Blind Source Separation of Nondisjoint Sources in The Time-Frequency Domain with Model-Based Determination of Source Contribution

Jalil Taghia, Timo Gerkmann, and Arne Leijon
KTH Royal Institute of Technology,
School of Electrical Engineering, Sound and Image Processing Lab., Stockholm, Sweden
{taghia, gerkmann, leijon}@kth.se

***Abstract*— *While most blind source separation (BSS) algorithms rely on the assumption that at most one source is dominant at each time-frequency (TF) point, recently, two BSS approaches, [1], [2], have been proposed that allow multiple active sources at time-frequency (TF) points under certain assumptions. In both algorithms, the active sources in every single TF point are found by an exhaustive search through an optimization procedure which is computationally expensive. In this work, we address this limitation and avoid the exhaustive search by determining the source contribution in every TF point. The source contributions are expressed by a set of posterior probabilities. Hereby, we propose a model-based blind source separation algorithm that allows sources to be nondisjoint in the TF domain while being computationally more tractable. The proposed BSS approach is shown to be robust with respect to different reverberation times and microphone spacings.*

***Keywords*— *Blind source separation, underdetermined BSS, model-based BSS, nondisjoint sources, convolutive mixtures*

I. INTRODUCTION

A general framework for underdetermined blind source separation (UBSS) is to exploit the sparsity of sources. The main assumption used in these approaches is that the sources are time-frequency (TF) disjoint. In other words, there is at most one source present at every point in the TF domain. This assumption is rather restrictive and cannot always be satisfied. In [1], it is proposed that the TF-disjoint condition can be relaxed by allowing sources to be nondisjoint in the TF domain under certain assumptions: the number of active sources at any TF point is strictly less than that of the observations, and the column vectors of the mixing matrix are pairwise linearly independent. The first assumption is still rather restrictive. In [2], a UBSS approach is proposed which only requires the number of active sources at any TF point to be no larger than the number of observations under three main assumptions: A1) any $M \times M$ sub-matrix of the mixing matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]$ is of full rank, where M and N indicate the number of observations and the number of sources, respectively. A2) The number of active sources L at any TF point does not exceed the number of observations. A3) For any matrix $\mathbf{H}^{(k)} \in \mathbf{H}$ (k^{th} sub-matirx of $\mathbf{H}$), all nonzero columns of $\rho(\mathbf{H}^{(k)})$ are linearly independent. The matrix set $\mathbf{H}$ is defined as $\mathbf{H}^{(k)} = \left\{\mathbf{H}^{(k)} \middle| \mathbf{H}^{(k)} = (\mathbf{T}^{(i)})^{-1}\mathbf{T}^{(j)}, \mathbf{T}^{(i)}, \mathbf{T}^{(j)} \in \mathbf{T}\right\}_{i \neq j}$, where $\mathbf{T}$ is a set of all $M \times M$ sub-matrices of $\mathbf{A}$ defined by $\mathbf{T} = \left\{\mathbf{T}^{(i)} \middle| \mathbf{T}^{(i)} = [\mathbf{a}_{\alpha_1}, \ldots, \mathbf{a}_{\alpha_M}]\right\}$. The matrix operator ρ for $\mathbf{Y} = [\mathbf{y}_1, \ldots \mathbf{y}_P]^T$ is defined as

$$\rho(\mathbf{Y}) = \begin{pmatrix} \mathbf{y}_1^T \odot \mathbf{y}_2^H \\ \cdots \\ \mathbf{y}_1^T \odot \mathbf{y}_P^H \\ \mathbf{y}_2^T \odot \mathbf{y}_3^H \\ \cdots \\ \mathbf{y}_2^T \odot \mathbf{y}_3^H \\ \vdots \\ \mathbf{y}_{P-1}^T \odot \mathbf{y}_P^H \end{pmatrix}, \tag{1}$$

where $\odot$, T and H are Hadamard product, transpose and Hermitian, respectively [2].

In both approaches, [1] and [2], an exhaustive search has to be done to find the indices of dominant sources in every TF point through a minimization or maximization procedure. To be specific, the number of trials is equal to the number of all $M \times M$ sub-matrices of the mixing matrix $\mathbf{A}$ which is equal to the size of set $\mathbf{T}$. Since this blind search is done for all TF points, it becomes computationally demanding with increasing the number of sources N and the number of observations M. In this work, we address this issue and propose a new approach to deal with this limitation. The idea is to find the indices of dominant sources by calculating the contribution of each source in every TF point instead of performing a blind search. This contribution is determined by a set of posterior probabilities. Here, we propose that by using this knowledge, we can avoid the blind search and relax the disjoint condition on sources while make it computationally tractable.

By finding indices of dominant sources in all TF points, a frequency-dependent unmixing matrix is identified by which we derive the TF-distribution values of separated sources in every frequency bin. Although a satisfactory separation is expected in all frequency bins, combining them to recover the original sources is a challenge because there exist unknown permutations associated with individual frequency bins. Hence, permutation ambiguity should be looked after properly so that the separated frequency components from the same source are grouped together. The efficiency of the proposed approach is correlated with the accuracy of the permutation alignment. In [3], a region-growing permutation alignment approach has been proposed in the frequency-domain. The method is based on an inter-frequency dependence measure:

This work was funded by the European Commission within the Marie Curie ITN AUDIS, grant PITNGA-2008-214699.

the power of separated signals. However, in this work, posterior probability sequences are used instead of the power of separated signals. Posterior probability sequences in a frequency bin are derived by making sequences of probabilities over time for every source. We show that posterior probability sequences are better candidates to measure the inter-frequency dependency.

This paper is organized as follows. Section II states the problem formulation. Section III and IV present the determination of the source contribution based on calculating the posterior probabilities in every TF point. Section V explains the proposed unmixing-matrix estimation based on the determined source contribution. Section VI reports experimental results. Section VII concludes this paper.

II. PROBLEM FORMULATION

Let $s_1(t), \dots, s_N(t)$ be desired sources and $x_1(t), \dots, x_M(t)$ be observation mixtures, where t is the time index. Assuming a convolutive mixture model, the observation $x_m(t)$ is given by

$$x_m(t) = \sum_{n=1}^{N} \sum_{l} a_{mn}(l) s_n(t-l), \qquad (2)$$

where $a_{mn}(l)$ represent the impulse response from source n to microphone m. The convolutive mixture model (2) is transformed to the time-frequency representation by using the short time Fourier transform (STFT). Assuming that the STFT frame size is long enough to cover the main part of the impulse response $a_{mn}(l)$, (2) can be approximated as an instantaneous mixture model at each frequency bin as $x_m(\tau, f) = \sum_{n=1}^{N} a_{mn}(f) s_n(\tau, f)$, where $a_{mn}(f)$ is the frequency response from source n to microphone m and $s_n(\tau, f)$ is the time-frequency representation of $s_n(t)$, where τ and f denote the time-frame index and the frequency bin, respectively.

In vector notation,

$$\mathbf{x}(\tau, f) = \sum\nolimits_{n=1}^{N} \mathbf{a}_n(f) s_n(\tau, f), \qquad (3)$$

where $\mathbf{x} = [x_1, ..., x_M]^T$ is the mixture STFT vector, $\mathbf{a}_n = [a_{1n}, ..., a_{Mn}]^T$, τ indicates the frame index and f indicates the frequency bin. Equation (3) under sparsity assumption can be expressed as

$$\mathbf{x}(\tau, f) = \mathbf{a}_{n^*(\tau,f)}(f) s_{n^*(\tau,f)}(\tau, f), \qquad (4)$$

where subscript $n^*(\tau, f)$ is the index of the dominant source in TF point (τ, f).

III. DETERMINATION OF SOURCE CONTRIBUTION

In this section we propose to determine the contribution of each source in every TF point by a set of posterior probabilities. Before calculating posterior probabilities, a unit-norm normalization and pre-whitening procedure is necessary, as a preprocessing stage. Unit-norm normalization removes the dependence on the source amplitude and can be achieved as

$$\bar{\mathbf{x}}(\tau, f) = \frac{\mathbf{x}(\tau, f)}{||\mathbf{x}(\tau, f)||}, \qquad (5)$$

where $\bar{\mathbf{x}}(\tau, f)$ is called spatial direction vectors of the mixture. A pre-whitening [4] is done by multiplying spatial direction vectors $\bar{\mathbf{x}}(\tau, f)$ by the whitening matrix $\mathbf{V}$, as

$$\bar{\mathbf{x}}(\tau, f) \leftarrow \mathbf{V}\bar{\mathbf{x}}(\tau, f), \qquad (6)$$

where $\mathbf{V} = \sqrt{\mathbf{D}}\mathbf{E}^H$, and $\mathbf{E}$ and $\mathbf{D}$ are calculated from eigenvalue decomposition of the correlation matrix of the mixtures $\bar{\mathbf{x}}(f)$ at frequency bin f across time. The normalizing procedure is done one more time after whitening.

To compute the posterior probabilities, we adopt the line orientation idea, [5], which has been employed in [6]. Considering (4), spatial direction vectors $\bar{\mathbf{x}}(\tau, f)$ are modeled by a mixture of complex Gaussian density functions. The complex Gaussian density function is given by

$$p(\bar{\mathbf{x}}(\tau, f)|\mathbf{m}_i, \sigma_i) = \frac{1}{(\pi\, \sigma_i^2)^{M-1}} e^{-\frac{||\bar{\mathbf{x}}(\tau,f) - (\mathbf{m}_i^H \bar{\mathbf{x}}(\tau,f))\mathbf{m}_i||^2}{\sigma_i^2}} \qquad (7)$$

where $\mathbf{m}$ is the centroid with unit norm and σ_i^2 is the variance. The distance $||\bar{\mathbf{x}}(\tau, f) - (\mathbf{m}_i^H \bar{\mathbf{x}}(\tau, f))\mathbf{m}_i||$ calculates the difference between point $\bar{\mathbf{x}}(\tau, f)$ and $(\mathbf{m}_i^H \bar{\mathbf{x}}(\tau, f))\mathbf{m}_i$ which is the orthogonal projection of $\bar{\mathbf{x}}(\tau, f)$ onto the subspace spanned by $\mathbf{m}_i$, and determines the probability that $\bar{\mathbf{x}}(\tau, f)$ belongs to the i^{th} class. The source subscript n is changed to i to highlight that there exist permutation ambiguities along the frequency bins.

The density function $p(\bar{\mathbf{x}}(\tau, f))$ can be determined by a multivariate mixture model

$$p(\bar{\mathbf{x}}(\tau, f)|\Theta) = \sum\nolimits_{i=1}^{N} \beta_i p(\bar{\mathbf{x}}(\tau, f)|\mathbf{m}_i, \sigma_i) \qquad (8)$$

with $\Theta = \{(\mathbf{m}_1, \sigma_1, \beta_1), \dots, (\mathbf{m}_N, \sigma_N, \beta_N)\}$ as the parameter set. β_i $(0 \leq \beta_i \leq 1; \sum_{i=1}^{N} \beta_i = 1)$ are the mixture ratios modeled by a Dirichlet distribution with prior hyperparameter ϕ. The expectation maximization (EM) algorithm [7] is employed to estimate posterior probabilities. Posterior probabilities are calculated for all sources, by

$$\gamma(C_i|\bar{\mathbf{x}}(\tau, f), \hat{\Theta}) = \frac{\hat{\beta}_i p(\bar{\mathbf{x}}(\tau, f)|\hat{\mathbf{m}}_i, \hat{\sigma}_i)}{\sum_{i=1}^{N} \hat{\beta}_i p(\bar{\mathbf{x}}(\tau, f)|\hat{\mathbf{m}}_i, \hat{\sigma}_i)} \qquad (9)$$

where $\hat{\Theta} = \left\{(\hat{\mathbf{m}}_1, \hat{\sigma}_1, \hat{\beta}_1), \dots, (\hat{\mathbf{m}}_N, \hat{\sigma}_N, \hat{\beta}_N)\right\}$ is the current parameter set, and C_i implies to the i^{th} source category. In the M-step, the parameter Θ is updated. As detailed in [8], each parameter is updated as follows. The new centroid $\mathbf{m}_i$ is given by the eigenvector corresponding to the maximum eigenvalue of

$$R = \sum\nolimits_{\tau=1}^{T} \gamma(C_i|\bar{\mathbf{x}}(\tau, f), \hat{\Theta}) \bar{\mathbf{x}}(\tau, f) \bar{\mathbf{x}}^H(\tau, f), \qquad (10)$$

where T is the number of frames. The variance σ_i^2 and mixture ratio β_i are updated by

$$\sigma_i^2 = \frac{\sum_{\tau=1}^{T} \gamma(C_i|\bar{\mathbf{x}}(\tau,f),\hat{\Theta})||\bar{\mathbf{x}}(\tau,f) - (\mathbf{m}_i^H \bar{\mathbf{x}}(\tau,f))\mathbf{m}_i||^2}{(M-1)\sum_{\tau=1}^{T} \gamma(C_i|\bar{\mathbf{x}}(\tau,f),\hat{\Theta})}, \tag{11}$$

and

$$\beta_i = \frac{\sum_{\tau=1}^{T} \gamma(C_i|\bar{\mathbf{x}}(\tau,f),\hat{\Theta}) + \phi - 1}{T + N(\phi - 1)}. \tag{12}$$

We found the initialization of the model parameters important for the EM algorithm. To initialize the model parameter $\hat{\mathbf{m}}_i$, we can use a random initialization by choosing N points $\tau_1, ..., \tau_N$ beforehand and set them by $\hat{\mathbf{m}}_i \leftarrow \bar{\mathbf{x}}(\tau_i, f)$ for $i = 1, ..., N$ or, alternatively, we can use k-means clustering algorithm [9]. In the later case, first the k-means algorithm is applied to $\bar{\mathbf{x}}(\cdot, f)$ while the number of clusters is set to the number of sources (the notation $(\cdot, f)$ means across all time indices τ for a particular frequency f). Denoting $\bar{y}_i(f)$ as the i^{th} centroid in the frequency bin f, the model parameter $\hat{\mathbf{m}}_i$ is set as $\hat{\mathbf{m}}_i \leftarrow \bar{\mathbf{y}}_i(f)$ for $i = 1, ..., N$. The two other parameters are set as $\hat{\sigma}_i^2 = 0.1$ and $\hat{\beta}_i = N^{-1}$. A large number is chosen for the hyper-parameter ϕ as $\phi = 60$. The EM process repeats until convergence and, in the end, we have the posterior probabilities $\gamma(C_i|\bar{\mathbf{x}}(\tau,f),\Theta)$ for all sources and all TF points. By having these posterior probabilities, we know the contribution of each source in every TF point. We use this knowledge later in Section 6 to find indices of dominant sources contributing in a particular TF point.

IV. PERMUTATION ALIGNMENT

There exists a disorder along frequency bins so that the class order $C_1, C_2, ..., C_N$ may be different from one frequency bin to another. We need to remove the permutation ambiguity such that the same index corresponds to the same source over all TF points. Spectral amplitude and power of separated signals are common measures for calculating dependency across frequency bins for separated signals. In [3], a permutation alignment approach called *region-growing permutation alignment* is proposed that, in a region-growing manner, determines the permutation along frequency bins by measuring the correlation among the power of separated signals. In this work we use the same permutation alignment procedure as [3], but posterior probability sequences, instead of power of separated signals, are used to determine the dependency along frequency bins. We show that posterior probability sequences are better candidates by which a more robust permutation alignment can be realized. The idea is that posterior probability sequences belonging to the same source generally have similar patterns among different frequency bins [8]. The correlation coefficient of the posterior probability sequences can be used to measure similarity among patterns or, in other words, measure the inter-frequency dependence.

The goal is to define a permutation

$$\Pi_f : \{1, \ldots, N\} \rightarrow \{1, \ldots, N\}, \tag{13}$$

for all frequencies f, and then update the posterior probabilities accordingly by

$$\gamma(C_n|\bar{\mathbf{x}}(\tau,f)) \longleftarrow \gamma(C_i|\bar{\mathbf{x}}(\tau,f))|_{i=\Pi_f(n)}. \tag{14}$$

Sequences of posterior probabilities, derived in section III, are defined in each frequency bin as

$$\zeta_i^f = \gamma(C_i|\bar{\mathbf{x}}(\cdot,f)), \tag{15}$$

where ζ_i^f describes the posterior probability sequence for the i^{th} class at frequency f. The correlation coefficient between two sequences $\zeta_{i_1}^{f_l}$ and $\zeta_{i_2}^{f_2}$ can be defined as

$$\rho(\zeta_{i_1}^{f_1}, \zeta_{i_2}^{f_2}) = \frac{r_{i_1 i_2}(f_1, f_2) - \mu_{i_1}(f_1)\mu_{i_2}(f_2)}{\sigma_{i_1}(f_1)\sigma_{i_2}(f_2)}, \tag{16}$$

where $i_1, i_2 \in N$, $f_1, f_2 \in K$. $r_{i_1 i_2}(f_1, f_2) = E\left\{\zeta_{i_1}^{f_1}\zeta_{i_2}^{f_2}\right\}$, $\mu_{i_1}(f_1) = E\left\{\zeta_{i_1}^{f_1}\right\}$, and $\sigma_{i_1}(f) = \sqrt{E\left\{(\zeta_{i_1}^{f})^2\right\} - \mu_{i_1}^2(f)}$ are the estimated correlation, mean, and standard deviation, respectively. $\rho(\zeta_{i_1}^{f_1}, \zeta_{i_2}^{f_2})$ is equal to 1 when two sequences are exactly similar and -1 when they are exactly dissimilar. A score function can be defined as

$$\text{score} = \text{trace}\left(\left[\rho(\zeta_i^{f_1}, \zeta_j^{f_2})\right]_{i,j \in N}\right), \tag{17}$$

where sequences $\rho(\zeta_1^{f_1}, \zeta_1^{f_2}), \rho(\zeta_2^{f_1}, \zeta_2^{f_2}), ..., \rho(\zeta_N^{f_1}, \zeta_N^{f_2})$ have the highest similarity since they are originating from the same source.

Here, we employ the region growing permutation alignment procedure and apply it to posterior probability sequences ζ_i^f), as follows.

Step 1: a bin-wise permutation alignment is applied across all frequency bins. For the current frequency bin f and previous frequency bin $f-1$, a permutation Π_f is exploited which maximizes the score function (17).

Step 2: dividing frequencies in 3 bands: low frequencies, middle frequencies, and high frequencies for the sake of robustness. Partitioning every band into D regions where the highly related frequency bins are assigned to the same region. Region R is identified based on the criteria introduced in [3].

Step 3: select a region with the largest number of elements as a seed; merge with its neighboring regions on both sides in a region-growing style until the new region covers the full frequency band.

Step 4: after permutation corrections in all frequency bands, the centroids for all bands are calculated and merged together.

Fig. 1 illustrates the score values (17) for every pair of frequencies. Every point corresponds to a pair of frequencies. Points with high score can be interpreted as points at which the permutation is most likely aligned. From this figure, we can see that posterior probability sequences show higher score values (specially at high frequencies but also at low frequencies) and, hence, they are better candidates for the permutation alignment procedure than the power of separated signals.

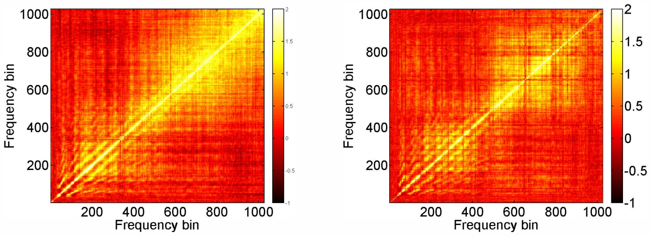

Fig. 1. Score values (17) calculated for every pair of frequencies after applying the permutation alignment procedure on posterior probability sequences (left) and the power of separated signals (right). Points with higher values (energy) interpret as higher confidence in the permutation alignment between the corresponding two frequencies.

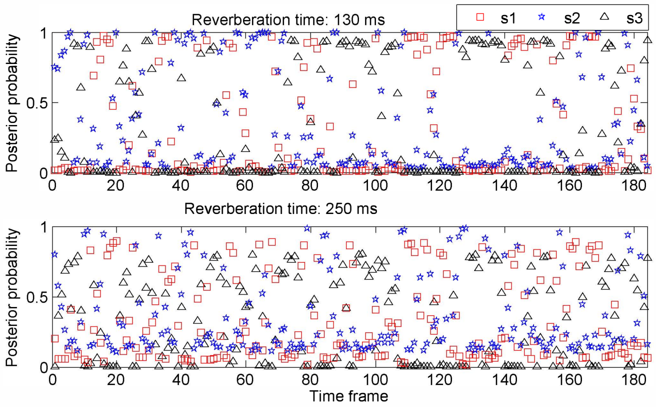

Fig. 2. Posterior probabilities at a certain frequency bin for a synthetic convolutive mixture of three male speakers in a room with 130 ms (top) and 250 ms reverberation time (bottom). Posterior probabilities determine the contribution of each source in every point. By increasing the reverberation time from 130 ms to 250 ms, in several points more than one sources have a high contribution.

V. ACTIVE SOURCE SELECTION AND SOURCE SEPARATION

Binary masking is a popular method for constructing separated signals in the frequency domain. Binary masking relies on the sparseness property of speech signals so that it assumes at most one source has a large contribution to each TF point. This assumption can be violated in highly reverberant environments. Fig. 2 illustrates the contribution of sources for a synthetic convolutive mixture of 3 male speakers in a room with 130 ms and 250 ms reverberation time. The contribution of each source is denoted with a certain color and shape. There are some points at which more than one source has a high contribution. By increasing the reverberation time, the situation becomes worse, as shown in Fig. 2. Here, after determining source contribution in evert TF point, we propose a procedure for active source selection and mixing matrix estimation. As mentioned before, the posterior probability $\gamma(C_n|\bar{\mathbf{x}}(\tau, f))$ determines the probability that the observation vector $\bar{\mathbf{x}}$ at time τ and frequency f belong to the n^{th} class. The collection of all points whose spatial direction vectors belongs to the class C_n form the time support $\Omega_n(f)$ of the source $s_n(t)$. Then, the frequency-dependent column vector $\mathbf{a}_n$ of the mixing matrix $\mathbf{A}$ is estimated as the centroid of this set of vectors

$$\mathbf{a}_n(f) = \frac{1}{|C_n|} \sum_{\tau \in \Omega_n(f)} \bar{\mathbf{x}}(\tau, f), \tag{18}$$

where $|C_n|$ is the number of vectors in this class and $\bar{\mathbf{x}}(\tau, f)$ is the spatial direction vector of the observed mixtures (3). Considering a TF point (τ^*, f^*), there are at most K sources active at this point where $K \leq N$. Let $\{\alpha_1, ..., \alpha_K\}_{(\tau^*,f^*)}$ be the indices of the active sources and

$$\hat{\mathbf{A}}_\alpha(\tau^*, f^*) = [\mathbf{a}_{\alpha_1}(f^*), \ldots, \mathbf{a}_{\alpha_K}(f^*)] \tag{19}$$

be a frequency-dependent matrix which contains the active column vectors of the mixing matrix $\mathbf{A}$. $\{\alpha_1, ..., \alpha_K\}_{(\tau^*,f^*)}$ is identified by

$$\{\alpha_1, ..., \alpha_K\}_{(\tau^*,f^*)} = \arg_j \left(\left(\frac{\gamma(C_j|\bar{\mathbf{x}}(\tau^*, f^*))}{\gamma_{\max}(C|\bar{\mathbf{x}}(\tau^*, f^*))} \right) > \lambda \right), \tag{20}$$

where $j \in [1, N]$. λ is a predefined threshold that we set to $\lambda = 0.55$ and $\gamma_{\max}(C|\bar{x}(\tau^*, f^*))$ indicates the maximum posterior probability at (τ^*, f^*). Given $\hat{\mathbf{A}}_\alpha(\tau^*, f^*)$ by (19), the TF-distribution values of the K active sources at (τ^*, f^*) are estimated by

$$\hat{\mathbf{s}}(\tau^*, f^*) \approx \hat{\mathbf{A}}_\alpha^\sharp(\tau^*, f^*)\, \mathbf{x}(\tau^*, f^*), \tag{21}$$

where $\mathbf{x}$ is given by (3) and $\hat{\mathbf{s}} = [s_{\alpha_1}, \ldots, s_{\alpha_K}]^T$. Superscript $(\cdot)^\sharp$ indicates the the Moore-Penroses pseudo-inversion operator. The above procedure is performed for all TF points. Finally, estimated source signals $\hat{\mathbf{s}}(\tau, f)$ are transformed to the time domain by the inverse short-time Fourier transform.

VI. EXPERIMENTAL RESULTS

The algorithm is evaluated on the test database used in the audio source separation campaign (SiSEC08) [10]. We consider live recording mixtures of three male speech signals (male3), three female speech signals (female3) and four female speech signals (female4), sampled at 16 kHz and with a 10 s duration. In order to evaluate the robustness of the proposed algorithm, two microphone spacings, 5 cm and 1 m, are considered in a room with 130 ms and 250 ms reverberation times. The algorithm uses a 2048 sample length Hann window with a 50% overlap and a random initialization of the EM algorithm. The separation performance is evaluated for each estimated source n by the signal to distortion ratio (SDR) [11]. The separation performance of the proposed approach is compared with 5 algorithms: mandel [12], chami [13], weiss [14], cobos [15] and araki [16], which participated in SiSEC08 [10]. In this work, we do not compare the result of our algorithm with the one in [8] because there are some unclear points which make it difficult to make a fair comparison.

Fig. 3 shows the comparison in terms of average SDRs of the estimated sources in different scenarios. In terms of

Table 1. Average standard deviation (SD) of the SDRs of the estimated sources for all scenarios in dB. A lower SD implies that source signals are recovered in the same range.

algorithm	proposed	chami	araki
average SD (dB)	1.74	2.23	3.57

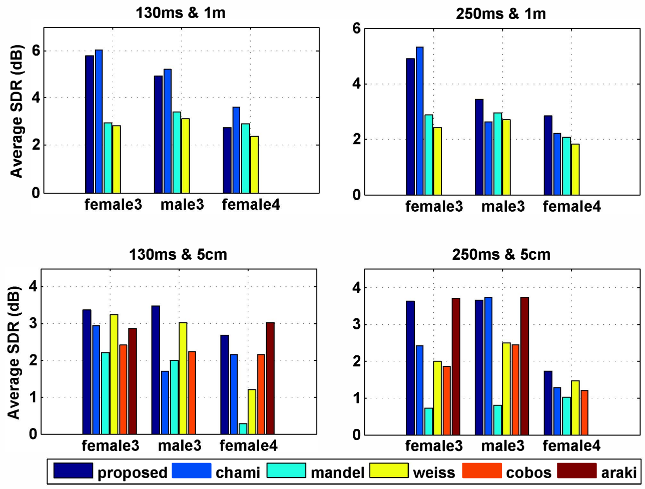

Fig. 3. Comparison between the proposed algorithm and selected algorithms that participated in the SiSEC2008 [10] in terms of average SDR (dB). Two microphone spacings 5 cm and 1 m are considered in a room with 130 ms and 250 ms reverberation times. (fe)male3 stands for mixtures of three (fe)male speech signals and female4 stands for a mixture of 4 female speech signals.

SDR, the proposed algorithm yields similar or better results as compared to the competing algorithms and is robust with respect to the different microphone spacing and reverberation times.

The standard deviation (SD) of SDRs should be also evaluated. Hence, Table 1 compares the average SD of the SDRs of estimated sources for all scenarios from our algorithm with those of [13] and [16] as the main competitors. The proposed algorithm yields the lowest SD among the competing algorithms, araki and chami, which means that the SDR of separated sources are rather similar while for the competing algorithms, some sources are recovered well and others rather poorly. The audio files of the experiments can be downloaded from [17].

VII. CONCLUSIONS

This paper proposed a model-based under-determined blind source separation algorithm that allows sources to be nondisjoint in the frequency domain, which means more than one source can be active in each time-frequency (TF) point. In contrast to the nondisjoint approaches [1], [2], in which the active sources are determined by an exhaustive search through an optimization procedure, we proposed a less computationally demanding procedure for selecting active sources based on determining the source contributions in every TF point, which avoids the exhaustive search. The source contributions in every TF point were determined by a set of posterior probabilities by which we found the indices of active sources. By knowing the indices of active sources, a frequency dependent unmixing matrix was identified to find the TF-distribution value of the estimated signals. For the permutation alignment along frequency bins, posterior probability sequences were used instead of the power of separated signals to measure the inter-frequency dependency. We showed that the proposed algorithm is robust to the different reverberation times and microphone spacing and yields a similar or better signal to distortion ratio (SDR) as compared to the competing algorithms that participated in SiSEC08 [10].

ACKNOWLEDGMENTS

The authors would like to thank Dr. E. Vincent for evaluation of our results on SiSEC test database.

REFERENCES

[1] A. Aissa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Y. Grenier, “Underdetermined blind separation of nondisjoint sources in the time-frequency domain,” *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 897 –907, 2007.

[2] D. Peng and Y. Xiang, “Underdetermined blind source separation based on relaxed sparsity condition of sources,” *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 809 –814, 2009.

[3] L. Wang, H. Ding, and F. Yin, “A region-growing permutation alignment approach in frequency-domain blind source separation of speech mixtures,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 3, pp. 549 –557, 2011.

[4] J. K. A. Hyvrinen and E. Oja, *Independent Component Analysis*. John Wiley and Sons, 2001.

[5] P. D. O’Grady and B. A. Pearlmutter, “The lost algorithm: Finding lines and separating speechmixtures,” *EURASIP Journal on Advances in Signal Processing*, vol. Article ID 784296, 2008.

[6] H. Sawada, S. Araki, and S. Makino, “A two-stage frequency-domain blind source separation method for underdetermined convolutive mixtures,” in *in Proc. WASPAA 2007*, 2007, pp. 139 –142.

[7] C. M. Bshop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[8] H. Sawada, S. Araki, and S. Makino, “Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 3, pp. 516 –527, 2011.

[9] G. Seber, *Multivariate Observations*. Wiley, 1984.

[10] http://sisec2008.wiki.irisa.fr/tiki-index.php.

[11] E. Vincent, H. Sawada, P. Bofill, S. Makino, and J. P. Rosca, “First stereo audio source separation evaluation campaign: Data, algorithms and results,” in *Proc. ICA 2007*, 2007, pp. 552 –559.

[12] M. I. Mandel and D. P. W. Ellis, “Em localization and separation using interaural level and phase cues,” in *Proc. WASPAA 2007*, 2007, pp. 275 –278.

[13] Z. E. Chami, A. Pham, and A. Serviere, Christine nd Guerin, “A new model based underdetermined source separation,” in *Proc. IWAENC 2008*, 2008.

[14] M. I. Mandel, P. W. D. Ellis, and T. Jebara, “An em algorithm for localizing multiple sound sources in reverberant environments,” in *Proc. NIPS 2006*, 2006.

[15] M. Cobos and J. J. Lpez, “Stereo audio source separation based on time-frequency masking and multilevel thresholding,” *Digital Signal Processing*, vol. 18, no. 6, pp. 960 – 976, 2008.

[16] S. Araki, T. Nakatani, H. Sawada, and S. Makino, “Blind sparse source separation for unknown number of sources using gaussian mixture model fitting with dirichlet prior,” *in proc. ICASSP 2009*, pp. 33–36, 2009.

[17] www.ee.kth.se\~taghia\ISSPIT11.zip.