NMF WITH TIME-FREQUENCY ACTIVATIONS TO MODEL NON STATIONARY AUDIO EVENTS

Romain Hennequin, Roland Badeau and Bertrand David

Institut TELECOM, TELECOM ParisTech, CNRS LTCI
46, rue Barrault - 75634 Paris Cedex 13 - France
<forename>.<surname>@telecom-paristech.fr

ABSTRACT

Real world sounds often exhibit non-stationary spectral characteristics such as those produced by a harpsichord or a guitar. The classical Non-negative Matrix Factorization (NMF) needs a number of atoms to accurately decompose the spectrogram of such sounds. An extension of NMF is proposed hereafter which includes time-frequency activations based on ARMA modeling. This leads to an efficient single-atom decomposition for a single audio event. The new algorithm is tested on real audio data and shows promising results.

Index Terms— music information retrieval, non-negative matrix factorization, unsupervised machine learning.

1. INTRODUCTION

Human cognition utilizes redundancies to understand visual and audio signals, and several techniques tend to mimic this behavior when decomposing and approximating signals (sounds or images for instance): Principal Component Analysis, Independent Component Analysis or NMF (Lee and Seung [1]) have been introduced both to reduce the dimensionality and to explain the whole data by a few meaningful elementary objects. Thanks to the non-negativity constraint, NMF is able to provide a significant picturing of the data: applied to musical spectrograms it will hopefully decompose them in elementary notes or impulses. The technique is widely used in audio signal processing, with a number of applications such as automatic music transcription [2] and sound source separation [3].

However, the standard NMF is shown to be efficient when the elementary components (notes) of the analyzed sound are nearly stationary. In case of a noticeable spectral variability, the standard NMF will likely need several non-meaningful atoms to decompose a single event, which often leads to a necessary post-processing. To override this drawback, Smaragdis [4] proposes a shift-invariant extension of NMF in which time/frequency templates are factorized from the original data: each atom then corresponds to a time-frequency musical event able to include spectral variations over time. This method gives good results, but does not permit any variation between different occurrences of the same event (atom), its duration and spectral content evolution being fixed.

Durrieu introduces in [5] a source/filter model in a NMF framework in order to model the main melody of musical pieces. This model permits to efficiently take the strong spectral variations of the human voice into account. The filter model is constrained to be a linear combination of pass-band templates while the source is a harmonic template derived from physical modeling. In this paper an extension of NMF is proposed which presents some similarity with Durrieu's model but which includes AutoRegressive Moving Average (ARMA) models estimated from the data, and learns the sources (atoms) in a totally unsupervised way.

In section 2, we introduce the source/filter decomposition as an extension of NMF. In section 3, we derive an iterative algorithm similar to those used for NMF to compute this decomposition. In section 4, we present experiments of source/filter decomposition of the spectrogram of two sounds, and compare this decomposition to the standard NMF. Conclusions are drawn in section 5.

The research leading to this paper was supported by the French GIP ANR under contract ANR-06-JCJC-0027-01, DESAM, and by the Quaero Programme, funded by OSEO, French State agency for innovation.

2. MODEL

2.1. NMF and extension

Given an $F \times T$ non-negative matrix $\mathbf{V}$ and an integer R such that $FR+RT \ll FT$, NMF approximates $\mathbf{V}$ by the product of an $F \times R$ non-negative matrix $\mathbf{W}$ and an $R \times T$ non-negative matrix $\mathbf{H}$:

$$\mathbf{V} \approx \mathbf{WH} \quad \left(\text{i.e. } V_{ft} \approx \sum_{r=1}^{R} w_{fr} h_{rt} \right) \tag{1}$$

This approximation is generally quantified by a cost function $\mathcal{C}(\mathbf{V}, \mathbf{W}, \mathbf{H})$ to be minimized with respect to (wrt) $\mathbf{W}$ and $\mathbf{H}$, which is generally designed element-wise. In this article we will focus on a cost function built with the β-divergence d_β (see [6] for its expression) which includes usual measures: Euclidian distance ($\beta = 2$), Kullback-Leibler divergence ($\beta = 1$) and Itakura-Saito divergence ($\beta = 0$).

When applied to power (squared magnitude) spectrograms, NMF factorizes data into a matrix (or basis) of frequency templates which are the R columns of $\mathbf{W}$ and a matrix $\mathbf{H}$ whose R rows are the temporal vectors of activations corresponding to each template. For a musical signal made of several notes played by the same instrument, it is hoped that the decomposition leads to spectral templates corresponding to single notes or percussive sounds. $\mathbf{H}$ will then display a representation similar to a "piano-roll" (cf. [2]).

This factorization however does not permit to well represent a sound presenting a noticeable spectral evolution. For instance a single note of a plucked string instrument most of the time shows high frequency components which decrease faster than low frequency components. This characteristic is not well modeled with a single frequency template. Several templates are needed which results in a less meaningful decomposition.

To address this issue, we propose an extension of NMF where temporal activations become time/frequency activations. The factorization (1) becomes:

$$V_{ft} \approx \hat{V}_{ft} = \sum_{r=1}^{R} w_{fr} h_{rt}(f) \tag{2}$$

where the activation coefficients are now frequency dependent. To avoid an increase of the problem dimensionality the $h_{rt}(f)$ coefficients are further parameterized by means of ARMA models (paragraph 2.2).

Equation (2) can be interpreted with the help of the source/filter paradigm: the spectrum of each frame of the signal results from the combination of filtered templates (sources). $h_{rt}(f)$ corresponds to the time varying filter associated to the source r. The decomposition thus benefits from the versatility of the source/filter model which is well suited for numerous sound objects.

2.2. AutoRegressive Moving Average (ARMA) Modeling

$h_{rt}(f)$ is chosen following the general ARMA model:

$$h_{rt}^{ARMA}(f) = \sigma_{rt}^2 \frac{\left|\sum_{q=0}^{Q} b_{rt}^q e^{-i2\pi\nu_f q}\right|^2}{\left|\sum_{p=0}^{P} a_{rt}^p e^{-i2\pi\nu_f p}\right|^2}$$

where $\nu_f = \frac{f-1}{2F}$ is the normalized frequency associated to frequency index $f \in \{1, ..., F\}$ (as audio signal are real valued, we only consider frequencies between 0 and the Nyquist frequency). b_{rt}^q are the coefficients of the MA part of the filter and a_{rt}^p those of the AR part. σ_{rt}^2 is the global gain of the filter. For $P = Q = 0$, $h_{rt}^{ARMA}(f)$ no longer depends on f and the decomposition corresponds to a standard NMF with temporal activations σ_{rt}^2.

Defining $\mathbf{a}_{rt} = (a_{rt}^0, \ldots, a_{rt}^P)^T$ and $\mathbf{b}_{rt} = (b_{rt}^0, \ldots, b_{rt}^Q)^T$, time/frequency activations can be rewritten as:

$$h_{rt}^{ARMA}(f) = \sigma_{rt}^2 \frac{\mathbf{b}_{rt}^T \mathbf{T}(\nu_f) \mathbf{b}_{rt}}{\mathbf{a}_{rt}^T \mathbf{U}(\nu_f) \mathbf{a}_{rt}}$$

where $\mathbf{T}(\nu)$ is the $(Q+1)\times(Q+1)$ Toeplitz matrix with $[\mathbf{T}(\nu)]_{pq} = \cos(2\pi\nu(p-q))$ and $\mathbf{U}(\nu)$ is similar to $\mathbf{T}(\nu)$ but of dimension $(P+1)\times(P+1)$. MA only and AR only models are included by respectively taking $P = 0$ and $Q = 0$. It is worth noting that $h_{rt}^{ARMA}(f)$ is always non-negative while there exists neither non-negativity constraint on $\mathbf{b}_{rt}^p$ nor on $\mathbf{a}_{rt}^q$.

The parameterized power spectrogram then becomes:

$$\hat{V}_{ft} = \sum_{r=1}^{R} w_{fr} \sigma_{rt}^2 \frac{\mathbf{b}_{rt}^T \mathbf{T}(\nu_f) \mathbf{b}_{rt}}{\mathbf{a}_{rt}^T \mathbf{U}(\nu_f) \mathbf{a}_{rt}} \tag{3}$$

3. ALGORITHM

We choose a general β-divergence cost function:

$$\mathcal{C}(\mathbf{W}, \mathbf{A}, \mathbf{B}, \mathbf{\Sigma}) = \sum_{f=1}^{F} \sum_{t=1}^{T} d_\beta(V_{ft}, \hat{V}_{ft})$$

with $[\mathbf{W}]_{fr} = w_{fr}$, $[\mathbf{\Sigma}]_{rt} = \sigma_{rt}^2$, $[\mathbf{A}]_{rtp} = a_{rt}^p$ and $[\mathbf{B}]_{rtq} = b_{rt}^q$ and the expression of d_β is given in [6].

The partial derivative of the cost function wrt any variable y (y being any coefficient of $\mathbf{W}$, $\mathbf{\Sigma}$, $\mathbf{A}$ or $\mathbf{B}$) is:

$$\frac{\partial \mathcal{C}(\mathbf{W}, \mathbf{A}, \mathbf{B}, \mathbf{\Sigma})}{\partial y} = \sum_{f=1}^{F} \sum_{t=1}^{T} \hat{V}_{ft}^{\beta-2} (\hat{V}_{ft} - V_{ft}) \frac{\partial \hat{V}_{ft}}{\partial y} \tag{4}$$

The expression of the gradient of $\mathcal{C}$ wrt a vector $\mathbf{y}$ of several coefficients of $\mathbf{A}$ or $\mathbf{B}$ is the same, replacing the partial derivative by a gradient $\nabla_{\mathbf{y}}$ in (4).

This leads to update rules for a multiplicative gradient descent algorithm similar to those used in [1, 6, 4]. In such an iterative algorithm, the update rule associated to one of the parameters is obtained from the partial derivative of the cost function wrt this parameter, written as a difference of two positive terms. In the particular case of a scalar parameter, $\frac{\partial \mathcal{C}}{\partial y} = G_y - F_y$ with $G_y = \sum \hat{V}_{ft}^{\beta-1} \frac{\partial \hat{V}_{ft}}{\partial y}$ and $F_y = \sum \hat{V}_{ft}^{\beta-2} V_{ft} \frac{\partial \hat{V}_{ft}}{\partial y}$ and the update rule for y is:

$$y \leftarrow y \times \frac{F_y}{G_y} \tag{5}$$

This rule particularly ensures that y remains non-negative and becomes constant if the partial derivative is zero.

3.1. Update of frequency templates and filter gains

Following equation (5), we obtain the update rules of $w_{f_0 r_0}$ and $\sigma_{r_0 t_0}^2$:

$$w_{f_0 r_0} \leftarrow w_{f_0 r_0} \frac{\sum_{t=1}^{T} h_{r_0 t}^{ARMA}(f_0) \hat{V}_{f_0 t}^{\beta-2} V_{f_0 t}}{\sum_{t=1}^{T} h_{r_0 t}^{ARMA}(f_0) \hat{V}_{f_0 t}^{\beta-1}} \tag{6}$$

$$\sigma_{r_0 t_0}^2 \leftarrow \sigma_{r_0 t_0}^2 \frac{\sum_{f=1}^{F} w_{f r_0} h_{r_0 t_0}^{ARMA}(f) \hat{V}_{f t_0}^{\beta-2} V_{f t_0}}{\sum_{f=1}^{F} w_{f r_0} h_{r_0 t_0}^{ARMA}(f) \hat{V}_{f t_0}^{\beta-1}} \tag{7}$$

3.2. Update of filters

The update rules of the coefficients of the filters are derived in a similar way, but the updates are not element-wise, but rather "vector-wise": we derive an update rule for each $\mathbf{b}_{rt}$ and for each $\mathbf{a}_{rt}$.

Update of $\mathbf{b}_{rt}$: The gradient of the parameterized spectrogram $\hat{V}_{ft}$ wrt $\mathbf{b}_{r_0 t_0}$ is:

$$\nabla_{\mathbf{b}_{r_0 t_0}} \hat{V}_{ft} = \delta_{t_0 t} \frac{2 w_{f r_0} \sigma_{r_0 t_0}^2}{\mathbf{a}_{r_0 t_0}^T \mathbf{U}(\nu_f) \mathbf{a}_{r_0 t_0}} \mathbf{T}(\nu_f) \mathbf{b}_{r_0 t_0}$$

Then, by substituting this expression into equation (4) with $\mathbf{y} = \mathbf{b}_{r_0 t_0}$, we obtain the gradient of the cost function wrt $\mathbf{b}_{r_0 t_0}$:

$$\begin{aligned} \nabla_{\mathbf{b}_{r_0 t_0}} \mathcal{C} &= 2 \sum_{f=1}^{F} \frac{w_{f r_0} \sigma_{r_0 t_0}^2 \hat{V}_{f t_0}^{\beta-2} (\hat{V}_{f t_0} - V_{f t_0})}{\mathbf{a}_{r_0 t_0}^T \mathbf{U}(\nu_f) \mathbf{a}_{r_0 t_0}} \mathbf{T}(\nu_f) \mathbf{b}_{r_0 t_0} \\ &= 2 \sigma_{r_0 t_0}^2 \left(\mathbf{R}_{r_0 t_0} - \mathbf{R}'_{r_0 t_0} \right) \mathbf{b}_{r_0 t_0} \end{aligned}$$

where:

$$\mathbf{R}_{r_0 t_0} = \sum_{f=1}^{F} \frac{w_{f r_0} \hat{V}_{f t_0}^{\beta-1}}{\mathbf{a}_{r_0 t_0}^T \mathbf{U}(\nu_f) \mathbf{a}_{r_0 t_0}} \mathbf{T}(\nu_f)$$

$$\mathbf{R}'_{r_0 t_0} = \sum_{f=1}^{F} \frac{w_{f r_0} \hat{V}_{f t_0}^{\beta-2} V_{f t_0}}{\mathbf{a}_{r_0 t_0}^T \mathbf{U}(\nu_f) \mathbf{a}_{r_0 t_0}} \mathbf{T}(\nu_f)$$

Both matrices $\mathbf{R}_{r_0 t_0}$ and $\mathbf{R}'_{r_0 t_0}$ are positive definite under mild assumptions. Then, we follow the approach given in [7] and derive the following update rule for the MA part of the filter:

$$\mathbf{b}_{r_0t_0} \leftarrow \mathbf{R}_{r_0t_0}^{-1}\mathbf{R}'_{r_0t_0}\mathbf{b}_{r_0t_0} \quad (8)$$

As $\mathbf{R}_{r_0t_0}$ and $\mathbf{R}'_{r_0t_0}$ are both non singular, $\mathbf{R}_{r_0t_0}^{-1}$ is well defined and $\mathbf{b}_{r_0t_0}$ is ensured to never be zero.

Update of $\mathbf{a}_{rt}$: The update rules of $\mathbf{a}_{rt}$ are derived in the same way as for $\mathbf{b}_{rt}$. Thus, defining:

$$\mathbf{S}_{r_0t_0} = \sum_{f=1}^{F} w_{fr_0}\hat{V}_{ft_0}^{\beta-1}\frac{\mathbf{b}_{r_0t_0}^T\mathbf{T}(\nu_f)\mathbf{b}_{r_0t_0}}{\left(\mathbf{a}_{r_0t_0}^T\mathbf{U}(\nu_f)\mathbf{a}_{r_0t_0}\right)^2}\mathbf{U}(\nu_f)$$

$$\text{and } \mathbf{S}'_{r_0t_0} = \sum_{f=1}^{F} w_{fr_0}\hat{V}_{ft_0}^{\beta-2}V_{ft_0}\frac{\mathbf{b}_{r_0t_0}^T\mathbf{T}(\nu_f)\mathbf{b}_{r_0t_0}}{\left(\mathbf{a}_{r_0t_0}^T\mathbf{U}(\nu_f)\mathbf{a}_{r_0t_0}\right)^2}\mathbf{U}(\nu_f)$$

we derive the following update rule for the AR part of the filter:

$$\mathbf{a}_{r_0t_0} \leftarrow \mathbf{S}'^{-1}_{r_0t_0}\mathbf{S}_{r_0t_0}\mathbf{a}_{r_0t_0} \quad (9)$$

3.3. Description of the algorithm

The update rules (6), (7), (8) and (9) are applied successively to all the coefficients of $\mathbf{W}$, all the coefficients of $\mathbf{\Sigma}$, all the coefficients of $\mathbf{B}$ and all the coefficients of $\mathbf{A}$. Between the updates of each of these matrices (and tensors), the parameterized spectrogram $\hat{\mathbf{V}}$ is recomputed.

Identification: As for the standard NMF, the decomposition (3) which minimizes the cost function is not unique. To reduce identification problems, we impose constraints on $\mathbf{W}$, $\mathbf{\Sigma}$, $\mathbf{B}$ and $\mathbf{A}$:

- for all r and t, we impose that $\mathbf{b}_{rt}$ and $\mathbf{a}_{rt}$ (considered as polynomials) have all their roots inside the unit circle.
- for all r, we impose $\|\mathbf{w}_r\| = 1$ for some norm $\|.\|$.
- for all r and t, we impose $b_{rt}^0 = 1$ and $a_{rt}^0 = 1$.

Thus, at the end of each iteration of our algorithm, we transform $\mathbf{b}_{r,t}$ and $\mathbf{a}_{r,t}$ by replacing roots outside the unit circle by the conjugate of their inverse and accordingly adapting the gain, normalize each column of $\mathbf{W}$, divide $\mathbf{b}_{r,t}$ and $\mathbf{a}_{r,t}$ by their first coefficient and update $\mathbf{\Sigma}$ in order not to change $\hat{V}_{ft}$ by these modifications. All these transformations have no influence on the values of the parameterized spectrogram.

In the remainder of the article, we will refer to this algorithm by the expression "source/filter factorization".

3.4. Dimensionality

In the standard NMF with R atoms, the dimension of the parameterized spectrogram is $\dim\mathbf{W} + \dim\mathbf{H} = R(F+T)$. With our algorithm, the dimension of the parameters is: $\dim\mathbf{W} + \dim\mathbf{\Sigma} + \dim\mathbf{A} + \dim\mathbf{B} = RF + RT(P+Q+1)$. Thus, one should have $RF + RT(P+Q+1) \ll FT$, so P and Q must remain small.

One should notice that our decomposition permits to considerably reduce the number of atoms R needed to accurately fit the data when the parts of the sounds present strong spectral variations, which keeps the dimension of the parameters small.

4. EXAMPLES

In this section several experiments are presented to show that our algorithm is well adapted to decompose sounds having strong spectral variations. It is quite difficult to objectively compare the algorithm with other decomposition algorithms like NMF, as signal models are different. Thus, in this paper, we just present illustrations of the proposed decomposition on real audio data.

All the spectrograms used in these experiments are power spectrograms obtained from recorded signals by means of a short time Fourier transform (STFT).

In lack of any theoretical proof, we chose $\beta = 0.5$ since we empirically observed the monotonic decrease of the cost function and the convergence of the algorithm for $0 \leq \beta \leq 0.5$ over a large set of tests. Moreover the results were more accurate with $\beta = 0.5$ than with $\beta = 0$ (Itakura-Saito divergence).

Algorithms (standard NMF and source/filter factorization) were initialized with random values (except for the filters which were initially flat) and were run until apparent convergence. The algorithms were reinitialized a hundred times (in order to come close to the "best" minimum possible).

4.1. Didgeridoo

4.1.1. Description of the excerpt

In this section our algorithm is applied to a short didgeridoo excerpt. The didgeridoo is an ethnic wind instrument of northern Australia. It makes a continuous modulated sound produced by the vibrations of the lips. The modulations result from the mouth and throat configuration with the help of which the player is able to control several resonances. Figure 1(a) represents the spectrogram of the excerpt: the sound produced is almost harmonic (with some noise) and a strong moving resonance appears in the spectrogram. We can thus consider that this signal is composed of a single event encompassing spectral variations, and try to decompose it with a single atom ($R = 1$). The sampling rate of the excerpt is $f_s = 11025Hz$. We chose a 1024 samples long Hann window with 75% overlap for the STFT.

4.1.2. Experiment and results

The spectrogram of the excerpt is decomposed with a standard NMF algorithm for $R = 1$ atom and $R = 5$ atoms, and with source/filter factorization for $R = 1$ atom, with an order 3 AR modeling ($Q = 0$ and $P = 3$). Reconstructed spectrograms are respectively represented in figures 1(b), 1(c) and 1(d).

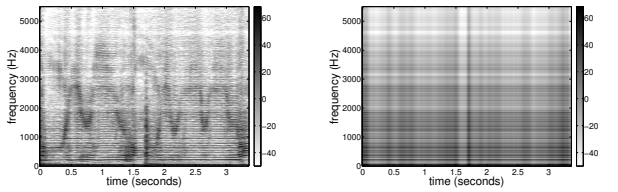
Albeit the didgeridoo is played alone in the analyzed spectrogram, the standard NMF needs many atoms to accurately decompose the power spectrogram. With 1 atom, NMF does not accurately represent the moving resonance (figure 1(b)). With 5 atoms, some spectral variations appears (figure 1(c)), but the resonance trajectory remains a bit unclear. Besides, the signal is not decomposed in a meaningful way (each atom is a part of the sound which has no perceptual meaning) and the dimensionality of the parameters is large ($FR + RT = 3290$).

In opposition to the standard NMF, source/filter factorization permits to accurately represent the spectral variability of the sound (figure 1(d)) with a single atom, keeping the dimensionality low ($FR + TR(Q+1) = 1093$): the moving resonance of the original sound is well tracked, and the total error $\mathcal{C}$ is smaller than that of the standard NMF with $R = 5$. In this case, the decomposition is more efficient and relevant than the standard NMF.

4.2. Harpsichord

4.2.1. Description of the excerpt

In this section our algorithm is applied to a short harpsichord excerpt, composed of two different notes ($C2$ and $E\flat2$): first, the $C2$ is played alone, then the $E\flat2$, and at last, both notes are played simultaneously. The spectrogram of the extract is represented in figure 2(a). As for most of plucked string instruments, high frequency partials of a harpsichord tone decay faster than low frequency partials.

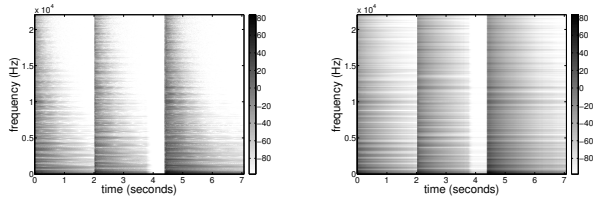


(a) Original spectrogram (b) Standard NMF, $R = 1$

(c) Standard NMF, $R = 5$ (d) Srce/filter factorization, $R = 1$

Fig. 1. Original power spectrogram of the extract of didgeridoo 1(a) and reconstructed spectrograms 1(b), 1(c) and 1(d)

This phenomenon clearly occurs in the L-shaped spectrograms of figure 2(a). The sampling rate of the excerpt is $f_s = 44100Hz$. We chose a 2048 samples long Hann window with 75% overlap for the STFT.



(a) Original spectrogram (b) Standard NMF, $R = 2$

(c) Standard NMF, $R = 6$ (d) Srce/filter factorization, $R = 2$

Fig. 2. Original power spectrogram of the extract of harpsichord 2(a) and reconstructed spectrograms 2(b), 2(c) and 2(d)

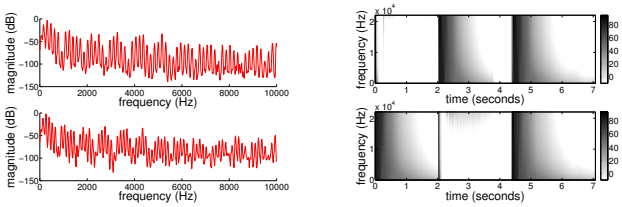
4.2.2. Experiment and results

The spectrogram of the excerpt was decomposed with a standard NMF algorithm for $R = 2$ atoms (1 atom per note) and $R = 6$ atoms, and with source/filter factorization for $R = 2$ atoms, with an ARMA modeling ($Q = 1$ and $P = 1$). Reconstructed spectrograms are respectively represented in figure 2(b), 2(c) and 2(d).

The standard NMF needs several atoms per note to accurately decompose the L-shaped power spectrograms: with only 2 atoms (1 per note played), the faster decay of high frequency content does not appear at all (figure 2(b)). With 6 atoms, the attenuation of high frequency partials appears (figure 2(c)), but each atom is a part of a note spectrum and has no real perceptual meaning.

The ARMA modeling included in our algorithm leads to a good description of the overall spectrogram shape. 2 atoms (1 per note) are enough to accurately fit the original short time spectrum: each atom is harmonic (figure 3(a)) and corresponds to one note while the decay of high frequency partials is clearly well described by the ARMA modeling (see time/frequency activations $h_{rt}^{ARMA}(f)$ in figure 3(b)). The dimensionality of the data provided by our algorithm ($FR + TR(Q + P + 1) = 5704$) remains lower than with a standard NMF with 6 atoms ($FR + RT = 9804$) and the global error $\mathcal{C}$ between the original and the reconstructed spectrogram is lower than that obtained with the standard NMF with $R = 6$.

Thus the decomposition provided by source/filter factorization seems to give a more meaningful representation of the given spectrogram than the one obtained with the standard NMF.



(a) Frequency templates (b) Time/frequency activations

Fig. 3. Source/filter decomposition ($R = 2$, $Q = 1$ and $P = 1$) of the power spectrogram of the harpsichord excerpt

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a new iterative algorithm which is an extended version of the Non-negative Matrix Factorization based on a source/filter representation. We showed that this representation is particularly suitable to efficiently and meaningfully decompose non-stationary sound objects including noticeable spectral variations.

In the future, this extended decomposition could be further developed to deal with small variations of the pitch (like vibrato), for instance by using a constant-Q spectrogram like in [8]. Moreover, quantitative evaluation of the method is under active consideration.

6. REFERENCES

[1] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, October 1999.

[2] P. Smaragdis and J.C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *WASPAA*, New Paltz, NY, October 2003, pp. 177 – 180.

[3] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity," *IEEE Trans. on ASLP*, vol. 15, no. 3, pp. 1066–1074, March 2007.

[4] P. Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs," in *ICA*, Grenada, Spain, September 2004, pp. 494–499.

[5] J.-L. Durrieu, G. Richard, and B. David, "An iterative approach to monaural musical mixture de-soloing," in *ICASSP*, Taipei, Taiwan, April 2009, pp. 105 – 108.

[6] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computation*, vol. 11, no. 3, pp. 793–830, March 2009.

[7] R. Badeau and B. David, "Weighted maximum likelihood autoregressive and moving average spectrum modeling," in *ICASSP*, Las Vegas, Nevada, USA, March 2008, pp. 3761 – 3764.

[8] M. Schmidt and M. Mørup, "Nonnegative matrix factor 2-D deconvolution for blind single channel source separation," in *ICA*, Paris, France, April 2006, vol. 3889 of *Lecture Notes in Computer Science (LNCS)*, pp. 700–707, Springer.