FastFCA: Joint Diagonalization Based Acceleration of Audio Source Separation Using a Full-Rank Spatial Covariance Model

Nobutaka Ito, Shoko Araki, and Tomohiro Nakatani
NTT Communication Science Laboratories, NTT Corporation, Kyoto, Japan
Email: {ito.nobutaka, araki.shoko, nakatani.tomohiro}@lab.ntt.co.jp

***Abstract*—Here we propose an accelerated version of one of the most promising methods for audio source separation proposed by Duong *et al.* ["Under-determined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Trans. ASLP*, vol. 18, no. 7, pp. 1830–1840, Sep. 2010]. We refer to this conventional method as *full-rank spatial covariance analysis (FCA)*, and the proposed method as *FastFCA*. A major drawback of the conventional FCA is computational complexity: inversion and multiplication of covariance matrices are required at each time-frequency point and each EM iteration. To overcome this drawback, the proposed FastFCA diagonalizes the covariance matrices jointly based on the generalized eigenvalue problem. This leads to significantly reduced computational complexity of the FastFCA, because the complexity of matrix inversion and matrix multiplication for diagonal matrices is $O(M)$ instead of $O(M^3)$ (M: matrix order). Furthermore, the FastFCA is rigorously equivalent to the FCA, and therefore the reduction in computational complexity is realized without degradation in source separation performance. An experiment showed that the FastFCA was over 250 times faster than the FCA with virtually no degradation in source separation performance. In this paper, we focus on the two-source case, while the case of more than two sources is treated in a separate paper.**

I. Introduction

The source separation method proposed by Duong *et al.* [1], which is called *full-rank spatial covariance analysis (FCA)* in this paper, can be considered one of the most promising source separation methods. In the FCA, the spatial characteristics of each source signal are modeled by a full-rank matrix called a *spatial covariance matrix*. The full-rank spatial covariance matrix enables the FCA to model not only point-source signals but also reverberant source signals and diffuse signals (*e.g.,* background noise). This contrasts markedly with conventional modeling of spatial characteristics with a steering vector, which is done, *e.g.,* in the independent component analysis (ICA) [2]. However, a major drawback of the FCA is computational complexity, which may be prohibitive in applications with restricted computational resources (*e.g.,* hearing aids) or to a large dataset (*e.g.,* the CHiME-3 dataset [3]). Indeed, the FCA requires matrix inversion and matrix multiplication (both of complexity $O(M^3)$ with M being the number of microphones) of covariance matrices at each time-frequency point and each EM iteration.

To accelerate the FCA based on sparseness of source signals, a time-varying *complex Gaussian mixture model (cGMM)* has recently been proposed [4], [5]. Various source signals such as speech possess sparseness, the property of having non-zero power only at a small percentage of the time-frequency points. Based on this property, the source signals are assumed to be *disjoint* in the sense that only one of them is present at each time-frequency point [6]. This implies that the observed signals at each time-frequency point are approximated by only one source signal. This simplified model leads to significantly reduced computational complexity of the cGMM compared to the FCA. However, the above disjointness assumption is not always true. For example, when a speech signal and background noise are to be separated, the latter may be non-sparse so that it contributes to all time-frequency points. Also, when music is to be separated into instrumental sounds, they may not be disjoint but rather have largely overlapped harmonic components. In such cases, the performance of cGMM-based source separation may be severely degraded.

Here we propose another way of accelerating the FCA, which relies on joint diagonalization of covariance matrices based on the generalized eigenvalue problem. This leads to an algorithm with significantly reduced computational complexity, because the complexity of inversion and multiplication of diagonal matrices is $O(M)$ instead of $O(M^3)$ (M: matrix order). We call this algorithm *FastFCA*. Importantly, the FastFCA is rigorously equivalent to the FCA, because it is based on *exact* joint diagonalization based on the generalized eigenvalue problem. Consequently, the reduction in computational complexity is realized without degradation in source separation performance. Also, free of disjointness assumption like the FCA, the FastFCA enables effective source separation even in the presence of non-sparse background noise or non-disjoint source signals.

The rest of this paper is organized as follows. Section II describes the conventional FCA. Section III elaborates on the proposed FastFCA. Section IV describes a source separation experiment. Section V concludes the paper.

II. Full-Rank Spatial Covariance Analysis (FCA)

A. Mixing Model

Suppose $N\,(\geq 2)$ source signals are mixed, and observed by an array of $M\,(\geq 2)$ microphones. Let the signal observed by the mth microphone in the short-time Fourier transform (STFT) domain be $y_m(t,f)$. Here, $m \in \{1,\ldots,M\}$ is the microphone index, $t \in \{1,\ldots,T\}$ the frame index, and

ISBN 978-90-827970-1-5

Algorithm 1. Full-rank spatial covariance analysis (FCA).
Input: $\mathbf{y}(t,f)$ $(t=1,\ldots,T; f=1,\ldots,F)$.
Output: $\boldsymbol{\mu}_\nu(t,f)$ $(\nu=1,\ldots,N; t=1,\ldots,T; f=1,\ldots,F)$.

Initialize $v_\nu(t,f)$ $(\nu=1,\ldots,N; t=1,\ldots,T; f=1,\ldots,F)$ and $\mathbf{R}_\nu(f)$ $(\nu=1,\ldots,N; f=1,\ldots,F)$.
for $l=1$ to L **do**
  % E step
  $\boldsymbol{\mu}_\nu(t,f) \leftarrow v_\nu(t,f)\mathbf{R}_\nu(f)\left(\sum_{\nu'=1}^{N} v_{\nu'}(t,f)\mathbf{R}_{\nu'}(f)\right)^{-1}\mathbf{y}(t,f)$ $(\nu=1,\ldots,N; t=1,\ldots,T; f=1,\ldots,F)$.
  $\boldsymbol{\Sigma}_\nu(t,f) \leftarrow v_\nu(t,f)\mathbf{R}_\nu(f)\left(\sum_{\nu'=1}^{N} v_{\nu'}(t,f)\mathbf{R}_{\nu'}(f)\right)^{-1}\left(\sum_{\nu'\neq\nu} v_{\nu'}(t,f)\mathbf{R}_{\nu'}(f)\right)$ $(\nu=1,\ldots,N; t=1,\ldots,T; f=1,\ldots,F)$.
  % M step
  $v_\nu(t,f) \leftarrow \frac{1}{M}\mathrm{tr}\left[\mathbf{R}_\nu(f)^{-1}(\boldsymbol{\mu}_\nu(t,f)\boldsymbol{\mu}_\nu(t,f)^H + \boldsymbol{\Sigma}_\nu(t,f))\right]$ $(\nu=1,\ldots,N; t=1,\ldots,T; f=1,\ldots,F)$.
  $\mathbf{R}_\nu(f) \leftarrow \frac{1}{T}\sum_{t=1}^{T}\frac{1}{v_\nu(t,f)}(\boldsymbol{\mu}_\nu(t,f)\boldsymbol{\mu}_\nu(t,f)^H + \boldsymbol{\Sigma}_\nu(t,f))$ $(\nu=1,\ldots,N; f=1,\ldots,F)$.
end for

$f \in \{1,\ldots,F\}$ the frequency-bin index. We refer to the vector $\mathbf{y}(t,f) = \begin{pmatrix} y_1(t,f) & \ldots & y_M(t,f)\end{pmatrix}^T$ composed of the signals observed by all microphones as the *observation vector*.

The observation vector $\mathbf{y}(t,f) \in \mathbb{C}^M$ is considered to be composed of N *source images* $\mathbf{x}_1(t,f),\ldots,\mathbf{x}_N(t,f) \in \mathbb{C}^M$ corresponding to the N source signals:

$$\mathbf{y}(t,f) = \sum_{\nu=1}^{N}\mathbf{x}_\nu(t,f). \tag{1}$$

The source separation problem we deal with in this paper is one of estimating each source image $\mathbf{x}_\nu(t,f)$ given the observation vector $\mathbf{y}(t,f)$.

B. Probabilistic Modeling with Full-rank Spatial Covariance Model

The probability distribution of each source image $\mathbf{x}_\nu(t,f)$ is modeled by a time-varying complex Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{S}_\nu(t,f)$:

$$p(\mathbf{x}_\nu(t,f)) = \mathcal{N}(\mathbf{x}_\nu(t,f); \mathbf{0}, \mathbf{S}_\nu(t,f)). \tag{2}$$

This is said to be time-varying, because the covariance matrix $\mathbf{S}_\nu(t,f)$ depends on t. Here, $\mathcal{N}(\boldsymbol{\eta}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

The covariance matrix $\mathbf{S}_\nu(t,f)$ is parametrized by a parameter $v_\nu(t,f)$ modeling the power spectrum of each source signal and a parameter $\mathbf{R}_\nu(f)$ modeling its spatial characteristics:

$$\mathbf{S}_\nu(t,f) = \underbrace{v_\nu(t,f)}_{\text{power spectrum}} \times \underbrace{\mathbf{R}_\nu(f)}_{\text{spatial characteristics}}. \tag{3}$$

The time-variant parameter $v_\nu(t,f)$ is called a *power parameter*, and assumed to be positive. The time-invariant parameter $\mathbf{R}_\nu(f)$ is called a *spatial covariance matrix*, and assumed to be Hermitian positive definite.

C. EM Algorithm for Parameter Estimation

As shown in (1) and (2), the observation vector $\mathbf{y}(t,f)$ equals the sum of the source images $\mathbf{x}_1(t,f),\ldots,\mathbf{x}_N(t,f)$, each of which follows a complex Gaussian distribution. Therefore, the reproduction property of the complex Gaussian distribution implies that $\mathbf{y}(t,f)$ follows a complex Gaussian distribution again (see also (3)):

$$p(\mathbf{y}(t,f)) = \mathcal{N}\left(\mathbf{y}(t,f); \mathbf{0}, \sum_{\nu=1}^{N} v_\nu(t,f)\mathbf{R}_\nu(f)\right). \tag{4}$$

The model parameters $v_\nu(t,f)$ and $\mathbf{R}_\nu(f)$ are estimated in the maximum-likelihood sense.

This can be done by the EM algorithm, which alternates the E step and the M step. The E step updates the posterior distribution $p(\mathbf{x}_\nu(t,f) \mid \mathbf{y}(t,f))$ of the source image $\mathbf{x}_\nu(t,f)$ based on the current estimates of the parameters $v_\nu(t,f)$ and $\mathbf{R}_\nu(f)$. Because the posterior distribution also turns out to be a complex Gaussian distribution, it suffices to update its mean $\boldsymbol{\mu}_\nu(t,f)$ and covariance matrix $\boldsymbol{\Sigma}_\nu(t,f)$. The M step updates the parameters so as to increase an auxiliary Q function, which is defined using the posterior distribution updated in the E step. This EM algorithm is theoretically guaranteed to increase the likelihood function monotonically. The algorithm is shown in Algorithm 1, where l denotes the iteration index, and L the number of iterations.

The mean $\boldsymbol{\mu}_\nu(t,f)$ of the posterior distribution obtained by Algorithm 1 can be regarded as an estimate of the source image $\mathbf{x}_\nu(t,f)$. Indeed, it is the minimum mean square error (MMSE) estimator of $\mathbf{x}_\nu(t,f)$.

ISBN 978-90-827970-1-5

TABLE I
COMPARISON OF THE NUMBER OF MATRIX OPERATIONS OF ORDER $O(M^3)$. NOTE THAT THE FASTFCA CAN BE APPLIED TO THE TWO-SOURCE CASE $N = 2$ ONLY.

	FCA	FastFCA
$\mathbf{A}^{-1}$	$(N+T)FL$	F
$\mathbf{AB}$	$2NTFL$	$F(L-1)$
$\mathbf{Ap} = \lambda\mathbf{Bp}$	0	FL
Total	$(N+T+2NT)FL$	$2FL$

D. Drawback

The main drawback of the FCA in Algorithm 1 is expensive computation. Especially, the E step of Algorithm 1 requires matrix inversion and matrix multiplication, which are of complexity $O(M^3)$, at each time-frequency point and each iteration. Table I shows the number of matrix operations of complexity $O(M^3)$ in Algorithm 1 (see the column labeled 'FCA') as a rough estimate of the computational complexity (see Section IV for evaluation in terms of the computation time).

For example, consider the experimental setting in Section IV: $N = 2$; $T = 249$; $F = 512$; $L = 10$. In this case, the number of matrix inversions is $(N+T)FL = 1285120$, the number of matrix multiplications is $2NTFL = 5099520$, and the total number of matrix operations of complexity $O(M^3)$ is $(N+T+2NT)FL = 6384640$.

III. FASTFCA

In this section, we propose a computationally more efficient algorithm for estimating the parameters of (4). This algorithm is called *FastFCA*.

A. Approach: Joint Diagonalization

The FastFCA exploits the well-known fact that the computational complexity of inversion and multiplication of diagonal matrices is just $O(M)$ instead of $O(M^3)$:

$$\begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_M \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\alpha_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\alpha_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\alpha_M} \end{pmatrix}, \tag{5}$$

$$\begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_M \end{pmatrix} \begin{pmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_M \end{pmatrix} = \begin{pmatrix} \alpha_1\beta_1 & 0 & \cdots & 0 \\ 0 & \alpha_2\beta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_M\beta_M \end{pmatrix}. \tag{6}$$

Now, suppose the covariance matrices $\mathbf{S}_\nu(t,f) = v_\nu(t,f)\mathbf{R}_\nu(f)$ are all diagonal. Then, the matrix inversion and the matrix multiplication in Algorithm 1 can be performed in $O(M)$. In practice, however, the covariance matrix $\mathbf{S}_\nu(t,f)$ usually has non-zero off-diagonal entries. This is because different elements of a source image $\mathbf{x}_\nu(t,f)$ correspond to different microphones, and are usually highly correlated.

This motivates us to consider diagonalizing N spatial covariance matrices $\mathbf{R}_1(f), \cdots, \mathbf{R}_N(f)$ jointly. That is, we consider transforming $\mathbf{R}_1(f), \cdots, \mathbf{R}_N(f)$ into some diagonal matrices $\mathbf{\Lambda}_1(f), \cdots, \mathbf{\Lambda}_N(f)$ by one non-singular matrix $\mathbf{P}(f)$ as follows:

$$\mathbf{P}(f)^H \mathbf{R}_\nu(f) \mathbf{P}(f) = \mathbf{\Lambda}_\nu(f) \ (\nu = 1, \ldots, N). \tag{7}$$

For $N = 2$, such $\mathbf{P}(f)$ and $\mathbf{\Lambda}_1(f), \mathbf{\Lambda}_2(f)$ are obtained based on the generalized eigenvalue problem. For $N \geq 3$, an exact solution does not exist in general, but joint approximate diagonalization [7] can still be performed. Here we focus on the former case, while we refer the readers to [8] for the latter case.

Let $N = 2$ in the following. By solving the generalized eigenvalue problem of $\mathbf{R}_1(f)$ and $\mathbf{R}_2(f)$, we obtain the M generalized eigenvalues $\lambda_1(f), \cdots, \lambda_M(f)$ and M generalized eigenvectors $\mathbf{p}_1(f), \cdots, \mathbf{p}_M(f)$ corresponding to $\lambda_1(f), \cdots, \lambda_M(f)$, respectively. They satisfy

$$\mathbf{R}_1(f)\mathbf{p}_m(f) = \lambda_m(f)\mathbf{R}_2(f)\mathbf{p}_m(f). \tag{8}$$

The generalized eigenvectors $\mathbf{p}_1(f), \cdots, \mathbf{p}_M(f)$ can be chosen so that they satisfy the following $\mathbf{R}_2(f)$-orthonormality:

$$\mathbf{p}_\mu(f)^H \mathbf{R}_2(f) \mathbf{p}_m(f) = \delta_{\mu m}, \tag{9}$$

where $\delta_{\mu m}$ denotes the Kronecker delta. The equations (8) and (9) can be rewritten in matrix form as follows:

$$\begin{cases} \mathbf{P}(f)^H \mathbf{R}_1(f) \mathbf{P}(f) = \mathbf{\Lambda}(f), \\ \mathbf{P}(f)^H \mathbf{R}_2(f) \mathbf{P}(f) = \mathbf{I}. \end{cases} \tag{10}$$

Here, $\mathbf{P}(f)$ denotes the matrix composed of the generalized eigenvectors $\mathbf{p}_1(f), \cdots, \mathbf{p}_M(f)$, and $\mathbf{\Lambda}(f)$ the diagonal matrix composed of the generalized eigenvalues $\lambda_1(f), \cdots, \lambda_M(f)$. Equation (10) is a special case of (7) with $N = 2$, $\mathbf{\Lambda}_1(f) = \mathbf{\Lambda}(f)$, and $\mathbf{\Lambda}_2(f) = \mathbf{I}$.

B. Derivation of FastFCA

Here we derive the FastFCA based on the joint diagonalization in (10).

1) E Step: Suppose we have obtained a matrix $\mathbf{P}(f)$ and a diagonal matrix $\mathbf{\Lambda}(f)$ satisfying (10) for the current estimates $\mathbf{R}_1(f)$ and $\mathbf{R}_2(f)$ of the spatial covariance matrices. By solving (10) for $\mathbf{R}_1(f)$ and $\mathbf{R}_2(f)$ and plugging it in the E step of Algorithm 1, we obtain

$$\begin{aligned} \boldsymbol{\mu}_\nu(t,f) \leftarrow{}& (\mathbf{P}(f)^H)^{-1}(v_\nu(t,f)\mathbf{\Lambda}_\nu(f))(v_1(t,f)\mathbf{\Lambda}(f) \\ &+ v_2(t,f)\mathbf{I})^{-1}\mathbf{P}(f)^H\mathbf{y}(t,f), \end{aligned} \tag{11}$$

$$\begin{aligned} \mathbf{\Sigma}(t,f) \leftarrow{}& (\mathbf{P}(f)^H)^{-1}(v_1(t,f)v_2(t,f)\mathbf{\Lambda}(f)) \\ &\times (v_1(t,f)\mathbf{\Lambda}(f) + v_2(t,f)\mathbf{I})^{-1}\mathbf{P}(f)^{-1}. \end{aligned} \tag{12}$$

ISBN 978-90-827970-1-5

Algorithm 2. FastFCA.

Input: $\mathbf{y}(t,f)$ $(t=1,\ldots,T; f=1,\ldots,F)$.

Output: $\boldsymbol{\mu}_\nu(t,f)$ $(\nu=1,\ldots,N; t=1,\ldots,T; f=1,\ldots,F)$.

Initialize $v_\nu(t,f)$ $(\nu=1,\ldots,N; t=1,\ldots,T; f=1,\ldots,F)$ and $\mathbf{R}_\nu(f)$ $(\nu=1,\ldots,N; f=1,\ldots,F)$.
for $l=1$ to L **do**
  % Joint diagonalization
  if $l=1$ **then**
    For each f, solve the generalized eigenvalue problem of $\mathbf{R}_1(f)$ and $\mathbf{R}_2(f)$ to obtain a diagonal matrix $\boldsymbol{\Lambda}(f)$ and a non-singular matrix $\mathbf{P}(f)$ satisfying (10).
  else
    For each f, solve the generalized eigenvalue problem of $\mathbf{R}'_1(f)$ and $\mathbf{R}'_2(f)$ to obtain a diagonal matrix $\boldsymbol{\Lambda}(f)$ and a non-singular matrix $\mathbf{Q}(f)$ satisfying $\mathbf{Q}(f)^H\mathbf{R}'_1(f)\mathbf{Q}(f)=\boldsymbol{\Lambda}(f)$ and $\mathbf{Q}(f)^H\mathbf{R}'_2(f)\mathbf{Q}(f)=\mathbf{I}$.
    $\mathbf{P}(f)\leftarrow\mathbf{P}(f)\mathbf{Q}(f)$ $(f=1,\ldots,F)$.
  end if
  % E step
  Update $\mathbf{y}'(t,f)$ $(t=1,\ldots,T; f=1,\ldots,F)$ by (15).
  Update $\boldsymbol{\mu}'_\nu(t,f)$ $(\nu=1,\ldots,N; t=1,\ldots,T; f=1,\ldots,F)$ by (16).
  Update $\boldsymbol{\Sigma}'(t,f)$ $(t=1,\ldots,T; f=1,\ldots,F)$ by (17).
  % M step
  Update $v_\nu(t,f)$ $(\nu=1,\ldots,N; t=1,\ldots,T; f=1,\ldots,F)$ by (18).
  Update $\mathbf{R}'_\nu(f)$ $(\nu=1,\ldots,N; f=1,\ldots,F)$ by (21).
end for
$\boldsymbol{\mu}_\nu(t,f)\leftarrow(\mathbf{P}(f)^H)^{-1}\boldsymbol{\mu}'_\nu(t,f)$ $(\nu=1,\ldots,N; t=1,\ldots,T; f=1,\ldots,F)$.

Here, $\boldsymbol{\Lambda}_\nu(f)$ in (11) is defined by $\boldsymbol{\Lambda}_1(f)\triangleq\boldsymbol{\Lambda}(f)$ and $\boldsymbol{\Lambda}_2(f)\triangleq\mathbf{I}$. It is straightforward to show $\boldsymbol{\Sigma}_1(t,f)=\boldsymbol{\Sigma}_2(t,f)$, and so we have defined $\boldsymbol{\Sigma}(t,f)\triangleq\boldsymbol{\Sigma}_1(t,f)=\boldsymbol{\Sigma}_2(t,f)$ in (12).

Defining $\boldsymbol{\mu}'_\nu(t,f)$, $\boldsymbol{\Sigma}'(t,f)$, and $\mathbf{y}'(t,f)$ by

$$\boldsymbol{\mu}'_\nu(t,f)\triangleq\mathbf{P}(f)^H\boldsymbol{\mu}_\nu(t,f), \tag{13}$$
$$\boldsymbol{\Sigma}'(t,f)\triangleq\mathbf{P}(f)^H\boldsymbol{\Sigma}(t,f)\mathbf{P}(f), \tag{14}$$
$$\mathbf{y}'(t,f)\triangleq\mathbf{P}(f)^H\mathbf{y}(t,f), \tag{15}$$

we have

$$\boldsymbol{\mu}'_\nu(t,f)\leftarrow v_\nu(t,f)\boldsymbol{\Lambda}_\nu(f)(v_1(t,f)\boldsymbol{\Lambda}(f)+v_2(t,f)\mathbf{I})^{-1}\times\mathbf{y}'(t,f), \tag{16}$$
$$\boldsymbol{\Sigma}'(t,f)\leftarrow v_1(t,f)v_2(t,f)\boldsymbol{\Lambda}(f)(v_1(t,f)\boldsymbol{\Lambda}(f)+v_2(t,f)\mathbf{I})^{-1}. \tag{17}$$

Note that the matrix inversion and the matrix multiplication in (16) and (17) can be performed in $O(M)$, owing to the joint diagonalization.

2) M Step: The M step of the FastFCA updates the parameter estimates based on $\boldsymbol{\mu}'_\nu(t,f)$ and $\boldsymbol{\Sigma}'(t,f)$ updated in the E step. By solving (13) and (14) for $\boldsymbol{\mu}_\nu(t,f)$ and $\boldsymbol{\Sigma}(t,f)$ and plugging them into the M step of Algorithm 1, we obtain

$$v_\nu(t,f)\leftarrow\frac{1}{M}\mathrm{tr}(\boldsymbol{\Lambda}_\nu(f)^{-1}(\boldsymbol{\Sigma}'(t,f)+\boldsymbol{\mu}'_\nu(t,f)\boldsymbol{\mu}'_\nu(t,f)^H)), \tag{18}$$

$$\mathbf{R}_\nu(f)\leftarrow\frac{1}{T}\sum_{t=1}^{T}\frac{1}{v_\nu(t,f)}(\mathbf{P}(f)^H)^{-1}(\boldsymbol{\Sigma}'(t,f)+\boldsymbol{\mu}'_\nu(t,f)\boldsymbol{\mu}'_\nu(t,f)^H)\mathbf{P}(f)^{-1}. \tag{19}$$

Defining $\mathbf{R}'_\nu(f)$ by

$$\mathbf{R}'_\nu(f)\triangleq\mathbf{P}(f)^H\mathbf{R}_\nu(f)\mathbf{P}(f), \tag{20}$$

we have

$$\mathbf{R}'_\nu(f)\leftarrow\frac{1}{T}\sum_{t=1}^{T}\frac{1}{v_\nu(t,f)}(\boldsymbol{\Sigma}'(t,f)+\boldsymbol{\mu}'_\nu(t,f)\boldsymbol{\mu}'_\nu(t,f)^H). \tag{21}$$

The resulting algorithm is shown in Algorithm 2. To further reduce the number of matrix inversions and matrix multiplications, the generalized eigenvalue problem for $\mathbf{R}'_1(f)$ and $\mathbf{R}'_2(f)$ is solved instead of that for $\mathbf{R}_1(f)$ and $\mathbf{R}_2(f)$ at the second iteration and forth. See [9] for more details.

C. Advantage of FastFCA

Table I compares the FCA and the FastFCA in terms of the number of matrix operations of complexity $O(M^3)$. We see

ISBN 978-90-827970-1-5 © EURASIP 2018

that the FastFCA requires a significantly reduced number of matrix operations of complexity $O(M^3)$ compared to the FCA. In Table I, inversions and multiplications of diagonal matrices were not counted because their complexity is $O(M)$ instead of $O(M^3)$. Although the FastFCA requires additional generalized eigenvalue decompositions (complexity $O(M^3)$), they are not required frame-wise. Consequently, the total number of matrix operations of complexity $O(M^3)$ is reduced significantly by the FastFCA. Indeed, the ratio

$$\frac{(N+T+2NT)FL}{2FL} = 1 + \frac{5}{2}T \ (\because N = 2) \qquad (22)$$

is always greater than one, and $\gg 1$ when $T \gg 1$.

For example, in the experimental setting in Section IV, the number of matrix inversions is $F = 512$, the number of matrix multiplications is $F(L-1) = 4608$, the number of generalized eigenvalue decompositions is $FL = 5120$, and the total number of matrix operations of complexity $O(M^3)$ is $2FL = 10240$.

IV. Source Separation Experiment

We conducted a source separation experiment to compare the proposed FastFCA (Algorithm 2) with the conventional FCA (Algorithm 1). Both methods were implemented in MATLAB (R2013a), and run on an Intel i7-2600 3.4-GHz octal-core CPU.

We generated source images by convolving 8-s English speech signals with room impulse responses measured in a room depicted in Fig. 1. Then, we generated observed signals by adding these source images. The reverberation time RT_{60} was 130, 200, 250, 300, 370, or 440 ms. The sampling frequency was 16 kHz. The observed signals were transformed into the time-frequency domain by the STFT, where the frame length was 1024 points (64 ms), the frame shift was 512 points (32 ms), and the window was the square root of Hann.

We conducted ten trials per reverberation time while changing speech signals. In both methods, the parameter estimates were initialized based on time-frequency masks estimated by the clustering method in [10]. The number of EM iterations was 10.

Figure 2 shows the computation time in terms of the real time factor (RTF) for the EM iterations averaged over all trials and all reverberation times. Figure 3 shows the source separation performance in terms of the signal-to-distortion ratio (SDR) [11] averaged over all sources and all trials. The input SDR was 0 dB.

V. Conclusion

In this paper, we proposed the FastFCA, joint diagonalization based acceleration of the FCA. In the source separation experiment, the FastFCA was over 250 times faster than the FCA with virtually the same SDR.

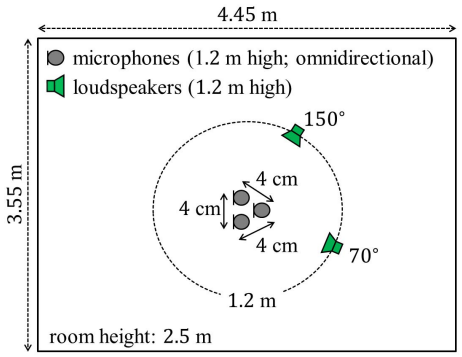

Fig. 1. Experimental setting.

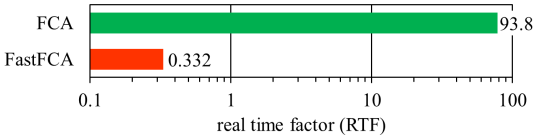

Fig. 2. Computation time in terms of the real time factor (RTF).

References

[1] N. Q. K. Duong, E. Vincent, and R. Gribonval, "Under-determined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Trans. ASLP*, vol. 18, no. 7, pp. 1830–1840, Sep. 2010.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, 2001.

[3] J. Barker, R. Marxer, E. Vincent, and S. Watanabe, "The third 'CHiME' speech separation and recognition challenge: Dataset, task and baselines," in *Proc. ASRU*, Dec. 2015, pp. 504–511.

[4] R. Sakanashi, S. Miyabe, T. Yamada, and S. Makino, "Comparison of superimposition and sparse models in blind source separation by multichannel Wiener filter," in *Proc. APSIPA*, Dec. 2012.

[5] N. Ito, S. Araki, T. Yoshioka, and T. Nakatani, "Relaxed disjointness based clustering for joint blind source separation and dereverberation," in *Proc. IWAENC*, Sept. 2014, pp. 268–272.

[6] Ö. Yılmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. SP*, vol. 52, no. 7, pp. 1830–1847, Jul. 2004.

[7] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. on SP*, vol. 50, no. 7, pp. 1545–1553, July 2002.

[8] N. Ito and T. Nakatani, "FastFCA-AS: Joint diagonalization based acceleration of full-rank spatial covariance analysis for separating any number of sources," *arXiv preprint*, May 2018, arXiv:1805.09498.

[9] N. Ito, S. Araki, and T. Nakatani, "FastFCA: A joint diagonalization based fast algorithm for audio source separation using a full-rank spatial covariance model," *arXiv preprint*, May 2018, arXiv: 1805.06572.

[10] H. Sawada, S. Araki, and S. Makino, "Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment," *IEEE Trans. ASLP*, vol. 19, no. 3, pp. 516–527, Mar. 2011.

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

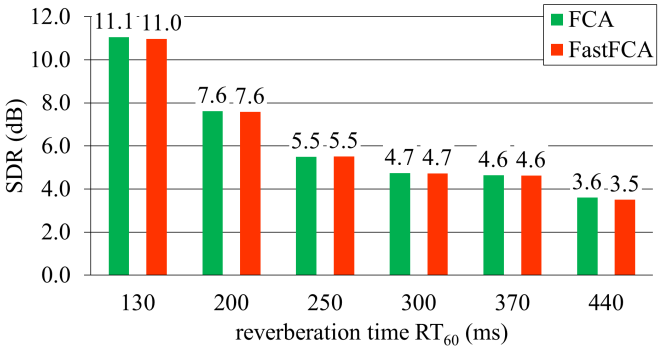

Fig. 3. Source separation performance in terms of the SDR.

ISBN 978-90-827970-1-5