

# JOINT SEPARATION AND DEREVERBERATION OF REVERBERANT MIXTURES WITH DETERMINED MULTICHANNEL NON-NEGATIVE MATRIX FACTORIZATION

Hideaki Kagami<sup>†</sup>, Hirokazu Kameoka<sup>‡</sup>, Masahiro Yukawa<sup>†</sup>

<sup>†</sup> Dept. Electronics and Electrical Engineering, Keio University, Japan.

<sup>‡</sup> NTT Communication Science Laboratories, NTT Corporation, Japan.

## ABSTRACT

This paper proposes an extension of multichannel non-negative matrix factorization (MNMF) that simultaneously solves source separation and dereverberation. While MNMF was originally formulated under an underdetermined problem setting where sources can outnumber microphones, a determined counterpart of MNMF, which we call the determined MNMF (DMNMF), has recently been proposed with notable success. This approach is particularly notable in that the optimization process can be more than 30 times faster than the underdetermined version owing to the fact that it involves no matrix inversion computations. One drawback as regards all methods based on instantaneous mixture models, including MNMF, is that they are weak against long reverberation. To overcome this drawback, this paper proposes an extension of DMNMF using a frequency-domain convolutive mixture model. The optimization process of the proposed method consists of iteratively updating (i) the spectral parameters of each source using the majorization-minimization algorithm, (ii) the separation matrix using iterative projection, and (iii) the dereverberation filters using multichannel linear prediction. Experimental results showed that the proposed method yielded higher separation performance and dereverberation performance than the baseline method under highly reverberant environments.

**Index Terms**— Blind source separation, blind dereverberation, non-negative matrix factorization, independent component analysis

## 1. INTRODUCTION

Blind source separation (BSS) is a technique for separating out individual source signals from microphone array inputs when the transfer characteristics between sources and microphones are unknown. The frequency-domain BSS approach provides the flexibility of allowing us to utilize various models for the time-frequency representations of source signals and/or array responses. For example, independent vector analysis (IVA) [1, 2] allows us to efficiently solve frequency-wise source separation and permutation alignment in a joint manner by assuming that the magnitudes of the frequency components originating from the same source tend to vary coherently over time.

With a different approach, multichannel extensions of non-negative matrix factorization (NMF) have attracted a lot of attention in recent years [3–8]. NMF was originally applied to monaural source separation tasks [9, 10]. The idea is to approximate the power (or magnitude) spectrogram of a mixture signal, interpreted as a non-negative matrix, as the product of two non-negative matrices. This amounts to assuming that the power spectrum of a mixture signal observed at each time frame can be approximated by the linear

sum of a limited number of basis spectra scaled by time-varying amplitudes. Multichannel NMF (MNMF) is an extension of this approach to a multichannel case in order to allow for the use of spatial information as an additional clue for separation. It can also be viewed as an extension of frequency-domain BSS that allows the use of spectral templates as a clue for both frequency-wise source separation and permutation alignment.

The original MNMF [3–5] was formulated under an underdetermined problem setting where sources can outnumber microphones. A determined version of MNMF (DMNMF) [6–8] specializes in the overdetermined case and can be implemented without using any matrix inversion, shortening the computation time more than 30 times than the underdetermined MNMF. In [7, 8], a theoretical relation of DMNMF to IVA is discussed, which has naturally allowed for the incorporation of a fast update rule of the separation matrix developed for IVA, called the “iterative projection (IP)” [11], into the parameter optimization process of DMNMF. It has been shown that this has contributed not only to further accelerating the entire optimization process but also improving the separation performance.

One drawback as regards methods using instantaneous mixture models is that they are weak against long reverberation. It has been shown in [6, 12] that source separation of highly reverberant mixtures can be solved effectively by using frequency-domain convolutive mixture models and that the parameters of frequency-domain convolutive mixture models can be efficiently estimated by iteratively updating the separation matrix, the dereverberation filters and the spectral parameters of each source. In this paper, we take this approach to develop an extension of DMNMF using a frequency-domain convolutive mixture model, where the optimization process consists of iteratively updating (i) the NMF parameters using the majorization-minimization (MM) algorithm, (ii) the separation matrix using iterative projection (IP), and (iii) the dereverberation filters using multichannel linear prediction.

## 2. PROBLEM FORMULATION

We start by introducing a frequency-domain convolutive mixture model employed in [6, 12]. Let us consider a situation in which signals emanate from  $M$  sources and are captured by  $M$  microphones. Let  $x_i(f, n)$  be the short-time Fourier transform (STFT) of the signal observed at the  $i$ -th microphone, where  $f$  and  $n$  are the frequency and time indices, respectively. Also let  $\mathbf{x}(f, n) = [x_1(f, n), \dots, x_M(f, n)]^T \in \mathbb{C}^M$  be a vector of observed data. We use  $s_j(f, n)$  to denote the  $j$ -th source components and  $\mathbf{s}(f, n) = [s_1(f, n), \dots, s_M(f, n)]^T \in \mathbb{C}^M$  to denote a vector of  $M$  source components. Under a determined problem setting, many conventional BSS systems employ frequency-domain instantaneous mixture model:

$$\mathbf{s}(f, n) = \mathbf{W}^H(f) \mathbf{x}(f, n) \quad (1)$$

This work was supported by JSPS Grants-in-Aid (15K06081, 15K13986, 15H02757, 17H01763).

where  $\mathbf{W}^H(f)$  is called the separation matrix. However, in a highly reverberant condition where the length of the room impulse responses can be longer than the STFT frame length, the ability of this model to separate sources will be limited. In this paper, we use a separation system that has a multichannel finite-impulse-response form in the time-frequency domain [6, 12] such that:

$$\mathbf{s}(f, n) = \mathbf{W}^H(f, 0)\mathbf{x}(f, n) + \sum_{n'=1}^{N'} \mathbf{W}^H(f, n')\mathbf{x}(f, n - n'), \quad (2)$$

where  $\mathbf{W}^H(f, n')$ ,  $0 \leq n' \leq N'$  are the coefficient matrices of size  $M \times M$ , and  $(\cdot)^H$  stands for Hermitian transpose. When  $\mathbf{W}^H(f, 0)$  is invertible, (2) can be written equivalently by the following process:

$$\mathbf{y}(f, n) = \mathbf{x}(f, n) - \sum_{n'=1}^{N'} \mathbf{G}^H(f, n')\mathbf{x}(f, n - n'), \quad (3)$$

$$\mathbf{s}(f, n) = \mathbf{W}^H(f, 0)\mathbf{y}(f, n), \quad (4)$$

where  $\mathbf{G}^H(f, n') := -(\mathbf{W}^H(f, 0))^{-1}\mathbf{W}^H(f, n')$ . Equation (3) can be seen as a dereverberation process of the observed mixture signal  $\mathbf{x}(f, n)$ , whereas (4) can be seen as an instantaneous demixing process of the dereverberated mixture signal  $\mathbf{y}(f, n)$ .

Let us assume that  $s_j(f, n)$  is a random variable:

$$s_j(f, n) \sim \mathcal{N}_{\mathbb{C}}(s_j(f, n) | 0, v_j(f, n)) \quad (5)$$

and that  $s_j(f, n)$  and  $s_{j'}(f', n')$  are statistically independent when  $(f, n, j) \neq (f', n', j')$ . Here,

$$\mathcal{N}_{\mathbb{C}}(s_j(f, n) | 0, v_j(f, n)) \propto \frac{1}{v_j(f, n)} \exp\left(-\frac{|s_j(f, n)|^2}{v_j(f, n)}\right) \quad (6)$$

denotes the complex Gaussian distribution. We further assume that the power spectral density  $v_j(f, n)$  can be modeled as

$$v_j(f, n) = \sum_{k=1}^K h_{j,k}(f)u_{j,k}(n), \quad (7)$$

where  $h_{j,k}(f) \geq 0$  is the  $(j, k)$  element of the basis matrix and  $u_{j,k}(n) \geq 0$  is the  $(j, k)$  element of the activation matrix for the  $j$ -th source. Multichannel source separation methods using the power spectrogram model (7) or its variants are called MNMF [3, 5–8]. The negative log-likelihood of the signal  $y_i(f, n)$  is given by

$$Q(\Theta_V, \Theta_W, \Theta_G) := -2N \sum_f \log \left| \det \mathbf{W}^H(f, 0) \right| + \sum_{f, n, j} \left( \log v_j(f, n) + \frac{|s_j(f, n)|^2}{v_j(f, n)} \right), \quad (8)$$

where  $\Theta_V := \{\mathbf{H}, \mathbf{U}\}$  with  $\mathbf{H} = \{h_{j,k}(f)\}$  and  $\mathbf{U} = \{u_{j,k}(n)\}$ ,  $\Theta_W := \{\mathbf{W}^H(f, 0)\}$ , and  $\Theta_G := \{\mathbf{G}^H(f, n')\}$ .

### 3. OPTIMIZATION PROCESS

The cost function (8) can be iteratively decreased by using a coordinate descent method in which each iteration comprises the following

three minimization steps:

$$\hat{\Theta}_V \leftarrow \underset{\Theta_V}{\operatorname{argmin}} Q(\Theta_V, \hat{\Theta}_W, \hat{\Theta}_G) \quad (9)$$

$$\hat{\Theta}_W \leftarrow \underset{\Theta_W}{\operatorname{argmin}} Q(\hat{\Theta}_V, \Theta_W, \hat{\Theta}_G) \quad (10)$$

$$\hat{\Theta}_G \leftarrow \underset{\Theta_G}{\operatorname{argmin}} Q(\hat{\Theta}_V, \hat{\Theta}_W, \Theta_G). \quad (11)$$

In the next subsections, we focus on each step and derive the update rules.

#### 3.1. Update procedure for $\hat{\Theta}_V$

The update rules for  $\hat{\Theta}_V$  can be derived by the MM algorithm. Dropping the constant terms with respect to  $\hat{\Theta}_V$  from (8), we obtain

$$C_1(\Theta_V) := \sum_{f, n, j} \left( \log \sum_{k=1}^K h_{j,k}(f)u_{j,k}(n) + \frac{|s_j(f, n)|^2}{\sum_{k=1}^K h_{j,k}(f)u_{j,k}(n)} \right), \quad (12)$$

where  $v_j(f, n)$  in (8) is replaced by (7). To minimize this function, we can design a majorization function of  $C_1$  with a convenient form [13–15]

$$C_1^+(\Theta_V, \bar{\Theta}) = \sum_{f, n, j} \left[ \frac{1}{\xi(f, n)} \left( \sum_{k=1}^K h_{j,k}(f)u_{j,k}(n) - \xi(f, n) \right) + \log \xi(f, n) + |s_j(f, n)|^2 \sum_{k=1}^K \frac{\lambda_k^2(f, n)}{h_{j,k}(f)u_{j,k}(n)} \right], \quad (13)$$

where  $\bar{\Theta} := \{\xi, \lambda\}$  with  $\xi = \{\xi(f, n)\}$  and  $\lambda = \{\lambda_k(f, n)\}$ . It can be verified that the majorization function  $C_1^+$  satisfies the following properties:

$$1. C_1(\Theta_V) \leq C_1^+(\Theta_V, \bar{\Theta}) \quad (14)$$

$$2. C_1(\Theta_V) = \min_{\bar{\Theta}} C_1^+(\Theta_V, \bar{\Theta}). \quad (15)$$

Here, (14) holds with equality and, at the same time, the minimum in (15) is achieved when

$$\xi(f, n) = \sum_{k'=1}^K h_{j,k'}(f)u_{j,k'}(n), \quad (16)$$

$$\lambda_k(f, n) = h_{j,k}(f)u_{j,k}(n) / \sum_{k'=1}^K h_{j,k'}(f)u_{j,k'}(n). \quad (17)$$

The function  $C_1$  is expected to be minimized indirectly by repeating the following two steps.

1. Minimize  $C_1^+$  with respect to  $\bar{\Theta}$  by (16) and (17), which makes  $C_1(\Theta_V) = C_1^+(\Theta_V, \bar{\Theta})$ .
2. Minimize  $C_1^+$  with respect to  $\Theta_V$ .

For the second step, making the partial derivatives of  $C_1^+$  with respect to each element of  $\mathbf{H}$  and  $\mathbf{U}$  to be zero yields the following update rules for  $\mathbf{H}$  and  $\mathbf{U}$ :

$$h_{j,k}(f) \leftarrow h_{j,k}(f) \sqrt{\frac{\sum_{j'} |s_{j'}(f, n)|^2 u_{j',k}(n) r_{j'}^{-2}(f, n)}{\sum_{j'} u_{j',k}(n) r_{j'}^{-1}(f, n)}}, \quad (18)$$

$$u_{j,k}(n) \leftarrow u_{j,k}(n) \sqrt{\frac{\sum_{j'} |s_{j'}(f, n)|^2 h_{j',k}(n) r_{j'}^{-2}(f, n)}{\sum_{j'} h_{j',k}(n) r_{j'}^{-1}(f, n)}}, \quad (19)$$

where  $r_{j'}(f, n) := \sum_{k'=1}^K h_{j',k'}(f)u_{j',k'}(n)$ .

### 3.2. Update procedure for $\hat{\Theta}_W$

Dropping the constant terms with respect to  $\hat{\Theta}_W$  from (8), we obtain

$$C_2(\Theta_W) := -2N \sum_f \log |\det \mathbf{W}^H(f, 0)| \\ + N \sum_{f,j} \mathbf{w}_j^H(f) \Sigma_{y/v_j(f)} \mathbf{w}_j(f), \quad (20)$$

where  $\mathbf{w}_j(f)$  is the  $j$ -th column vector of  $\mathbf{W}(f, 0)$  and  $\Sigma_{y/v_j(f)} = \frac{1}{N} \sum_n \frac{\mathbf{y}(f,n) \mathbf{y}^H(f,n)}{v_j(f,n)}$ , which is assumed positive definite (i.e.,  $\{\mathbf{y}(f,n)\}_n$  is a linearly independent set). As noted above, when the parameters  $\hat{\Theta}_V$  and  $\hat{\Theta}_G$  are fixed, the cost function (8), which is equal to (20), can be seen as an instantaneous demixing process of the dereverberated mixture signal  $\mathbf{y}(f,n)$ . To update  $\hat{\Theta}_W$ , several determined BSS techniques are available, such as the natural gradient method [16], FastICA (FICA) [17] and IP [2].

IP [2] is a block coordinate descent algorithm, consisting of sequentially optimizing one column vector of  $\mathbf{W}^H(f, 0)$  at a time while keeping the other column vectors fixed. Differentiating  $C_2(\Theta_W)$  with respect to the conjugate  $\mathbf{w}_j^*(f)$  of  $\mathbf{w}_j(f)$ , and setting the result to zero, we have

$$\Sigma_{y/v_j(f)} \mathbf{w}_j(f) - 2 \frac{\partial}{\partial \mathbf{w}_j^*(f)} \log \det \mathbf{W}^H(f, 0) = \mathbf{0}. \quad (21)$$

By using the matrix derivatives formula  $\frac{\partial}{\partial \mathbf{A}} \det(\mathbf{A}) = \mathbf{A}^{-T} \det(\mathbf{A})$ , (21) can be rearranged in the following simultaneous vector equations:

$$\mathbf{w}_j^H(f) \Sigma_{y/v_j(f)} \mathbf{w}_j(f) = 1, \quad (22)$$

$$\mathbf{w}_{j'}^H(f) \Sigma_{y/v_j(f)} \mathbf{w}_j(f) = 0 \text{ for } j' \neq j, \quad (23)$$

A solution to (22) and (23) can be found by performing the following updates

$$\mathbf{w}_j(f) \leftarrow (\mathbf{W}^H(f, 0) \Sigma_{y/v_j(f)})^{-1} \mathbf{e}_j, \quad (24)$$

$$\mathbf{w}_j(f) \leftarrow \frac{\mathbf{w}_j(f)}{\sqrt{\mathbf{w}_j^H(f) \Sigma_{y/v_j(f)} \mathbf{w}_j(f)}}, \quad (25)$$

for all  $f$  and  $j$ . Here,  $\mathbf{e}_j$  denotes the  $j$ -th column of the  $M \times M$  identity matrix  $\mathbf{I}$ .

### 3.3. Update procedure for $\hat{\Theta}_G$

Dropping the constant terms with respect to  $\hat{\Theta}_G$  from (8), we obtain

$$C_3(\Theta_G) := \sum_{f,n} \left\| \mathbf{x}(f,n) - \sum_{n'=1}^{N'} \mathbf{G}^H(f,n') \mathbf{x}(f,n-n') \right\|_{\Sigma_{w/v(f,n)}}^2, \quad (26)$$

where  $\|\mathbf{x}\|_{\Sigma_{w/v(f,n)}} := \sqrt{\mathbf{x}^H \Sigma_{w/v(f,n)} \mathbf{x}}$  with  $\Sigma_{w/v(f,n)} = \sum_j \frac{\mathbf{w}(f) \mathbf{w}^H(f)}{v_j(f,n)}$  which is assumed positive definite, and  $\mathbf{O}$  denotes the zero matrix. It is clear that, for each  $f$ , all elements of the matrices  $\{\mathbf{G}(f,n')\}$  minimizing (26) are mutually dependent. To update  $\mathbf{G}(f,n')$  for each  $f$  independently, we vectorize  $\{\mathbf{G}(f,n')\}$  as follows:

$$\mathbf{g}(f) := \text{vec}(\{\mathbf{G}(f,n')\}) \\ := [\mathbf{g}_1^T(f,1), \dots, \mathbf{g}_M^T(f,1), \mathbf{g}_1^T(f,2), \dots, \mathbf{g}_M^T(f,2), \dots, \\ \mathbf{g}_M^T(f,N'-1), \mathbf{g}_1^T(f,N'), \dots, \mathbf{g}_M^T(f,N')]^T \in \mathbb{C}^{M^2 N'}, \quad (27)$$

**Table 1.** Input average SIRs and DRRs (dB)

	Male + Male	Male + Female	Female + Female
SIR	0.2190	0.1160	0.0738
DRR	2.9172	2.2031	4.4501

**Table 2.** Average computational time (sec.)

Proposed (IP)	371.1974
Proposed (FICA)	396.7370
Baseline [7, 8]	303.5906
Sequential	321.8167

where  $\mathbf{g}_m(f,n')$  is the  $m$ -th column of  $\mathbf{G}(f,n')$ . By using  $\mathbf{g}(f)$ , we can rewrite the term  $\sum_{n'=1}^{N'} \mathbf{G}^H(f,n') \mathbf{x}(f,n-n')$  in (26) as

$$\sum_{n'=1}^{N'} \mathbf{G}^H(f,n') \mathbf{x}(f,n-n') = \mathbf{X}(f,n) \mathbf{g}^*(f), \quad (28)$$

where

$$\mathbf{X}(f,n) = [\mathbf{I} \otimes \mathbf{x}^T(f,n-1) \ \mathbf{I} \otimes \mathbf{x}^T(f,n-2) \\ \dots \ \mathbf{I} \otimes \mathbf{x}^T(f,n-N')] \in \mathbb{C}^{M \times M^2 N'}. \quad (29)$$

Here,  $\otimes$  stands for Kronecker product. Substituting (28) into (26), the cost function is rewritten as

$$C_3(\Theta_G) = \sum_{f,n} (\mathbf{x}(f,n) - \mathbf{X}(f,n) \mathbf{g}^*(f))^H \\ \times \Sigma_{w/v(f,n)} (\mathbf{x}(f,n) - \mathbf{X}(f,n) \mathbf{g}^*(f)). \quad (30)$$

Our objective in this part is now to minimize this function with respect to  $\mathbf{g}^*(f)$ .

Equation (30) is readily solved because the cost function is quadratic with respect to  $\mathbf{g}^*(f)$ . We calculate the partial derivatives of  $C_3(\Theta_G)$  to be zero, the update rule for  $\mathbf{g}^*(f)$  is obtained as

$$\mathbf{g}^*(f) \leftarrow \left( \sum_n \mathbf{X}^H(f,n) \Sigma_{w/v(f,n)} \mathbf{X}(f,n) \right)^{-1} \\ \times \left( \sum_n \mathbf{X}^H(f,n) \Sigma_{w/v(f,n)} \mathbf{x}(f,n) \right), \quad (31)$$

for all  $f$ .

### 3.4. Summary of optimization process

Overall, the proposed algorithm is summarized as follows.

1. Initialize  $\mathbf{H}, \mathbf{U}, \mathbf{W}^H(f, 0)$  and  $\mathbf{G}^H(f, n')$ .
2. Update  $\mathbf{H}, \mathbf{U}$  using (18), (19) for all  $f, n, j$ .
3. Update  $\mathbf{W}^H(f, 0)$  using (24), (25) for all  $f, j$ .
4. Update  $\mathbf{G}^H(f, n')$  using (31) for all  $f$ .  
Repeat 2–4 until convergence.

## 4. EXPERIMENTAL RESULTS

We evaluate the effect of the convolutive model under highly reverberant environments and the difference between computational time by separating matrix update method. We quantitatively compared the source separation performance of the proposed approach using IP and FICA, conventional DMNMF and sequential method (Sequential). All of the examples use the same two-input four-output impulse response, which was measured in an anechoic chamber where

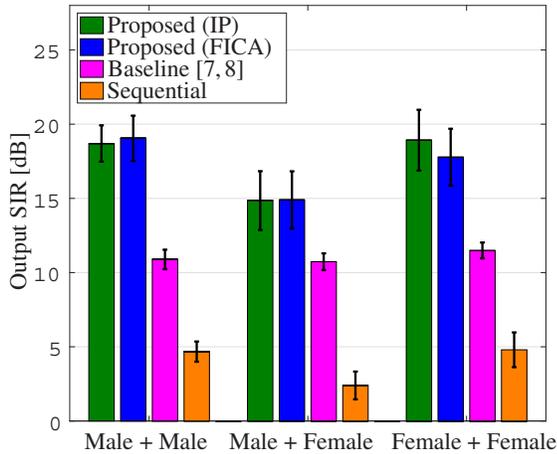


Fig. 1. Average SIR for each gender pair.

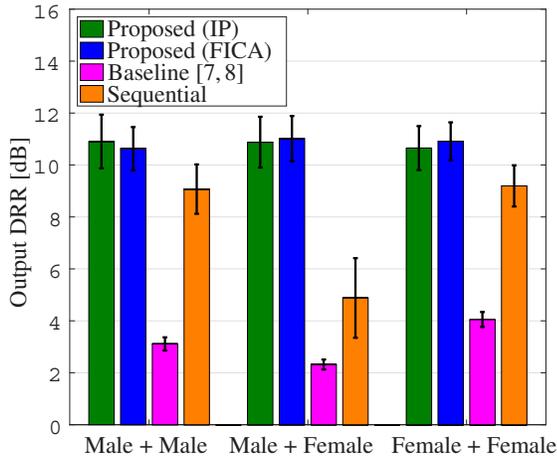


Fig. 2. Average DRR for each gender pair.

the reverberation time was 0.6 sec. With this impulse response, we mixed two speech signals into four mixtures. The two speech signals of male or female speakers, taken from the ATR speech database, were sampled at 16 kHz and band limited to the 50 Hz to 7 kHz frequency range. We generate 10 speech combinations for each gender pair, male+male, male+female and female+female. The average input Signal-to-Interference ratios (SIRs) and Direct-to-Reverberation ratios (DRRs) are shown in Table 1. Time-frequency representations were obtained using the polyphase filterbank analysis with a frame length of 32 ms and a hop size of 8 ms. For the DMNMF, we performed time-frequency analysis with a frame length of 256 ms and a hop size of 64 ms. The filter length  $N'$  was set as follows:  $N' = 25$  for  $P_f < 0.8$ ;  $N' = 20$  for  $0.8 \leq P_f < 1.5$ ;  $N' = 15$  for  $1.5 \leq P_f < 3$ ;  $N' = 10$  for  $P_f \geq 3$ , where  $P_f$  is the frequency in kHz of the  $f$ -th frequency bin. The NMF basis number  $K$  was set at 20. The iterative algorithm was run for 20 iterations. For each step of update the separation matrix  $\mathbf{W}^H(f, 0)$ , the FICA algorithm was run for 100 iterations.

Table 2 shows IP is faster than FICA algorithm, where Baseline refers to conventional DMNMF [7, 8]. The ‘‘Sequential’’ is a method

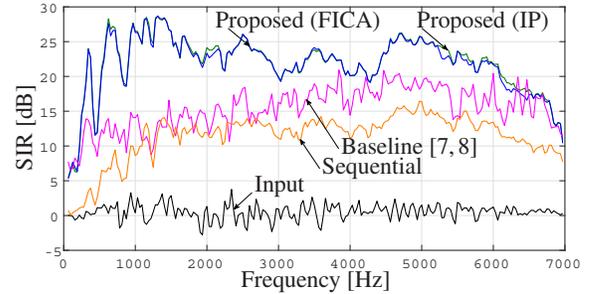


Fig. 3. Average SIR for each frequency.

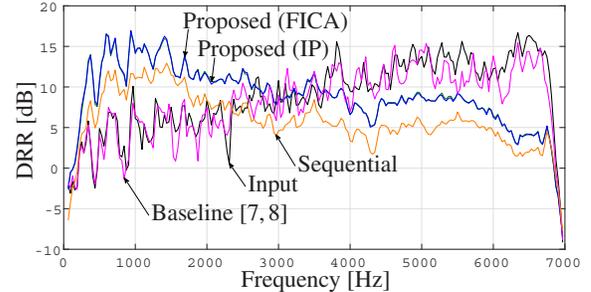


Fig. 4. Average DRR for each frequency.

that performs dereverberation and DMNMF sequentially. Figs. 1 and 2 show average output SIRs and DRRs and the 95% confidence interval for each gender pair. One factor of the performance difference is that the filter length  $N'$  was fixed as described above in all cases. Figs. 3 and 4 show average output SIRs and DRRs for each frequency. Fig. 3 shows that the proposed approach attains higher SIR for all frequencies. Referring to Fig. 4, the proposed approach attains significantly higher DRRs than the baseline method over the wide range below 3 kHz, although its DRRs are lower (significantly only for frequency between 6.2 – 6.7 kHz) above 3 kHz. Overall, these results show that the proposed approaches perform better than the conventional DMNMF and sequential methods.

## 5. CONCLUSION

One drawback as regards all methods based on instantaneous mixture models, including MNMF, is that they are weak against long reverberation. To overcome this drawback, this paper proposed an extension of DMNMF using a frequency-domain convolutive mixture model, which allows us to solve source separation and dereverberation simultaneously. The optimization process of the proposed method consists of iteratively updating (i) the spectral parameters of each source using the majorization-minimization algorithm, (ii) the separation matrix using IP, and (iii) the dereverberation filters using multichannel linear prediction. Experimental results showed that the proposed method yielded higher separation performance and dereverberation performance than the baseline method under highly reverberant environments.

## 6. REFERENCES

- [1] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ica to multivariate components," in *International Conference on Independent Component Analysis and Signal Separation*. Springer, 2006, pp. 165–172.
- [2] N. Ono and S. Miyabe, "Auxiliary-function-based independent component analysis for super-gaussian sources," in *LVA/ICA*. Springer, 2010, pp. 165–172.
- [3] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "New formulations and efficient algorithms for multichannel nmf," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*. IEEE, 2011, pp. 153–156.
- [4] N.Q.K. Duong, E. Vincent, and R. Gribonval, "Underdetermined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1830–1840, 2010.
- [5] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 550–563, 2010.
- [6] H. Kameoka, T. Yoshioka, M. Hamamura, J. Le Roux, and K. Kashino, "Statistical model of speech signals based on composite autoregressive system with application to blind source separation," in *LVA/ICA*. Springer, 2010, pp. 245–253.
- [7] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 276–280.
- [8] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1626–1641, 2016.
- [9] P. Smaragdīs, "Non-negative matrix factorization for polyphonic music transcription," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2003, pp. 177–180.
- [10] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.
- [11] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*. IEEE, 2011, pp. 189–192.
- [12] T. Yoshioka, T. Nakatani, M. Miyoshi, and H. G. Okuno, "Blind separation and dereverberation of speech mixtures by joint optimization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 69–84, 2011.
- [13] H. Kameoka, M. Goto, and S. Sagayama, "Selective amplifier of periodic and non-periodic components in concurrent audio signals with spectral control envelopes," *Proc. SIG Tech. Reports on Music and Computer of IPSJ*, pp. 77–84, 2006.
- [14] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama, "Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with  $\beta$ -divergence," in *Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on*. IEEE, 2010, pp. 283–288.
- [15] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the  $\beta$ -divergence," *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011.
- [16] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in neural information processing systems*, 1996, pp. 757–763.
- [17] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, vol. 46, John Wiley & Sons, 2004.