276–280, 2015.

[11] S. Araki, R. Mukai, S. Makino, T. Nishikawa and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. SAP*, vol.11, no.2, pp.109–116, 2003.

[12] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," *Proc. WASPAA*, pp.189–192, 2011.

[13] H. Kameoka, T. Nakatani and T. Yoshioka, "Robust speech dereverberation based on non-negativity and sparse nature of speech spectrograms," *Proc. ICASSP*, pp.45–48, 2009.

[14] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," *Proc. LREC*, pp.965–968, 2000.

[15] S. Araki, F. Nesta, E. Vincent, Z. Koldovský, G. Nolte, A. Ziehe and A. Benichoux, "The 2011 signal separation evaluation campaign (SiSEC2011):-audio source separation," *Proc. LVA/SS*, pp.414–422, 2012.

[16] H. L. Van Trees, "Detection, Estimation, and Modulation Theory, Optimum Array Processing (Part IV)," *Wiley Interscience*, 2002.

[17] N. Murata, S. Ikeda and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol.41, no.1–4, pp.1–24, 2001.

[18] E. Vincent, R. Gribonval and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. ASLP*, vol.14, no.4, pp.1462–1469, 2006.