Gaussian modeling of mixtures of non-stationary signals in the time-frequency domain (HR-NMF)

Roland Badeau

GAUSSIAN MODELING OF MIXTURES OF NON-STATIONARY SIGNALS IN THE TIME-FREQUENCY DOMAIN (HR-NMF)

Roland Badeau

Institut Telecom, Telecom ParisTech, CNRS LTCI
46 rue Barrault, 75634 Paris Cedex 13
roland.badeau@telecom-paristech.fr

ABSTRACT

Nonnegative Matrix Factorization (NMF) is a powerful tool for decomposing mixtures of non-stationary signals in the Time-Frequency (TF) domain. However, unlike the High Resolution (HR) methods dedicated to mixtures of exponentials, its spectral resolution is limited by that of the underlying TF representation. In this paper, we propose a unified probabilistic model called HR-NMF, that permits to overcome this limit by taking both phases and local correlations in each frequency band into account. This model is estimated with a recursive implementation of the EM algorithm, that is successfully applied to source separation and audio inpainting.

Index Terms— Nonnegative Matrix Factorization, High Resolution methods, Expectation-Maximization algorithm, source separation, audio inpainting.

1. INTRODUCTION

Some very powerful tools have been recently introduced for modeling mixtures of non-stationary signal components in the TF domain. Among them, NMF [1] and Probabilistic Latent Component Analysis (PLCA) [2] compute an approximate factorization of a magnitude or power TF representation, such as the spectrogram.

Since phases are generally discarded in these models, reconstructing the phase field requires employing ad-hoc methods [3]. To the best of our knowledge, apart the complex NMF which was designed in a deterministic framework [4], the only probabilistic model that takes the phase field into account is the Itakura-Saito (IS)-NMF [5]. Separating the signal components is then proven equivalent to Wiener filtering. The spectral resolution of IS-NMF is thus limited by that of the TF representation (sinusoids in the same frequency band cannot be properly separated).

In other respects, IS-NMF assumes that all TF coefficients are independent, which is not suitable for modeling sinusoids or transients for instance. Markov models have thus been proposed for taking the local dependencies between contiguous TF coefficients of a magnitude or power TF representation into account [6, 7].

In this paper, we introduce a unified model called HR-NMF, which natively takes both phases and local correlations in each frequency band into account. This approach avoids using a phase reconstruction algorithm, and we show that it overcomes the spectral resolution of the TF representation. It can be used with both complex-valued and real-valued TF representations (like the short-time Fourier transform or cosine modulated filterbanks).

This work is supported by the French National Research Agency (ANR) as a part of the DReaM project (ANR-09-CORD-006-03) and partly supported by the Quaero Program, funded by OSEO.

This paper is organized as follows: HR-NMF is introduced in section 2, then our recursive implementation of the EM algorithm for estimating this model is presented in section 3. Section 4 is devoted to experimental results, and conclusions are drawn in section 5. The following notation will be used throughout the paper:

- x: scalar (normal letter),
- $\boldsymbol{v} = [v_1; \ldots; v_K]$: column vector (bold lower case letter),
- $\boldsymbol{S}$: matrix (bold upper case letter),
- $S_{(t,p)}$: (t,p)-th entry of matrix $\boldsymbol{S}$ (indexed upper case letter),
- $\boldsymbol{S}^*$ (resp. $\boldsymbol{S}^H$): conjugate (resp. conjugate transpose) of $\boldsymbol{S}$,
- $\mathrm{diag}(.)$: (block)-diagonal matrix whose diagonal blocks are $(.)$,
- $\mathbf{1}$ (resp. $\mathbf{0}$): vector whose entries are all equal to 1 (resp. 0),
- $\mathbb{E}[.]_{/x}$: conditional expectation of $(.)$ given the observation x,
- $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{R})$: real or circular complex normal distribution of mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{R}$,
- for a given vector $\overline{\boldsymbol{v}}$ of dimension $\overline{D}$, and any subvector $\boldsymbol{v}$ of dimension $D \leq \overline{D}$ (whose entries are a subset of those of $\overline{\boldsymbol{v}}$), $\boldsymbol{J}_v^{\overline{v}}$ denotes the $\overline{D} \times D$ selection matrix such that $\boldsymbol{v} = \boldsymbol{J}_v^{\overline{v}H} \overline{\boldsymbol{v}}$.

2. TIME-FREQUENCY MIXTURE MODEL

The mixture model $x(f,t)$ is defined for all frequencies $1 \leq f \leq F$ and times $1 \leq t \leq T$ as the sum of K latent components $c_k(f,t)$ plus a white noise $n(f,t) \sim \mathcal{N}(0, \sigma^2)$:

$$x(f,t) = n(f,t) + \sum_{k=1}^{K} c_k(f,t) \tag{1}$$

where

- $c_k(f,t) = \sum_{p=1}^{P(k,f)} a(p,k,f)\, c_k(f, t-p) + b_k(f,t)$ is obtained by autoregressive filtering of a non-stationary signal $b_k(f,t)$ (and $P(k,f) \in \mathbb{N}$ is such that $a(P(k,f),k,f) \neq 0$),
- $b_k(f,t) \sim \mathcal{N}(0, v_k(f,t))$ where $v_k(f,t)$ is defined as
$$v_k(f,t) = w(k,f)\, h(k,t), \tag{2}$$
with $w(k,f) \geq 0$ and $h(k,t) \geq 0$,
- processes n and $b_1 \ldots b_K$ are mutually independent.

Since $\mathcal{N}$ denotes either the real or the circular complex normal distribution, model (1) is either real or complex-valued. Moreover, we assume that $\forall k \in \{1 \ldots K\}$, $\forall f \in \{1 \ldots F\}$, $\forall t \leq 0$, $c_k(f,t)$ are independent and identically distributed random variables: $c_k(f,t) \sim \mathcal{N}(0,1)$, and $x(f,t)$ are unobserved. The parameters to be estimated are σ^2, $a(p,k,f)$, $w(k,f)$, and $h(k,t)$.

This time-frequency model generalizes some very popular models, widely used in various signal processing communities:

- If $\sigma^2 = 0$ and $\forall k, f, P(k,f) = 0$, (1) becomes $x(f,t) = \sum_{k=1}^{K} b_k(f,t)$, thus $x(f,t) \sim \mathcal{N}(0, \widehat{V}_{ft})$, where $\widehat{\boldsymbol{V}}$ is defined by the NMF $\widehat{\boldsymbol{V}} = \boldsymbol{W}\,\boldsymbol{H}$ with $W_{fk} = w(k,f)$ and $H_{kt} = h(k,t)$. Maximum likelihood estimation of $\boldsymbol{W}$ and $\boldsymbol{H}$ is then equivalent to the minimization of the IS-divergence between the matrix model $\widehat{\boldsymbol{V}}$ and the spectrogram $\boldsymbol{V}$ (where $V_{ft} = |x(f,t)|^2$), that is why this model is referred to as IS-NMF [5].
- For given values of k and f, if $\forall t$, $h(k,t) = 1$, then $c_k(f,t)$ is an autoregressive process of order $P(k,f)$.
- For given values of k and f, if $P(k,f) \geq 1$ and $\forall t \geq P(k,f)+1$, $h(k,t) = 0$, then $c_k(f,t)$ can be written in the form $c_k(f,t) = \sum_{p=1}^{P(k,f)} \alpha_p\, z_p^t$, where $z_1 \dots z_{P(k,f)}$ are the roots of the polynomial $z^{P(k,f)} - \sum_{p=1}^{P(k,f)} a(p,k,f) z^{P(k,f)-p}$. This corresponds to the Exponential Sinusoidal Model (ESM)[1] commonly used in HR spectral analysis of time series [8].

  For these reasons, we refer to model (1) as HR-NMF.

3. EXPECTATION-MAXIMIZATION (EM) ALGORITHM

In order to estimate the model parameters, we apply the EM algorithm[2] to the observed data x and the latent components $c_1 \dots c_K$. In order to handle the case of missing data, we define $\delta(f,t) = 1$ if $x(f,t)$ is observed, and $\delta(f,t) = 0$ else.

3.1. Maximization Step (M-step)

The conditional expectation (given the observations) of the log-likelihood of the complete data is $Q = \mathbb{E}_{/x}\left[\ln(p(c_1 \dots c_K, x))\right] = \mathbb{E}_{/x}\left[\ln(p(x/c_1 \dots c_K))\right] + \sum_{k=1}^{K} \mathbb{E}_{/x}\left[\ln(p(c_k))\right]$.

We can thus write[3] $Q \stackrel{c}{=} Q_0 + \sum_{k=1}^{K} Q_k$ where

$$Q_0 = -\sum_{f=1}^{F}\sum_{t=1}^{T} \delta(f,t)\ln(\sigma^2) + e(f,t)/\sigma^2,$$
$$Q_k = -\sum_{f=1}^{F}\sum_{t=1}^{T} \ln(w(k,f)h(k,t)) + \frac{\boldsymbol{a}(k,f)^H \boldsymbol{S}(k,f,t)\,\boldsymbol{a}(k,f)}{w(k,f)h(k,t)},$$

$$e(f,t) = \delta(f,t)\,\mathbb{E}_{/x}\Big[\Big|x(f,t) - \sum_{k=1}^{K} c_k(f,t)\Big|^2\Big], \quad (3)$$

$$\boldsymbol{a}(k,f) = [1; -a(1,k,f); \dots; -a(P(k,f),k,f)], \quad (4)$$

and $\forall k, f$, for all $0 \leq p_1, p_2 \leq P(k,f)$,

$$S_{(p_1,p_2)}(k,f,t) = \mathbb{E}_{/x}[c_k(f,t-p_1)^*\, c_k(f,t-p_2)]. \quad (5)$$

Maximizing Q is thus equivalent to independently maximizing Q_0 with respect to (w.r.t.) σ^2 and each Q_k w.r.t. $h(k,t)$, $w(k,f)$ and $\boldsymbol{a}(k,f)$. Since the maximization of Q_k does not admit a closed form solution, we propose to recursively maximize Q_k w.r.t. $(w(k,f), \boldsymbol{a}(k,f))$ and w.r.t. $h(k,t)$. We cannot provide here the full mathematical derivation of the M-step because of the page limit, but its pseudo-code is summarized in Table 1 (according to the notation introduced in section 1, $\boldsymbol{J}_1^{\boldsymbol{a}(k,f)} = [1;0;\dots;0]$). Its complexity is $O(FTKP^2)$, where $P = \max_{k,f} P(k,f)$.

Inputs after the E-step: $\delta(f,t),\ e(f,t),\ \boldsymbol{S}(k,f,t), h(k,t)$
$\sigma^2 = \sum_{f=1}^{F}\sum_{t=1}^{T} e(f,t) \Big/ \sum_{f=1}^{F}\sum_{t=1}^{T} \delta(f,t)$
For $k = 1$ to K [4],
Repeat (as many times as wanted) [5]:
For $f = 1$ to F [4],
$\boldsymbol{\Sigma}(k,f) = \frac{1}{T}\sum_{t=1}^{T} \frac{\boldsymbol{S}(k,f,t)}{h(k,t)}$
$\boldsymbol{\alpha}(k,f) = \boldsymbol{\Sigma}(k,f)^{-1}\,\boldsymbol{J}_1^{\boldsymbol{a}(k,f)}$
$w(k,f) = 1/(\boldsymbol{J}_1^{\boldsymbol{a}(k,f)^H}\boldsymbol{\alpha}(k,f))$
$\boldsymbol{a}(k,f) = w(k,f)\,\boldsymbol{\alpha}(k,f)$
End for f;
For $t = 1$ to T [4],
$h(k,t) = \frac{1}{F}\sum_{f=1}^{F} \frac{\boldsymbol{a}(k,f)^H \boldsymbol{S}(k,f,t)\boldsymbol{a}(k,f)}{w(k,f)}$
End for t;
Normalization of the NMF: $H_k = \max_t(h(k,t))$,
$w(k,f) = H_k\, w(k,f),\ h(k,t) = h(k,t)/H_k$
End repeat;
End for k;
Outputs: $\sigma^2,\ \boldsymbol{a}(k,f), w(k,f), h(k,t)$.

Table 1: Pseudo-code of the M-step

3.2. Expectation Step (E-step)

The purpose of the E-step is to determine the a posteriori distribution of the latent components $c_k(f,t)$ given the observations $x(f,t)$ (and more precisely, $e(f,t)$ and $\boldsymbol{S}(k,f,t)$). Since these random variables are mutually independent for different values of f, the E-step can process each f separately. As our recursive implementation of the E-step is inspired from Kalman filtering theory [9], we first introduce the Kalman representation of the HR-NMF model:

$$\boldsymbol{\gamma}(f,t) = \boldsymbol{A}(f)\,\boldsymbol{\gamma}(f,t-1) + \boldsymbol{b}'(f,t), \quad (6)$$
$$x(f,t) = \boldsymbol{u}(f)^H \boldsymbol{\gamma}(f,t) + n'(f,t), \quad (7)$$

where

- $\mathcal{K}(f)$ denotes the set $\{k \in \{1 \dots K\}/P(k,f) \geq 1\}$;
- $\forall k \in \mathcal{K}(f)$, $\boldsymbol{c}(k,f,t) = [c_k(f,t); \dots; c_k(f,t\text{-}P(k,f)\text{+}1)]$;
- the state vector $\boldsymbol{\gamma}(f,t)$ contains $\boldsymbol{c}(k,f,t)$ for all $k \in \mathcal{K}(f)$;
- the state transition matrix is $\boldsymbol{A}(f) = \mathrm{diag}(\{\boldsymbol{A}(k,f)\}_{k\in\mathcal{K}(f)})$, where[6] $\boldsymbol{A}(k,f) = a(P(k,f),k,f)$ if $P(k,f) = 1$, otherwise

$$\boldsymbol{A}(k,f) = \begin{bmatrix} a(1,k,f) \dots a(P(k)\text{-}1,k,f) & a(P(k,f),k,f) \\ \mathrm{diag}(\mathbf{1}) & \mathbf{0} \end{bmatrix}$$

- $\forall f, t$, vector $\boldsymbol{c}(f,t)$ contains $c_k(f,t)$ for all $k \in \mathcal{K}(f)$,

[1] Actually HR-NMF also encompasses the more general Polynomial Amplitude Complex Exponentials (PACE) model introduced in [8].

[2] The EM algorithm is an iterative method which alternates two steps called Expectation and Maximization. It is proved to increase the likelihood of the observed data at each iteration. In section 3, we assume that $w(k,f)$, $h(k,t)$, σ^2 and $a(P(k,f),k,f)$ are non-zero. However, these parameters might become zero after some iterations, depending on the input data. Such singular cases should be addressed in a rigorous implementation of EM.

[3] $\stackrel{c}{=}$ denotes equality up to additive and multiplicative constants which do not depend on the model parameters to be estimated.

[4] This loop can be processed in parallel.

[5] This loop has to be processed sequentially.

[6] Note that since $\forall k \in \mathcal{K}(f), a(P(k,f),k,f) \neq 0$, $\boldsymbol{A}(f)$ is invertible.

- the white process noise is $\boldsymbol{b}'(f,t) = \boldsymbol{J}^{\gamma(f,t)}_{c(f,t)}\,\boldsymbol{b}(f,t)$, where notation $\boldsymbol{J}^{\gamma(f,t)}_{c(f,t)}$ was introduced in section 1, and $\boldsymbol{b}(f,t)$ contains $b_k(f,t)\ \forall k \in \mathcal{K}(f)$; thus $\boldsymbol{b}(f,t) \sim \mathcal{N}(\boldsymbol{0}, \mathrm{diag}(\boldsymbol{v}(f,t)))$, where $\boldsymbol{v}(f,t)$ contains $v_k(f,t)\ \forall k \in \mathcal{K}(f)$;
- the observation matrix is $\boldsymbol{u}(f)^H$, where $\boldsymbol{u}(f) = \boldsymbol{J}^{\gamma(f,t)}_{c(f,t)}\,\boldsymbol{1}$;
- the white observation noise is $n'(f,t) = n(f,t) + \sum_{k/P(k,f)=0} b_k(f,t) \sim \mathcal{N}(0, \sigma^2(f,t))$, where

$$\sigma^2(f,t) = \sigma^2 + \sum_{k/P(k,f)=0} v_k(f,t). \tag{8}$$

We cannot provide the full mathematical derivation of the E-step in this paper because of the page limit, but its pseudo-code is summarized in Table 2. Its overall computational complexity is $O(FTK^3P^3)$. In Table 2, we have used the following notation:

- $\forall f,t$, vector $\boldsymbol{d}(f,t)$ contains $c_k(f,t-P(k,f))\ \forall k \in \mathcal{K}(f)$,
- $\forall f,t$, vector $\boldsymbol{c}'(f,t)$ contains $c_k(f,t)$ for all $k \notin \mathcal{K}(f)$,
- $\forall f,t$, vector $\boldsymbol{v}'(f,t)$ contains $v_k(f,t)$ for all $k \notin \mathcal{K}(f)$,
- $\forall f,t, \forall k \in \mathcal{K}(f), \overline{\boldsymbol{c}}(k,f,t) = [c_k(f,t); \ldots; c_k(f,t-P(k,f))]$,
- $\forall f,t, \overline{\boldsymbol{\gamma}}(f,t)$ contains $\overline{\boldsymbol{c}}(k,f,t)$ for all $k \in \mathcal{K}(f)$,
- $\forall f,t$, and for any random vector $\boldsymbol{v}$, $\boldsymbol{v}^{f,t}$ is the conditional expectation of $\boldsymbol{v}$ given $\{x(f,1)\ldots x(f,t)\}$. Besides, $\boldsymbol{R}^{f,t}_{v}$ is the conditional expectation of $(\widetilde{\boldsymbol{v}}^{f,t})\,(\widetilde{\boldsymbol{v}}^{f,t})^H$ given $\{x(f,1)\ldots x(f,t)\}$, where $\widetilde{\boldsymbol{v}}^{f,t} = \boldsymbol{v} - \boldsymbol{v}^{f,t}$. Similarly, for any vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, $\boldsymbol{R}^{f,t}_{v_1,v_2}$ is the conditional expectation of $(\widetilde{\boldsymbol{v}}^{f,t}_1)\,(\widetilde{\boldsymbol{v}}^{f,t}_2)^H$ given $\{x(f,1)\ldots x(f,t)\}$.

Other letters denote temporary variables used in the computations.

4. APPLICATIONS

In order to illustrate the capabilities of HR-NMF, we consider two examples of straightforward applications. First, noticing that the E-step estimates $c_k(f,t)$, source separation will be addressed in section 4.1. Moreover, since $c_k(f,t)$ is estimated even at time-frequencies where the observation $x(f,t)$ is missing, audio inpainting will be addressed in section 4.2. The following experiments deal with a real piano sound, composed of a C4 tone played alone at $t = 0$ ms, and a C3 tone played at $t = 680$ ms while the C4 tone is maintained. The sampling frequency is 8600 Hz, and $x(f,t)$ is obtained by computing the STFT of the input signal with dimensions $F = 400$ and $T = 60$, using 90 ms-long Hann windows with 75% overlap (the corresponding spectrogram is plotted in Figure 1). HR-NMF is then computed by running 5 iterations of the EM algorithm with $P(k,f) = 1$, after initialization with IS-NMF.

4.1. Source separation

In this first experiment, the whole STFT $x(f,t)$ is observed, and we aim at separating $K = 2$ components $c_k(f,t)$ in the frequency band f which corresponds to the first harmonic of C4 and to the second harmonic of C3 (around 270 Hz). These two sinusoidal components (whose real parts are represented as red solid lines in Figure 2) have very close frequencies, which makes them hardly separable. As expected, IS-NMF, which involves Wiener filtering, is not able to properly separate the components when they overlap, from $t = 680$ ms to 1.36 s (the estimated components are represented as black dash-dotted lines). As a comparison, the components estimated by HR-NMF (blue dashed lines) better fit the ground truth.

Inputs after the M-step: $x(f,t)$, $\delta(f,t)$, σ^2, $\boldsymbol{A}(f)$, $v_k(f,t)$

Initialization: $\forall f, \boldsymbol{\gamma}^{f,0}(f,0) = \boldsymbol{0},\ \boldsymbol{R}^{f,0}_{\gamma(f,0)} = \mathrm{diag}(\boldsymbol{1})$

$\forall f,t, \sigma^2(f,t) = \sigma^2 + \sum_{k/P(k,f)=0} v_k(f,t)$

For $f = 1$ to F, [4]

For $t = 1$ to T (forward pass): [5]

Predict:

$\boldsymbol{R}^{f,t-1}_{\gamma(f,t),d(f,t)} = \boldsymbol{A}(f)\boldsymbol{R}^{f,t-1}_{\gamma(f,t-1)}\boldsymbol{J}^{\gamma(f,t-1)}_{d(f,t)}$

$\boldsymbol{R}^{f,t-1}_{d(f,t)} = \boldsymbol{J}^{\gamma(f,t-1)H}_{d(f,t)}\boldsymbol{R}^{f,t-1}_{\gamma(f,t-1)}\boldsymbol{J}^{\gamma(f,t-1)}_{d(f,t)}$

$\boldsymbol{R}^{f,t-1}_{b'(f,t)} = \boldsymbol{J}^{\gamma(f,t)}_{c(f,t)}\,\mathrm{diag}(\boldsymbol{v}(f,t))\,\boldsymbol{J}^{\gamma(f,t)H}_{c(f,t)}$

$\boldsymbol{R}^{f,t-1}_{\gamma(f,t)} = \boldsymbol{A}(f)\boldsymbol{R}^{f,t-1}_{\gamma(f,t-1)}\boldsymbol{A}(f)^H + \boldsymbol{R}^{f,t-1}_{b'(f,t)}$

$\boldsymbol{d}^{f,t-1}(f,t) = \boldsymbol{J}^{\gamma(f,t-1)H}_{d(f,t)}\boldsymbol{\gamma}^{f,t-1}(f,t-1)$

$\boldsymbol{\gamma}^{f,t-1}(f,t) = \boldsymbol{A}(f)\,\boldsymbol{\gamma}^{f,t-1}(f,t-1)$

$\boldsymbol{\Phi}^{f,t-1}_{\gamma(f,t),d(f,t)} = \left(\boldsymbol{R}^{f,t-1}_{\gamma(f,t)}\right)^{-1}\boldsymbol{R}^{f,t-1}_{\gamma(f,t),d(f,t)}$

$\boldsymbol{\Psi}^{f,t-1}_{d(f,t)} = \boldsymbol{R}^{f,t-1}_{d(f,t)} - \boldsymbol{R}^{f,t-1}_{d(f,t),\gamma(f,t)}\boldsymbol{\Phi}^{f,t-1}_{\gamma(f,t),d(f,t)}$

$\boldsymbol{\phi}^{f,t-1}_{d(f,t)} = \boldsymbol{d}^{f,t-1}(f,t) - \boldsymbol{\Phi}^{f,t-1H}_{\gamma(f,t),d(f,t)}\boldsymbol{\gamma}^{f,t-1}(f,t)$

Update:

$\boldsymbol{\mu}(f,t) = \boldsymbol{R}^{f,t-1}_{\gamma(f,t)}\,\boldsymbol{u}(f)$

$\varepsilon(f,t) = \sigma^2(f,t) + \boldsymbol{u}(f)^H\boldsymbol{\mu}(f,t)$

$\boldsymbol{\lambda}(f,t) = \frac{\delta(f,t)}{\varepsilon(f,t)}\boldsymbol{\mu}(f,t)$

$\boldsymbol{R}^{f,t}_{\gamma(f,t)} = \boldsymbol{R}^{f,t-1}_{\gamma(f,t)} - \boldsymbol{\lambda}(f,t)\,\boldsymbol{\mu}(f,t)^H$

$\epsilon^{f,t}(f,t) = x(f,t) - \boldsymbol{u}(f)^H\boldsymbol{\gamma}^{f,t-1}(f,t)$

$\boldsymbol{\gamma}^{f,t}(f,t) = \boldsymbol{\gamma}^{f,t-1}(f,t) + \boldsymbol{\lambda}(f,t)\,\epsilon^{f,t}(f,t)$

End for t;

For $t = T$ down to 1 (backward pass): [5]

Wiener filtering:

$\epsilon^{f,T}(f,t) = x(f,t) - \boldsymbol{u}(f)^H\boldsymbol{\gamma}^{f,T}(f,t)$

$\boldsymbol{c}'^{f,T}(f,t) = \frac{\delta(f,t)}{\sigma^2(f,t)}\,\boldsymbol{v}'(f,t)\,\epsilon^{f,T}(f,t)$

$e'(f,t) = \frac{|\epsilon^{f,T}(f,t)|^2 + \boldsymbol{u}(f)^H\boldsymbol{R}^{f,T}_{\gamma(f,t)}\boldsymbol{u}(f)}{\sigma^2(f,t)} - 1$

$e(f,t) = \delta(f,t)\,\sigma^2\left(\frac{\sigma^2}{\sigma^2(f,t)}e'(f,t) + 1\right)$

$\forall k \notin \mathcal{K}(f), \boldsymbol{S}(k,f,t) = v_k(f,t) + \delta(f,t)\frac{v_k(f,t)^2}{\sigma^2(f,t)}e'(f,t)$

Smoothing:

$\boldsymbol{R}^{f,T}_{\gamma(f,t),d(f,t)} = \boldsymbol{R}^{f,T}_{\gamma(f,t)}\boldsymbol{\Phi}^{f,t-1}_{\gamma(f,t),d(f,t)}$

$\boldsymbol{R}^{f,T}_{d(f,t)} = \boldsymbol{\Psi}^{f,t-1}_{d(f,t)} + \boldsymbol{R}^{f,T}_{d(f,t),\gamma(f,t)}\boldsymbol{\Phi}^{f,t-1}_{\gamma(f,t),d(f,t)}$

$\boldsymbol{R}^{f,T}_{\overline{\gamma}(f,t)} = \boldsymbol{J}^{\overline{\gamma}(f,t)}_{\gamma(f,t)}\boldsymbol{R}^{f,T}_{\gamma(f,t)}\boldsymbol{J}^{\overline{\gamma}(f,t)H}_{\gamma(f,t)} + \boldsymbol{J}^{\overline{\gamma}(f,t)}_{d(f,t)}\boldsymbol{R}^{f,T}_{d(f,t)}\boldsymbol{J}^{\overline{\gamma}(f,t)H}_{d(f,t)}$
$+\boldsymbol{J}^{\overline{\gamma}(f,t)}_{\gamma(f,t)}\boldsymbol{R}^{f,T}_{\gamma(f,t),d(f,t)}\boldsymbol{J}^{\overline{\gamma}(f,t)H}_{d(f,t)} + \boldsymbol{J}^{\overline{\gamma}(f,t)}_{d(f,t)}\boldsymbol{R}^{f,T}_{d(f,t),\gamma(f,t)}\boldsymbol{J}^{\overline{\gamma}(f,t)H}_{\gamma(f,t)}$

$\boldsymbol{R}^{f,T}_{\gamma(f,t-1)} = \boldsymbol{J}^{\overline{\gamma}(f,t)H}_{\gamma(f,t-1)}\boldsymbol{R}^{f,T}_{\overline{\gamma}(f,t)}\boldsymbol{J}^{\overline{\gamma}(f,t)}_{\gamma(f,t-1)}$

$\boldsymbol{d}^{f,T}(f,t) = \boldsymbol{\phi}^{f,t-1}_{d(f,t)} + \boldsymbol{\Phi}^{f,t-1H}_{\gamma(f,t),d(f,t)}\boldsymbol{\gamma}^{f,T}(f,t)$

$\overline{\boldsymbol{\gamma}}^{f,T}(f,t) = \boldsymbol{J}^{\overline{\gamma}(f,t)}_{\gamma(f,t)}\boldsymbol{\gamma}^{f,T}(f,t) + \boldsymbol{J}^{\overline{\gamma}(f,t)}_{d(f,t)}\boldsymbol{d}^{f,T}(f,t)$

$\boldsymbol{\gamma}^{f,T}(f,t-1) = \boldsymbol{J}^{\overline{\gamma}(f,t)H}_{\gamma(f,t-1)}\overline{\boldsymbol{\gamma}}^{f,T}(f,t)$

$\overline{\boldsymbol{S}}(f,t) = (\boldsymbol{R}^{f,T}_{\overline{\gamma}(f,t)} + \overline{\boldsymbol{\gamma}}^{f,T}(f,t)\,\overline{\boldsymbol{\gamma}}^{f,T}(f,t)^H)^*$

$\forall k \in \mathcal{K}(f), \boldsymbol{S}(k,f,t) = \boldsymbol{J}^{\overline{\gamma}(f,t)H}_{\overline{c}(k,f,t)}\overline{\boldsymbol{S}}(f,t)\boldsymbol{J}^{\overline{\gamma}(f,t)}_{\overline{c}(k,f,t)}$

End for t;

End for f;

Outputs: $e(f,t)$, $\boldsymbol{S}(k,f,t)$, and $c^{f,T}_k(f,t)$ if wanted.

Table 2: Pseudo-code of the E-step

4.2. Audio inpainting

In this second experiment, the second part of the STFT (from $t = 680$ ms to 1.36 s) is unobserved, and in the first part (from

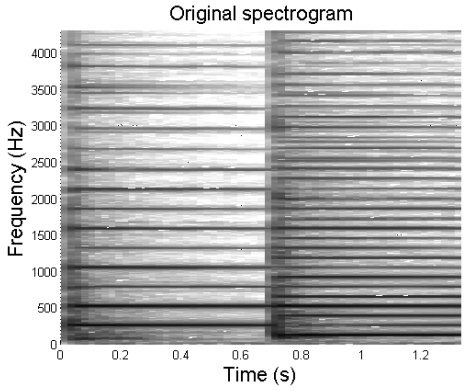

Figure 1: Spectrogram of the input piano sound

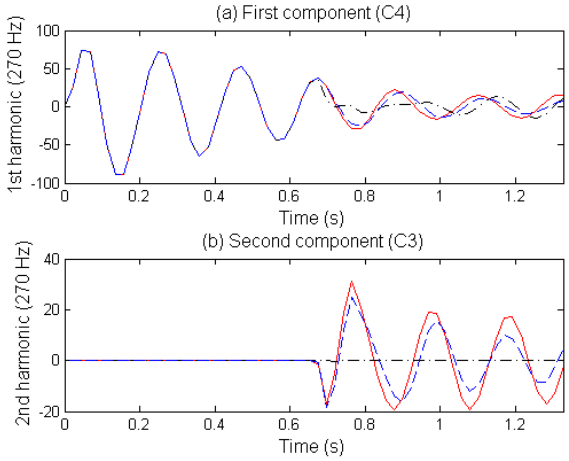

Figure 2: Separation of two sinusoidal components

$t = 0$ ms to $t = 680$ ms), only 50% of the TF coefficients $x(f,t)$ are randomly observed. HR-NMF is computed with $K = 1$, and the estimated component $c_1(f,t)$ is represented in Figure 3. Of course the C3 tone, which was unobserved, could not be recovered, but the C4 tone is correctly estimated. Moreover, the noise in the unobserved part has been removed. Actually we observed that a listening test does not permit to perceive any artifact in the signal synthesized from $c_1(f,t)$ by a standard overlap-add technique. As a comparison, note that IS-NMF is not capable of audio inpainting, because it does not take the correlations between contiguous TF coefficients into account (the missing coefficients are estimated as zeros).

5. CONCLUSIONS

In this paper, we introduced a new method for modeling mixtures of non-stationary signals in the time-frequency domain, which was successfully applied to source separation and audio inpainting. Compared to standard IS-NMF, the proposed approach natively takes both phases and local correlations in each frequency band into account. We showed that it achieves high resolution, which means that two sinusoids of different frequencies can be properly separated within the same frequency band[7]. Besides, HR-NMF is also suitable for modeling stationary and non-stationary noise, as well as transients. In future work, an alternative algorithm faster than EM could be developed for estimating the HR-NMF model. The basic NMF that we used for modeling the non-stationarities in the distribution of $b_k(f,t)$ could be replaced by any non-stationary model, such as one of the many variants of NMF. The model could also be extended in several ways, for instance by taking the correlations across frequencies and/or across components into account, or by representing multichannel signals.

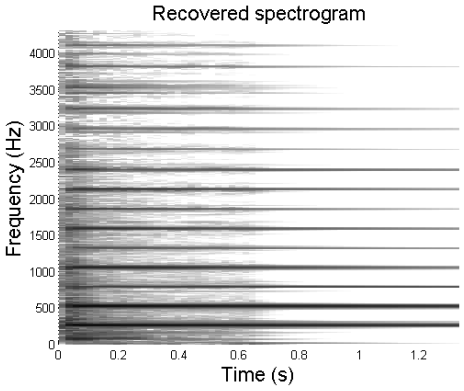

Figure 3: Recovery of the full C4 piano tone

6. REFERENCES

[1] T. Virtanen, A. Cemgil, and S. Godsill, "Bayesian extensions to non-negative matrix factorisation for audio signal modelling," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, Nevada, USA, Apr. 2008, pp. 1825–1828.

[2] P. Smaragdis, *Blind Speech Separation*. Springer, 2007, ch. Probabilistic decompositions of spectra for sound separation, pp. 365–386.

[3] D. Griffin and J. Lim, "Signal reconstruction from short-time Fourier transform magnitude," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 2, pp. 236–243, 1984.

[4] J. Le Roux, H. Kameoka, E. Vincent, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF under spectrogram consistency constraints," in *Proceedings of the Acoustical Society of Japan Autumn Meeting*, no. 2-4-5, Sept. 2009.

[5] 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.

[6] O. Dikmen and A. T. Cemgil, "Gamma Markov random fields for audio source modeling," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 589–601, Mar. 2010.

[7] G. Mysore, P. Smaragdis, and B. Raj, "Non-negative hidden Markov modeling of audio with application to source separation," in *9th international conference on Latent Variable Analysis and Signal Separation (LCA/ICA)*, St. Malo, France, Sept. 2010.

[8] R. Badeau, B. David, and G. Richard, "High resolution spectral analysis of mixtures of complex exponentials modulated by polynomials," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1341–1350, Apr. 2006.

[9] G. Bierman, *Factorization methods for discrete sequential estimation*. Academic Press, 1977.

[7] Note that contrary to standard high resolution methods, the proposed approach is able to handle mixtures of amplitude-modulated sinusoids starting at different times; it also performs the clustering of these sinusoids into several components according to their temporal dynamics.