HAL open science

Blind Source Separation Using Mixtures of Alpha-Stable Distributions

Nicolas Keriven, Antoine Deleforge, Antoine Liutkus

To cite this version:

Nicolas Keriven, Antoine Deleforge, Antoine Liutkus. Blind Source Separation Using Mixtures of Alpha-Stable Distributions. 2017. hal-01633215v1

HAL Id: hal-01633215

https://inria.hal.science/hal-01633215v1

Preprint submitted on 11 Nov 2017 (v1), last revised 9 Feb 2018 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BLIND SOURCE SEPARATION USING MIXTURES OF ALPHA-STABLE DISTRIBUTIONS

Nicolas Keriven, Antoine Deleforge* and Antoine Liutkus*†
*Inria Rennes - Bretagne Atlantique, France
†Inria and LIRMM, Montpellier, France

ABSTRACT

We propose a new blind source separation algorithm based on mixtures of alpha-stable distributions. Complex symmetric alpha-stable distributions have been recently showed to better model audio signals in the time-frequency domain than classical Gaussian distributions thanks to their larger dynamic range. However, inference of these models is notoriously hard to perform because their probability density functions do not have a closed-form expression in general. Here, we introduce a novel method for estimating mixture of alpha-stable distributions based on random moment matching. We apply this to the blind estimation of binary masks in individual frequency bands from multichannel convolutive audio mixes. We show that the proposed method yields better separation performance than Gaussian-based binary-masking methods.

***Index Terms*—** Blind Source Separation, Binary Masking, Alpha-Stable, Generalized Method of Moments

1. INTRODUCTION

This paper is concerned with source separation, which is a topic in applied mathematics that aims at processing *mixture* signals so as to recover their constitutive components, called *sources* [1]. It is a field of important and widespread practical applications, notably in biological signal processing [2] and in audio. In audio, it is traditionally exemplified by the *cocktail party problem*, which consists in isolating some specific discussion within the recording of a crowd [3, 4]. Apart from such speech processing scenarios, source separation also enjoyed much interest in the music processing literature, due to its important applications in the entertainment industry [5]

From the perspective of this paper, it is worth mentioning that a significant portion of the research on source separation first makes some *assumptions* on the sources signals and then picks some *mixing model*. While the former usually stands on probabilistic grounds, the latter often comes from physical assumptions and explains how the observed mixtures are generated from the sources.

Historically, the *overdetermined* linear case was considered, for which more mixtures than sources are available [1]. The interesting fact about such mixing models is they can be inverted easily, allowing to recover the sources from the mixtures, provided their parameters are known. The breakthrough brought in by source separation is to allow identification of such mixing parameters with only very general assumptions about the sources. These assumptions are mostly either that sources are both independent, identically distributed (i.i.d.) and non-Gaussian, as in Independent Components Analysis (ICA, [6]), or that they are Gaussian but not i.i.d. as in Second-Order Blind Identification (SOBI [7]). Going in the frequency domain allowed to extend such approaches to *convolutive* mixtures, i.e. for which the sensors capture the sources after some acoustic propagation whose duration is not negligible.

The validity of the mixing model and its invertibility is crucial for applying separation methods that make only broad assumptions on the sources. When such assumptions are violated, those approaches are not applicable. This typically happens in the underdetermined scenario, where fewer mixtures than sources are available, which is classical in audio. In that case, separation may only be achieved through more involved source models and *time-varying* filtering procedures [5]. For this reason, it is natural that research in underdetermined separation focused on highly parameterized and tractable source models. In short, a huge part of the models proposed in the literature stands on Gaussian grounds, where one wants to estimate time-varying power-spectral densities and steering vectors for building the corresponding multichannel Wiener filters [8, 9]In that framework, estimation is typically achieved through maximization of likelihoods, including the celebrated Expectation-Maximization algorithm [10]. This line of thought leaves room for much flexibility and a large community strived to provide effective audio spectrogram models, from sophisticated linear factorization [11] to recent developments in deep learning [12, 13].

An intrinsic weakness of the Gaussian processes for modeling audio sources is to require many parameters to faithfully represent sophisticated signals. This is made unavoidable by their light-tails, which only allow for small explorations around averages and standard deviations. One typically has to pick a different Gaussian distribution for *each* time and frequency to obtain a good model [9], and precise estimation of *all* parameters is required for good performance. This inevitably makes all related estimation methods very sensitive to initialization. Using distributions with heavier-tails than the Gaussian for underdetermined separation is yet at its infancy, although it is common practice in the overdetermined case [14, 15]. Among such distributions, the α-stable distribution [16] enjoyed some interest in signal processing [17]and more particularly in source separation recently, because it was shown to straightforwardly yield effective filters with better perceived audio quality than the standard Wiener [18, 19].

However, the delicate question of how to estimate the parameters of α-stable source models remains quite an open issue. It appears to be very challenging because such distributions do not provide an analytical expression for their likelihood, which prevents standard methodologies. Two alternative options were considered so far. First, Markov Chain Monte Carlo methods [20] are applicable and effective at yet a high computational cost. Second, classical moment-matching methods were proposed [21] that are effective, but somehow ad-hoc and hard to translate into the *multichannel* case of several mixtures.

In this paper, we use a variant of the recent algorithm introduced in [22] for the estimation of mixture models by generalized moment

matching, to exploit mixtures of *multivariate* α-stable distributions in the context of audio source separation. This algorithm, referred to as Compressive Learning-Orthogonal Matching Pursuit with Replacement (CL-OMPR), is a greedy, heuristic method that was initially used in the context of *sketching* [22], to estimate mixture models on large-scale databases using only a collection of generalized moments computed in one pass. Sketching thus enjoyed several successful applications in machine learning [23], but also for source localization [24]. Here, we exploit instead the capacity of CL-OMPR to successfully estimate exotic mixture models from a collection of generalized moments, and show for the first time how it may be useful to devise time-varying filters able to achieve good performance for source separation.

2. ALPHA-STABLE UNMIXING

2.1. Alpha-stable mixture model

Let us consider a mixture of K sound sources observed through M channels. We denote by $\{s_k(f,t)\}_{k=1}^K$ the emitted source spectrograms and by $\{x_m(f,t)\}_{m=1}^M$ the observed channel spectrograms in the complex short-time Fourier domain, where $f \in [1 \dots F]$ and $t \in [1 \dots T]$ denote the discrete frequency and time indexes. Assuming time-domain convolutive filters from sources to channels which are short compared to the Fourier windows, the mixing model at (f,t) can be written

$$\boldsymbol{x}(f,t) = \sum_{k=1}^{K} \boldsymbol{a}_k(f) s_k(f,t) \tag{1}$$

where $\boldsymbol{x}(f,t) \in \mathbb{C}^M$ is the observed vector, $\boldsymbol{s}(f,t) \in \mathbb{C}^K$ the source vector and $\boldsymbol{a}_k(f) \in \mathbb{C}^M$ source k's steering vector.

Now, we choose an original probabilistic model for the source signals, inspired by recent research on α-harmonizable processes [18, 24]. All $\{s_k(f,t)\}_{t=1}^T$ are assumed independent and identically distributed (i.i.d.) with respect to (wrt.) a symmetric complex and centered α-stable distribution of unit scale parameter and characteristic exponent $\alpha_k(f)$, which we write:

$$p(s_k(f,t); \alpha_k(f)) = \mathcal{S}_c(s_k(f,t); \alpha_{kf}). \tag{2}$$

In short, the symmetric centered α-stable distribution generalizes the Gaussian isotropic [25], while providing significantly heavier tails as its characteristic exponent $\alpha_{kf} \in]0,2]$ gets small, $\alpha_{kf} = 2$ corresponding to the Gaussian case. An interesting feature of our model is to be time-invariant, contrasting with classical Gaussian models [8, 9]. This is because the $\mathcal{S}_c$ distribution enables important dynamics for $s_k(f,t)$. In short, (2) corresponds to a model for the *marginal* distribution of the sources. Such ideas have already been considered in [24]. The particularity of our approach in this regard is to feature a *frequency-dependent* characteristic exponent α_{kf} for increased expressive power. The choice of a unit scale for the distribution comes with no loss of generality: any frequency-dependent scaling is incorporated in the steering vectors $\boldsymbol{a}_k(f)$.

We highlight that the probability density function (pdf) of $s_k(f,t)$ in (2) does not have a closed-form expression except for $\alpha_{kf} = 1$ (Cauchy) and $\alpha_{kf} = 2$ (Gaussian). However, its *characteristic function*, defined as the Fourier transform of its pdf does. We have [16, 18]:

$$\forall \omega \in \mathbb{C}, \mathbb{E}\{\exp(i \mathrm{Re}\,[\omega^\star s_k(f,t)])\} = \exp(-|\omega|^{\alpha_{kf}}). \tag{3}$$

At this point, we make one important simplifying assumption: we suppose *only one source* is significantly active at each time-frequency (TF) point. More specifically, let $w(f,t)$ be the index of the source that has the strongest magnitude $|s_k(f,t)|$ at TF bin (f,t). Our assumption is: all other sources have a magnitude *close to* 0. This is less strong than the so called W-disjoint orthogonality assumption [4] where a single source is assumed to be active. This allows us to assume weak sources are approximately distributed wrt a Gaussian distribution. Indeed, even if it lacks an analytical expression, the pdf for a symmetric α-stable distribution is infinitely derivable close to the origin [16], justifying this second order approximation for weak sources.

As a result of these assumptions, we take our mixture as:

$$\boldsymbol{x}(f,t) = \sum_{k=1}^{K} \mathbb{I}(z(f,t)=k)\{\boldsymbol{a}_k(f)s_k(f,t) + \boldsymbol{e}_k(f,t)\}, \tag{4}$$

where $\mathbb{I}$ is the indicator function and $\boldsymbol{e}_k(f,t) \in \mathbb{C}^M$ is a residual Gaussian term containing all non-dominating signals (other than k) and possible additional noise. For convenience, we neglect the inter-channel correlations coming from weak sources, to simply assume that e_k is composed of iid entries with variance σ_{kf}:

$$p(\boldsymbol{e}_k(f,t)|z(f,t)=k; \sigma_{kf}^2) = \mathcal{N}_c(\boldsymbol{e}_k(f,t); \boldsymbol{0}, \sigma_{kf}^2 \mathbf{I}_M) \tag{5}$$

where $\mathcal{N}_c$ denotes the multivariate complex circular-symmetric Gaussian distribution [25], $\mathbf{I}_M$ is the $M-$dimensional identity matrix and σ_{kf}^2 is the residual variance at frequency f when source k dominates. Furthermore, the indexes $w(f,t)$ of the strongest source for each TF bin are modelled as iid multinomial variables:

$$p(z(f,t)=k; \boldsymbol{\pi}_f) = \pi_{kf} \tag{6}$$

where π_{kf} is the probability of source k dominating at frequency f, and $\sum_k \pi_{kf} = 1$.

From all the preceding assumptions and dropping the indexes (f,t) for convenience, we deduce the characteristic functions of $\boldsymbol{a}_k s_k$, $\boldsymbol{e}_k$ and $\boldsymbol{x}|z=k$, where $\boldsymbol{\omega} \in \mathbb{C}^M$:

$$\psi_{\boldsymbol{a}_k s_k}(\boldsymbol{\omega}) = \exp(-|\boldsymbol{a}_k^\star \boldsymbol{\omega}|^{\alpha_k}) \tag{7}$$

$$\psi_{\boldsymbol{e}_k}(\boldsymbol{\omega}) = \exp(-\sigma_k^2 \|\boldsymbol{\omega}\|_2^2) \tag{8}$$

$$\psi_{\boldsymbol{x}|z=k}(\boldsymbol{\omega}) = \exp(-|\boldsymbol{a}_k^\star \boldsymbol{\omega}|^{\alpha_k} - \sigma_k^2 \|\boldsymbol{\omega}\|_2^2) \tag{9}$$

Combining (6) and (9), we deduce that $\{\boldsymbol{x}(f,t)\}_t$ follows a mixture model parameterized by

$$\boldsymbol{\theta}_f = \{\alpha_{kf}, \sigma_{kf}^2, \boldsymbol{a}_k(f), \pi_{kf}\}_{k=1}^K. \tag{10}$$

Following the two-stage approach of [26], the proposed blind source separation method consists in clustering observations $\boldsymbol{x}(f,t)$ independently at each frequency according to this mixture model. The resulting classical source permutation ambiguity across frequencies is left aside here (section 2.4), and a binary mask is then obtained for each source. The special Gaussian case $\alpha_{fk} = 2$ is discussed in section 2.2 while a parameter estimation method for the general case is given in section 2.3.

2.2. Special case $\alpha_{fk} = 2$

Let us consider the special Gaussian case where $\alpha_{fk} = 2$ for all f, k. The observation model at each frequency becomes

$$p(\boldsymbol{x}_t | w_t = k; \boldsymbol{\theta}) = \mathcal{N}_c(\boldsymbol{x}_t; \mathbf{0}, \boldsymbol{a}_k \boldsymbol{a}_k^\star + \sigma_k^2 \mathbf{I}_M) \tag{11}$$

where frequency indexes have been dropped for convenience. The parameters $\boldsymbol{\theta}$ of this mixture model can be straightforwardly estimated via an expectation-maximization (EM) procedure [27]. Interestingly, using the re-parameterization $\boldsymbol{a}_k \leftarrow \sigma_k \boldsymbol{a}_k$ and $\sigma_k^2 \leftarrow 2\sigma_k^2$, it turns out that these EM updates match those of the blind source separation model proposed in [26], up to a small additive constant for σ_k^2. A key difference is that in [26], the observations are normalized so that $\|\boldsymbol{x}_t\|_2^2 = 1$. As such, [26] belongs to the class of spatial-feature clustering-based methods, similarly to DUET [4], while our method operates in the signal domain.

2.3. General parameter estimation via moment matching

The estimation of the general α-stable model is done by generalized moment matching, that is, minimizing the difference between the empirical and theoretical values of a finite number of generalized moments. Since the characteristic function of our model (9) has a closed-form expression, these moments are selected as a sampling the characteristic function at some frequency vectors $\boldsymbol{\omega}_j \in \mathbb{C}^M$, $j \in [1...m]$. Following the methodology in [22], these frequencies are drawn randomly according to some probability distribution Λ, in practice designed automatically from the data using the method prescribed in [22].

More precisely: given the data points to cluster $\boldsymbol{x}_1, ..., \boldsymbol{x}_T \in \mathbb{C}^M$ (where the index f has been dropped), the estimation is performed as follows:

1. Draw m random frequency vectors $\boldsymbol{\omega}_j \overset{i.i.d.}{\sim} \Lambda$ for $j \in [1...m]$;
2. Compute the empirical characteristic function at these frequencies $\boldsymbol{y} = \left[\frac{1}{T}\sum_{t=1}^{T} e^{i\mathrm{Re}(\boldsymbol{\omega}_j^\star \boldsymbol{x}_t)}\right]_{j=1}^{m} \in \mathbb{C}^m$;
3. Estimate the model parameters (10) by (approximately) solving

$$\min_{\boldsymbol{\theta}} \left\| \boldsymbol{y} - \left[\psi_{\boldsymbol{x}|z=k}(\boldsymbol{\omega}_j)\right]_{j=1}^{m} \right\|_2^2 \tag{12}$$

where $\psi_{\boldsymbol{x}|z=k}(\boldsymbol{\omega})$ is defined by (9), parameterized by $\boldsymbol{\theta}$.

CL-OMPR. The moment matching minimization (12) is carried out by a modified version of the CL-OMPR algorithm [22] adapted to our model. It is a greedy, heuristic algorithm precisely designed to perform mixture model estimation by generalized moment matching. Although it offers limited theoretical guarantees except for very particular settings [28], it has been empirically shown to perform well for a large variety of models [22]. In particular, it is applicable as soon as the considered mixture model has a closed-form characteristic function with respect to the parameters of the model. Initially designed for performing mixture model estimation for large databases, here we advocate that it is also efficient for estimating more exotic, problem-tailored models such as the one proposed in this paper.

The CL-OMPR algorithm is a variant of Orthogonal Matching Pursuit (OMP), a classical greedy algorithm in compressive sensing. It iteratively adds a component to the mixture model by maximizing its correlation to the residual signal, and alternates with a non-convex descent on (12). It also performs more iterations than OMP and includes a Hard Thresholding step to maintain the number of components at K. This allows for replacing spurious components, and has been demonstrated to greatly enhance the algorithm over similar approaches that do not integrate the Hard Thresholding step [22].

The CL-OMPR algorithm is described in detail in previous papers [22], where it is applied to GMM estimation. Replacing the GMM by our alpha-stable model is easily implemented and only requires computation of the gradient of $\psi(\boldsymbol{x}|z = k)$ with respect to the different parameters.

Approximate clustering. A drawback of the alpha-stable model, and major lead for future work, is that the pdf $p(\boldsymbol{x}|z = k)$ does not have an explicit expression, and therefore the clustering of the data points $\boldsymbol{x}_t$ cannot be done by exactly maximizing the posterior $p(\boldsymbol{x}_t|z = k)$ with respect to k.

Although a few methods may exist to approximately compute this posterior using approximate numerical integration [29], in practice we found them to be extremely unstable and time consuming. Instead, we decided to cluster the data *as if the model was Gaussian*, $i.e.$ with $\alpha_k = 2$, since the likelihood is then computable. Hence, the "clustering" part (and therefore the final source separation results) of both EM (Section 2.2) and the alpha-stable model is in fact *the same*: it uses only the parameters $(\boldsymbol{a}_k, \sigma_k, \pi_k)$. The difference between the two lies in the *parameters estimation*: our hope is that, by using the more realistic α-stable source model, steering vectors $\boldsymbol{a}_k$ will be estimated more precisely.

2.4. Frequency permutation ambiguity

Once clustering is performed at each frequency, a permutation ambiguity remains as the assignment of frequency masks to sources is not known. This is a classical problem in blind source separation referred to as *permutation alignment*. It notably occurs when using ICA [6] and clustering-based methods [9, 26]. A number of techniques have been proposed to tackle it, based on temporal activation patterns [26], steering vector models [9] or adjacent frequency bands similarity [30]. The selection and tuning of a specific permutation technique highly depends on the type of signal and mixing model considered, which is out of the scope of this study. For this reason and for fairness, all methods evaluated in the next section benefited from the same *oracle permutation* scheme. At each frequency, the permutation minimizing the mean-squared error between estimated and true source images is selected.

3. EVALUATION AND RESULTS

We use two datasets for evaluation. First, the QUASI database[1], which consists in 10 musical excerpts of 30s. For each excerpt, we produced stereo ($M = 2$) mixes of $K = 4$ musical tracks (vocals, bass, drums, electric guitar, keyboard,...) using random pure gains and delays. Second, the TIMIT speech database[2], from which we created 10 tracks of 30s. For each experiment we mix $K = 3$ of them selected at random into $M = 2$ channels, again with random pure gains and delays. In all cases, the gain differences between the two channels are at most 5dB and the delay is at most 20 samples. Note that none of the tested methods make assumption on the specific convolutive filters used for mixing, as long as they are relatively

[1] www.tsi.telecom-paristech.fr/aao/en/2012/03/12/quasi/
[2] catalog.ldc.upenn.edu/ldc93s1

	SDR (dB)	SIR (dB)	MER (dB)
Mix	-5.96 ± 4.96	-5.49 ± 4.85	N/A
Oracle	8.33 ± 3.16	18.3 ± 4.13	N/A
[26]	1.26 ± 2.44	2.88 ± 3.82	10.5 ± 9.84
EM	3.50 ± 2.87	9.04 ± 4.92	12.3 ± 11.0
CF-GMM	3.80 ± 2.53	8.60 ± 3.62	12.3 ± 9.90
CF-α	$\mathbf{4.11 \pm 2.59}$	$\mathbf{9.17 \pm 3.51}$	$\mathbf{12.65 \pm 9.73}$

(a) QUASI database (music), $K = 4$

	SDR (dB)	SIR (dB)	MER (dB)
Mix	-3.14 ± 1.91	-3.13 ± 1.90	N/A
Oracle	11.9 ± 0.980	25.9 ± 1.05	N/A
[26]	2.16 ± 1.33	4.90 ± 2.54	$\mathbf{22.0 \pm 6.57}$
EM	0.541 ± 0.504	1.44 ± 1.21	12.0 ± 3.64
CF-GMM	1.60 ± 1.10	4.13 ± 2.46	14.8 ± 3.32
CF-α	$\mathbf{2.70 \pm 1.74}$	$\mathbf{6.11 \pm 3.31}$	$\mathbf{18.9 \pm 2.72}$

(b) TIMIT database (speech), $K = 3$

Table 1: Separation results with K sources and $M = 2$ channels, for the four clustering algorithms as well as oracle and mixture results. Each slot contains the mean and standard deviation over the 100 trials and K sources, $i.e.$ over $100K$ values.

small compared to the Fourier analysis window. The STFT parameters were fixed to 64ms windows at 16kHz with 75% overlap.

Each experiment is averaged over 100 trials: each of the 10 songs in the QUASI is selected 10 times, and at each trial random speech tracks are picked from TIMIT and mixed. The results are evaluated using the classical `bss_eval` toolbox [31], and expressed in terms of Source to Distorsion Ratio (SDR), Source to Interferences Ratio (SIR), which evaluate the quality of the reconstructed source signals, and Mixing Error Ratio (MER), which evaluates the estimation of the steering vectors $\boldsymbol{a}_k$, all in dB.

We compare the four following clustering algorithms (recall that in each case the binary masks are then created using the oracle permutation method of Sec. 2.4):

- **EM**: the clustering to form the binary mask is done with a GMM as described in Sec. 2.2. The EM algorithm is repeated 10 times and the result that yields the best log-likelihood is kept.
- **[26]**: This is our implementation of the method of Sawada et al., as described in section 2.2. The EM is also repeated 10 times.
- **CF-GMM**: the clustering is formed with the moment matching method of Sec. 2.3, but with all the α_k fixed to 2. Hence the estimated model is Gaussian. Note that EM and CF-GMM thus estimate models belonging to the *same* family, however with a different cost function: EM through maximization of likelihood and CF-GMM through moment matching of the characteristic function.
- **CF-α**: the clustering is done with the mixture of α-stable distributions of Sec. 2.3. As mentioned before, recall that the *clustering* part is done by approximating the model is Gaussian, only the *estimation* of the parameters is different.

To put the results in context, we also outline the "best" and "worst" possible results, denoted respectively: **oracle**: the separation is performed with the binary mask formed by considering the source that has the highest energy at each TF bin (with oracle knowledge of each source signal); and **mix**: the result are obtained by directly feeding the mixture signal into the function `bss_eval_images`.

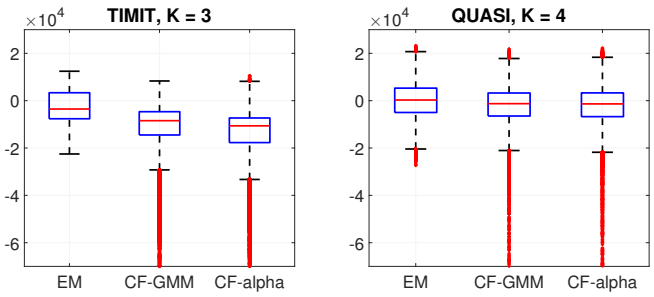

Fig. 1: Log-likelihood of the data at each frequency index for each trial (i.e. $100F$ values), for the EM, CF-GMM and CF-α. For readability the low end of the y-axis has been cut at -7.10^4, the CF-GMM and CF-α algorithms have outliers that go down to, respectively, approximately -2.10^{10} and -3.10^{10}.

Separation results. In Table 1 we show the separation results for all algorithms. We see that CF-α outperforms the other approaches, both on musical and speech signal. In particular, it seems to be especially efficient on speech signal compared to the other algorithms. Since the clustering of CF-α is done using the same Gaussian posteriors as with CF-GMM, the superiority of CF-α must come from a better estimation of the steering vectors, as indicated by the MER. The algorithm of [26] performs well on speech data (it even outperforms CF-α in terms of MER on speech but has a high standard deviation, while CF-α is very stable), but generally fails on musical signals. On the contrary, the EM algorithm is untroubled by musical signals, but fails on speech data.

Relevance of log-likelihood. This last observation is somewhat surprising: on speech data in particular, CF-GMM is seen to outperform EM by a non-negligible margin, despite the fact that both estimate a GMM. In Fig. 1 we compare the log-likelihood results obtained with the three algorithms during the clustering phase subsequent to the estimation of the parameters (recall that all three algorithms have the same clustering phase). As expected, EM significantly outperforms the two other algorithms on this criterion. This is not surprising since EM precisely aims at maximizing the log-likelihood while the two CF algorithms consider only the characteristic function. Since the CF approaches outperforms EM in terms of separation results, we conclude that maximization of the log-likelihood, while natural, might not be the most appropriate approach to estimate the mixture parameters in this case, which is an interesting lead for future work.

4. CONCLUSION

We presented a novel method for multichannel the blind separation of audio sources using an $\alpha-$stable model for source signals, combined with the assumption that only one source dominates each (t, f) point. The parameters of the proposed model, including distinct scale and α values for each source, are estimated at each frequency using a novel method based on random-moment matching. Results show that using oracle permutations, the proposed model performs better than Gaussian models, and that the proposed estimation method outperforms EM even with the same Gaussian model. Future work will further investigate the α and scale values estimated by our method. In particular, it would be interesting to see if they can be constrained or exploited to resolve permutation ambiguities. The potential of random moment matching versus maximum likelihood methods in source separation should also be further studied.

5. REFERENCES

[1] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation: Independent Component Analysis and Blind Deconvolution*, Academic Press, 2010.

[2] T-P Jung, S. Makeig, C. Humphries, T-W Lee, M. Mckeown, V. Iragui, and T. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, 2000.

[3] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, July 2004.

[4] Scott Rickard, "The duet blind source separation algorithm," *Blind Speech Separation*, pp. 217–237, 2007.

[5] E. Vincent, N. Bertin, R. Gribonval, and F. Bimbot, "From blind to guided audio source separation: How models and side information can improve the separation of sound," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 107–115, May 2014.

[6] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, vol. 46, John Wiley & Sons, 2004.

[7] A. Belouchrani, K. Abed-Meraim, J-F Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on signal processing*, vol. 45, no. 2, pp. 434–444, 1997.

[8] L. Benaroya, F. Bimbot, and R. Gribonval, "Audio source separation with a single sensor," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 191–199, Jan. 2006.

[9] 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, sept. 2010.

[10] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Transactions on Acoustics*, vol. 36, pp. 477–489, 1988.

[11] A. Ozerov, E. Vincent, and F. Bimbot, "A general flexible framework for the handling of prior information in audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. PP, no. 99, pp. 1, 2011.

[12] P-S Huang, M. Kim, M. Hasegawa-Johnson, and P. Smaragdis, "Deep learning for monaural speech separation," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1562–1566.

[13] A. Nugraha, A. Liutkus, and E. Vincent, "Multichannel audio source separation with deep neural networks," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1652–1664, 2016.

[14] P. Kidmose, *Blind separation of heavy tail signals*, Ph.D. thesis, Technical University of Denmark, Lyngby, Denmark, 2001.

[15] P. Kidmose, "Independent component analysis using the spectral measure for alpha-stable distributions," in *IEEE-EURASIP 2001 Workshop on Nonlinear Signal and Image Processing*, 2001, vol. 400.

[16] G. Samoradnitsky and M. Taqqu, *Stable non-Gaussian random processes: stochastic models with infinite variance*, vol. 1, CRC Press, 1994.

[17] C. Nikias and M. Shao, *Signal processing with alpha-stable distributions and applications*, Wiley-Interscience, 1995.

[18] A. Liutkus and R. Badeau, "Generalized Wiener filtering with fractional power spectrograms," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, April 2015.

[19] M. Fontaine, A. Liutkus, L. Girin, and R. Badeau, "Explaining the parameterized wiener filter with alpha-stable processes," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2017.

[20] U. Şimşekli, A. Liutkus, and A.T. Cemgil, "Alpha-stable matrix factorization," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2289–2293, 2015.

[21] A. Liutkus, T. Olubanjo, E. Moore, and M. Ghovanloo, "Source Separation for Target Enhancement of Food Intake Acoustics from Noisy Recordings," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, United States, Oct. 2015.

[22] N. Keriven, A. Bourrier, R. Gribonval, and P. Pérez, "Sketching for large-scale learning of mixture models," *arXiv:1606.02838, Information and Inference: A Journal of the IMA*, 2017.

[23] R. Gribonval, G. Blanchard, N. Keriven, and Y. Traonmilin, "Compressive statistical learning with random feature moments," *arXiv preprint arXiv:1706.07180*, 2017.

[24] M. Fontaine, C. Vanwynsberghe, A. Liutkus, and R. Badeau, "Scalable source localization with multichannel alpha-stable distributions," in *25th European Signal Processing Conference (EUSIPCO 2017)*, 2017.

[25] R. Gallager, "Circularly Symmetric Complex Gaussian Random Vectors - A Tutorial," Tech. Rep., Massachusetts Institute of Technology, 2008.

[26] Hiroshi Sawada, Shoko Araki, and Shoji Makino, "A two-stage frequency-domain blind source separation method for underdetermined convolutive mixtures," in *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop on*. IEEE, 2007, pp. 139–142.

[27] Michael E Tipping and Christopher M Bishop, "Mixtures of probabilistic principal component analyzers," *Neural computation*, vol. 11, no. 2, pp. 443–482, 1999.

[28] Nicholas Boyd, Geoffrey Schiebinger, and Benjamin Recht, "The Alternating Descent Conditional Gradient Method for Sparse Inverse Problems," pp. 1–21, 2015.

[29] John P. Nolan, "Multivariate elliptically contoured stable distributions: Theory and estimation," *Computational Statistics*, vol. 28, no. 5, pp. 2067–2089, 2013.

[30] Leandro E Di Persia and Diego H Milone, "Using multiple frequency bins for stabilization of fd-ica algorithms," *Signal Processing*, vol. 119, pp. 162–168, 2016.

[31] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462 –1469, July 2006.