MULTICHANNEL HR-NMF FOR MODELLING CONVOLUTIVE MIXTURES OF NON-STATIONARY SIGNALS IN THE TIME-FREQUENCY DOMAIN

Roland Badeau[*]

Institut Mines-Telecom
Telecom ParisTech
CNRS LTCI

Mark D. Plumbley[†]

Centre for Digital Music, School of
Electronic Engineering & Computer Science
Queen Mary University of London

ABSTRACT

Several probabilistic models involving latent components have been proposed for modelling time-frequency (TF) representations of audio signals (such as spectrograms), notably in the nonnegative matrix factorization (NMF) literature. Among them, the recent high resolution NMF (HR-NMF) model is able to take both phases and local correlations in each frequency band into account, and its potential has been illustrated in applications such as source separation and audio inpainting. In this paper, HR-NMF is extended to multichannel signals and to convolutive mixtures. A fast variational expectation-maximization (EM) algorithm is proposed to estimate the enhanced model. This algorithm is applied to a stereophonic piano signal, and proves capable of accurately modelling reverberation and restoring missing observations.

***Index Terms*—** Non-stationary signal modelling, Time-frequency analysis, Separation of convolutive mixtures, Multichannel signal analysis, Variational EM algorithm.

1. INTRODUCTION

Following previous works which aimed to provide a probabilistic framework for NMF [1–4], the HR-NMF model introduced in [5,6] offers an improved frequency resolution, able to separate sinusoids within the same frequency band, and an improved synthesis capability, able to restore missing TF observations. It is suitable for both complex-valued and real-valued TF representations, such as the short-time Fourier transform (STFT) and the modified discrete cosine transform (MDCT). It also generalizes some popular models, such as the Itakura-Saito NMF model (IS-NMF) [4], autoregressive (AR) processes [7], and the exponential sinusoidal model (ESM), commonly used in HR spectral analysis of time series [7].

In this paper, HR-NMF is extended to multichannel signals and to convolutive mixtures. Contrary to the multichannel NMF [8] where convolution was approximated, convolution is here accurately implemented in the TF domain by following the exact approach proposed in [9]. Consequently, correlations *within* and *between* frequency bands are both taken into account. In order to estimate this multichannel HR-NMF model, we propose a variational EM algorithm, which has a reduced computational complexity and a parallel implementation compared to [10].

[*]This work was undertaken while Roland Badeau was visiting the Centre for Digital Music, partly funded by EPSRC Platform Grant EP/K009559/1, and by the French National Research Agency (ANR) as a part of the DReaM project (ANR-09-CORD-006-03).

[†]Mark D. Plumbley is funded by EPSRC Leadership Fellowship EP/G007144/1.

The paper is structured as follows. The multichannel HR-NMF model is introduced in section 2, and the variational EM algorithm is derived in section 3. An application to a stereophonic piano signal is presented in section 4. Finally, conclusions are drawn in section 5.

2. MULTICHANNEL HR-NMF MODEL

The multichannel HR-NMF model of TF data $y_m(f,t) \in \mathbb{F}$ (where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) is defined for all channels $m \in [0 \dots M-1]$, discrete frequencies $f \in [0 \dots F-1]$, and times $t \in [0 \dots T-1]$, as the sum of S components $y_{ms}(f,t) \in \mathbb{F}$ plus a white noise

$$n_m(f,t) \sim \mathcal{N}_{\mathbb{F}}(0, \sigma_y^2), \tag{1}$$

where $\mathcal{N}_{\mathbb{F}}(0, \sigma_y^2)$ denotes a real (if $\mathbb{F} = \mathbb{R}$) or circular complex (if $\mathbb{F} = \mathbb{C}$) normal distribution of mean 0 and variance σ_y^2:

$$y_m(f,t) = n_m(f,t) + \textstyle\sum_{s=0}^{S-1} y_{ms}(f,t). \tag{2}$$

Each component $y_{ms}(f,t)$ for any $s \in [0 \dots S-1]$ is defined as

$$y_{ms}(f,t) = \textstyle\sum_{\varphi=-P_b}^{P_b} \sum_{\tau=0}^{Q_b} b_{ms}(f,\varphi,\tau)\, z_s(f-\varphi, t-\tau)$$

where $P_b, Q_b \in \mathbb{N}$, $b_{ms}(f,\varphi,\tau) = 0$ if $f - \varphi \notin [0 \dots F-1]$, and the latent components $z_s(f,t) \in \mathbb{F}$ are defined as follows:

- $\forall t \in [0 \dots T-1]$, $x_s(f,t) \sim \mathcal{N}_{\mathbb{F}}(0, \sigma_{x_s}^2(t))$ and

$$z_s(f,t) = x_s(f,t) - \textstyle\sum_{\tau=1}^{Q_a} a_s(f,\tau) z_s(f, t-\tau) \tag{3}$$

  where $Q_a \in \mathbb{N}$ and $a_s(f,\tau)$ defines a stable autoregressive filter,
- $\forall t \in [-Q_z \dots -1]$ where $Q_z = \max(Q_b, Q_a)$,

$$z_s(f,t) \sim \mathcal{N}(\mu_s(f,t), 1/\rho_s(f,t)). \tag{4}$$

Moreover, the random variables $n_m(f,t)$ and $x_s(f,t)$ for all s, m, f, t are assumed mutually independent. Besides, $\forall m \in [0 \dots M-1]$, $\forall f \in [0 \dots F-1]$, $\forall t \in [-Q_z \dots -1]$, $y_m(f,t)$ is unobserved, and $\forall s \in [0 \dots S-1]$, the prior mean $\mu_s(f,t) \in \mathbb{F}$ and the prior precision (inverse variance) $\rho_s(f,t) > 0$ of the latent variable $z_s(f,t)$ are considered to be known (fixed) parameters.

The set θ of parameters to be estimated consists of:

- the **autoregressive parameters** $a_s(f,\tau) \in \mathbb{F}$ for $s \in [0 \dots S\text{-}1]$, $f \in [0 \dots F-1]$, $\tau \in [1 \dots Q_a]$ (we further define $a_s(f,0)=1$),
- the **moving average parameters** $b_{ms}(f,\varphi,\tau) \in \mathbb{F}$ for $m \in [0 \dots M-1]$, $s \in [0 \dots S-1]$, $f \in [0 \dots F-1]$, $\varphi \in [-P_b \dots P_b]$, and $\tau \in [0 \dots Q_b]$,
- the **variance parameters** $\sigma_y^2 > 0$ and $\sigma_{x_s}^2(t) > 0$ for $s \in [0 \dots S-1]$ and $t \in [0 \dots T-1]$.

This model encompasses the following special cases:

- If $M = 1$, $\sigma_y^2 = 0$ and $P_b = Q_b = Q_a = 0$, equation (2) reduces to $y_0(f,t) = \sum_{s=0}^{S-1} b_{0s}(f,0,0)x_s(f,t)$, thus $y_0(f,t) \sim \mathcal{N}_{\mathbb{F}}(0, \widehat{V}_{ft})$, where matrix $\widehat{\boldsymbol{V}}$ is defined by the NMF $\widehat{\boldsymbol{V}} = \boldsymbol{W}\,\boldsymbol{H}$ with $W_{fs} = |b_{0s}(f,0,0)|^2$ and $H_{st} = \sigma_{x_s}^2(t)$. The maximum likelihood estimation of $\boldsymbol{W}$ and $\boldsymbol{H}$ is then equivalent to the minimization of the Itakura-Saito (IS) divergence between matrix $\widehat{\boldsymbol{V}}$ and spectrogram $\boldsymbol{V}$ (where $V_{ft} = |y_0(f,t)|^2$), hence this model is referred to as **IS-NMF** [4].
- If $M = 1$ and $P_b = Q_b = 0$, $y_0(f,t)$ follows the **HR-NMF** model [5,6,10] involving variance σ_y^2, autoregressive parameters $a_s(f,\tau)$ for all $f \in [0 \dots F-1]$ and $\tau \in [1 \dots Q_a]$, and the same NMF $\widehat{\boldsymbol{V}} = \boldsymbol{W}\,\boldsymbol{H}$.
- If $S = 1, \sigma_y^2 = 0, P_b = 0, \sigma_{x_0}^2(t) = 1\ \forall t \in [0 \dots T-1]$, and $\mu_s(f,t) = 0$ and $\rho_s(f,t) = 1\ \forall t \in [-Q_z \dots -1]$, then $\forall m \in [0 \dots M-1], \forall f \in [0 \dots F-1]$, $y_m(f,t)$ is an autoregressive moving average **(ARMA) process** [7].
- If $S = 1$, $\sigma_y^2 = 0$, $P_b = 0$, $Q_a > 0$, $Q_b = Q_a - 1$, $\forall t \in [-Q_z \dots -1]$, $\mu_0(f,t) = 0$, $\rho_0(f,t) \to +\infty$, and $\sigma_0^2(t) = \mathbb{1}_{\{t=0\}}$, then $\forall m \in [0 \dots M-1]$, $\forall f \in [0 \dots F-1]$, $y_m(f,t)$ can be written in the form $y_m(f,t) = \sum_{\tau=1}^{Q_a} \alpha_{m\tau}\, z_\tau^t$, where $z_1 \dots z_{Q_a}$ are the roots of the polynomial $z^{Q_a} + \sum_{\tau=1}^{Q_a} a_s(f,\tau) z^{Q_a-\tau}$. This corresponds to the **Exponential Sinusoidal Model (ESM)** commonly used in HR spectral analysis of time series [7].

For these reasons, model (2) is called multichannel HR-NMF.

3. VARIATIONAL EM ALGORITHM

In order to estimate the multichannel HR-NMF model introduced in section 2, we derive below a variational EM algorithm.

3.1. Review of variational EM algorithm

Variational inference [11] is now a classical approach for estimating a probabilistic model involving both observed variables y and latent variables z, determined by parameters θ. Let $\mathcal{F}$ be a set of probability density functions (PDFs) over the latent variables z. For any PDF $q \in \mathcal{F}$ and any function $f(z)$, we note $\langle f \rangle_q = \int f(z) q(z) dz$. Then for any parameter θ, the *variational free energy* is defined as

$$\mathcal{L}(q;\theta) = \left\langle \ln\left(\frac{p(y,z;\theta)}{q(z)}\right)\right\rangle_q . \tag{5}$$

The variational EM algorithm is a recursive algorithm for estimating θ. It consists of the two following steps at each iteration i:

- Expectation (E)-step (update q):

$$q^\star = \underset{q\in\mathcal{F}}{\operatorname{argmax}}\ \mathcal{L}(q;\theta_{i-1}) \tag{6}$$

- Maximization (E)-step (update θ):

$$\theta_i = \underset{\theta}{\operatorname{argmax}}\ \mathcal{L}(q^\star;\theta). \tag{7}$$

In the case of multichannel HR-NMF, θ has been defined in section 2, and we define $\delta_m(f,t) = 1$ if $y_m(f,t)$ is observed, otherwise $\delta_m(f,t) = 0$, in particular $\delta_m(f,t) = 0\ \forall (f,t) \notin [0 \dots F-1] \times [0 \dots T-1]$. The complete set of variables thus consists of:

- the set y of **observed variables** $y_m(f,t)$ for $m \in [0 \dots M-1]$ and for all f and t such that $\delta_m(f,t) = 1$,
- the set z of **latent variables** $z_s(f,t)$ for $s \in [0 \dots S-1]$, $f \in [0 \dots F-1]$, and $t \in [-Q_z \dots T-1]$.

We use a *mean field approximation* [11]: $\mathcal{F}$ is defined as the set of PDFs which can be factorized in the form

$$q(z) = \prod_{s=0}^{S-1} \prod_{f=0}^{F-1} \prod_{t=-Q_z}^{T-1} q_{sft}(z_s(f,t)). \tag{8}$$

With this particular factorization of $q(z)$, the solution of (6) is such that each PDF q_{sft} is Gaussian: $z_s(f,t) \sim \mathcal{N}_{\mathbb{F}}(\overline{z_s}(f,t), \gamma_{z_s}(f,t))$.

3.2. Variational free energy

Let $\alpha = 1$ if $\mathbb{F} = \mathbb{C}$, and $\alpha = 2$ if $\mathbb{F} = \mathbb{R}$. Let $\mathbb{1}_{\mathcal{S}}$ denote the indicator function of a set $\mathcal{S}$, and

$$\begin{aligned}
D &= \sum_{m=0}^{M-1}\sum_{f=0}^{F-1}\sum_{t=0}^{T-1} \delta_m(f,t),\\
I(f,t) &= \mathbb{1}_{\{0\le f<F,\, 0\le t<T\}},\\
e_{y_m}(f,t) &= \delta_m(f,t)\left(y_m(f,t) - \sum_{s=0}^{S-1} y_{ms}(f,t)\right),\\
e_{x_s}(f,t) &= I(f,t)\Big(\sum_{\tau=0}^{Q_a} a_s(f,\tau) z_s(f,t-\tau)\Big).
\end{aligned}$$

Then using equations (1) to (4), the joint log-probability distribution $L = \log(p(y,z;\theta))$ of the complete set of variables satisfies

$$\begin{aligned}
-\alpha L = {} & -\alpha\left(\ln(p(y|z;\theta)) + \ln(p(z;\theta))\right)\\
= {} & (D + SF(T+Q_z))\ln(\alpha\pi)\\
& + D\ln(\sigma_y^2) + \tfrac{1}{\sigma_y^2}\sum_{m=0}^{M-1}\sum_{f=0}^{F-1}\sum_{t=0}^{T-1} |e_{y_m}(f,t)|^2\\
& + \sum_{s=0}^{S-1}\sum_{f=0}^{F-1}\sum_{t=-Q_z}^{-1} \ln(\tfrac{1}{\rho_s(f,t)}) + \rho_s(f,t)|z_s(f,t)\text{-}\mu_s(f,t)|^2\\
& + \sum_{s=0}^{S-1}\sum_{f=0}^{F-1}\sum_{t=0}^{T-1} \ln(\sigma_{x_s}^2(t)) + \tfrac{1}{\sigma_{x_s}^2(t)}\,|e_{x_s}(f,t)|^2
\end{aligned}$$

Thus the variational free energy defined in (5) satisfies

$$\begin{aligned}
-\alpha\mathcal{L}(q;\theta) = {} & D\ln(\alpha\pi) - SF(T+Q_z)\\
& + D\ln(\sigma_y^2) + \sum_{m=0}^{M-1}\sum_{f=0}^{F-1}\sum_{t=0}^{T-1} \tfrac{\gamma_{e_{y_m}}(f,t) + \left|\overline{e}_{y_m}(f,t)\right|^2}{\sigma_y^2}\\
& + \sum_{s=0}^{S-1}\sum_{f=0}^{F-1}\sum_{t=-Q_z}^{-1} -\ln(\rho_s(f,t)\gamma_{z_s}(f,t))\\
& + \rho_s(f,t)\left(\gamma_{z_s}(f,t) + |\overline{z}_s(f,t) - \mu_s(f,t)|^2\right)\\
& + \sum_{s=0}^{S-1}\sum_{f=0}^{F-1}\sum_{t=0}^{T-1} \ln\left(\tfrac{\sigma_{x_s}^2(t)}{\gamma_{z_s}(f,t)}\right) + \tfrac{\gamma_{x_s}(f,t) + \left|\overline{e}_{x_s}(f,t)\right|^2}{\sigma_{x_s}^2(t)}
\end{aligned}$$

where $\forall f \in [0 \dots F-1], \forall t \in [0 \dots T-1]$,

$$\gamma_{e_{y_m}}(f,t) = \delta_m(f,t)\sum_{s=0}^{S-1}\sum_{\varphi=-P_b}^{P_b}\sum_{\tau=0}^{Q_b} |b_{ms}(f,\varphi,\tau)|^2 \gamma_{z_s}(f-\varphi, t-\tau),$$

$$\overline{e}_{y_m}(f,t) = \delta_m(f,t)\left(y_m(f,t) - \sum_{s=0}^{S-1}\overline{y}_{ms}(f,t)\right),$$

$$\overline{y}_{ms}(f,t) = \sum_{\varphi=-P_b}^{P_b}\sum_{\tau=0}^{Q_b} b_{ms}(f,\varphi,\tau)\,\overline{z}_s(f-\varphi, t-\tau),$$

$$\gamma_{x_s}(f,t) = I(f,t)\Big(\sum_{\tau=0}^{Q_a} |a_s(f,\tau)|^2 \gamma_{z_s}(f,t-\tau)\Big),$$

$$\overline{e}_{x_s}(f,t) = I(f,t)\Big(\sum_{\tau=0}^{Q_a} a_s(f,\tau)\overline{z}_s(f,t-\tau)\Big).$$

3.3. Variational EM algorithm for multichannel HR-NMF

According to the mean field approximation, the maximizations in equations (6) and (7) are performed for each scalar parameter in turn [11]. The resulting dominant complexity of each iteration of this variational EM algorithm is $4MFST\Delta f\Delta t$, where $\Delta f = 1+2P_b$ and $\Delta t = 1+Q_z$. However we highlight a possible parallel implementation, by making a difference between **parfor** loops which can be implemented in parallel, and **for** loops which have to be implemented sequentially.

E-step: For all $s \in [0 \dots S-1]$, $f \in [0 \dots F-1]$, $t \notin [-Q_z, -1]$, let $\rho_s(f,t) = 0$. Considering the mean field approximation (8), the E-step defined in (6) leads to the updates described in Table 1 (where $*$ denotes complex conjugation). Note that the updates of $\gamma_{z_s}(f,t)$ and $\overline{z}_s(f,t)$ have to be processed sequentially.

M-step: The M-step defined in (7) leads to the updates described in Table 2. The four parameters can be processed in parallel.

---

parfor $s \in [0 \dots S\text{-}1], f \in [0 \dots F\text{-}1], t \in [\text{-}Q_z \dots T\text{-}1]$ **do**

$$\gamma_{z_s}(f,t)^{-1} = \rho_s(f,t) + \sum_{\tau=0}^{Q_a} \frac{I(f,t+\tau)|a_s(f,\tau)|^2}{\sigma^2_{x_s}(t+\tau)} + \sum_{m=0}^{M-1}\sum_{\varphi=-P_b}^{P_b}\sum_{\tau=0}^{Q_b} \frac{\delta_m(f+\varphi,t+\tau)|b_{ms}(f+\varphi,\varphi,\tau)|^2}{\sigma^2_y}$$

end parfor

for $s \in [0 \dots S\text{-}1], f_0 \in [0 \dots \Delta f\text{-}1], t_0 \in [\text{-}Q_z \dots \text{-}Q_z\text{+}\Delta t\text{-}1]$ **do**

parfor $\frac{f-f_0}{\Delta f} \in [0 \dots \lfloor \frac{F-1-f_0}{\Delta_f} \rfloor], \frac{t-t_0}{\Delta t} \in [0 \dots \lfloor \frac{T-1-t_0}{\Delta t} \rfloor]$ **do**

$$\overline{z}_s(f,t) = \overline{z}_s(f,t) - \gamma_{z_s}(f,t)\Big(\rho_s(f,t)(\overline{z}_s(f,t) - \mu_s(f,t)) + \sum_{\tau=0}^{Q_a} \frac{a_s(f,\tau)^*\,\overline{e}_{x_s}(f,t+\tau)}{\sigma^2_{x_s}(t+\tau)} - \sum_{m=0}^{M-1}\sum_{\varphi=-P_b}^{P_b}\sum_{\tau=0}^{Q_b} \frac{b_{ms}(f+\varphi,\varphi,\tau)^*\,\overline{e}_{y_m}(f+\varphi,t+\tau)}{\sigma^2_y}\Big)$$

end parfor

end for

Table 1: E-step of the variational EM algorithm

$$\sigma^2_y = \frac{1}{D}\sum_{m=0}^{M-1}\sum_{f=0}^{F-1}\sum_{t=0}^{T-1} \gamma_{e_{y_m}}(f,t) + |\overline{e}_{y_m}(f,t)|^2$$

parfor $s \in [0 \dots S-1], t \in [0 \dots T-1]$ **do**

$$\sigma^2_{x_s}(t) = \frac{1}{F}\sum_{f=0}^{F-1} \gamma_{x_s}(f,t) + |\overline{e}_{x_s}(f,t)|^2$$

end parfor

for $\tau \in [1 \dots Q_a]$ **do**

parfor $s \in [0 \dots S-1], f \in [0 \dots F-1]$ **do**

$$a_s(f,\tau) = \frac{\sum_{t=0}^{T-1} \frac{1}{\sigma^2_{x_s}(t)}\left(\overline{z}_s(f,t-\tau)^*\left(a_s(f,\tau)\overline{z}_s(f,t-\tau) - \overline{e}_{x_s}(f,t)\right)\right)}{\sum_{t=0}^{T-1} \frac{1}{\sigma^2_{x_s}(t)}\left(\gamma_{z_s}(f,t-\tau) + |\overline{z}_s(f,t-\tau)|^2\right)}$$

end parfor

end for

for $s \in [0 \dots S-1], \varphi \in [-P_b \dots P_b], \tau \in [0 \dots Q_b]$ **do**

parfor $m \in [0 \dots M\text{-}1], f \in [\max(0,\varphi) \dots F\text{-}1\text{+}\min(0,\varphi)]$ **do**

$$b_{ms}(f,\varphi,\tau) = \frac{\sum_{t=0}^{T-1} \overline{z}_s(f\text{-}\varphi,t\text{-}\tau)^*\left(\delta_m(f,t)b_{ms}(f,\varphi,\tau)\overline{z}_s(f\text{-}\varphi,t\text{-}\tau) + \overline{e}_{y_m}(f,t)\right)}{\sum_{t=0}^{T-1} \delta_m(f,t)\left(\gamma_{z_s}(f\text{-}\varphi,t\text{-}\tau) + |\overline{z}_s(f\text{-}\varphi,t\text{-}\tau)|^2\right)}$$

end parfor

end for

Table 2: M-step of the variational EM algorithm

4. SIMULATION RESULTS

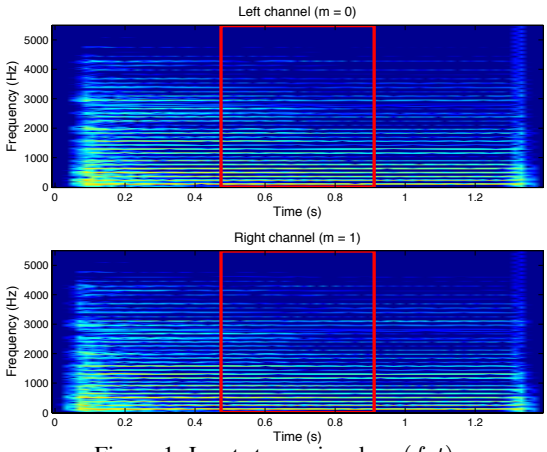

Figure 1: Input stereo signal $y_m(f,t)$.

In this section, we present a basic proof of concept of the proposed multichannel HR-NMF model. The following experiments deal with a single source ($S = 1$) formed of a real piano sound sampled at 11025 Hz. A 1.25ms-short stereophonic signal ($M = 2$) has been synthesized by filtering the monophonic recording of a C3 piano note with two room impulse responses simulated using the Matlab code presented in [12]. The TF representation $y_m(f,t)$ of this signal has then been computed by applying a critically sampled perfect reconstruction cosine modulated filter bank ($\mathbb{F} = \mathbb{R}$) with $F = 201$ frequency bands, involving filters of length $8F = 1608$ samples[1]. The resulting TF representation, of dimension $F \times T$ with $T = 77$, is displayed in Figure 1. In particular, it can be noticed that the two channels are not synchronous, which suggests that the order Q_b of filters $b_{ms}(f,\varphi,\tau)$ should be chosen greater than zero.

In the following experiments, we have set $\mu_s(f,t) = 0$ and $\rho_s(f,t) = 10^5$ (these values force $\overline{z}_s(f,t)$ to be close to zero $\forall t \in [-Q_z \dots -1]$, which is relevant if the observed sound is preceded by silence). The variational EM algorithm is initialized with $\overline{z}_s(f,t) = 0, \gamma_{z_s}(f,t) = \sigma^2_y = \sigma^2_{x_s}(t) = 1, a_s(f,\tau) = \mathbb{1}_{\{\tau=0\}}$, and $b_{ms}(f,\varphi,\tau) = \mathbb{1}_{\{\varphi=0,\tau=0\}}$. In order to illustrate the capability of the multichannel HR-NMF model to synthesize realistic audio data, we address the case of missing observations. We suppose that all TF points within the red frame in Figure 1 are unobserved ($\delta_m(f,t) = 0\ \forall t \in [26 \dots 50]$, and $\delta_m(f,t) = 1$ for all other t in $[0 \dots T-1]$). In each experiment, 100 iterations of the algorithm are performed, and the restored signal is returned as $\overline{y}_{ms}(f,t)$.

In the first experiment, a multichannel HR-NMF with $Q_a = Q_b = P_b = 0$ is estimated. Similarly to the example provided in section 2, this is equivalent to modelling the two channels by two IS-NMF models [4] having distinct spectral atoms and sharing the same temporal activation, or by a multichannel NMF of rank 1 [8]. The resulting TF representation $\overline{y}_{ms}(f,t)$ is displayed in Figure 2. It can be noticed that wherever $y_m(f,t)$ is observed ($\delta_m(f,t) = 1$), $\overline{y}_{ms}(f,t)$ does not accurately fit $y_m(f,t)$, because the length Q_b of filters $b_{ms}(f,\varphi,\tau)$ has been underestimated: the source to distortion ratio (SDR) in the observed area is 11.7dB. In other respects, the missing observations ($\delta_m(f,t) = 0$) could not

[1] Note that the case $\mathbb{F} = \mathbb{C}$ has already been addressed in [5,6,10].

be restored ($\overline{y}_{ms}(f,t)$ is zero inside the red frame, resulting in an SDR of 0dB in this area), because the correlations between contiguous TF coefficients in $y_m(f,t)$ have not been taken into account.

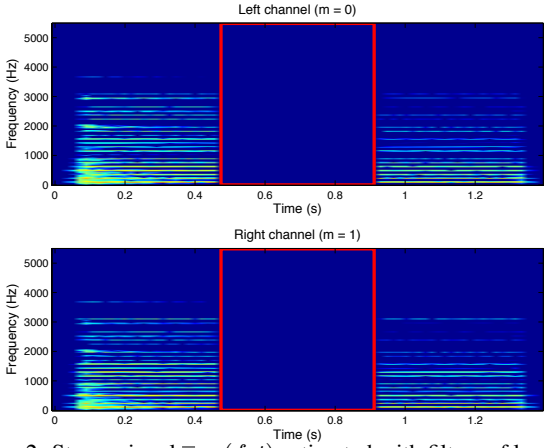

Figure 2: Stereo signal $\overline{y}_{ms}(f,t)$ estimated with filters of length 1.

In the second experiment, a multichannel HR-NMF model with $Q_a = 2$, $Q_b = 3$, and $P_b = 1$ is estimated. The resulting TF representation $\overline{y}_{ms}(f,t)$ is displayed in Figure 3. The SDR is increased to 36.8dB in the observed area, and to 4.8dB inside the red frame.

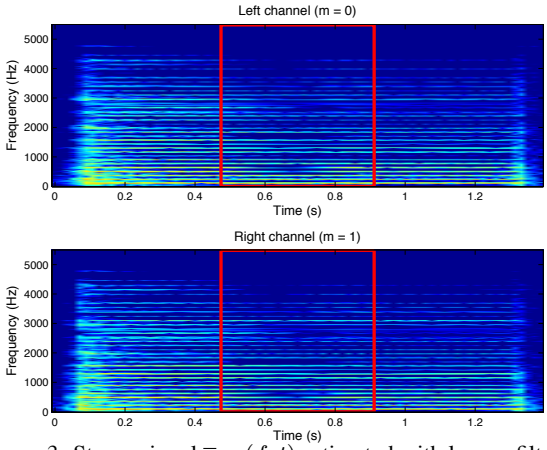

Figure 3: Stereo signal $\overline{y}_{ms}(f,t)$ estimated with longer filters.

The two experiments were runned with a 1.80GHz processor and 1Go RAM. In the first experiment the CPU time required to run 100 iterations of the variational EM algorithm in Matlab was 0.82s; in the second experiment it was 23,63s. The improved performance thus comes at the expense of an increased computational cost.

5. CONCLUSIONS

In this paper, the HR-NMF model [5,6] has been extended to multichannel signals and to convolutive mixtures. The new multichannel HR-NMF model accurately represents convolution in the TF domain [9], and also takes the correlations over frequencies into account. In order to estimate this model from real audio data, a variational EM algorithm has been proposed, which has a reduced computational complexity (by updating the model parameters in turn rather than jointly, the complexity of the M-step has been divided by a factor $(\Delta t)^2$) and a parallel implementation compared to [10]. This algorithm has been successfully applied to a stereophonic piano signal, and has been capable of modelling reverberation and restoring missing observations.

In future work, the sparsity of the model parameters in the TF domain could be enforced by using sparse Bayesian learning [13]. Some other desirable properties such as harmonicity and temporal or spectral smoothness could also be enforced by introducing some prior distributions of the parameters. Similarly to the high spectral resolution, a high temporal resolution could be achieved by extending the model as proposed in [9]. Other Bayesian estimation techniques such as Markov chain Monte Carlo (MCMC) methods and message passing algorithms [11] might prove more effective than the variational EM algorithm. Lastly, the proposed approach could be used in a variety of applications, such as source separation, source coding, audio inpainting, and automatic music transcription.

6. REFERENCES

[1] M. N. Schmidt and H. Laurberg, "Non-negative matrix factorization with Gaussian process priors," *Computational Intelligence and Neuroscience*, 2008, Article ID 361705, 10 pages.

[2] P. Smaragdis, "Probabilistic decompositions of spectra for sound separation," in *Blind Speech Separation*, S. Makino, T.-W. Lee, and H. Sawada, Eds. Springer, 2007, pp. 365–386.

[3] T. Virtanen, A. Cemgil, and S. Godsill, "Bayesian extensions to non-negative matrix factorisation for audio signal modelling," in *Proc. IEEE ICASSP*, Las Vegas, Nevada, USA, Apr. 2008, pp. 1825–1828.

[4] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, Mar. 2009.

[5] R. Badeau, "Gaussian modeling of mixtures of non-stationary signals in the time-frequency domain (HR-NMF)," in *Proc. IEEE WASPAA*, New York, USA, Oct. 2011, pp. 253–256.

[6] ——, "High resolution NMF for modeling mixtures of non-stationary signals in the time-frequency domain," Télécom ParisTech, Paris, France, Tech. Rep. 2012D004, July 2012.

[7] M. H. Hayes, *Statistical Digital Signal Processing And Modeling*. Wiley, Aug. 2009.

[8] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 550–563, Mar. 2010.

[9] R. Badeau and M. D. Plumbley, "Probabilistic time-frequency source-filter decomposition of non-stationary signals," in *Proc. EUSIPCO*, Marrakech, Morocco, Sept. 2013.

[10] R. Badeau and A. Drémeau, "Variational Bayesian EM algorithm for modeling mixtures of non-stationary signals in the time-frequency domain (HR-NMF)," in *Proc. IEEE ICASSP*, Vancouver, Canada, May 2013, pp. 6171–6175.

[11] D. J. MacKay, *Information Theory, Inference, And Learning Algorithms*. Cambridge University Press, 2003.

[12] S. G. McGovern, "A model for room acoustics," http://www.sgm-audio.com/research/rir/rir.html.

[13] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2154, Aug. 2004.