New Approaches for Solving Permutation Indeterminacy and Scaling Ambiguity in Frequency Domain Separation of Convolved Mixtures

Zhitang Chen and Laiwan Chan

Abstract—Permutation indeterminacy and scaling ambiguity occur in ICA and they are particularly problematic in time-frequency domain separation of convolutive mixtures. The quality of separation is severely degraded if these two problems are not well addressed. In this paper, we propose new approaches to solve the permutation indeterminacy and scaling ambiguity in the separation of convolutive mixture in frequency domain. We first apply Short Time Fourier Transform to the observed signals in order to transform the convolutive mixing in time domain to instantaneous mixing in time-frequency domain. A fixed-point algorithm with test of saddle point is adopted to derive the separated components in each frequency bin. To solve the permutation problem, we first use discrete Haar Wavelet Transform to extract the feature vectors from the magnitude waveforms of the separated components and use Singular Value Decomposition to achieve dimension reduction. The permutation problem is solved by clustering the feature vectors using Hybrid K-means Clustering algorithm which is a combination of basic K-means and Hungarian algorithm. To solve the scaling ambiguity problem, we treat it as an overcomplete problem and realize it by maximizing the posterior of the scaling factor. Finally, experiments are conducted using benchmark data to present the effectiveness and performance of our proposed algorithms.

I. INTRODUCTION

BLIND-SOURCE-SEPARATION (BSS) using Independent Component Analysis (ICA) has received extensive attention in the recent years. The basic problem in BSS is to recover the original source signals given the mixed signals, of which the mixing procedure is unknown to us. Our objective is to extract the source signals from the mixtures using some techniques. ICA is one of the most widely used method. The basic assumption of ICA is that the sources are statistically independent. Under this assumption, we can separate the mixtures up to the permutation and scaling ambiguity.

Many algorithms have been developed using ICA to solve the BSS in the case where the observations are the instantaneous mixtures of the original sources. However, due to the environmental reflection and propagation delays through the medium, the mixed signals are not simple instantaneous mixtures but convolutive mixtures of sources.

ICA cannot be applied to the convolutive mixtures directly because the basic ICA algorithm is based on the instantaneous mixing model. There are mainly two categories of approaches proposed to solve the convolutive mixtures. One method is to separate the convolutive mixture in time domain [1][2]. This method is often referred to as blind source deconvolution. The advantage of time domain separation by blind deconvolution is that there is no permutation indeterminacy and scaling ambiguity unlike its frequency domain counterpart. However, time domain approach has the disadvantage that the algorithm is complicated and computationally expensive because it involves too many deconvolutions. The other approach solves the convolutive mixtures in frequency domain [3]-[11]. This approach transforms the convolved mixtures from time domain to time-frequency domain so that the convolutive mixing becomes instantaneous mixing in each frequency bin. The advantage of the frequency domain approach is that we can simply apply ICA to each frequency bin to derive the separated components. However, as we apply ICA to each frequency bin independently, the permutation indeterminacies and scaling ambiguities in different frequency bins are different. It remains a difficult task to solve the permutation indeterminacy and scaling ambiguity. The permutation problem is much more severe because if it fails to be addressed the final results still remain mixtures.

The rest paper is organized as follows: In section II, we briefly introduce some related work for solving the permutation indeterminacy and scaling ambiguity in the most cited literatures in the past decades. In section III, we describe the blind source separation problem and we will focus on the convolutive mixture problem. We also introduce an advanced FastICA algorithm by Zbynek Koldovsky et al. [12]. In section IV, we present a new approach for solving permutation indeterminacy using Hybrid K-means Clustering (HKC) algorithm. In section V, we show that the scaling ambiguity can be treated as an overcomplete problem and we solve the problem by Maximizing A Posterior (MAP). In section VI, some experiments are carried out with benchmark data and the results show the effectiveness of our proposed methods. In section VII, some conclusions are drawn on our algorithms.

II. RELATED WORK

In the past decades, many algorithms have been proposed to solve the permutation indeterminacy. Smoothing the estimated separating matrices is one solution. To enforce training converging to the same permutation, Smaragdis [3] proposed a method called influence factor, of which the basic idea is to couple the separating matrices. This is achieved by adding a weighted separating matrix of the previous frequency bin.
to that of the current frequency bin. This method is heuristic and it works well only in some cases [3].

Employing the correlations of the output signal envelopes is another possible solution [4][5]. Dapena et al.[4] proposed to solve the permutation indeterminacy by computing the cross-correlation between each separated components in the first frequency bin and those in the following bins and then put them into different groups according to the cross-correlation. The algorithm by Dapena et al. does not necessarily solve the permutation indeterminacy. This is due to the fact that the spectrogram of a signal varies gradually along the frequency axis. The waveforms in the later frequency bins may not have the maximal correlation with that in the first frequency bin even though they are from the same source signal. Ikeda et al. [5] also proposed a correlation based method to solve the permutation indeterminacy utilizing the non-stationary of the source signals. They first extracted the frequency axis. The waveforms in the later frequency bins may not have the maximal correlation with that in the first frequency bin even though they are from the same source signal

Several papers by Sawada et al.[6][7] and Kim et al.[8] proposed approaches to solve the permutation problem by clustering the basic vectors or estimated directions/distance. The approaches proposed by them are quite similar to our approach (to be proposed in this paper) as we all resort to clustering. However, the main difference lies in that they solved the problem by clustering either basic vectors or directions (both of basic vectors and directions are estimated from the separating matrix) whereas our approach solves the permutation by clustering the feature vectors extracted from the separated components. Our approach is more precise and robust as it is not sensitive to the relative locations of the sensors and sources. Moreover, our approach does not require any prior knowledge about the locations and distances of sources and sensors.

Recently Kim et al.[9] and Lee et al.[10] employed Independent Vector Analysis (IVA) to avoid permutation indeterminacy. The main idea of the approaches is to estimate the separating matrices in all the frequency bins in batch so that the permutation indeterminacy is avoided. The coupling of the separating matrices is realized by constructing multivariate score functions. This kind of approaches has the advantage of avoiding the permutation indeterminacy. However, the main difficulty still lies in how to estimate the multivariate score function precisely. If the multivariate score functions cannot model the dependence between the frequencies accurately, it is possible that the estimated separating matrices fail to achieve the optimal directions due to the coupling effect introduced by this kind of approaches.

Other method such as APDP was proposed by Ciaramella et al. [11] to solve the permutation problem. The authors reduced the permutation problem to an Assignment Problem (AP) and used Hungarian algorithm to solve it. In order to obtain global matching, they coupled the Hungarian algorithm with the Dynamic Programming (DP) algorithm.

Compared to permutation indeterminacy, there is little work contributing to the scaling ambiguity. In [4], Dapena forced all the separated components in the same group to the same amplitude. However, this is not the case as many real word signals especially the speech and audio signals are colored. Ikeda et al. [5] proposed an algorithm called split spectrograms to solve the scaling ambiguity. Their method is useful in eliminating the scaling ambiguity but it leads to another problem, which is by this method the separated component in each frequency bin is weighted by a complex constant. The constant is the frequency response of the filters. Ciaramella et al. [11] used the same but slightly modified method as Ikeda et al. to solve the scaling ambiguity.

III. CONVOLUTIVE BLIND SOURCE SEPARATION IN TIME-FREQUENCY DOMAIN

In this section, we will describe the convolutive mixture signal model and discuss our main approach to deal with this model.

Let $s(t)$ be the vector denoting the source signals

$$s(t) = [s_1(t), s_2(t), \cdots, s_N(t)]^T \quad t = 0, 1, 2, \cdots \quad (1)$$

The source signals and their exact probability density function are unknown to us. We assume that these source signals are statistically independent to each other. Mathematically, independence of the source signals means that given any time instants, the joint probability density function can be factorized by the marginal probability density functions.

$$p(s_1(t_1), s_2(t_2), \cdots, s_N(t_N)) = \prod_{i=1}^{N} p(s_i(t_i)), \forall t_i \quad (2)$$

The source signals propagate through the medium and they are picked up by $M$ sensors. The observed signals by the sensors can denoted by the vector

$$x(t) = [x_1(t), x_2(t), \cdots, x_M(t)]^T \quad (3)$$

For simplicity, here we only consider the case where $M = N$ and we do not consider the noise effect.

Basic blind source separation model assumes that the observed signals are a linear instantaneous mixture of the source signals:

$$x(t) = As(t) \quad (4)$$

Here, the elements of matrix $A$ are real scalar value. This model is extensively studied in many ICA literatures. However this instantaneous model cannot fully describe the mixing procedure in real world. Due to the delays and reflection of the real environment [5], the mixed signals are
not instantaneous mixtures but convolutive mixtures, which can be described by the following model:

$$x(t) = \sum_{\tau=0}^{\infty} A(\tau)s(t - \tau) \quad (5)$$

Applying L points Short Time Fourier Transform to both sides of equation (5), we obtain:

$$X(\omega, t_s) = A(\omega)S(\omega, t_s) \quad \omega = 0, 1/L, 2\pi/L, \ldots; \quad t_s = 0, \Delta T, 2\Delta T, \ldots \quad (6)$$

$$\Delta T$$ is the step size of the moving window.

Equation (6) shows that for a fixed frequency \( \omega \), \( X(\omega, t_s) \) is the instantaneous mixture of \( S(\omega, t_s) \). We use subscripts to denote the frequency. Hence equation (6) can be rewritten as follows:

$$X_\omega(t_s) = A_\omega S_\omega(t_s) \quad (7)$$

Equation (7) shows that by transforming the signals to time-frequency domain, the original convolutive mixing problem reduces to an instantaneous mixing problem in each frequency bin. We can apply ICA algorithm to derive the independent components in each frequency bin. However, special attention should be paid to the fact that in this case both the mixing matrices and the components are complex-valued so we should use complex-valued ICA to solve the problem. In this paper, we will adopt the algorithm developed by Zbynek Koldovskhy et al. This algorithm involves saddle point testing to enhance the chance that the algorithm converges to the global maxima [12]. After separating the components in all frequency bins, we can transform the signals in time-frequency domain back to time domain and we get the separated signals.

IV. SOLVING THE PERMUTATION INDETERMINACY

We apply complex-valued ICA to each frequency bin and obtain the separated components independently. The permutations in each frequency bin are not exactly in the same order. The permutation needs to be well addressed otherwise the final results will still be mixtures containing frequency components from different sources. In this section, we propose a new approach to solve this problem. The basic idea behind the proposed approach is as follows: we exploit the nonstationary of the signals as [5]. The nonstationary property of the signal in time domain reveals that the frequency components in the close bins are correlated. Based on this fact, it is natural to assume that the magnitude waveforms of components in the frequency bins which are close to each other are similar. Hence the permutation problem can be solved by finding a matching so that those components belonging to the same source signal are correctly clustered into one group. This is exactly a time series matching problem. It is unrealistic and computationally expensive to use the whole time series to do the matching due to the high dimension. In this paper, we propose to use Wavelet Transform to extract the features from the waveforms of the components for all frequency bins. Singular Value Decomposition is then adopted to further reduce the dimension and increase the separability of the feature vectors. A Hybrid K-means Clustering (HKC) algorithm is developed and applied to cluster the feature vectors so that those vectors from the same source signals are correctly grouped.

A. Feature Extraction by Haar Wavelet Transform

Wavelet transform is a powerful tool in signal processing dating back to work of Karl Weierstrass in 1873. It performs the inner products of the signal and a family of wavelets, decomposing the signal to different scales or resolutions. It has been widely used in area of feature extraction. Kin-pong Chan et al. [13] first proposed to use Haar Wavelet Transform for time series matching. In their work, they showed that Euclidean distance is preserved in Haar transformed domain and no false dismissal will occur. It is also shown in their work that Haar transform outperforms Discrete Fourier Transform in that Wavelet Transform encodes a coarse resolution of the original time sequence with its preceding coefficients [13]. In this paper, we will use Haar Wavelet Transform to extract the features of the waveforms.

Let \( \psi(t) \) be the mother wavelet, the family of the wavelets generating from the mother wavelet is shown as follows:

$$\psi_{\alpha, \tau}(t) = \frac{1}{\sqrt{\alpha}} \psi\left(\frac{t - \tau}{\alpha}\right) \quad (8)$$

The Continuous Wavelet Transform of a time sequence \( x(t) \) is defined as follows:

$$X(\alpha, \tau) = \frac{1}{\sqrt{\alpha}} \int x(t) \psi^*(\frac{t - \tau}{\alpha}) dt \quad (9)$$

Here, the asterisk denotes complex conjugate. The Discrete Wavelet Transform (DWT) can be realized by sampling the wavelet on a dyadic lattice [14].

$$\psi_{n,m}(k) = 2^{-\frac{n}{2}} \psi(2^{-\frac{n}{2}} k - mb_0) \quad (10)$$

The \( p \) coefficients of the DWT form the feature vector representing the pattern of the envelop of the separated component. Let \( v_{\omega_j,i} \) be the feature vector with size \( 1 \times p \) extracted from the \( i \)th component in the frequency bin \( \omega_j \).

Put all together, we get the feature-object matrix \( V_{\omega_j} \) in the frequency bin \( \omega_j \).

$$V_{\omega_j} = \begin{bmatrix} v_{\omega_j,1} & \vdots & v_{\omega_j,N} \end{bmatrix} \quad \forall \omega_j \quad (11)$$

B. Dimension Reduction by Singular Value Decomposition

Dimension reduction can be beneficial not only for reasons of computational efficiency but also because it can improve the accuracy of the analysis [15]. In most of the cases, the time cost of training or learning algorithm increase dramatically as the dimension of the feature vector increases. This is often referred to as dimensionality curse problem. What is more, some of the extracted features of the feature vector maybe noisy and irrelevant [15], which may reduce
the accuracy of analysis. It also has been reported that things look more similar on average the more features used to describe them [15]. This is probably because in high dimension feature space, the difference between the concept of the most similar and the least similar is not significant. Considering the negative effect of having large dimensions, we apply Singular Value Decomposition to the feature-object to reduce the dimension.

Let $X_{p \times n}$ be the feature-object matrix, where $p$ is the dimension of the features and $n$ is the number of objects. By Singular Value Decomposition, $X_{p \times n}$ can be decomposed as follows:

$$X = UDV^T$$

where $U_{p \times m}$ is the matrix of eigenvectors of $XX^T$, $m$ is the rank of $XX^T$; $D_{m \times m}$ is the diagonal matrix of square root of the eigenvalues of $XX^T$; $V_{n \times m}$ is the matrix of eigenvectors of $X^TX$.

Dimension reduction is achieved by dropping all but $k$ largest eigenvalues in $D$. Now $D$ is a $k \times k$ matrix, $V_{n \times k}$ is from matrix $V$ dropping the corresponding eigenvectors.

$$\hat{X}^T = \hat{V}D$$

$\hat{X}^T$ is a $n \times k$ matrix that contains the coordinates of the $n$ objects in the new $k$-dimension feature space. More details are available in [15].

C. Hybrid K-means Clustering for Solving Permutation

We propose a new approach by K-means clustering to solve the permutation problem. The basic idea is that components in the entire frequency bins are now represented by feature vectors extracted by Discrete Wavelet Transform. It is natural to assume that all the feature vectors form different clusters corresponding to different source signals. This assumption inspires us to apply clustering algorithm to find out the clusters and the corresponding membership so that the permutation problem is addressed. In this paper, we adopt a simple but very effective algorithm (K-means clustering) to do the clustering.

However there are still two major problems we need to tackle:

1) Even though we assume the feature vectors from the same source signal should form a cluster, for real world signals (especially the audio and speech signals), the spectrogram changes gradually along the frequency axis. The waveform in the later frequency bins may show significant difference with those in the first bins.

2) In the basic K-means algorithm, each point is assigned to the centroid which is closest to it. There is no further constrain for the assignment. However in our problem we should avoid assigning points in the same frequency bin to the same cluster. In other words, we should do a matching between the points in the same bin and the cluster centroids.

To deal with the first problem, we first divide the whole spectrogram into several segments. We apply K-means algorithm to each segment to realize intra-segment clustering. After the intra-segment clustering, we apply the algorithm to the centroids of the segments and realize the inter-segment clustering.

To deal with the second problem, inspired by Ciaramella's work [11], we introduce the Hungarian algorithm to the K-means algorithm to realize the matching. Unlike basic K-means clustering which treats each point independently, we process the points in the same frequency bin all at one. We do a matching between points and centroids instead of simply assigning the points to the nearest centroid.

Let $\delta(A, C)$ be the sum of distances between the elements in $A$ and their corresponding mappings in $C$ due to a one-to-one mapping $f: A \rightarrow C$. Here $A$ is a set containing points in the same frequency bin and $C$ is the centroids set corresponding to $N$ source signals. Our objective is to find an optimal matching so that $\delta(A, C)$ is minimized.

$$f_{opt} = \min_f \delta(A, C)$$

Let $x_{ij}$ be a dummy variable which takes the value 1 if there is a mapping from element $a_i$ of set $A$ to element $c_j$ of set $C$ and 0 otherwise. $d_{ij}$ is the distance between element $a_i$ and element $c_j$.

$$\delta(A, C) = \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} x_{ij}$$

The model, see [11], for solving the matching problem is given as follows:

$$\begin{align*}
\min & \quad \delta(A, C) \\
\text{subject to} & \quad \sum_{i=1}^{N} x_{ij} = 1 \quad i = 1, 2, \ldots, N \\
& \quad \sum_{j=1}^{N} x_{ij} = 1 \quad j = 1, 2, \ldots, N \\
& \quad x_{ij} = 0/1 \quad i, j = 1, 2, \ldots, N
\end{align*}$$

This model can be easily solved by Hungarian algorithm [6]. We combine K-means clustering and Hungarian algorithm to solve the permutation problem. We call this combined algorithm Hybrid K-means Clustering (HKC) algorithm. The pseudo-code of this two phase HKC algorithm is summarized as follows:

Phase 1
- Divide the spectrogram into $k$ segments
- FOR each segment $S_i$
  - Let $V_{S_i} = [V^{1}_{\omega_{r}}, \ldots, V^{m}_{\omega_{s}}]^T$ \quad $\omega_{r}, \ldots, s \in S_i$
  - Reduce the dimension of $V_{S_i}$ by SVD and get $\hat{V}_{S_i}$
  - WHILE Iter < Maxiter
    - FOR frequency bin $\omega_j \in S_i$
      - Find matching $f: \hat{V}_{\omega_j} \rightarrow C_{S_i}$ \quad ($C_{S_i}$ is the centroid set of segment $S_i$)
      - END
  - Update $C_{S_i}$ by the means of the points in each cluster
  - END
- END

Phase 2
- Use the same algorithm to cluster the centroids of all the segments into $N$ clusters
V. SOLVING THE SCALING AMBIGUITY

We know that ICA can separate the mixture up to the permutation indeterminacy and scaling ambiguity. In the previous section, we propose an approach to solve the permutation. In this section, we focus our attention on solving the scaling ambiguity. As a matter of fact, the scaling ambiguity is caused by forcing the separated components to uniform energy. Unlike the time domain separation, in time-frequency domain separation scaling ambiguity needs to be well addressed otherwise there will be a magnitude distortion of the spectrogram.

Below we will present our approach to solve the scaling ambiguity. Let $W_f$ be the real separating matrix while $\hat{W}_f$ be the estimated separating matrix obtained by any ICA algorithm and $A_f$ be the diagonal matrix corresponding to the scaling ambiguity in frequency bin $f$. we have:

$$\hat{W}_f = A_f W_f$$ \hspace{1cm} (17)

For illustration, we have $\hat{w}_{f,ij} = \lambda_{f,ii} w_{f,ij}$. We make the following assumption:

**Assumption 1:** All the diagonal elements of $A_f$ are real positive.

Based on the assumption, we have $|\hat{w}_{f,ij}| = |\lambda_{f,ii} w_{f,ij}|$. Take log of both sides, we have:

$$log(|\hat{w}_{f,ij}|) = log(\lambda_{f,ii}) + log(|w_{f,ij}|)$$

To solve the scaling ambiguity, let us drop the subscript $f$ and rewrite all the variable with subscript $f$ as a function of $f$. Let

$$X = \begin{bmatrix} log(|\hat{w}_{11}(f_0)|) & \cdots & log(|\hat{w}_{11}(f_{L-1})|) \\ \vdots & \cdots & \vdots \\ log(|\hat{w}_{N1}(f_0)|) & \cdots & log(|\hat{w}_{N1}(f_{L-1})|) \end{bmatrix}$$

$$S = \begin{bmatrix} log(\lambda_{i1}(f_0)) & \cdots & log(\lambda_{i1}(f_{L-1})) \\ \vdots & \cdots & \vdots \\ log(w_{i1}(f_0)) & \cdots & log(w_{i1}(f_{L-1})) \\ \vdots & \cdots & \vdots \\ log(w_{iN}(f_0)) & \cdots & log(w_{iN}(f_{L-1})) \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 1 \end{bmatrix}_{N \times (N+1)}$$

It is easy to find that

$$X = \text{AS}$$ \hspace{1cm} (18)

Since $A$ is not invertible, there is no unique solution for $S$. Actually, this is a overcomplete problem. For more details about overcomplete problem, see [16]. The internal states $S$ can be solved by maximizing the posterior distribution of $S$.

$$\hat{S} = \arg \max_S p(S|X, A) = \arg \max_S p(X|S, A)p(S)$$ \hspace{1cm} (19)

In the case that no noise are considered, $p(S)$ are Gaussian and the components in $S$ are statistically independent, maximizing equation (19) is equivalent to:

$$\min_S \| S \|_2 \text{ subject to } X = \text{AS}$$ \hspace{1cm} (20)

The solution to equation (20) can be obtained by pseudo-inverse [16].

$$S = A^\dagger X$$ \hspace{1cm} (21)

where $A^\dagger = (A^T A)^+ A^T$ and $(A^T A)^+$ is the Moore-Penrose inverse of $A^T A$.

However, assuming the internal states have Gaussian prior is a simple but by no means the best way to solve the problem. It would be better if we can estimate the probability density from the data. Mixture of Gaussian (MOG) is a widely used method to estimate the probability density function from the data. It is reported that MOG can approximate any probability density given enough number of functions. The basic idea of MOG is to approximate a probability density function by the sum of several Gaussian distribution function with different weights, means and variances.

$$p(s) = \Sigma_k \alpha_k e^{-\frac{(s-\mu_k)^2}{\sigma_k^2}}$$ \hspace{1cm} (22)

The parameter set $\{\alpha_k, \mu_k, \sigma_k^2\}$ can be learnt from data using Expectation Maximization (EM) algorithm. Back to our problem, we would like to use MOG to approximate the probability densities of the internal states. To make the problem tractable, we make the following assumption.

**Assumption 2:** The components of the internal state $S$ are statistically independent and identically distributed (i.i.d.) at any frequency bin, that is $p(S) = \Pi_f \Pi_m p(s_m(f))$

Due to the assumption of i.i.d., we can solve the problem at each frequency bin independently. Suppose we use $k$ Gaussian functions to approximate the probability density of each internal state $s_r$. Now the problem reduces to the following optimization problem:

$$\max_S \Pi_r (\Sigma_k \alpha_{rk} e^{-\frac{(s_r-\mu_{rk})^2}{2\sigma_{rk}^2}}) \text{ subject to } X = \text{AS}$$ \hspace{1cm} (23)

Take log of the contrast function, we have:

$$L = \sum_r \log (\Sigma_k \alpha_{rk} e^{-\frac{(s_r-\mu_{rk})^2}{2\sigma_{rk}^2}})$$

Substitute $s_r = x_{r-1} - s_1 (r > 1)$ into the log contrast function, we have:

$$L = \log (\Sigma_k \alpha_{1k} e^{-\frac{(s_1-\mu_{1k})^2}{2\sigma_{1k}^2}}) + \sum_{r>1} \log (\Sigma_k \alpha_{rk} e^{-\frac{(s_r-\mu_{rk})^2}{2\sigma_{rk}^2}})$$

Now, the contrast function becomes a function of $s_1$. In order to maximize the contrast function, we take derivation of $L$.
the sound quality of the recovered signals are much better. Our algorithm preserves more details of the higher frequency components so that compared to the results by Ciaramella, two results are quite different. As Ciaramella et al. used more or less the same approach to solve the scaling ambiguity, they suffered the same problem that the magnitudes of the medium and higher frequency components are severely attenuated. What is more, our algorithms costs much less time than that by Ciaramella et al.

The second dataset consists of two convolutive mixtures of two speeches with sampling rate of 16 kHz. Both speeches are digits from “one” to “ten” but in two different language. The figures are organized as the same order as the first dataset. Note that there is a serious problem in Ikeda’s results. We can find that actually the frequency components from around 3700 Hz to 8000 Hz are completely mis-placed. This is probably because two signals are digits from “one” to “ten” so that the spectrograms of two signals are very similar to each other. Ikeda’s algorithm fails to distinguish them. Both Ciaramella’s results and ours are free from such problem. But as the first dataset, the scaling ambiguity in Ciaramella’s results is not well addressed. Experimental results using benchmark datasets are showed as follows:

VI. Experiment Results

In this section, we show the results obtained by our approach and compare the results with those obtained by Ikeda et al. and Ciaramella et al. The benchmark data are available at [17].

The first set of data consists of two convolutive mixtures of two signals with sampling rate of 16 kHz. One signal is music and the other one is a speech signal of digits “one” to “ten” in English. Fig.1 shows the scaling factors we estimate from the separating matrix. In order to show the details of the signal spectrograms in the higher frequency bins, we take log of the magnitude. Fig.2-4 show the spectrograms of the separated signals by three algorithms. Fig.5-6 shows the convolutive mixtures and the separated signals in time domain by three algorithms. Fig.2 shows the results by Ikeda’s algorithms. Note that there are several frequency bins (around 3500Hz, 5000Hz, 5500Hz) where the permutation problem is not exactly solved. What is more, the magnitudes of the medium and higher frequency components are severely weakened due to their approach to solve the scaling ambiguity. Fig.3 shows the results by Ciaramella’s algorithms. Fig.4 shows results by our algorithms. Note that both results by Ciaramella’s and our algorithm are free from permutation problem. However, the magnitudes of the spectrograms of two results are quite different. As Ciaramella et al. used more or less the same approach to solve the scaling ambiguity, they suffered the same problem that the magnitudes of the medium and higher frequency components are severely attenuated. Our algorithm preserves more details of the higher frequency components so that compared to the results by Ciaramella, the sound quality of the recovered signals are much better.

VII. Conclusion

In this paper, we propose new approaches to solve the permutation indeterminacy and scaling ambiguity in the separation of convolutive mixture in frequency domain. The permutation indeterminacy is solved mainly by clustering the feature vectors extracted from the separated components in each frequency bin using a Hybrid K-means Clustering algorithm. By combining matching and clustering, HKC shows strong power as it makes full use of local and global
Fig. 3. Spectrogram of separated signals by Ciaramella et al.

Fig. 4. Spectrogram of separated signals by Chen et al.

Fig. 5. Convolutive mixture and separated signal of channel 1

Fig. 6. Convolutive mixture and separated signal of channel 2

Fig. 7. Estimated scaling factors

Fig. 8. Spectrogram of separated signals by Ikeda et al.

Fig. 9. Spectrogram of separated signals by Ciaramella et al.

Fig. 10. Spectrogram of separated signals by Chen et al.
information to do the clustering. In this paper, we also show that scaling ambiguity can be treated as an overcomplete problem and we solve it by maximizing a posterior. Experimental results show that our algorithm outperforms Ikeda’s algorithm and performs even not better but at least as well as Ciaramella’s algorithm in solving permutation indeterminacy. For the scaling ambiguity, our results show much better quality compared with the results by Ikeda et al. and Ciaramella et al. as we preserve the energy of the signals in higher frequency bins which in turn preserve more details of the signals.

Acknowledgment The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administration Region, China. The first author would like to thank Prof. Angelo Ciaramella for providing the code and helpful discussion. The first author also would like to thank Dr. Zhang Kun for helpful discussion and comments.

REFERENCES