A BAYESIAN APPROACH TO TIME-FREQUENCY BASED BLIND SOURCE SEPARATION

Cédric Févotte and Simon J. Godsill*

Signal Processing Lab
Cambridge University Engineering Dept
Cambridge CB2 1PZ, UK
cf269@cam.ac.uk

ABSTRACT

In this paper we propose a bayesian approach for time-frequency (t-f) based source separation. We propose a Gibbs sampler, a standard Markov Chain Monte Carlo (MCMC) simulation method, to sample from the mixing matrix, the source t-f coefficients and the input noise variance, under two models for the sources. In the first one the t-f coefficients of the sources are assumed i.i.d, while a frequency dependent modeling of the coefficients is proposed in the second one, which provides improved interference and noise rejection. Audio results are presented over several time resolutions of the t-f transform.

1. INTRODUCTION

Blind Source Separation (BSS) consists in estimating n signals (the sources) from the sole observation of m mixtures of them (the observations). 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.

In [1, 2], we described a Bayesian approach to source separation using the assumption of source sparsity on a chosen basis. Sparsity means that only "few" expansion coefficients of the sources on the basis are significantly different from zero and its use to handle source separation problem (possibly underdetermined) was introduced in the seminal papers [3, 4]. In [1, 2] we modeled the expansion coefficients of the sources on the chosen basis by identically and independently distributed (i.i.d) Student t processes and a Gibbs sampler (a standard MCMC simulation method) 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. The method was successfully applied to determined and underdetermined noisy audio mixtures, decomposed on a MDCT basis (a local cosine basis).

In this paper, we propose two contributions to the latter work. In the first half of this paper we study the impact of time resolution of the MDCT on the separation results and we provide many audio examples. The i.i.d Student t model of the sources coefficients is however shown to have rather limited denoising performance. To deal with this issue we propose in the second half of this paper an improved model of the sources coefficients which takes into account the non-uniformity of the energy distribution of audio signals along frequency. The proposed model consists of modeling each source frequency subband by a Student t i.i.d process with its own (frequency dependent) scale parameter. This model is shown to give better Signal to Interference and Signal to Noise Ratios of the source estimates.

The paper is organized as follows: Section 2 introduces notations and the dual problem of source separation in the t-f plane. Section 3 briefly summarizes the Gibbs sampler proposed in [1, 2] for the i.i.d source coefficients model, and provides separation results on a 2×3 mixture for several time resolutions of the t-f transform. Section 4 presents the improved model of the sources and illustrates its better separation performances in terms of interference and noise rejection. Finally Section 5 draws conclusions and perspectives.

*Cédric Févotte acknowledges support of the European Commission funded Research Training Network HASSIP (HPRN-CT-2002-00285)

2. NOTATIONS

2.1. Mixture and aim

We consider the following standard linear instantaneous model, $\forall t = 1, \ldots, N$:

$$\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

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 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 $\hat{\mathbf{A}} = \mathbf{A}\,\mathbf{D}\,\mathbf{P}$ 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

We propose to solve the problem defined by Eq. (1) using a t-f model of the sources. Let $x \in \mathbb{R}^{1\times N} \rightarrow \tilde{x} \in \mathbb{R}^{l_{frame}\times n_{frame}}$ denote a bijective linear t-f transform, preferably orthonormal, with time resolution l_{frame}/f_s and frequency resolution $\frac{f_s}{2}/l_{frame}$, where f_s is the sampling frequency and $n_{frame} = N/l_{frame}$. Such transforms includes the Modulated Lapped Transforms family to which the Modified Discrete Cosine Transform (MDCT) belongs (see [5] for a survey). Denoting $\forall (q,k) \in [\![1, l_{frame}]\!] \times [\![1, n_{frame}]\!]$, $\tilde{\mathbf{x}}(q,k) = [\tilde{x}_1(q,k), \ldots, \tilde{x}_m(q,k)]^T$ and $\tilde{\mathbf{n}}(q,k)$, $\tilde{\mathbf{s}}(q,k)$ similarly, by linearity of the t-f transform we have

$$\tilde{\mathbf{x}}(q,k) = \mathbf{A}\,\tilde{\mathbf{s}}(q,k) + \tilde{\mathbf{n}}(q,k) \tag{2}$$

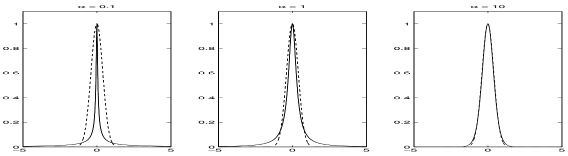

Figure 1: Student *t* densities for $\alpha \in \{0.1, 1, 10\}$ with equal value at the mode - The dash-lined plot is the Gaussian density with variance $1/2\,\pi$.

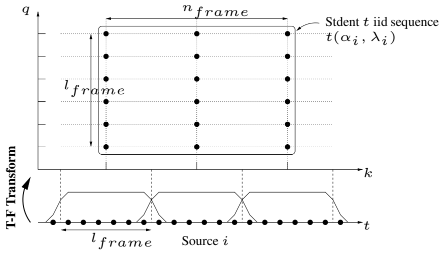

Figure 2: i.i.d model of the sources

Furthermore, the t-f transform being bijective, solving the problem defined by Eq. (1) in the time domain is equivalent to solving Eq. (2) in the transform domain.

3. I.I.D MODEL OF THE SOURCES

In [1, 2] we exploited sparsity of the t-f coefficients of the sources: $\forall i$, the coefficients $\{\tilde{s}_i(q,k), (q,k) \in [\![1, l_{frame}]\!] \times [\![1, n_{frame}]\!]\}$ were modeled by a Student t i.i.d sequence $t(\alpha, \lambda)$ with density given by

$$t(x|\alpha,\lambda) = \frac{\Gamma(\frac{\alpha+1}{2})}{\lambda\sqrt{\alpha\,\pi}\,\Gamma(\frac{\alpha}{2})}\left(1+\frac{1}{\alpha}\left(\frac{x}{\lambda}\right)^2\right)^{-\frac{\alpha+1}{2}} \quad (3)$$

where α is the degree of freedom and λ is a scale parameter. Fig. 1 shows that for small α, the Student *t* density gathers most of its probability mass around zero and exhibits "fatter tails" than the normal distribution. The Student t distribution is thus a relevant model for sparsity. A useful feature of the Student t distribution is its formulation as a Scale Mixture of Gaussians [6]: $x \sim t(\alpha, \lambda)$ can be drawn by 1) sampling a variance $v \sim \mathcal{IG}\left(v|\frac{\alpha}{2}, \frac{2}{\alpha\,\lambda^2}\right)$, 2) sampling $x \sim \mathcal{N}(0, v)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution and $\mathcal{IG}(\gamma, \beta)$ is the inverted-Gamma distribution defined by $\mathcal{IG}(x|\gamma,\beta) = \frac{x^{-(\gamma+1)}}{\Gamma(\gamma)\,\beta^\gamma}\exp(-\frac{1}{\beta\,x})\,\mathbb{I}_{[0,+\infty)}(x)$.[1]

The source model is illustrated on Fig 2. We furthermore assumed the sequences of coefficients of the several sources to be mutually independent, and assumed the noise $\tilde{\mathbf{n}}(q,k)$ to be i.i.d Gaussian with power $\sigma^2\,\mathbf{I}_m$. A Gibbs sampler was presented to sample from the posterior distribution $p(\boldsymbol{\theta}|\tilde{\mathbf{x}})$ of the set of parameters $\boldsymbol{\theta} = \{\mathbf{A}, \sigma^2, \tilde{\mathbf{s}}, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}\}$, where $\mathbf{v}$ contains the variances of the source coefficients in their Scale Mixture of Gaussians, and where $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_n\}$ and $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_n\}$. Minimum Mean Square Estimates of the parameters are computed by averaging the samples obtained from the Gibbs sampler after the burnin period, time-domain source estimates are reconstructed by inverse transform of the estimated coefficients.

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

Initialize $\boldsymbol{\theta}^{(0)} = \{\tilde{\mathbf{s}}^{(0)}, \mathbf{A}^{(0)}, \sigma^{(0)}, \mathbf{v}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\lambda}^{(0)}\}$
for $k = 1 : K + K_{Burnin}$ **do**
  $\mathbf{A}^{(k)} \sim p(\mathbf{A}|\tilde{\mathbf{s}}^{(k-1)}, \sigma^{(k-1)}, \tilde{\mathbf{x}})$
  $\sigma^{(k)} \sim p(\sigma|\tilde{\mathbf{s}}^{(k-1)}, \tilde{\mathbf{x}})$
  $\tilde{\mathbf{s}}^{(k)} \sim p(\tilde{\mathbf{s}}|\mathbf{A}^{(k)}, \sigma^{(k)}, \mathbf{v}^{(k-1)}, \tilde{\mathbf{x}})$
  $\mathbf{v}^{(k)} \sim p(\mathbf{v}|\tilde{\mathbf{s}}^{(k)}, \boldsymbol{\alpha}^{(k-1)}, \boldsymbol{\lambda}^{(k-1)})$
  $\boldsymbol{\alpha}^{(k)} \sim p(\boldsymbol{\alpha}|\mathbf{v}^{(k)}, \boldsymbol{\lambda}^{(k-1)})$
  $\boldsymbol{\lambda}^{(k)} \sim p(\boldsymbol{\lambda}|\mathbf{v}^{(k)}, \boldsymbol{\alpha}^{(k)})$
end for

Table 1: Gibbs sampler for source separation of linear instantaneous mixtures.

3.1. Gibbs sampler

We recall in this section the principal steps of the Gibbs sampler presented in [1, 2]. The sampler requires sampling from the distributions of each (subset of) parameters conditionally upon the others and the data. The steps are summarized in Table 1 and the conditional distributions are given below.

• *Sampling* $\mathbf{A}$ *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 (Jeffrey's) uninformative uniform prior $p(\mathbf{A}) \propto 1$, the rows of $\mathbf{A}$ are a posteriori mutually independent with

$$\mathbf{r}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{r}_i}, \boldsymbol{\Sigma}_{\mathbf{r}}) \quad (4)$$

where $\boldsymbol{\Sigma}_{\mathbf{r}} = \sigma^2\left(\sum_{q,k}\tilde{\mathbf{s}}(q,k)\,\tilde{\mathbf{s}}(q,k)^T\right)^{-1}$ and $\boldsymbol{\mu}_{\mathbf{r}_i} = \frac{1}{\sigma^2}\,\boldsymbol{\Sigma}_{\mathbf{r}}\sum_{q,k}\tilde{x}_i(q,k)\,\tilde{\mathbf{s}}(q,k)$. [2]

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

$$\sigma^2 \sim \mathcal{IG}(\alpha_\sigma, \beta_\sigma) \quad (5)$$

with $\alpha_\sigma = \frac{(N-n)\,m}{2}$ and $2/\beta_\sigma = \sum_{i=1}^m\left(\sum_{q,k}\tilde{x}_i^2(q,k)\right) - \left(\sum_{q,k}\tilde{x}_i(q,k)\tilde{\mathbf{s}}(q,k)^T\right)\left(\sum_{q,k}\tilde{\mathbf{s}}(q,k)\tilde{\mathbf{s}}(q,k)^T\right)^{-1} \times \left(\sum_{q,k}\tilde{x}_i(q,k)\tilde{\mathbf{s}}(q,k)\right)$.

• *Sampling* $\tilde{\mathbf{s}}$: $\forall (q,k) \in [\![1, l_{frame}]\!] \times [\![1, n_{frame}]\!]$

$$\tilde{\mathbf{s}}(q,k) \sim \mathcal{N}(\boldsymbol{\mu}_{\tilde{\mathbf{s}}(q,k)}, \boldsymbol{\Sigma}_{\tilde{\mathbf{s}}(q,k)}) \quad (6)$$

where $\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}(q,k)} = \left(\frac{1}{\sigma^2}\,\mathbf{A}^T\,\mathbf{A} + \text{diag}\,(\mathbf{v}(q,k))^{-1}\right)^{-1}$ and $\boldsymbol{\mu}_{\tilde{\mathbf{s}}(q,k)} = \frac{1}{\sigma^2}\,\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}(q,k)}\,\mathbf{A}^T\,\tilde{\mathbf{x}}(q,k)$ (and where diag $(\mathbf{u})$ is the diagonal matrix whose main diagonal is given by $\mathbf{u}$).

• *Sampling* $\mathbf{v}$: Since the likelihood does not depend on the parameters $\{\mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}\}$, their posterior distributions are conditionally independent of $\tilde{\mathbf{x}}$. $\forall (q,k) \in [\![1, l_{frame}]\!] \times [\![1, n_{frame}]\!]$

$$v_i(q,k) \sim \mathcal{IG}(\gamma_{v_i}, \beta_{v_i}(q,k)) \quad (7)$$

with $\gamma_{v_i} = (\alpha_i + 1)/2$ and $\beta_{v_i}(q,k) = 2/(\tilde{s}_i^2(q,k) + \alpha_i\,\lambda_i^2)$.

• *Sampling* $\boldsymbol{\lambda}$: With the uninformative Jeffreys prior $p(\lambda_i) = 1/\lambda_i$, we have:

$$\lambda_i^2 \sim \mathcal{G}\left(\gamma_{\lambda_i}, \beta_{\lambda_i}\right) \quad (8)$$

with $\gamma_{\lambda_i} = (\alpha_i\,N)/2$ and $\beta_{\lambda_i} = 2/(\alpha_i\,R_i)$, where $R_i = \sum_{q,k} 1/v_i(q,k)$ and $\mathcal{G}(\gamma, \beta)$ is the Gamma distribution defined by $\mathcal{G}(x|\gamma,\beta) =$

[2]In practice $\mathbf{r}_1$ is clamped to ones to solve the BSS indeterminacy on gain.

$\frac{x^{\gamma-1}}{\Gamma(\gamma)\,\beta^{\gamma}}\,\exp(-\frac{x}{\beta})\,\mathbb{I}_{[0,+\infty)}(x)$.

• *Sampling* $\boldsymbol{\alpha}$: We have

$$p(\alpha_i|\boldsymbol{\theta}_{-\boldsymbol{\alpha}},\tilde{\mathbf{x}}) \propto \frac{P_i^{-(\frac{\alpha_i}{2}+1)}}{\Gamma(\frac{\alpha_i}{2})^N}\left(\frac{\alpha_i\,\lambda_i^2}{2}\right)^{\frac{\alpha_i\,N}{2}} \exp\left(-\frac{\alpha_i\,\lambda_i^2}{2}\,R_i\right)\,p(\alpha_i) \tag{9}$$

with $P_i = \prod_{q,k} v_i(q,k)$. The conditional posterior distribution of $\boldsymbol{\alpha}$ is not straightforward to sample from but since the precise value of α_i for each source is unlikely to be important provided it is within an appropriate small range, in practice we sample $\boldsymbol{\alpha}$ from a uniform grid of discrete values with probability masses proportional to Eq. (9) (with the uninformative uniform prior $p(\boldsymbol{\alpha}) \propto 1$). However, the Student t being a special case of the Generalized Hyperbolic distribution, sampling from α_i can also be done exactly by rejection sampling as in [7].

3.2. Results

We study in this section the quality of the separation results over several time resolutions l_{frame}. We apply the previous Gibbs sampler to a mixture of $n = 3$ sources with $m = 2$ observations. The sources are $\approx 8s$ audio signals sampled at $8kHz$ ($N = 65536$). s_1 is an excerpt of speech, s_2 an excerpt of piano solo and s_3 an excerpt of rhythmic guitar. All sound samples, including sources, observations and source estimates can be listened to at `http://www-sigproc.eng.cam.ac.uk/~cf269/waspaa05_2/sound_files`. The mixing matrix is chosen as

$$\mathbf{A} = \begin{bmatrix} 0.7071 & 0.9239 & 0.3827 \\ 0.7071 & 0.3827 & 0.9239 \end{bmatrix} \tag{10}$$

which, if x_1 and x_2 are respectively the left and right channels of a stereo excerpt, provides a mixture where the speech s_1 is in the middle, the piano s_2 originates at 45^o on the left and the guitar s_3 at 45^o on the right. Gaussian noise was added to the observations with $\sigma = 0.05$, resulting in $\approx 12dB$ input SNR on each channel.

The observations were decomposed on a MDCT basis with a sine bell analysis window, and several time resolutions were tried. Table 2 presents separation evaluation criteria for the sources for 4 different values of l_{frame}, ranging from $32ms$ to $256ms$. The criteria are described in [8], 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.

Table 2 also shows the estimated values of the degrees of freedom α estimated by the Gibbs sampler (and sampled from a grid of linearly spaced values between 0.5 and 5 with step size 0.1), together with an index of sparsity computed on the *original* source coefficients sequences. The index of sparsity was chosen as $c_{\tilde{x}} = \|\tilde{x}\|_1/\|\tilde{x}\|_2$. [3] The smaller $c_{\tilde{x}}$ is, the sparser $\tilde{x}$ is.

3.3. Discussion

Table 2 shows that the best separation results are obtained for time resolutions of $64ms$ and $128ms$. Table 2 also shows that the correlation between the sparsity index and the separation quality is not completely straightforward, in particular for the speech. This suggests that the specifications of the basis are to be chosen accordingly to the nature of the signals. In every case, estimated values of α are near one, which corresponds to the Cauchy distribution. We ran the Gibbs sampler by clamping α to 1 for every source and very similar results were obtained (with generally less than $1dB$ difference on the SDRs). Clamping α to 1 has the advantage of accelerating the convergence of the Gibbs sampler and also the update step of the variances $v_i(q,k)$. Indeed, when $\alpha = 1$, the inverted-Gamma distribution can be efficiently sampled as $1/X$ where X is a simple exponential distribution (while the general Gamma distribution usually involves rejection sampling). However it decreases the total computational burden involved by the Gibbs sampler by only 14%, and 1000 iterations of the sampler still take $2.5h$ on a Mac G4 cadenced at 1.25GHz (and convergence was usually obtained after $\approx$ 1500 iterations). The bottleneck of the algorithm is the update step of the sources which requires sampling N times from a multivariate Gaussian vector of size n at each iteration.

From a subjective point of view, the sound quality of the source estimates obtained for the various time resolutions appears quite similar, except for $256ms$ because of the presence of (pre)echo on the speech. It appears that the estimates are contaminated by disturbing remains of the additive noise on the observations. We believe that the reason for this comes from the scale parameter λ being chosen equal for all the frequencies subbands in the modeling of the sources, which thus emphasizes additive noise in the high frequencies domain, where the energy of the sources is normally low. To alleviate this problem, we propose in the next section a frequency-dependent model of the sources.

[3] Note that since the MDCT is an orthonormal transformation, the norm-2 of $\tilde{x}$ is identical whatever time resolution is used.

	$\hat{s}_1$ (speech)					
l_{frame}	SDR	SIR	SAR	SNR	$\hat{\alpha}$	c_{s_1}
$32ms$	1.6	9.2	3.1	21.8	1.1	63.5
$64ms$	3.1	13.6	3.9	21.0	1.2	64.8
$128ms$	3.2	14.1	3.9	19.8	1.2	70.3
$256ms$	-0.3	8.6	1.0	19.0	1.2	78.0
	$\hat{s}_2$ (piano)					
l_{frame}	SDR	SIR	SAR	SNR	$\hat{\alpha}$	c_{s_2}
$32ms$	8.5	18.9	9.5	18.6	1.2	60.1
$64ms$	10.0	20.5	10.9	20.8	1.1	50.1
$128ms$	10.2	21.9	10.9	21.7	1.2	44.1
$256ms$	7.1	20.8	7.5	21.1	1.2	42.3
	$\hat{s}_3$ (guitar)					
l_{frame}	SDR	SIR	SAR	SNR	$\hat{\alpha}$	c_{s_3}
$32ms$	7.5	16.9	8.7	17.6	1.0	74.7
$64ms$	8.7	19.8	9.8	18.3	1.1	64.6
$128ms$	9.0	20.0	9.8	20.0	1.0	59.7
$256ms$	6.2	16.0	7.0	19.8	1.1	59.3

Table 2: Performance criteria for the i.i.d model.

4. FREQUENCY-DEPENDENT MODEL

Instead of modeling each sequence $\{\tilde{s}_i(q,k), (q,k) \in [\![1, l_{frame}]\!] \times [\![1, n_{frame}]\!]\}$ by a i.i.d sequence with a fixed scale parameter λ_i, we now propose to model each source frequency subband q $\{s_i(q,k), k \in [\![1, n_{frame}]\!]\}$ by a i.i.d Student t distributed sequence with its own scale parameter $\lambda_i(q)$ and degree of freedom $\alpha_i(q)$. The model is illustrated on Fig. 3.

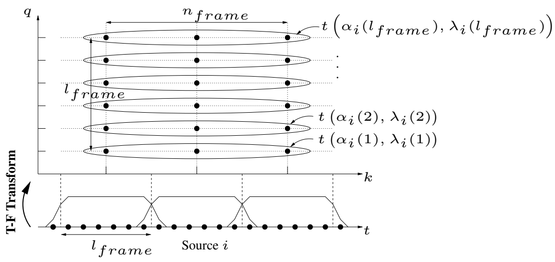

Figure 3: Frequency-dependent scale model of the sources

4.1. Gibbs sampler

The sampling steps for $\mathbf{A}$, σ and $\mathbf{s}$ remain unchanged, but one now has to update l_{frame} scale parameters and degrees of freedom for each source, and γ_{v_i} is now frequency dependent. This is of course at the cost of extra computation, but is still reasonable compared to the update of $\tilde{\mathbf{s}}$.

• *Sampling* $\mathbf{v}$: The variances of the sources coefficients are now sampled as:

$$v_i(q,k) \sim \mathcal{IG}(\gamma_{v_i}(q), \beta_{v_i}(q,k)) \tag{11}$$

with $\gamma_{v_i}(q) = (\alpha_i(q)+1)/2$ and $\beta_{v_i}(q,k) = 2/(\tilde{s}_i^2(q,k) + \alpha_i(q)\,\lambda_i(q)^2)$.

• *Sampling* $\boldsymbol{\lambda}$: With the uninformative Jeffreys prior $p(\lambda_i(q)) = 1/\lambda_i$, we have, $\forall q \in [\![1, l_{frame}]\!]$:

$$\lambda_i^2(q) \sim \mathcal{G}\left(\gamma_{\lambda_i}(q), \beta_{\lambda_i}(q)\right) \tag{12}$$

with $\gamma_{\lambda_i}(q) = (\alpha_i(q)\, n_{frame})/2$ and $\beta_{\lambda_i}(q) = 2/(\alpha_i(q)\, R_i(q))$, where $R_i(q) = \sum_{k=1}^{n_{frame}} 1/v_i(q,k)$.

• *Sampling* $\boldsymbol{\alpha}$: We have, $\forall q \in [\![1, l_{frame}]\!]$:

$$p(\alpha_i(q)|\boldsymbol{\theta}_{-\boldsymbol{\alpha}}, \tilde{\mathbf{x}}) \propto$$
$$\frac{P_i(q)^{-(\frac{\alpha_i(q)}{2}+1)}}{\Gamma(\frac{\alpha_i(q)}{2})^{n_{frame}}} \left(\frac{\alpha_i(q)\,\lambda_i(q)^2}{2}\right)^{\frac{\alpha_i(q)\, n_{frames}}{2}} \exp\left(-\frac{\alpha_i(q)\,\lambda_i(q)^2}{2} R_i(q)\right)$$

with $P_i(q) = \prod_{k=1}^{n_{frames}} v_i(q,k)$ and where we chose $p(\alpha_i(q)) \propto 1$. As before, we sample $\alpha_i(q)$ from a grid of discrete values.

4.2. Results and discussion

The Gibbs sampler is applied to the same mixture as in Section 3.2, and several time resolutions were also tried. The results are given in Table 3. Table 3 shows the important improvements brought by the frequency-dependent scale modeling of the sources when comparing the values with those of Table 2. All the SDRs values were improved by at least $1dB$ (which mostly reflects the improvement by at least $1dB$ of the SARs), but most importantly, the SIRs for the speech were all improved by $\approx 5dB$ and the SNRs for the piano and the guitar were nearly all improved by $4dB$. These improvements can be clearly heard when listening to the sound samples, and from a subjective point of view sound quality is much better than with the i.i.d model. Best results are obtained with $l_{frame} = 64ms$.

	$\hat{s}_1$			
l_{frame}	SDR	SIR	SAR	SNR
$32ms$	4.1	14.1	4.9	20.7
$64ms$	4.6	19.4	5.0	20.9
$128ms$	4.1	18.9	4.5	20.8
$256ms$	3.5	19.2	3.8	20.5
	$\hat{s}_2$			
	SDR	SIR	SAR	SNR
$32ms$	10.8	19.4	11.9	22.6
$64ms$	11.5	20.8	12.4	24.3
$128ms$	11.2	23.4	11.7	25.7
$256ms$	10.9	21.4	11.6	25.6
	$\hat{s}_3$			
	SDR	SIR	SAR	SNR
$32ms$	9.8	18.6	11.0	20.5
$64ms$	10.4	19.9	11.3	22.1
$128ms$	10.1	21.2	10.8	22.5
$256ms$	10.0	19.4	10.8	23.0

Table 3: Performance criteria for the frequency dependent model.

5. CONCLUSIONS

In this paper we described a Bayesian approach to t-f based source separation. We illustrated how time resolution relates to the quality of the separation and then proposed a frequency dependent scale model of the sources which significantly improves the results obtained with the i.i.d model.

Further work will include the comparison of the MDCT basis with other bases, such as wavelets bases. An interesting line of research will be the extension of the current work to the use of overcomplete dictionaries (such as unions of basis). In order to further improve the models of the sources, we would also like to model persistencies over time and correlations over frequencies of the t-f coefficients.

6. REFERENCES

[1] C. Févotte, S. J. Godsill, and P. J. Wolfe, "Bayesian approach for blind separation of underdetermined mixtures of sparse sources," in *Proc. 5th International Conference on Independent Component Analysis and Blind Source Separation (ICA 2004)*, Granada, Spain, 2004, pp. 398–405.

[2] C. Févotte and S. J. Godsill, "A bayesian approach for blind separation of sparse sources," *Technical Report of Cambridge University Engineering Dept. CUED/F-INFENG/TR.511*, Jan. 2005, http://www-sigproc.eng.cam.ac.uk/~cf269/TechRep/techrep05.pdf.

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

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

[5] S. Mallat, *A wavelet tour of signal processing*. Academic Press, 1998.

[6] D. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *J. R. Statist. Soc. Series B*, vol. B, no. 36, pp. 99–102, 1974.

[7] H. Snoussi and J. Idier, "Blind separation of generalized hyperbolic processes: Unifying approach to stationary non gaussianity and gaussian non stationarity," in *Proc. ICASSP*, Philadelphia, Pennsylvania, Mar. 2005.

[8] R. Gribonval, L. Benaroya, E. Vincent, and C. Févotte, "Proposals for performance measurement in source separation," in *Proc. 4th Symposium on Independent Component Analysis and Blind Source Separation (ICA'03)*, Nara, Japan, Apr. 2003.