BAYESIAN NON-PARAMETRIC MULTI-SOURCE MODELLING BASED DETERMINED BLIND SOURCE SEPARATION

Chaitanya Narisetty, Tatsuya Komatsu, and *Reishi Kondo*

Data Science Research Laboratories, NEC Corporation, Japan

ABSTRACT

This paper proposes a determined blind source separation method using Bayesian non-parametric modelling of sources. Conventionally source signals are separated from a given set of mixture signals by modelling them using non-negative matrix factorization (NMF). However in NMF, a latent variable signifying model complexity must be appropriately specified to avoid over-fitting or under-fitting. As real-world sources can be of varying and unknown complexities, we propose a Bayesian non-parametric framework which is invariant to such latent variables. We show that our proposed method adapts to different source complexities, while conventional methods require parameter tuning for optimal separation.

Index Terms— Blind Source Separation, Non-Negative Matrix Factorization, Bayesian Non-Parametrics, Inference

1. INTRODUCTION

The ubiquity of smart devices in our present day provide a diverse range of audio based applications like detecting threat indicating sounds, transcribing speech, building AI assistants etc. to facilitate our day-to-day activities [1]. Depending on the application, our desired information extracted from the audio signals changes. Real world audio signals are often obscured by undesired information and necessitate the development of source separation methods [2]. As most smart devices are equipped with two or more microphones, the recorded multi-channel data can be leveraged to improve the separation performance [3]. However the arrangement of microphones varies across devices, as do the source characteristics and locations. Separating sources without such device or source information is termed as blind source separation (BSS) [4,5].

Fundamental BSS techniques include matrix factorization methods like independent component analysis (ICA) [6], non-negative matrix factorization (NMF) [7] etc. These techniques formulate the audio signals captured by microphones as linear mixtures of two or more source signals. In other words, the spectra of mixture signals are treated as linear combinations of spectra of source signals, and solved by assuming independence among the distributions of source spectra. This assumption of independence is often not true because the spectra of a typical audio source is highly corre-lated with itself across different time intervals. Such correlations or fundamental patterns in each source are efficiently extracted using NMF [8], which are referred to as bases. The way these bases are linearly combined to reconstruct spectra, are referred to as activations and are extracted simultaneously by NMF. BSS methods are therefore extended by modelling such correlations using multi-channel NMF [7,9]. It is intuitive that source separation becomes more reliable as the number of available microphones increases. When there are as many number of microphones as there are sources, the situation is said to be determined. Ono et. al. [10] proposed independent vector analysis (IVA) for determined BSS using an auxiliary function technique to derive stable and fast update rules for demixing parameters. Kitamura et. al. [11] proposed an independent low-rank matrix analysis method (ILRMA) by unifying IVA and multi-source NMF modelling.

In the above discussed NMF-based BSS methods, it is required to provide an appropriate value for NMF's model complexity i.e. the number of bases extracted for each source model. This latent variable when too low or too high, leads to under-fitting or over-fitting respectively. Its value is often chosen depending on source characteristics or by parameter tuning. However real world source characteristics are unknown, and their complexities range from simple keyboard clicking to moderate sounds of drums to complex music pieces. Although non-parametric methods exist for estimating number of sources, to the best of our knowledge, there aren't any NMF-based BSS techniques which can adaptively model sources having different complexities.

We propose a non-parametric framework of multi-source modelling unified with IVA for determined BSS, and overcome the problem of tuning NMF's model complexity parameter. Our proposed method utilizes the concepts of variational Bayesian inference to statistically estimate each source's complexity, thereby optimally separating the sources. Proposed method therefore serves as a generalization of ILRMA by adapting to varying source complexities.

The remainder of this paper is organized as follows. Section 2 details the conventional method: ILRMA. Section 3 formulates our proposed method. Section 4 details the experimental results and discusses a few potential implications of the proposed method. Finally section 5 concludes our work.

2. CONVENTIONAL METHOD

Most BSS methods model the given mixture spectra as linear combinations of source spectra and are formulated as

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij}, \qquad (1)$$

where $\boldsymbol{x}_{ij} = (x_{ij,1}, \ldots, x_{ij,M})^{\mathsf{T}}$ and $\boldsymbol{s}_{ij} = (s_{ij,1}, \ldots, s_{ij,N})^{\mathsf{T}}$ denote the mixture and source spectra respectively for each index $i \in \{1, 2, \ldots, I\}$, $j \in \{1, 2, \ldots, J\}$. $(.)^{\mathsf{T}}$ denotes a transpose and I, J, M, N denote the number of frequency bins, time frames, microphones and sources respectively. $\boldsymbol{A}_i$ is an $M \times N$ mixing matrix comprising of N steering vectors for the N respective sources. In a determined case, $M = N$ and the square matrix $\boldsymbol{A}_i$ has a valid inverse matrix. Therefore [10] proposed IVA by defining a demixing matrix $\boldsymbol{W}_i = \boldsymbol{A}_i^{-1} = (\boldsymbol{w}_{i,1}, \ldots, \boldsymbol{w}_{i,M})^{\mathsf{H}}$ and reformulated Eq. (1) as

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij}, \qquad (2)$$

where $\boldsymbol{y}_{ij} = (y_{ij,1}, \ldots, y_{ij,M})^{\mathsf{T}}$ denotes the estimated source spectra and $(.)^{\mathsf{H}}$ denotes a hermitian transpose.

2.1. ILRMA: Unifying IVA and NMF

This method extends IVA by independently modelling each source with an isotropic complex Gaussian distribution. For each source index $m \in \{1, \ldots, M\}$, the distribution variance denoted as $r_{ij,m}$ is non-negative and modelled using NMF as

$$r_{ij,m} = \sum_{k=1}^{K_m} t_{ik,m} v_{kj,m}, \qquad (3)$$

where $t_{ik,m}$ and $v_{kj,m}$ are the elements of basis and activation vectors respectively. K_m is NMF's model complexity parameter, signifying the number of basis vectors for m^{th} source. The cost function Q of ILRMA is derived in [11] as

$$Q = -2J \sum_i |\det \boldsymbol{W}_i| + \sum_{i,j,m} \left[\log r_{ij,m} + \frac{|y_{ij,m}|^2}{r_{ij,m}} \right]. \quad (4)$$

ILRMA estimates the demixing parameters by maximizing cost function Q using multi-source NMF modelling in Eq. (3).

2.2. Limitation: Number of Source Bases

In the formulation of ILRMA, note that a set of complexity parameters $\{K_m, 1 \le m \le M\}$ is to be specified by the user. Failing to provide a reasonable estimate of each source's complexity will lead to over-fitting or under-fitting. This is discussed in [11] by comparing the cases of separating speech signals with that of music signals. The former requires only 2 bases for optimal modelling as compared to the latter requiring more than 30. Therefore the model complexity parameter can effect separation performance, depending on the types of sources being separated. For simplicity, ILRMA assumes equal number of bases for each source i.e.

$\{K_1 = \cdots = K_M = K\}$. This further limits ILRMA's ability to optimally separate a low complexity source and a high complexity source from their mixture signals.

3. PROPOSED MULTI-SOURCE MODELLING

We overcome the limitation of ILRMA by proposing a probabilistic modelling of the variance of source distributions. In such techniques, it is common to introduce hidden variables to capture the structure of given observed data, and utilize inference algorithms to estimate the posterior distribution. Accordingly, our proposed method flexibly models each source using a large number of basis vectors K and incorporates a *reliability* value for each basis vector. We denote each reliability value as $z_{k,m}$ which can be interpreted as a quantified contribution of kth basis towards m^{th} source.

3.1. Model Formulation

Contrast to the NMF-based source modelling in Eq. (3), we propose a probabilistic model for source variance $r_{ij,m}$ as

$$r_{ij,m} = \sum_{k=1}^{K} z_{k,m} t_{ik,m} v_{kj,m}. \qquad (5)$$

where the prior distributions for each of $t_{ik,m}$, $v_{kj,m}$ and $z_{k,m}$ are drawn from a random process as

$$\begin{aligned}
p(t_{ik,m}) &\sim \text{Gamma}(a_0, a_0), \\
p(v_{kj,m}) &\sim \text{Gamma}(b_0, b_0), \\
p(z_{k,m}) &\sim \text{Gamma}(c_0, c_m), \quad (6)
\end{aligned}$$

where a_0, b_0, c_0 are positive constants, Gamma$(.,.)$ is a gamma distribution defined over a shape parameter and a rate (inverse-scale) parameter. When $c_0 \ll 1$, a sparse prior is set over the reliability values and therefore adapts to different model complexities depending on each source's characteristics [12]. As each source's expected variance should correspond to the expectation of its power, the choice of prior parameters require that $\mathbb{E}_p[|y_{ij,m}|^2] = \mathbb{E}_p[r_{ij,m}]$,

$$\Rightarrow \mathbb{E}_p[|y_{ij,m}|^2] = \sum_k \mathbb{E}_p[z_{k,m}] \mathbb{E}_p[t_{ik,m} v_{kj,m}] = \sum_k (c_0/c_m),$$

$$\Rightarrow \quad c_m = c_0 K \big[\sum_i \sum_j |y_{ij,m}|^2 / (IJ) \big]^{-1}. \qquad (7)$$

Maximizing the cost function in Eq. (4), as updated by Eq. (5) is the central focus of our approach.

3.2. Variational Bayesian Inference

The isotropic gaussian distribution of sources is not conjugate to the gamma distribution of source parameters t, v, z. This non-conjugacy precludes the use of Markov chain Monte Carlo (MCMC) based Metropolis-Hastings, Gibbs sampling for inferring our desired posterior [13,14]. However variational approaches have provided alternatives for non-conjugate models that are also faster for large amounts of

data. We adopt a fully factorized mean-field variational inference technique as it assumes conditional independence among the hidden variables [15] and approximates them from a family of conditional distributions over variational parameters [16]. Generalized inverse Gaussian (GIG) distributions are chosen as our variational family and are expressed as

$$\text{GIG}(\theta|\gamma,\rho,\tau) = \frac{\exp\{(\gamma-1)\log\theta - \rho\theta - \tau/\theta\}}{2(\tau/\rho)^{\gamma/2}\mathcal{K}_\gamma(2\sqrt{\rho\tau})}, \quad (8)$$

where $\mathcal{K}(.)$ is a modified Bessel function of the second kind and γ, ρ, τ are the variational hyper-parameters. GIG's sufficient statistics includes $(1/\theta)$ which eases the optimization of our cost function's $(1/r_{ij,m})$ term and motivates our choice of GIG [17]. We define the conditional distributions as

$$q(t_{ik,m}|\Theta_{\backslash t_{ik,m}}) \sim \text{GIG}(a_0, \rho_{ik,m}^{(t)}, \tau_{ik,m}^{(t)}),$$
$$q(v_{kj,m}|\Theta_{\backslash v_{kj,m}}) \sim \text{GIG}(b_0, \rho_{kj,m}^{(v)}, \tau_{kj,m}^{(v)}),$$
$$q(z_{k,m}|\Theta_{\backslash z_{k,m}}) \sim \text{GIG}(c_0, \rho_{k,m}^{(z)}, \tau_{k,m}^{(z)}). \quad (9)$$

We now derive update equations from the cost function in Eq. (4) using first-order Taylor expansion and Jensen's inequality [18] by introducing their respective auxiliary positive constants $\alpha_{ij,m}$ and $\beta_{ijk,m}$ as

$$Q + 2J\sum_i |\det \boldsymbol{W}_i| = \sum_{i,j,m}\left[\log r_{ij,m} + \frac{|y_{ij,m}|^2}{r_{ij,m}}\right],$$

$$\leq \sum_{i,j,m}\mathbb{E}_q\left[\log r_{ij,m} + \frac{|y_{ij,m}|^2}{r_{ij,m}}\right] + \mathbb{E}_q\left[\log\frac{q(t|\Theta_{\backslash t})}{p(t|a_0)}\right]$$

$$+ \mathbb{E}_q\left[\log\frac{q(v|\Theta_{\backslash v})}{p(v|b_0)}\right] + \mathbb{E}_q\left[\log\frac{q(z|\Theta_{\backslash z})}{p(z|c_0,c_m)}\right],$$

$$\leq \sum_{i,j}\left[\sum_{k,m}|y_{ij,m}|^2\beta_{ijk,m}^2\mathbb{E}_q\left[z_{k,m}^{-1}t_{ik,m}^{-1}v_{kj,m}^{-1}\right]\right.$$

$$\left. -1 + \log\alpha_{ij,m} + \frac{1}{\alpha_{ij,m}}\sum_k\mathbb{E}_q[z_{k,m}t_{ik,m}v_{kj,m}]\right]$$

$$+ \mathbb{E}_q[-\rho_{ik,m}^{(t)}t_{ik,m} - \tau_{ik,m}^{(t)}/t_{ik,m} + a_0 t_{ik,m}]$$

$$+ \mathbb{E}_q[-\rho_{kj,m}^{(v)}v_{kj,m} - \tau_{kj,m}^{(v)}/v_{kj,m} + b_0 v_{kj,m}]$$

$$+ \mathbb{E}_q[-\rho_{k,m}^{(z)}z_{k,m} - \tau_{k,m}^{(z)}/z_{k,m} + c_m z_{k,m}] + C, \quad (10)$$

where C denotes a leftover constant. Note that the pairs $(t, t^{-1}), (v, v^{-1})$ and (z, z^{-1}) are sufficient statistics for their respective GIG distributions. This allows us to avoid taking partial derivatives and directly derive the analytic coordinate ascent updates for our hyper-parameters by comparing the coefficients of sufficient statistics [17]. Constants $\alpha_{ij,m}$ and $\beta_{ijk,m}$ re-tighten the above inequality (10) when:

$$\alpha_{ij,m} = \sum_k \mathbb{E}_q[z_{k,m}]\mathbb{E}_q[t_{ik,m}]\mathbb{E}_q[v_{kj,m}], \quad (11)$$

$$\beta_{ijk,m} = \frac{\mathbb{E}_q\left[z_{k,m}^{-1}\right]\mathbb{E}_q\left[t_{ik,m}^{-1}\right]\mathbb{E}_q\left[v_{kj,m}^{-1}\right]}{\sum_k \mathbb{E}_q\left[z_{k,m}^{-1}\right]\mathbb{E}_q\left[t_{ik,m}^{-1}\right]\mathbb{E}_q\left[v_{kj,m}^{-1}\right]}. \quad (12)$$

Expectation of source parameters in Eqs. (11) and (12) can be obtained similar to the expectation of a random variable θ defined over a GIG distribution in Eq. (8) as

$$\mathbb{E}[\theta] = \frac{\mathcal{K}_{\gamma+1}(2\sqrt{\rho\tau})\sqrt{\tau}}{\mathcal{K}_\gamma(2\sqrt{\rho\tau})\sqrt{\rho}}, \mathbb{E}\left[\frac{1}{\theta}\right] = \frac{\mathcal{K}_{\gamma-1}(2\sqrt{\rho\tau})\sqrt{\rho}}{\mathcal{K}_\gamma(2\sqrt{\rho\tau})\sqrt{\tau}}. \quad (13)$$

The update equations for our hyper-parameters are derived as

$$\rho_{ik,m}^{(t)} = a_0 + \mathbb{E}_q[z_{k,m}]\sum_j \mathbb{E}_q[v_{kj,m}]\alpha_{ij,m}^{-1}, \quad (14)$$

$$\tau_{ik,m}^{(t)} = \sum_j |y_{ij,m}|^2\beta_{ijk,m}^2\mathbb{E}_q\left[z_{k,m}^{-1}\right]\mathbb{E}_q\left[v_{kj,m}^{-1}\right], \quad (15)$$

$$\rho_{kj,m}^{(v)} = b_0 + \mathbb{E}_q[z_{k,m}]\sum_i \mathbb{E}_q[t_{ik,m}]\alpha_{ij,m}^{-1}, \quad (16)$$

$$\tau_{kj,m}^{(v)} = \sum_i |y_{ij,m}|^2\beta_{ijk,m}^2\mathbb{E}_q\left[z_{k,m}^{-1}\right]\mathbb{E}_q\left[t_{ik,m}^{-1}\right], \quad (17)$$

$$\rho_{k,m}^{(z)} = c_m + \sum_i\sum_j \mathbb{E}_q[t_{ik,m}]\mathbb{E}_q[v_{kj,m}]\alpha_{ij,m}^{-1}, \quad (18)$$

$$\tau_{k,m}^{(z)} = \sum_i\sum_j |y_{ij,m}|^2\beta_{ijk,m}^2\mathbb{E}_q\left[t_{ik,m}^{-1}\right]\mathbb{E}_q\left[v_{kj,m}^{-1}\right]. \quad (19)$$

3.3. Update Rules for Demixing Matrix

Given each source's variance, the proposed method does not alter the partial derivatives of cost function Q over the demixing matrix $\boldsymbol{W}_i$. Hence its update equations coincide with those described for IVA using an auxiliary function technique [10] and are derived as follows:

$$V_{i,m} = \frac{1}{J}\sum_j \frac{1}{r_{ij,m}}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^h, \quad (20)$$

$$\boldsymbol{w}_{i,m} \leftarrow (\boldsymbol{W}_i V_{i,m})^{-1}\boldsymbol{e}_m, \quad (21)$$

$$\boldsymbol{w}_{i,m} \leftarrow \boldsymbol{w}_{i,m}(\boldsymbol{w}_{i,m}^h V_{i,m}\boldsymbol{w}_{i,m})^{-1/2}, \quad (22)$$

where $\boldsymbol{e}_m$ is a unit vector whose m^{th} element equals one. After the elements of demixing matrix are estimated, the separated source spectra can be extracted as

$$y_{ij,m} \leftarrow \boldsymbol{w}_{i,m}^h\boldsymbol{x}_{ij}. \quad (23)$$

In each iteration of the proposed method, hyper-parameters $\rho^{(\cdot)}$ and $\tau^{(\cdot)}$ are updated using Eq. (14)-(19). Demixing matrices $\boldsymbol{W}_i$ and the separated source spectra are then updated using Eq. (20)-(23). As each source's modelling begins with a large K, our variational inference is computationally exacting. However over a few iterations, the sparse prior placed over z estimates a small number of bases which are reliable. So we employ a thresholding technique [19] to skip the optimization of less reliable bases in subsequent iterations.

4. SIMULATIONS AND RESULTS

4.1. Experimental Conditions

We evaluate our proposed method on the DSD100 dataset containing 100 professionally produced songs from the 2016
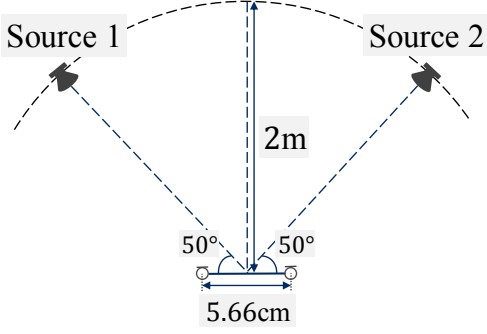
Fig. 1: E2A recording conditions (reverberation time: 300ms)

SiSEC challenge [20]. Each song data consists of clean sounds for vocals, drums, bass and other accompaniments, and lasts for more than 2-3 minutes. So we only choose a 30 second portion (30s-60s time interval) of clean sources and down-sample them to $16\,$kHz for creating mixture signals. For each of these 100 songs, we randomly choose between the pairs (drums, vocals) or (bass, vocals) and create synthetic two-channel reverberant mixture signals using the recoding conditions shown in Fig. 1. Room impulse responses E2A ($T_{60} = 300\,$ms) for above recording conditions were obtained from the RWCP Sound Scene Database [21].

Mixture spectra are estimated from the time domain signals using a Hamming window of length $512\,$ms shifted every $128\,$ms, which were found to be optimal by [22]. Each demixing matrix $\boldsymbol{W}_i$ is initialized with an identity matrix. We model each source's variance with $K = 30$ bases and set $a_0 = b_0 = 0.1, c_0 = 1/K$. Hyper-parameters $\rho^{(\cdot)}$ and $\tau^{(\cdot)}$ are initialized randomly from gamma distributions with shape and rate parameters set to 1000. Parameters are optimized for 100 iterations and then the separated spectra are converted to time domain using a back-projection technique [23].

4.2. Evaluations and Comparisons

Three metrics: signal to distortion ratio (SDR), signal to interference ratio (SIR) and signal to artifacts ratio (SAR) [24] are used to evaluate the quality of separated sources. Separation performance of the proposed method is compared with three NMF-based BSS method i.e. IVA [10], MNMF [9] and IL-RMA [11] (with 5 source bases). Each separation is repeated for 10 different random initializations, and the average of the above performance metrics are reported in Table 1.

Table 1: Comparison of Separation Performance

Methods	SDR	SIR	SAR
IVA [10]	2.6 dB	7.4 dB	5.7 dB
MNMF [9]	3.5 dB	8.6 dB	6.2 dB
ILRMA [11]	8.7 dB	16.2 dB	12.3 dB
Proposed	**9.9** dB	**17.4** dB	**13.2** dB

Although the proposed method outperforms ILRMA, it is important to verify that we overcome its limitation of having to tune the complexity parameter K. This limitation can be seen in Fig. 2 where ILRMA's SDR averaged over all the mixture signals containing (bass, vocals) increases as K increases, while an opposite trend is seen for mixture signals containing (drums, vocals). The proposed method on the other hand starts with a large value for K ($= 30$ in this case) and tunes itself depending on each source's characteristics. Hence it optimally separates the sources from both types of mixture signals. The choice of K, if sufficiently large, does not to significantly impact proposed method's performance.
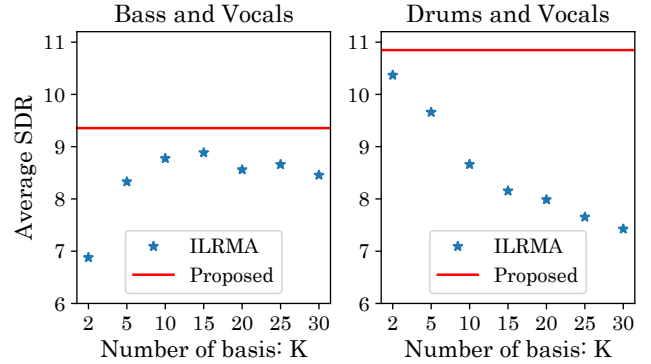
Fig. 2: SDR for ILRMA using different number of bases as compared with proposed method's SDR for mixture signals of bass and vocals (left), and of drums and vocals (right).

4.3. Possible Extension

As each source is modelled individually using NMF, it is possible to extend our approach by considering a common basis and activation matrix for capturing the variance of all sources i.e. $r_{ij,m} = \sum_k z_{k,m} t_{ik} v_{kj}$. This requires fewer parameters to be estimated and its computational complexity becomes at least half as compared to that of the proposed method. We note that this extension is similar to Ozerov's MNMF [7] and has also been considered for ILRMA [11]. However the key difference is that we do not restrict the overall contribution of each basis to be 1, i.e. $\sum_m z_{k,m} = 1 \,\forall\, k$. Due to space constraints, this extension will be explored as a future work.

5. CONCLUSIONS

This work proposes a Bayesian generalization of ILRMA for determined blind source separation by performing multi-source modelling using non-parametric NMF. Our formulation for individual source modelling is able to overcome the limitation of conventional method, whose separation performance is effected by NMF's model complexity parameter. Proposed approach is flexible in modelling sources of different complexities and is therefore able to optimally separate them. We further show that our approach outperforms the state-of-the-art NMF based techniques.

6. REFERENCES

[1] J. Foote, "An overview of audio information retrieval," *Multimedia systems*, vol.7, no.1, pp.2–10, 1999.

[2] M. Davies, "Audio source separation," in *Institute of mathematics and its applications conference series*, 2002, vol. 71, pp. 57–68.

[3] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multi-channel signal separation by decorrelation," *IEEE transactions on Speech and Audio Processing*, vol.1, no.4, pp.405–413, 1993.

[4] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic press, 2010.

[5] X. Cao and R. Liu, "General approach to blind source separation," *IEEE Transactions on signal Processing*, vol.44, no.3, pp.562–571, 1996.

[6] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol.36, no.3, pp.287–314, 1994.

[7] A. Ozerov, C. Févotte, R. Blouet, and J. Durrieu, "Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 257–260.

[8] C. Narisetty, T. Komatsu, and R. Kondo, "Modelling of sound events with hidden imbalances based on clustering and separate Sub-Dictionary learning," in *26th European Signal Processing Conference (EUSIPCO)*, 2018.

[9] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complex-valued data," *IEEE Transactions on Audio, Speech, and Language Processing*, vol.21, no.5, pp.971–982, 2013.

[10] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 189–192.

[11] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol.24, no.9, pp.1622–1637, 2016.

[12] V. Y. Tan and C. Févotte, "Automatic relevance determination in nonnegative matrix factorization,"

in *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[13] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," 1970.

[14] D. M. Blei, J. D. Lafferty, et al., "A correlated topic model of science," *The Annals of Applied Statistics*, vol.1, no.1, pp.17–35, 2007.

[15] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference," *The Journal of Machine Learning Research*, vol.14, no.1, pp.1303–1347, 2013.

[16] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Machine learning*, vol.37, no.2, pp.183–233, 1999.

[17] D. M. Blei, P. R. Cook, and M. Hoffman, "Bayesian nonparametric matrix factorization for recorded music," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 439–446.

[18] J. D. Lafferty and D. M. Blei, "Correlated topic models," in *Advances in neural information processing systems*, 2006, pp. 147–154.

[19] J. Paisley and L. Carin, "Nonparametric factor analysis with beta process priors," in *ACM Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 777–784.

[20] A. Liutkus, F. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, "The 2016 signal separation evaluation campaign," in *Latent Variable Analysis and Signal Separation: 13th International Conference, LVA/ICA, Grenoble, France*, 2017, pp. 323–332.

[21] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and Hands-Free speech recognition.," in *LREC*, 2000, [Online; accessed 29-Oct-2018] Available: http://www.openslr.org/13/.

[22] D. Kitamura, N. Ono, and H. Saruwatari, "Experimental analysis of optimal window length for independent low-rank matrix analysis," in *25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 1170–1174.

[23] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol.41, no.1-4, pp.1–24, 2001.

[24] 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, 2006.