

Title	Unsupervised Singing Voice Separation Based on Robust Principal Component Analysis Exploiting Rank-1 Constraint
Author(s)	Li, Feng; Akagi, Masato
Citation	2018 26th European Signal Processing Conference (EUSIPCO): 1934-1938
Issue Date	2018
Type	Conference Paper
Text version	author
URL	<a href="http://hdl.handle.net/10119/15513">http://hdl.handle.net/10119/15513</a>
Rights	This is the author's version of a work. Copyright (C) 2018 EURASIP. Feng Li and Masato Akagi, 2018 26th European Signal Processing Conference (EUSIPCO), 2018, pp.1934-1938. DOI:10.23919/EUSIPCO.2018.8553584. First published in the Proceedings of the 26th European Signal Processing Conference (EUSIPCO-2018) in 2018, published by EURASIP.
Description	

# Unsupervised Singing Voice Separation Based on Robust Principal Component Analysis Exploiting Rank-1 Constraint

Feng Li and Masato Akagi

Japan Advanced Institute of Science and Technology

1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan

Email: {lifeng, akagi}@jaist.ac.jp

**Abstract**—In this paper, we address the singing voice separation problem and propose a novel unsupervised approach based on robust principal component analysis (RPCA) exploiting rank-1 constraint (CRPCA). RPCA is a recently proposed singing voice separation algorithm that can separate singing voice from the monaural recordings. Although RPCA has been successfully applied to singing voice separation task, it ignores the different characteristic values of singular value decomposition and computational complexity to minimize the nuclear norm for separating singing voice. Since rank-1 constraint in the background music, as the background music has a large variation in richness than singing voice among different songs. Furthermore, rank-1 constraint can utilize a prior target rank to separate singing voice and background music from the mixture music signal. Accordingly, the proposed CRPCA method utilizes rank-1 constraint minimization of singular values in RPCA instead of minimizing the whole nuclear norm, which can not only describe the different values of singular value decomposition, but also the computation complexity is reduced. The experiment evaluation results reveal that CRPCA can achieve better separation performance than the previous methods, especially with regard to use time frequency masking on ccMixer and DSD100 datasets. In addition, the running time on CRPCA is shorter than others under the same conditions.

## I. INTRODUCTION

Singing voice separation has aroused considerable attention and interest in recent years. It attempts to separate singing voice and background music parts of a music recording, which is a significant technology for chord recognition [1], music information retrieval (MIR) [2], and karaoke applications [3]. However, the current separation results of many popular methods are still far behind human hearing capability. The existing problems of singing voice separation are faced with serious challenges [4].

Many previous approaches have been proposed with the goal of overcoming the difficulty in separation task. Although the approaches based on deep neural network (DNN) [5]-[9] have recently proved to be powerful tools for singing voice separation task, they need a large number of training samples to be available in advanced. Unsupervised approach therefore still remains attractive for singing voice separation particularly where only a limited amount of singing voice data is available or without using any additional information. Rafii et al. [10] proposed a repeating accompaniment idea about

background music and used repeating pattern extraction technique (REPET) approach for separating the repeating music part from the non-repeating singing voice in a mixture signal. The main method is to identify the periodically repeating parts in the audio data, then compare them to a repeating segment model, and finally extract the repeating patterns with time frequency masking. Huang et al. [11] proposed a robust principal component analysis (RPCA) algorithm for singing voice separation, which decomposed an input matrix part into a low-rank matrix part and a sparse matrix part. Inspired by a low-rank and sparse model, Yang [12] proposed a novel low-rank and sparse matrices for incorporating harmonic priors and a back-end drum removal procedure. Also, he [13] presented a multiple low-rank representation to decompose a magnitude spectrogram (matrix) into two low-rank matrices. Yukara et al. [14] presented mutually-dependent singing voice separation and vocal fundamental frequency (F0) estimation. Yu et al. [15] proposed low-rank and sparse matrices representation with pre-learned dictionaries under the Alternating Direction Method of Multipliers framework model.

As mentioned above, RPCA can be used as an efficient strategy to separate singing voice in a mixture music signal, which decomposes the given amplitude spectrogram (matrix) of a mixture music signal into the sum of a sparse matrix (singing voice) and a low-rank matrix (music accompaniment). Owing to the part of background music can reproduce the same signal in the same song, the magnitude spectrogram therefore can be considered as a part of low-rank matrix. Singing voice, on the other hand, varies significantly and has a sparse distribution since its harmonic structure part in the spectrogram domain, resulting in a spectrogram with a sparse matrix part. Although RPCA has been successfully applied to singing voice separation, it ignores the different characteristic values of singular value decomposition (SVD) and computational complexity to minimize the nuclear norm for separating singing voice. Thus the separation performance decreases due to drums may lie in the sparse subspace instead of being low-rank. In our previous work, we proposed a weighted robust principal component analysis (WRPCA) [16], [17], which chose the different weighted values to describe the low-rank matrix for singing voice separation. However, it suffers from high computational cost, as it requires an

SVD at each iteration. Hence the running time of WRPCA is slower than RPCA. Recently a partial sum minimization of singular values instead of minimizing the nuclear norm in RPCA [18] was published, which used the minimized rank way to solve the different values of SVD in image processing, especially for background subtraction under the condition of rank-1 constraint.

To solve the above-mentioned problems, the partial sum minimization of SVD and computation complexity are significant for separating singing voice from the mixture music signal. In this paper, we combine the idea in [18] and propose a novel extension of RPCA exploiting rank-1 constraint (CRPCA), which utilizes the rank-1 constraint minimization singular values in RPCA instead of minimizing the nuclear norm for singing voice separation. Owing to rank-1 constraint in the background music, which is very similar to background subtraction, as the background music has a large variation in richness than singing voice among different songs. In addition, rank-1 constraint can utilize a prior target rank to separate background music and singing voice from the mixture music signal, which leads to a reduction in computation complexity. Therefore, the proposed CRPCA can not only describe the different values of SVD under the rank-1 constraint information, but also the computation complexity is reduced. Finally, we apply time frequency masking to further improve the separation results.

The remainder of this paper is organized as follows: In Section II, we review the conventional RPCA, WRPCA and WRPCA for singing voice separation. In Section III, we explain the proposed CRPCA and its application to singing voice separation. Experiments are conducted in Section IV and we finally draw conclusions in Section V.

## II. CONVENTIONAL METHODS

In this section, we briefly explain the overview of RPCA and WRPCA methods. We also introduce WRPCA for singing voice separation.

### A. Overview of RPCA

Candés et al. [19] proposed a convex model RPCA, which decomposed an input matrix  $M \in \mathbb{R}_{m \times n}$  into the sum of a low-rank matrix  $L \in \mathbb{R}_{m \times n}$  and a sparse matrix  $S \in \mathbb{R}_{m \times n}$ . The model can be defined as follows:

$$\begin{aligned} & \text{minimize } |L|_* + \lambda |S|_1, \\ & \text{subject to } M = L + S. \end{aligned} \quad (1)$$

where  $|\cdot|_1$  is the  $L_1$ -norm, which is the sum of absolute values of matrix entries.  $|\cdot|_*$  denotes the nuclear norm (sum of singular values), and  $\lambda > 0$  is a positive constant parameter between the parts of sparsity matrix  $S$  and low-rank matrix  $L$ . Moreover, this convex model can be solved by accelerated proximal gradient (APG) or augmented Lagrange multipliers (ALM) [20]. In this study, we used an inexact version of ALM (iALM) as a baseline for comparison.

### B. Overview of WRPCA

In our previous work, we presented an extension of RPCA called WRPCA [16], [17], which can decompose into the different scale values between low-rank and sparse matrices. WRPCA is an extension of RPCA, which is inspired from Huang et al. [11]. We define the model as follows:

$$\begin{aligned} & \text{minimize } |L|_{w,*} + \lambda |S|_1, \\ & \text{subject to } M = L + S. \end{aligned} \quad (2)$$

where  $|L|_{w,*}$  is weighted value of low-rank matrix,  $S$  is the value of sparse matrix.  $M \in \mathbb{R}_{m \times n}$  is the value of an input matrix, which consists of two values of  $S \in \mathbb{R}_{m \times n}$  and  $L \in \mathbb{R}_{m \times n}$ , and  $\lambda > 0$  is a positive constant parameter. We set  $\lambda = 1/\sqrt{\max(m, n)}$  as suggested in [19]. And we used an efficient iALM [20] method to solve this convex program model. The corresponding  $J$  is the Lagrange multiplier and  $\mu$  is a positive scalar value. We define the augmented Lagrangian function as follows:

$$\begin{aligned} J(M, L, S, \mu) = & |L|_{w,*} + \lambda |S|_1 + \langle J, M - L - S \rangle \\ & + \frac{\mu}{2} |M - L - S|_F^2. \end{aligned} \quad (3)$$

### C. WRPCA for singing voice separation

In previous studies [16], [17], WRPCA was used to separate singing voice from mixture music signal. On account of the part of background music can reproduce the same signal in the same song, although it has different values, the magnitude spectrogram can still be considered as a part of low-rank matrix. Singing voice signal, on the other hand, varies significantly and has a sparse distribution since its harmonic structure part in the spectrogram domain, resulting in a spectrogram with a sparse matrix part. So we can use WRPCA method to decompose an input matrix structure part into a low-rank matrix part and a sparse matrix part. The separated results outperform RPCA in the different audio data. However, it suffers from high computational cost during computing an SVD at each iteration, which leads to the running time to be slow.

## III. PROPOSED METHOD

In this section, we explain the proposed CRPCA, CRPCA with time frequency masking and describe a block diagram of singing voice separation system.

### A. Motivations

Although RPCA has been successfully applied to singing voice separation task, it ignores the different characteristic values of SVD and the computational complexity to minimize the nuclear norm. To overcome these two problems, we adopt a partial sum minimization of singular value based on rank-1 constraint called CRPCA. The aim is to fully utilize a prior rank-1 constraint to minimize the partial sum of singular values in RPCA. Since rank-1 constraint in the background music, as the background music has a large variation in richness than singing voice among different songs. Rank-1 constraint can utilize a prior target rank to separate background music and

---

**Algorithm 1** CRPCA for singing voice separation
 

---

**Input:** Mixture signal  $M \in \mathbb{R}_{m \times n}$ .

1: **Initialize:**  $\rho > 1, \mu_0 > 0, k = 0, L_0 = S_0 = 0$ .

2: **While** not convergence,

3: **do** :

4:  $L_{k+1} = P_{1, \mu_k^{-1}}(M - S_k + \mu_k^{-1} J_k)$ .

5:  $S_{k+1} = Q_{\lambda \mu_k^{-1}}(M - L_{k+1} + \mu_k^{-1} J_k)$ .

6:  $J_{k+1} = J_k + \mu_k(M - L_{k+1} - S_{k+1})$ .

7:  $\mu_{k+1} = \rho * \mu_k$ .

8:  $k = k + 1$ .

9: **end while.**

**Output:**  $L_{m \times n}, S_{m \times n}$ .

---

singing voice from the mixture music signal. As the result, the separation performance can be improved. Moreover, the computation complexity is also reduced due to minimize the partial sum of singular value instead of minimizing the whole nuclear norm. Finally, in order to obtain better separation performance, we apply time frequency masking to further improve the separation results.

### B. Principle of CRPCA

CRPCA is a novel extension of RPCA, which exploiting rank-1 constraint for singing voice separation. We define the model as follows:

$$\begin{aligned} & \text{minimize} \quad \sum_{i=2}^{\min(m,n)} \delta_i(L) + \lambda |S|_1, \\ & \text{subject to} \quad M = L + S. \end{aligned} \quad (4)$$

where  $L$  is the value of low-rank matrix,  $S$  is the value of sparse matrix.  $M \in \mathbb{R}_{m \times n}$  is the value of an input matrix, which consists of  $L \in \mathbb{R}_{m \times n}$  and  $S \in \mathbb{R}_{m \times n}$ , and  $\lambda > 0$  is a positive constant parameter between the sparse matrix  $S$  and the low-rank matrix  $L$ . And  $\delta_i(L)$  is the  $i$ -th singular value of  $L$ . We used  $\lambda = 1/\sqrt{\max(m, n)}$  as suggested in [19]. We also used an efficient iALM [20] method to solve this convex model in this work. The augmented Lagrangian function can be defined as follows:

$$\begin{aligned} J(M, L, S, \mu) = & \min \sum_{i=2}^{\min(m,n)} \delta_i(L) + \lambda |S|_1 \\ & + \langle J, M - L - S \rangle + \frac{\mu}{2} |M - L - S|_F^2. \end{aligned} \quad (5)$$

where  $J$  is the Lagrange multiplier and  $\mu$  is a positive scaler. The process of separating singing voice from the mixture music signal can be seen in **Algorithm 1** CRPCA for singing voice separation. The value of  $M$  is a mixture music signal from the observed audio data. After separated by using CRPCA, we can get a low-rank matrix  $L$  (music accompaniment) and a sparse matrix  $S$  (singing voice).

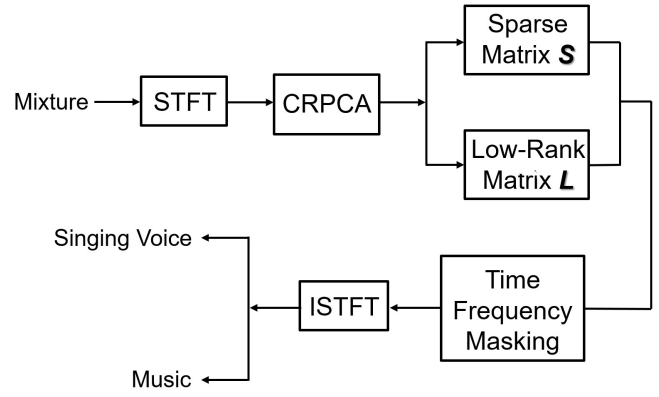


Fig. 1. Block diagram of singing voice separation system.

From the augmented Lagrangian function, we solve the following two sub-problems about  $L$  and  $S$ :

$$\begin{aligned} L_{k+1} = & \min_L \sum_{i=2}^{\min(m,n)} \delta_i(L) + \langle J_k, M - L - S_k \rangle \\ & + \frac{\mu_k}{2} |M - L - S_k|_F^2. \end{aligned} \quad (6)$$

$$\begin{aligned} S_{k+1} = & \min_S \lambda |S|_1 + \langle J_k, M - L_k - S \rangle \\ & + \frac{\mu_k}{2} |M - L_k - S|_F^2. \end{aligned} \quad (7)$$

### C. Update rules based on rank-1 constraint

As suggested by Oh et al. [18], the update rules of  $L$  and  $S$  are equivalent to solve the above two sub-problems as the following two equations (8) and (9):

$$L_{k+1} = P_{1, \mu_k^{-1}}(M - S_k + \mu_k^{-1} J_k) \quad (8)$$

$$S_{k+1} = Q_{\lambda \mu_k^{-1}}(M - L_{k+1} + \mu_k^{-1} J_k) \quad (9)$$

$P_{1, \mu_k^{-1}}(\cdot)$  can be defined as follows:

$$P_{1, \mu_k^{-1}}(Y) = U_Y (D_{Y_1} + Q_{\mu_k^{-1}}(D_{Y_2})) V_Y^T \quad (10)$$

where  $Y = Y_1 + Y_2$  ( $Y \in \mathbb{R}_{m \times n}$ ),  $D_{Y_1} = \text{diag}(\delta_1, 0, \dots, 0)$ ,  $Q_{\mu_k^{-1}}(D_{Y_2}) = \text{sign}(D_{Y_2}) \cdot \max(|D_{Y_2}| - \mu_k^{-1}, 0)$  is the soft-thresholding operator [21],  $D_{Y_2} = \text{diag}(0, \delta_2, \dots, \delta_{\min(m,n)})$ ,  $\delta_1$  and  $\delta_2$  are the first and second singular values.

### D. CRPCA for singing voice separation

In order to improve the separation performance, after separated by using CRPCA, we apply ideal binary time frequency masking (IBM) estimation to further improve the separation results. We define  $M_{ibm}$  as follows:

$$M_{ibm} = \begin{cases} 1 & S_{ij} \geq L_{ij} \\ 0 & S_{ij} < L_{ij} \end{cases} \quad (11)$$

where  $S_{ij}$  and  $L_{ij}$  are the values of sparse and low-rank matrices.

A block diagram of our proposed singing voice separation system can be seen in Fig. 1. For each mixture audio in

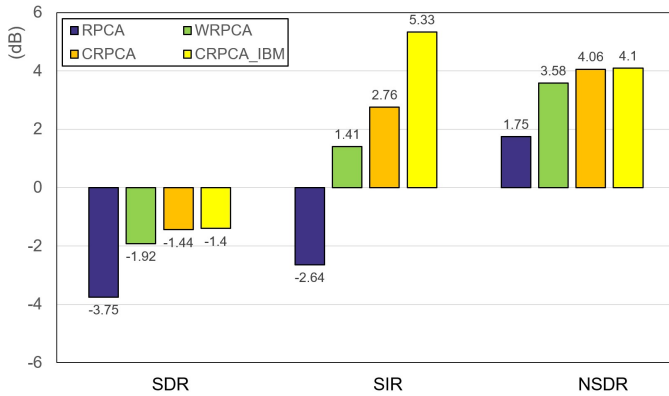


Fig. 2. Comparison of singing voice separation results on **ccMixer** dataset among RPCA, WRPCA, CRPCA and CRPCA with IBM on SDR, SIR, and NSDR, respectively.

the test dataset, we apply a short-time Fourier transform (STFT) and inverse STFT (ISTFT) based on being separated by using CRPCA. Furthermore, we utilize IBM method to further improve the separation results. And finally, we can obtain low-rank matrix  $L$  (music accompaniment) and sparse matrix  $S$  (singing voice), respectively.

#### IV. EXPERIMENTAL EVALUATION

In this section, we evaluate the proposed CRPCA and CRPCA with IBM by using two different datasets: ccMixer [22] and DSD100 [4], and compare with RPCA and WRPCA.

##### A. Experimental datasets

In our experiments, to evaluate the performance of the proposed CRPCA method, we used two different datasets to compare with RPCA, WRPCA, CRPCA and CRPCA with IBM methods. The first one was the ccMixer dataset, which contains 50 full songs with durations ranging from 1'17" to 7'36". Each audio data contains three parts: music accompaniment, singing voice, and a mixture of them, respectively.

The other one was DSD100 dataset. It contains 100 full stereo songs of different styles with durations ranging from 2'21" to 7'15" as also used in the 2016 Signal Separation Evaluation Campaign (SiSEC) [4], which is divided into 50 development and 50 test songs. We considered the sum of drums, bass and other as music accompaniment part. The target was to separate singing voice from the music accompaniment in the mixture music signal.

In this work, we evaluated the proposed CRPCA and CRPCA with IBM methods mainly concentrate on monaural source separation task. It was even more difficult than multichannel source separation due to only one single channel was available. The two-channel stereo mixture datasets we used were downmixed into a single channel and obtained an average value of each channel. We dealt with the whole audio data instead of part length on two datasets.

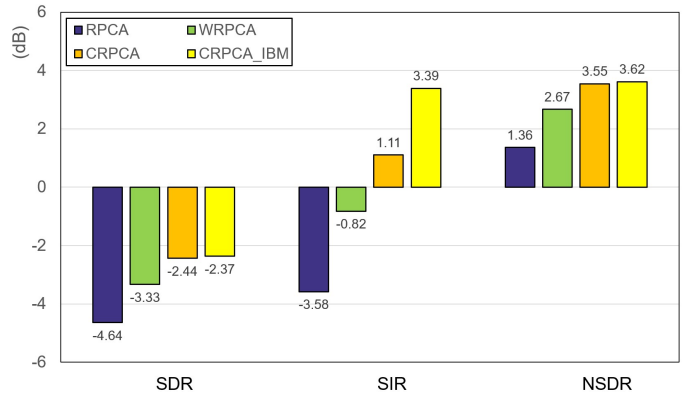


Fig. 3. Comparison of singing voice separation results on **DSD100** dataset among RPCA, WRPCA, CRPCA and CRPCA with IBM on SDR, SIR, and NSDR, respectively.

##### B. Experiment conditions

All data were sampled at 44.1 kHz. The input feature we used was calculated using STFT and ISTFT. A window size of 1024 samples and a hop size of 256 samples. All the experiments were run by using MATLAB R2015a, on a PC win10, X64-based processor, RAM32GB with i7-6700K CPU@4.00 GHz.

To evaluate the effectiveness of CRPCA and CRPCA with IBM, in the experiments, we compared with RPCA and WRPCA, assessed its quality of separation performance in terms of Source-to-Distortion Ratio (SDR) and Source-to-Interference Ratio (SIR) using BSS Eval Toolbox available at [23] and the normalized SDR (NSDR). The NSDR is defined as follows:

$$NSDR(\hat{v}, v, x) = SDR(\hat{v}, v) - SDR(x, v). \quad (12)$$

where  $v$  is the original clean signal part,  $\hat{v}$  is the separated voice, and  $x$  is the original mixture structure. The NSDR can be used to estimate the overall improvement of separation performance.

The SDR stands for the quality of the separated target sound signal. The SIR represents the degree of separation between the target. The obtained higher values of SDR, SIR and NSDR indicate that the method has better separation performance in source separation task. All evaluation metrics are calculated in dB.

##### C. Experimental results

To confirm CRPCA and CRPCA with IBM methods, we first evaluated the proposed method on ccMixer dataset. Fig. 2 shows the comparison of singing voice separation results on RPCA, WRPCA, CRPCA and CRPCA with IBM (CRPCA\_IBM). The experiment results are obtained with SDR, SIR and NSDR, we can clearly see that CRPCA obtains better separation results than RPCA and WRPCA, especially for using IBM estimation on ccMixer. With regard to SIR, CRPCA with IBM has a significant improved result among

TABLE I  
RUNNING TIME (HH:MM:SS)

Dataset	RPCA	WRPCA	CRPCA
ccMixer	02:04:40	03:03:31	<b>00:52:10</b>
DSD100	04:34:30	06:49:28	<b>01:54:17</b>

them. The degree of separation result values of SIR, RPCA and CRPCA with IBM, are -2.64dB and 5.33dB, respectively.

We also evaluated CRPCA and CRPCA with IBM methods on DSD100 dataset. Fig. 3 shows the comparison results with RPCA, WRPCA, CRPCA and CRPCA with IBM. The experiment results clearly reveal that CRPCA with IBM also obtains better separation performance on DSD100. These two result figures indicate that rank-1 constraint minimization can improve the separation performance than minimizing the nuclear norm in RPCA, even exceed the previous proposed WRPCA method. In addition, after using IBM estimation by CRPCA, we can further improved the separation results, especially with regard to SIR values among them. RPCA and CRPCA with IBM, are -3.58dB and 3.39dB, respectively.

Table I shows the running time of each method on two datasets under the same conditions, according to the experiment results, we can see that the running time of CRPCA is much shorter than that of RPCA and WRPCA. On the contrary, WRPCA has the worst results in all of them.

As the above results shown, although WRPCA can get better separation results, the running time is longer than RPCA on two datasets. Owing to CRPCA can utilize a prior target rank to separate singing voice from the mixture music signal, no matter separation performance or running time, the rank-1 constraint minimization singular values in RPCA is better than minimization the nuclear norm for separating singing voice. Furthermore, we applied IBM method to improve the separation performance. In terms of running time, CRPCA is preferable under the same conditions on two datasets.

## V. CONCLUSIONS

In this paper, we proposed a novel unsupervised approach that extends RPCA exploiting rank-1 constraint for singing voice separation task. The experiment evaluation results on ccMixer and DSD100 datasets indicated that CRPCA outperforms the conventional RPCA and WRPCA, especially for using time frequency masking. In addition, with regard to the running time, CRPCA is shorter than others under the same conditions while WRPCA is the worst among them. For future work, since F0 estimation and melody extraction are very crucial for separating singing voice from the mixture music signal, we therefore will unify them to improve the separation results.

## ACKNOWLEDGMENTS

This work was partly supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan Scholarship and the China Scholarship Council (CSC) of China Scholarship.

## REFERENCES

- [1] T. Fujishima, "Realtime chord recognition of musical sound: a system using common lisp music," in *Proc. ICMC*, pp. 464-467, 1999.
- [2] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, "Content-based music information retrieval: current directions and future challenges," *IEEE*, vol. 96, no. 4, pp. 668-696, 2008.
- [3] A. J. R. Simpson, G. Roma, and M. D. Plumbley, "Deep karaoke: extracting vocals from musical mixtures using a convolutional deep neural network," in *Proc. LVA/ICA*, pp. 429-436, 2015.
- [4] A. Liutkus, F. R. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, "The 2016 signal separation evaluation campaign," in *Proc. LVA/ICA*, pp. 323-332, 2017.
- [5] P. S. Huang, M. Kim, M. Hasegawa-Johnson, and P. Smaragdis, "Singing-voice separation from monaural recordings using deep recurrent neural networks," in *Proc. ISMIR*, pp. 477-482, 2014.
- [6] S. Uhlich, M. Porcu, F. Giron, M. Enekl, T. Kemp, N. Takahashi, and Y. Mitsufuji, "Deep neural network based instrument extraction from music," in *Proc. ICASSP*, pp. 2135-2139, 2015.
- [7] J. R. Hershey, Z. Chen, J. L. Roux, and S. Watanabe, "Deep clustering: discriminative embeddings for segmentation and separation," in *Proc. ICASSP*, pp. 31-35, 2016.
- [8] Y. Luo, Z. Chen, J. R. Hershey, J. L. Roux, and N. Mesgarani, "Deep clustering and conventional networks for music separation: stronger together," in *Proc. ICASSP*, pp. 61-65, 2017.
- [9] S. Uhlich, M. Porcu, F. Giron, M. Enekl, T. Kemp, N. Takahashi, and Y. Mitsufuji, "Improving music source separation based on deep neural networks through data augmentation and network blending," in *Proc. ICASSP*, pp. 261-265, 2017.
- [10] Z. Rafii and B. Pardo, "Repeating pattern extraction technique (REPET): a simple method for music/voice separation," *IEEE Trans. ALSP*, vol. 21, no. 1, pp. 73-84, 2013.
- [11] P. S. Huang, S. D. Chen, P. Smaragdis, and M. Hasegawa-Johnson, "Singing-voice separation from monaural recordings using robust principal component analysis," in *Proc. ICASSP*, pp. 57-60, 2012.
- [12] Y. H. Yang, "On sparse and low-rank matrix decomposition for singing voice separation," in *Proc. ACM Multimedia*, pp. 757-760, 2012.
- [13] Y. H. Yang, "Low-rank representation of both singing voice and music accompaniment via learned dictionaries," in *Proc. ISMIR*, pp. 427-432, 2013.
- [14] Y. Ikemiya, K. Itoyama, and K. Yoshii, "Singing voice separation and vocal F0 estimation based on mutual combination of robust principal component analysis and subharmonic summation," *IEEE/ACM Trans. ALSP*, vol. 24, no. 11, pp. 2084-2095, 2016.
- [15] S. Yu, H. Zhang, and Z. Duan, "Singing voice separation by low-rank and sparse spectrogram decomposition with prelearned dictionaries," *Journal of the Audio Engineering Society*, vol. 65, no. 5, pp. 377-388, 2017.
- [16] F. Li and M. Akagi, "Singing voice separation using weighted robust principal component analysis," in *Proc. ASJ Autumn Meeting*, pp. 559-562, 2017.
- [17] F. Li and M. Akagi, "Weighted robust principal component analysis with gammatone auditory filterbank for singing voice separation," in *Proc. ICONIP*, pp. 849-858, 2017.
- [18] T. H. Oh, Y. W. Tai, J. C. Bazin, H. Kim, and I. S. Kweon, "Partial sum minimization of singular values in robust PCA: algorithm and applications," *IEEE Trans. PAMI*, vol. 38, no. 4, pp. 744-758, 2016.
- [19] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, 2011.
- [20] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.
- [21] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for  $\ell_1$ -minimization: methodology and convergence," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1107-1130, 2008.
- [22] A. Liutkus, D. Fitzgerald, and Z. Rafii, "Scalable audio separation with light kernel additive modelling," in *Proc. ICASSP*, pp. 76-80, 2015.
- [23] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. ALSP*, vol. 14, no. 4, pp. 1462-1469, 2006.