14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

BAYESIAN BLIND SEPARATION OF AUDIO MIXTURES WITH STRUCTURED PRIORS

Cédric Févotte

Mist-Technologies
204, rue de Crimée, 75019 Paris, France
Email: cedric.fevotte@mist-technologies.com
Web: www.mist-technologies.com

ABSTRACT

In this paper we describe a Bayesian approach for separation of linear instantaneous mixtures of audio sources. Our method exploits the sparsity of the source expansion coefficients on a time-frequency basis, chosen here to be a MDCT. Conditionally upon an indicator variable which is 0 or 1, one source coefficient is either set to zero or given a Student *t* prior. Structured priors can be considered for the indicator variables, such as horizontal structures in the time-frequency plane, in order to model temporal persistency. A Gibbs sampler (a standard Markov chain Monte Carlo technique) is used to sample from the posterior distribution of the indicator variables, the source coefficients (corresponding to nonzero indicator variables), the hyperparameters of the Student *t* priors, the mixing matrix and the variance of the noise. We give results for separation of a musical stereo mixture of 3 sources.

1. INTRODUCTION

Blind Source Separation (BSS) consists in estimating n signals (the sources) from the sole observation of m mixtures of them (the observations). In this paper we consider linear instantaneous mixtures of time series: at each time index, the observations are a linear combination of the sources at the same time index. If many efficient approaches exist for (over)determined ($m \geq n$) non-noisy linear instantaneous, in particular within the field of Independent Component Analysis, the general linear instantaneous case, with mixtures possibly noisy and/or underdetermined ($m < n$) is still a very challenging problem.

A now common approach to the latter problem is the use of source sparsity assumptions, as introduced in the seminal papers [1, 2]. The assumption of sparsity means that only a few coefficients of the sources are significantly non-zero. If the sources are not sparse in their original domain (e.g, the time domain for audio signals), they might be sparse in a transformed domain (e.g, the Fourier domain, wavelet transform). Within a Bayesian framework, we modeled in [3] the expansion coefficients of the sources on a chosen basis by identically and independently distributed (i.i.d) Student t processes with low degrees of freedom; a Gibbs sampler was proposed to sample from the posterior distribution of the mixing matrix, the input noise variance, the source coefficients and hyperparameters of the Student t distributions. An extension of this approach was proposed in [4], where a frequency-dependent (instead of i.i.d) model of the sources was considered. The method was successfully applied to determined and underdetermined noisy audio mixtures, decomposed on a MDCT basis (a local cosine basis).

This paper presents further developments of the above mentioned Bayesian approach. The coefficients of the sources are now given a “strict” sparse prior: conditionally upon an indicator variable which is 0 or 1, one source coefficient is either set to zero or given a Student t distribution (which does not need to have a low degree of freedom anymore). The indicator variable can be given an independent Bernoulli prior or, more interestingly, *structured priors*. For example, when using a time-frequency basis such as a MDCT, horizontal structures can be favored in the time-frequency plane to model tonals of musical sources. This model was successfully applied to audio denoising in [5].[1] Temporal Markov chain source models have also been used for BSS purposes in [6, 7]. The scope of these papers is however slightly different than ours. Reference [6] deals with convolutive mixtures, but assumes the mixing filters known, and relies on prior training of the Markov transition probabilities. Reference [7] addresses more specifically musical (non-percussive) source separation, exploiting prior information about the microphones spatial configuration and relying on thorough note models training. Our approach, though limited at the moment to instantaneous mixtures, is in contrast completely adaptive: we are able to estimate both the mixing matrix and the sources, and do not need any prior model training.

This paper is organized as follows. Section 2 introduces notations and presents the source models. Section 3 gives the update equations of the Gibbs sampler. Separation results of a linear instantaneous stereo mixture of 3 audio sources are given in Section 4, with comparison to our previous work. Conclusions and perspectives are given in Section 5.

2. MODEL

2.1 Linear instantaneous mixture model

We consider the following standard linear instantaneous model, $\forall t \in [\![1, N]\!]$:

$$\mathbf{x}_t = \mathbf{A}\, \mathbf{s}_t + \mathbf{n}_t \tag{1}$$

where $\mathbf{x}_t = [x_{1,t}, \ldots, x_{m,t}]^T$ is a vector of size m containing the observations, $\mathbf{s}_t = [s_{1,t}, \ldots, s_{n,t}]^T$ is a vector of size n containing the sources and $\mathbf{n}_t = [n_{1,t}, \ldots, n_{m,t}]^T$ is a vector of size m

Most of this work was done while the author was a Research Associate with University of Cambridge (Signal Processing Lab, Engineering Dept). The author acknowledges the European Commission funded Research Training Network HASSIP (HPRN-CT-2002-00285) for financial support.

[1] Reference [5] even considers overcomplete dictionaries; however in this paper we limit ourselves to an orthonormal basis.

containing additive noise. Variables without time index t denote whole sequences of samples, *e.g*, $\mathbf{x} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ and $x_1 = [x_{1,1}, \ldots, x_{1,N}]$.

The aim of the following work is to estimate the sources $\mathbf{s}$ and the mixing matrix $\mathbf{A}$ up to the standard BSS indeterminacies on gain and order, that is, compute $\hat{\mathbf{s}}$ and $\hat{\mathbf{A}}$ such that ideally $\hat{\mathbf{A}} = \mathbf{A}\,\mathbf{D}\,\mathbf{P}$ and $\hat{\mathbf{s}} = \mathbf{P}^T\,\mathbf{D}^{-1}\,\mathbf{s}$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{P}$ is a permutation matrix.

2.2 Time domain / Transform domain

Let $x \in \mathbb{R}^{1\times N} \rightarrow \tilde{x} \in \mathbb{R}^{1\times N}$ denote a bijective linear transform, preferably orthonormal. Denoting for $k \in [\![1,N]\!]$, $\tilde{\mathbf{x}}_k = [\tilde{x}_{1,k}, \ldots, \tilde{x}_{m,k}]^T$ and $\tilde{\mathbf{n}}_k$, $\tilde{\mathbf{s}}_k$ similarly, by linearity of the t-f transform we have

$$\tilde{\mathbf{x}}_k = \mathbf{A}\,\tilde{\mathbf{s}}_k + \tilde{\mathbf{n}}_k \tag{2}$$

Furthermore, the transform being bijective, solving the problem defined by Eq. (1) in the time domain is equivalent to solving Eq. (2) in the transform domain. In the rest of this paper we will more specifically use a MDCT with time resolution l_{frame}/f_s and frequency resolution $\frac{f_s}{2}/l_{\text{frame}}$, where f_s is the sampling frequency. When required, the index $k \in [\![1,N]\!]$ will be more conveniently rewritten $k = (q,p) \in [\![1,l_{\text{frame}}]\!] \times [\![1,n_{\text{frame}}]\!]$, with $n_{\text{frame}} = N/l_{\text{frame}}$ and where q is a frequency index and p is a frame index.

2.3 Priors

2.3.1 Coefficients priors

The coefficients $\tilde{s}_{i,k}$, $i \in [\![1,n]\!]$, $k \in [\![1,N]\!]$ are given the following hierarchical prior:

$$\begin{aligned} p(\tilde{s}_{i,k}|\gamma_{i,k}, v_{i,k}) &= (1-\gamma_{i,k})\,\delta_0(\tilde{s}_{i,k}) + \gamma_{i,k}\,N(\tilde{s}_{i,k}|0, v_{i,k}) && (3)\\ p(v_{i,k}|\alpha_i, \lambda_i) &= IG(v_{i,k}|\alpha_i, \lambda_i(q)) && (4)\end{aligned}$$

where $N(u|\mu, v)$ and $IG(u|\alpha,\beta)$ are the normal and inverted-Gamma distributions as defined in Appendix, $\delta_0(u)$ is the Dirac delta function and $\gamma_{i,k} \in \{0,1\}$ is an indicator variable. When $\gamma_{i,k} = 0$, $\tilde{s}_{i,k}$ is set to zero; when $\gamma_{i,k} = 1$, $\tilde{s}_{i,k}$ has a normal distribution with zero mean and variance $v_{i,k}$, which is in turn assigned a conjugate inverted-Gamma prior. $\lambda_i(q)$ (where q is the frequency index in $k = (q,p)$) is a frequency dependent scale parameter. $\lambda_i(q)$ should decrease with frequency, modeling the non-uniform energy distribution of audio signals. In practice we used $\lambda_i(q) = \lambda_i\, f(q)$ with

$$f(q) = \frac{1}{(1 + ((q-1)/q_0)^2)}, \quad q \in [\![1, l_{\text{frame}}]\!] \tag{5}$$

Integrating out $v_{i,k}$, the prior of $\tilde{s}_{i,k}$ conditionally upon $\gamma_{i,k} = 1$ is simply $t(\tilde{s}_{i,k}|2\,\alpha_i, \sqrt{\lambda_i(q)/\alpha_i})$, where $t(u|\alpha,\lambda)$ is the Student t distribution defined in the Appendix. The hierarchical formulation of the prior (3)-(4) is preferred because it allows for easy Gibbs sampling.

2.3.2 Indicator variable priors

We consider two scenarios for the indicator variables $\gamma_{i,k}$:

1. *Bernoulli priors*: no structure is imposed on the indicator variables, which are assigned the following independent Bernoulli priors:

$$P(\gamma_{i,k} = 1|P_i) = P_i \quad P(\gamma_{i,k} = 0|P_i) = 1 - P_i \tag{6}$$

2. *"Horizontal" Markov models*: in order to model temporal persistency of the t-f coefficients, we give a prior horizontal structure to the indicator variable. More precisely, when a MDCT basis is used and $k = (q,p)$, for a fixed frequency index q the sequence $\{\gamma_{i,q,p}\}_{p=1,\ldots,n_{\text{frame}}}$ is modeled by a 2-state first order Markov chain with transition probabilities $P_{i,0\rightarrow 0}$ and $P_{i,1\rightarrow 1}$.

2.3.3 Noise variance prior

The noise variance σ^2 is given an inverted-Gamma (conjugate) prior $p(\sigma^2|\alpha_\sigma, \beta_\sigma) = IG(\sigma^2|\alpha_\sigma, \beta_\sigma)$.

2.3.4 Hyperparameters priors

The scale constants λ_i of each source are given independent Gamma (conjugate) priors $p(\lambda_i|\alpha_{\lambda_i}, \beta_{\lambda_i}) = G(\lambda_i|\alpha_{\lambda_i}, \beta_{\lambda_i})$, allowing an automatic adaptation to the scaling of the coefficients in each basis. The degrees of freedom α_i can be fixed to a certain value or estimated like in [3]. In our simulations the value of α_i happened to have little influence on the results and in practice we fixed it to 1. The probabilities P_i in the Bernoulli models and $P_{i,0\rightarrow 0}$, $P_{i,1\rightarrow 1}$ in the Markov models are given uniform priors on $[0,1]$, which may be routinely extended to Beta priors if required to favor certain values over others.

3. METHOD

We propose to sample from the posterior distribution of the parameters $\theta = \{\tilde{s}_i, v_i, \alpha_i, \lambda_i\}_{i=1,n} \cup \sigma^2$, using a Gibbs sampler. The Gibbs sampler is a standard Markov Chain Monte Carlo technique which simply requires to sample from the conditional distributions of each parameter upon the others [8]. Point estimates can then be computed from the obtained samples of the posterior distribution $p(\theta|\tilde{\mathbf{x}})$. In contrast with EM-like methods which aim directly at point estimates (ML or MAP), MCMC approaches are very robust because they scan the full posterior distribution and are thus unlikely to fall into local minima. This is however at the cost of a higher computational burden. We now give the expression for the update steps of the parameters. In the following most of the derivations have been skipped, further details can be found in [3, 9, 5]. Note that all the conditional posterior distributions of all the parameters can be easily sampled from.

3.1 Update of $\mathbf{A}$ and σ^2

Let $\mathbf{r}_1, \ldots, \mathbf{r}_m$ be the $n \times 1$ vectors denoting the transposed rows of $\mathbf{A}$, such that $\mathbf{A}^T = [\mathbf{r}_1 \; \ldots \; \mathbf{r}_m]$. With uninformative uniform prior $p(\mathbf{A}) \propto 1$, the rows of $\mathbf{A}$ are a posteriori mutually independent with

$$\mathbf{r}_i \sim N(\mu_{\mathbf{r}_i}, \Sigma_{\mathbf{r}}) \tag{7}$$

where $\Sigma_{\mathbf{r}} = \sigma^2\,(\sum_k \tilde{\mathbf{s}}_k\,\tilde{\mathbf{s}}_k^T)^{-1}$ and $\mu_{\mathbf{r}_i} = \frac{1}{\sigma^2}\,\Sigma_{\mathbf{r}}\,\sum_k \tilde{x}_{i,k}\,\tilde{\mathbf{s}}_k$.[2]

$\mathbf{A}$ can be integrated out in the posterior distribution of σ, resulting in

$$\sigma^2 \sim IG(\alpha_\sigma, \beta_\sigma) \tag{8}$$

with $\alpha_\sigma = \frac{(N-n)\,m}{2}$ and $2\,\beta_\sigma =$
$\sum_{j=1}^m \left(\sum_k \tilde{x}_{j,k}^2 - \left(\sum_k \tilde{x}_{j,k}\tilde{\mathbf{s}}_k^T\right)\,\left(\sum_k \tilde{\mathbf{s}}_k\tilde{\mathbf{s}}_k^T\right)^{-1}\,\left(\sum_k \tilde{x}_{j,k}\tilde{\mathbf{s}}_k\right)\right)$.

[2] In practice the columns of $\mathbf{A}$ are normalized to 1 to solve the BSS indeterminacy on gain.

3.2 Update of $(\gamma_i, \tilde{s}_i)$

Two main sampling strategies can be considered for $(\gamma_i, \tilde{s}_i)$. The first option is make block draws from the vectors $\gamma_k = [\gamma_{1,k}, \ldots, \gamma_{n,k}]^T$ and $\tilde{\mathbf{s}}_k$. The second option is to update $(\gamma_{i,k}, \tilde{s}_{i,k})$ individually, conditionally upon $(\gamma_{-i,k}, \tilde{s}_{-i,k})$, where $-i$ denotes $[\![1,n]\!] \setminus i$. In theory, the first option is better as sampling as many parameters as possible together is supposed to improve the rate of convergence of the Gibbs sampler [10]. However, in practice, the second option can lead to a faster implementation. Indeed, the first option requires 1) sampling γ_k, which requires computing the posterior probabilities of its 2^n possible values, 2) sampling $\tilde{\mathbf{s}}_k$, whose posterior distribution is multivariate Gaussian (see [3]) and thus involves inverting a $n \times n$ matrix for each $k \in [\![1,N]\!]$ and at each iteration of the Gibbs sampler. The second option involves 1) computing the two posterior probabilities $p(\gamma_{i,k} = 1|\gamma_{-i,k}, \ldots, \tilde{\mathbf{x}})$ and $p(\gamma_{i,k} = 0|\gamma_{-i,k}, \ldots, \tilde{\mathbf{x}})$, 2) sampling from $p(\tilde{s}_{i,k}|\tilde{s}_{-i,k}, \ldots, \tilde{\mathbf{x}})$, which is simply a univariate Gaussian distribution. In practice, the two latter steps can be efficiently vectorized (along k), avoiding long loops and significantly reducing the computation times (in particular when using MATLAB). In this paper we only consider the second option. To do so, note that Eq. (2) can be rewritten

$$\tilde{\mathbf{x}}_k = \tilde{s}_{i,k}\,\mathbf{a}_i + \sum_{j \neq i} \tilde{s}_{j,k}\,\mathbf{a}_j + \tilde{\mathbf{n}}_k \tag{9}$$

where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ and thus

$$\frac{\mathbf{a}_i^T\,\tilde{\mathbf{x}}_k}{\mathbf{a}_i^T\,\mathbf{a}_i} - \sum_{j \neq i} \frac{\mathbf{a}_i^T\,\mathbf{a}_j}{\mathbf{a}_i^T\,\mathbf{a}_i}\,\tilde{s}_{j,k} = \tilde{s}_{i,k} + \frac{\mathbf{a}_i^T\,\tilde{\mathbf{n}}_k}{\mathbf{a}_i^T\,\mathbf{a}_i}. \tag{10}$$

Hence, inferring $\tilde{s}_{i,k}$ conditionally upon the other source coefficients can be regarded as a simple regression problem, with unknown $\tilde{s}_{i,k}$ and data $\tilde{x}_{i|-i,k}$, where $\tilde{x}_{i|-i,k}$ denotes the left term of Eq. (10).

As pointed out in [11], an implementation of the Gibbs sampler consisting of sampling alternatively $\tilde{s}_{i,k}|\gamma_{i,k}$ and $\gamma_{i,k}|\tilde{s}_{i,k}$ cannot be used as it leads to a nonconvergent Markov chain (the Gibbs sampler gets stuck when it generates a value $\tilde{s}_{i,k} = 0$). Thus, as in [11], we jointly draw from $(\gamma_{i,k}, \tilde{s}_{i,k})$ by marginalizing $\tilde{s}_{i,k}$ from the posterior conditional distribution of $\gamma_{i,k}$, leading to

$$p(\gamma_{i,k} = 0|\sigma^2, v_{i,k}, \tilde{x}_{i|-i,k}) = 1/(1+\tau_{i,k}) \tag{11}$$

$$p(\gamma_{i,k} = 1|\sigma^2, v_{i,k}, \tilde{x}_{i|-i,k}) = \tau_{i,k}/(1+\tau_{i,k}) \tag{12}$$

with

$$\tau_{i,k} = \sqrt{\frac{\sigma^2}{\sigma^2 + v_{i,k}}}\,\exp\left(\frac{\tilde{x}_{i|-i,k}^2\,v_{i,k}}{2\,\sigma^2(\sigma^2+v_{i,k})}\right)\frac{p(\gamma_{i,k}=1|\gamma_{i,-k})}{p(\gamma_{i,k}=0|\gamma_{i,-k})} \tag{13}$$

where $\gamma_{i,-k}$ denotes the set of all indicator variables $\{\gamma_{i,l}\}_{l=1,\ldots,N}$ except $\gamma_{i,k}$. The expression of the ratio $p(\gamma_{i,k} = 1|\gamma_{i,-k})/p(\gamma_{i,k} = 0|\gamma_{i,-k})$ changes according to the chosen prior for the indicator variables. When $\gamma_{i,k}$ has a Bernoulli prior, this ratio is simply $P_i/(1-P_i)$. When $\gamma_{i,k}$ has a Markov horizontal structure and $k = (q,p)$, this ratio depends on the values of $\gamma_{i,q,p-1}$ and $\gamma_{i,q,p+1}$. The exact expressions are standard results from the Markov chain literature (see *e.g*, [12]).

The posterior distribution of $\tilde{s}_{i,k}$ is written as

$$p(\tilde{s}_{i,k}|\gamma_{i,k}, v_{i,k}, \sigma^2, \tilde{x}_{i|-i,k}) = (1-\gamma_{i,k})\,\delta_0(\tilde{s}_{i,k}) + \gamma_{i,k}\,N(\tilde{s}_{i,k}|\mu_{\tilde{s}_{i,k}}, \sigma^2_{\tilde{s}_{i,k}}) \tag{14}$$

with $\sigma^2_{\tilde{s}_{i,k}} = \left(1/\sigma^2 + 1/v_{i,k}\right)^{-1}$ and $\mu_{\tilde{s}_{i,k}} = (\sigma^2_{\tilde{s}_{i,k}}/\sigma^2)\,\tilde{x}_{i|-i,k}$.

3.2.1 Update of v_i

The conditional posterior distribution of $v_{i,k}$ is

$$p(v_{i,k}|\gamma_{i,k}, \tilde{s}_{i,k}, \alpha_i, \lambda_i) = (1-\gamma_{i,k})\,IG\left(v_{i,k}|\alpha_i, \lambda_i(q)\right) + \gamma_{i,k}\,IG\left(v_{i,k}|\frac{1}{2}+\alpha_i, \frac{\tilde{s}_{i,k}^2}{2}+\lambda_i(q)\right) \tag{15}$$

3.2.2 Update of the hyperparameters

- The posterior distribution of the scale parameters is $p(\lambda_i|v_i) = G\left(\lambda_i|N\,\alpha_i + \alpha_{\lambda_i}, \sum_k f(q)/v_{i,k} + \beta_{\lambda_i}\right)$. However, because we are looking for sparse representations, most of the indicator variables $\gamma_{i,k}$ take the value 0 and thus most of the variances $v_{i,k}$ are sampled from their prior (see Eq. (15)). Thus, the influence of the data in the full posterior distribution of λ_i becomes small, and the convergence of λ_i can be very slow. A faster scheme, employed in [13, 9], consists of making one draw from $p(\{v_{i,k} : \gamma_{i,k} = 1\}|\{\tilde{s}_{i,k} : \gamma_{i,k} = 1\}, \lambda_i, \alpha_i)$, then one draw from $p(\lambda_i|\{v_{i,k} : \gamma_{i,k} = 1\}, \alpha_i)$ and finally one draw from $p(\{v_{i,k} : \gamma_{i,k} = 0\}|\lambda_i, \alpha_i)$.
- When the indicator variables are given Bernoulli priors, the posterior distribution of P_i is simply $p(P_i|\gamma_i) = B(P_i|\#\gamma_i + 1, N - \#\gamma_i + 1)$, where $B(x|\alpha,\beta)$ is the Beta distribution defined in the Appendix and $\#\gamma_i$ is the number of values of $\gamma_{i,k}$ equal to 1. Similarly, when the indicator variables are given Markov priors, the posterior distributions of the transition probabilities can be sampled using a Metropolis-Hasting step as in [13]. In this work we simply update them as the number of transitions from 0 to 0 and 1 to 1 divided by N.

4. RESULTS

We present results for blind separation of a stereo mixture ($m = 2$) of $n = 3$ musical sources (voice, acoustic guitar, bass guitar). The sources were obtained from the BASS-dB database [14]. They consist of excerpts of original tracks from the song *Anabelle Lee* (Alex Q), published under a Creative Commons Licence. The signals are sampled at $f_s = 22.5kHz$ with length $N = 131072$ ($\approx 6s$). The mixing matrix is given in Table 1; it provides a mixture where the voice s_1 is in the middle, the acoustic guitar s_2 originates at 67.5^o on the left and the bass guitar s_3 at 67.5^o on the right. Gaussian noise was added to the observations with $\sigma = 0.01$, resulting in respectively $25dB$ and $27dB$ input SNR on each channel. We applied a MDCT to the observations using a sine bell and 50% overlap, with time resolution (half the window length) $l_{\text{frame}} = 512$ (22ms). We present the following results:

a) We apply the method in [3], in which the source coefficients are given a Student t distribution. This amounts to set $\gamma_{i,k} = 1$ for all k, but the scale parameters and the

Original matrix
$\mathbf{A} = \begin{bmatrix} 0.7071 & 0.9808 & 0.1951 \\ 0.7071 & 0.1951 & 0.9808 \end{bmatrix}$
Method (a)
$\hat{\mathbf{A}} = \begin{bmatrix} 0.7077 & 0.9810 & 0.1949 \\ (\pm 0.0037) & (\pm 0.0017) & (\pm 0.0048) \\ 0.7065 & 0.1941 & 0.9808 \\ (\pm 0.0037) & (\pm 0.0086) & (\pm 0.0009) \end{bmatrix}$
Method (b)
$\hat{\mathbf{A}} = \begin{bmatrix} 0.7070 & 0.9811 & 0.1946 \\ (\pm 0.0035) & (\pm 0.0012) & (\pm 0.0046) \\ 0.7073 & 0.1935 & 0.9809 \\ (\pm 0.0035) & (\pm 0.0060) & (\pm 0.0009) \end{bmatrix}$
Method (c)
$\hat{\mathbf{A}} = \begin{bmatrix} 0.7044 & 0.9821 & 0.1943 \\ (\pm 0.0039) & (\pm 0.0023) & (\pm 0.0060) \\ 0.7098 & 0.1881 & 0.9809 \\ (\pm 0.0039) & (\pm 0.0118) & (\pm 0.0012) \end{bmatrix}$
Method (d)
$\hat{\mathbf{A}} = \begin{bmatrix} 0.7079 & 0.9811 & 0.1946 \\ (\pm 0.0003) & (\pm 0.0001) & (\pm 0.0006) \\ 0.7064 & 0.1933 & 0.9809 \\ (\pm 0.0003) & (\pm 0.0007) & (\pm 0.0001) \end{bmatrix}$

Table 1: Estimates of **A**.

degrees of freedom are both updated. The sources are updated with block draws of $\tilde{\mathbf{s}}_k$. Using a MATLAB implementation running on a 1.25GHz Powerbook G4, 1000 iterations of the sampler take 6.6 hours. Approximate convergence was usually observed after 1500 iterations.

b) We apply the approach of a), but the sources are updated one by one, conditionally upon the others. 1000 iterations of the sampler take 1.1 hours. Approximate convergence was usually observed after 2000 iterations.

c) We apply the method described in this paper, with a Bernoulli prior on $\gamma_{i,k}$. 1000 iterations of the sampler take 50min. Approximate convergence was usually observed after 5000 iterations.

d) We apply the method described in this paper, with horizontal Markov priors on $\gamma_{i,k}$. The computational burden is unchanged and 1000 iterations of the sampler still take 50min. Approximate convergence was usually observed after 5000 iterations. Fig. 1 shows the sampled values of the parameters over 10000 iterations of the sampler.

In the four cases the values of α_σ, β_σ, α_{λ_i}, β_{λ_i} were chosen as to yield Jeffreys noninformative priors, **A** was initialised to $[1\ 1\ 1;\ 0\ 0\ 0]$, $\tilde{s}_i$ to $\tilde{x}_1/3$, v_i to ones, λ_i to 0.1. The samplers were run for 2500 iterations in case (a) and for 10000 iterations in the other cases. σ^2 was annealed to its true posterior distribution during the first 500 iterations in case (a) and during the first 1000 iterations in the other cases (see [3]). Minimum Mean Square Error estimates of the source coefficients were computed in each case by averaging the last 1000 samples. Table 2 presents separation evaluation criteria for the estimated sources in each case. The criteria are described in [15], but basically, the SDR (Source to Distortion Ratio) provides an overall separation performance criterion, the SIR (Source to Interferences Ratio) measures the level of interferences from the other sources in each source estimate, SNR (Source to Noise Ratio) measures the error due to the additive noise on the sensors and the SAR (Source to Artifacts Ratio) measures the level of artifacts in the source estimates. Source estimates can be listened to at `http://www-sigproc.eng.cam.ac.uk/~cf269/eusipco06/`, which is perhaps the best way to assess the quality of the results. Fig. 2 presents the *significance maps* of the source coefficients, i.e the Maximum A Posteriori estimates of γ_1, γ_2 and γ_3, in the Bernoulli and Markov cases.

Method	$\hat{s}_1$ (voice)			
	SDR	SIR	SAR	SNR
a)	4.3	14.7	4.9	28.9
b)	-4.0	7.6	-3.0	24.2
c)	0.1	5.6	2.6	28.4
d)	-0.75	12.0	-0.23	28.3

Method	$\hat{s}_2$ (acoustic guitar)			
	SDR	SIR	SAR	SNR
a)	6.0	17.4	6.4	28.8
b)	1.7	6.9	4.1	27.9
c)	0.7	7.7	2.4	27.4
d)	3.1	10.3	4.5	38.6

Method	$\hat{s}_3$ (bass guitar)			
	SDR	SIR	SAR	SNR
a)	10.7	22.2	11.1	39.8
b)	5.7	11.4	7.4	38.9
c)	5.9	14.4	6.7	51.5
d)	7.4	15.1	8.3	50.1

Table 2: Performance criteria.

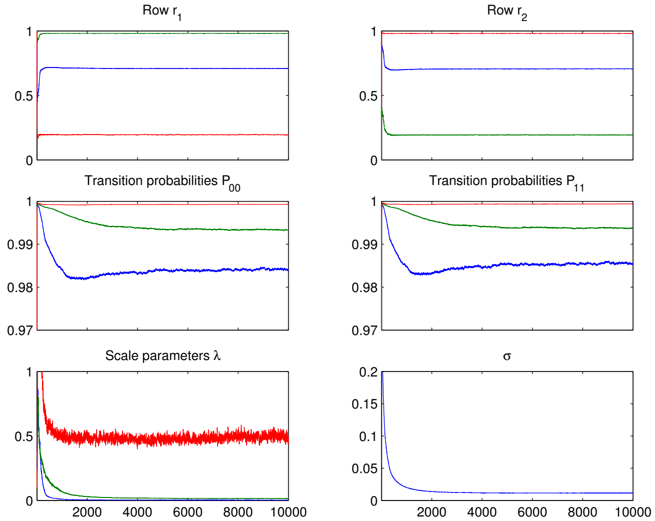

Figure 1: Gibbs sampler updates of the various model parameters; blue = source 1, green = source 2, red = source 3.

5. CONCLUSIONS

We have described in this paper a Bayesian approach to source separation in which the source coefficients in a transform domain are given an exact sparse prior: conditionally upon an indicator variable, the coefficients are either set to zero or given a hierarchical prior. The advantage of this framework over other sparse priors is the ability to favor structures in the time-frequency plane by choosing relevant priors for the indicator variables. Our method is also completely adaptive, none of the model parameters need to be trained.

An interesting issue of this paper is the individual update of each source conditionally upon the others, as compared to

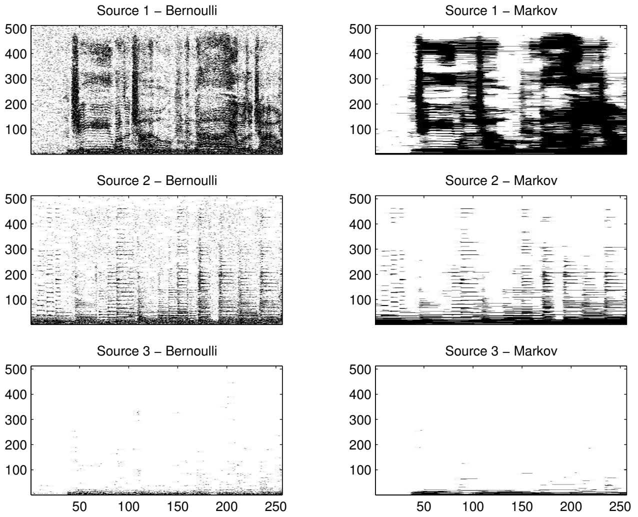

Figure 2: Significance maps of the estimated sources, obtained with Bernoulli priors (left) and horizontal markovian priors (right).

block draws. Though the two methods should theoretically yield similar source estimates after a "large enough" number of iterations, in practice, over an horizon of 10000 iterations, method (a) still yields better estimates, in particular in terms of SIRs. We believe this is because the individual update of each source conditionally upon the others creates some correlation between the sources. If the amount of correlation should theoretically fade away when averaging a large number of samples, well after the burn-in period, in practice this seems to be a problem over our limited horizon. We also noticed that, depending on the initializations and the random sequence seeds, method (b) could get stuck for long periods in some irrelevant areas of the posterior distribution of the mixing matrix, and that full exploration of the posterior could be tedious. In contrast, method (a) reliably explores the modal areas of $p(\mathbf{A}|\tilde{\mathbf{x}})$, and convergence is rather fast when σ^2 is annealed.

Table 1 shows that the four methods give very good estimates of the mixing matrix, with best results obtained with method (d). Table 2 shows that methods (c) and (d), described in this paper, do not beat method (a) in terms of SIRs and SARs, but they yield source estimates with subjectively good audio properties. They perform very well in terms of denoising (see the SNRs), in particular for the acoustic and bass guitars. Poorer results are obtained for the voice with method (d) rather than (c) because horizontal Markov structures are better suited for the latter instruments rather than for voice (which is more inharmonic).

Future work will involve building a framework allowing efficient block draws of $(\gamma_k, \tilde{\mathbf{s}}_k)$ as well as using audio models involving overcomplete dictionaries, in the fashion of [5].

A. STANDARD DISTRIBUTIONS

Normal	$N(x\|u,\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp -\frac{(x-u)^2}{2\sigma^2}$
Beta	$B(x\|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}\,(1-x)^{\beta-1}, \quad x \in [0,1]$
Gamma	$G(x\|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1} \exp(-\beta x), x \geq 0$
inv-Gamma	$IG(x\|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{-(\alpha+1)} \exp(-\frac{\beta}{x}), x \geq 0$

The inverted-Gamma distribution is the distribution of $1/X$ when X is Gamma distributed.

REFERENCES

[1] M. Zibulevsky, B. A. Pearlmutter, P. Bofill, and P. Kisilev. Blind source separation by sparse decomposition. In S. J. Roberts and R. M. Everson, editors, *Independent Component Analysis: Principles and Practice*. Cambridge University Press, 2001.

[2] A. Jourjine, S. Rickard, and O. Yilmaz. Blind separation of disjoint orthogonal signals: Demixing n sources from 2 mixtures. In *Proc. ICASSP*, volume 5, pages 2985–2988, Istanbul, Turkey, Jun. 2000.

[3] C. Févotte and S. J. Godsill. A Bayesian approach for blind separation of sparse sources. *IEEE Trans. Speech and Audio Processing*, In press. Preprint available at `http://www-sigproc.eng.cam.ac.uk/˜cf269/`.

[4] C. Févotte and S. J. Godsill. A bayesian approach to time-frequency based blind source separation. In *Proc. 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'05), Mohonk, NY, Oct. 2005*, Mohonk, NY, Oct 2005.

[5] C. Févotte, L. Daudet, S. J. Godsill, and B. Torrésani. Sparse regression with structured priors: Application to audio denoising. In *Proc. ICASSP'06*, Toulouse, France, 2006.

[6] R. Balan and J. Rosca. Sparse source separation using discrete prior models. In *Proc. Workshop on Signal Processing with Adaptative Sparse Structured Representations (SPARS'05)*, Rennes, France, Nov. 2005.

[7] E. Vincent. Musical source separation using time-frequency source priors. *Trans. on Speech and Audio Processing*, (Special issue on Statistical and Perceptual Audio Processing), In press.

[8] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov 1984.

[9] C. Févotte and S. J. Godsill. Sparse linear regression in unions of bases via Bayesian variable selection. *IEEE Signal Processing Letters*, 2005. In press - Preprint available at `http://www-sigproc.eng.cam.ac.uk/˜cf269/`.

[10] J. S. Liu, W. H. Wong, and A. Kong. Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, 81(1):27–40, Mar. 1994.

[11] J. Geweke. *Variable selection and model comparison in regression*, pages 609–620. Oxford Press, 5 edition, 1996. Edited by J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Swith.

[12] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, Feb. 1989.

[13] P. J. Wolfe, S. J. Godsill, and W.-J. Ng. Bayesian variable selection and regularisation for time-frequency surface estimation. *J. R. Statist. Soc. Series B*, 2004.

[14] E. Vincent, R. Gribonval, C. Févotte, and al. BASS-dB: the blind audio source separation evaluation database. Available on-line. `http://www.irisa.fr/metiss/BASS-dB/`.

[15] E. Vincent, R. Gribonval, and C. Févotte. Performance measurement in blind audio source separation. *IEEE Trans. Speech and Audio Processing*. In press - Preprint available at `http://www-sigproc.eng.cam.ac.uk/˜cf269/`.