DOES INHARMONICITY IMPROVE AN NMF-BASED PIANO TRANSCRIPTION MODEL ?

François Rigaud [1]*, Antoine Falaize* [2]*, Bertrand David* [1]*, and Laurent Daudet* [3]

[1] Institut Télécom; Télécom ParisTech; CNRS LTCI; Paris, France

[2] Sound Analysis/Synthesis Team; STMS IRCAM-CNRS-UPMC; Paris, France

[3] Institut Langevin; Paris Diderot Univ.; ESPCI ParisTech; CNRS; Paris, France and Institut Universitaire de France

ABSTRACT

This paper investigates how precise a model should be for a robust model-based NMF analysis of piano recordings. While inharmonicity is an essential feature of piano tones from a perceptual point of view, its explicit inclusion in sound models is not straightforward and may even damage the quality of the analysis. Here, we assess the quality of the analysis with a transcription task, and compare three different models for the spectra of the dictionary : one strictly harmonic, one following the theoretical inharmonicity law, and one with relaxed inharmonicity constraints. Experimental results show that both inharmonic models can indeed significantly enhance the results, but only in the case when a good initialization is provided.

Index Terms— non-negative matrix factorization, music transcription, piano, inharmonicity

1. INTRODUCTION

Methods based on non-negative signal representations (such as Non-negative Matrix Factorization - NMF [1] - or Probabilistic Latent Component Analysis - PLCA [2]) have been widely applied to audio analysis in the last decade, giving promising results in transcription [3, 4] or source separation [5, 6], to name only but a few applications. These approaches mainly target the decomposition of time-frequency representations of musical pieces into two non-negative matrices: one dictionary containing the spectra/atoms of the notes/instruments, and one activation matrix containing their temporal activations. Besides the generic "blind" approaches, in order to better fit the decomposition to specific properties of audio data (these can be physics-based or signal-based properties), prior information can also be used explicitly. For instance, when isolated note recordings are available, on the same instrument and with the same recording conditions, the spectra of the dictionary can be learned independently [7, 4]. In that case, since the dictionary is learned on monophonic data and fixed at the learning step, high transcription performances can be obtained. Another approach consists in including the information as a parametrization of the model. For instance, harmonicity [8, 9], temporal evolution of spectral envelop [10], sparsity of simultaneously activated notes [11], beat structure [12], *etc.*, have been considered in modelling the dictionary or the time-activation matrices.

In the case of piano tones, a perceptually important feature is the inharmonicity of their tones : due to the string stiffness [13], the frequencies of the partials slightly deviate from a purely harmonic relation (the frequency f_n of the n-th partial is above nf_0, where f_0 is the fundamental frequency). Taking into account the inharmonicity of the tones in an NMF-based model has been proposed [14], but surprisingly the transcription results were found slightly below those obtained by a simpler harmonic model. These results seem in contradiction with a naive intuition that inharmonicity should help lifting typical transcription ambiguities such as with harmonically-related notes (octave or fifth relations, for instance, where partials fully overlap).

The goal of this paper is to have a better understanding about these issues. For this, two types of inclusion of the inharmonicity, with different degrees of constraints, are proposed and compared to an harmonic model. The first inharmonic model (later called *Inh*) forces the partial's frequency to strictly follow the theoretical inharmonicity law. The second inharmonic model (later called *InhR*) relaxes this constraint, and enhances inharmonicity through a weighted penalty term. We discuss in particular the influence of the initialization on the optimization process. Since it is difficult to gauge intrinsically the quality of NMF decompositions, these are here evaluated on a transcription task, on a large database of piano recordings where we have a ground-truth transcription at hand. It should be emphasized that the proposed algorithms do not target to be competitive with state-of-the-art fully dedicated piano transcription algorithms, since the only information that is taken into account is the inharmonicity of piano tones (for instance, no model of smooth spectral envelop or temporal continuity of the activations is considered). On the contrary, the simple post-processing of the data should allow one to better highlight the differences in the core model.

2. NMF MODELLING FOR PIANO TRANSCRIPTION

NMF consists in solving the approximate low rank factorization:

$$V \approx WH \quad \Leftrightarrow \quad V_{kt} \approx \widehat{V}_{kt} = \sum_{r=1}^{R} W_{kr} H_{rt}, \tag{1}$$

where V, W, and H are respectively $K \times T$, $K \times R$ and $R \times T$ non-negative matrices. In audio application, V usually corresponds to a magnitude or power spectrogram, K being the number of frequency bins and T the number of time-frames. W is then expected to be a dictionary containing the spectra/atoms of R sources or notes, and H a time activation matrix [3].

In the following, a parametric model for the spectra of the dictionary W is introduced. It is based on an additive model for which three different constraints on the partial frequencies are introduced: these are constrained to follow a strict harmonic, a strict inharmonic and a relaxed inharmonic relation.

This work is supported by the DReaM project of the French Agence Nationale de la Recherche (ANR-09-CORD-006, under CONTINT program).

2.1. General additive model for the dictionary of spectra

Each spectrum of a note, indexed by $r \in [1, R]$, is composed of the sum of N_r partials, the partial rank being denoted by $n \in [1, N_r]$. Each partial is parametrized by its amplitude a_{nr} and its frequency f_{nr}. Thus, the set of parameters for a single atom is denoted by $\theta_r = \{a_{nr}, f_{nr} \mid n \in [1, N_r]\}$ and the set of parameters for the dictionary by $\theta = \{\theta_r \mid r \in [1, R]\}$. Then, the expression of a parametric atom indexed by r (as similarly proposed in [9]) is set to:

$$W_{kr}^{\theta_r} = \sum_{n=1}^{N_r} a_{nr} \cdot g_\tau(f_k - f_{nr}), \tag{2}$$

where f_k corresponds to the frequency associated to the kth bin, and g_τ to the magnitude of the Fourier Transform of the τ-length analysis window. In order to limit the computational time and to obtain simple optimization rules, the spectral support of g_τ is restricted to its main lobe. For the experiments proposed in this paper, a Hann window was used to compute the spectrograms. Its main lobe magnitude spectrum (normalized to a maximal magnitude of 1) expression is given by $g_\tau(f_k) = \frac{1}{\pi\tau} \cdot \frac{\sin(\pi f_k \tau)}{f_k - \tau^2 f_k^3}$, for $f_k \in [-2/\tau, 2/\tau]$.

In order to estimate the parameters, the reconstruction cost-function is chosen as the β-divergence between V and $W^\theta H$:

$$C_0(\theta, H) = \sum_{k=1}^{K} \sum_{t=1}^{T} d_\beta \left(V_{kt} \mid \sum_{r=1}^{R} W_{kr}^{\theta_r} \cdot H_{rt} \right), \tag{3}$$

with

$$d_\beta(x \mid y) = (x^\beta + (\beta - 1).y^\beta - \beta.x.y^{\beta-1})/(\beta.(\beta - 1)). \tag{4}$$

The family of β-divergences is widely used in audio application [15] because it encompass three common metrics: an Euclidean distance ($\beta = 2$), the Kullback-Leibler ($\beta \to 1$) and the Itakura-Saito ($\beta \to 0$) divergence.

2.2. Constraints on partial frequencies

• *Strictly harmonic / Ha-NMF*

The strict harmonic constraint consists in fixing:

$$f_{nr} = nF_{0r}, \quad n \in \mathbb{N}^*, \tag{5}$$

directly in the parametric model (eq. (2)). Then, the set of parameters for a single atom reduces to $\theta_r^{\text{Ha}} = \{a_{nr}, F_{0r} \mid n \in [1, N_r]\}$.

• *Strictly inharmonic / Inh-NMF*

The strict inharmonic constraint consists in setting [13]:

$$f_{nr} = nF_{0r}\sqrt{1 + B_r n^2}, \quad n \in \mathbb{N}^*, \tag{6}$$

that rules the partials frequencies of stiff strings with clamped ends. Here, $\theta_r^{\text{Inh}} = \{a_{nr}, F_{0r}, B_r \mid n \in [1, N_r]\}$, and B_r is called the inharmonicity coefficient. It is dependent on the piano string design and then differs from one piano to another but also from one key to another. From the low bass to the high treble range it can vary from around 10^{-5} to 10^{-2} [16].

• *Inharmonic relaxed / InhR-NMF*

An alternative way of enforcing inharmonicity is through an extra penalty term to the reconstruction cost function C_0 [17]. Thus, the global cost function can be expressed as:

$$C^{\text{InhR}}(\theta, \gamma, H) = C_0(\theta, H) + \lambda \cdot C_1(f_{nr}, \gamma), \tag{7}$$

where the set of parameters of the constraint is denoted by $\gamma = \{F_{0r}, B_r \mid r \in [1, R]\}$. λ is a parameter that sets the weight between the reconstruction cost error and the inharmonicity constraint. The constraint cost function C_1 is chosen as the sum on each note of the mean square error between the estimated partial frequencies f_{nr} and those given by the inharmonicity relation:

$$C_1(f_{nr}, \gamma_r) = \sum_{r=1}^{R} \sum_{n=1}^{N_r} \left(f_{nr} - nF_{0r}\sqrt{1 + B_r n^2} \right)^2. \tag{8}$$

A potential benefit of this relaxed formulation is to allow a slight deviation of the partial frequencies around the theoretical inharmonicity relation, that is typically observed in the low frequency range due to the coupling between the strings and the soundboard [13, 18].

3. TRANSCRIPTION ALGORITHM

3.1. NMF optimization

3.1.1. Multiplicative update rules

As commonly proposed in NMF modelling, the optimization is performed iteratively, using multiplicative update rules. These are obtained in a similar way to [9]. In the following, $P(\theta^*)$ and $Q(\theta^*)$ refer to positive quantities obtained by decomposing the partial derivative of a cost function $C(\theta)$ with relation to a particular parameter θ^* so that $\frac{\partial C(\theta)}{\partial \theta^*} = P(\theta^*) - Q(\theta^*)$. The parameter is then updated as $\theta^* \leftarrow \theta^* \cdot Q(\theta^*)/P(\theta^*)$.

The updates for H and a_{nr} are identical for the 3 models (*Ha / Inh / InhR*), since these parameters only appear in C_0 cost function. The rule for H is similar to standard NMF with β-divergence [15]:

$$H \leftarrow H \otimes \frac{{}^t W^\theta \left(\widehat{V}^{\cdot[\beta-2]} \otimes V \right)}{{}^t W^\theta \, \widehat{V}^{\cdot[\beta-1]}}, \tag{9}$$

where ${}^t.$ corresponds to the transpose operator, and $\otimes$, $\cdot^{[]}$, $\frac{\cdot}{\cdot}$, to entry-wise multiplication, exponentiation and division operators, respectively. $\widehat{V} = W^\theta H$ is the spectrogram model.
The rule for a_{nr} is obtained by a decomposition similar to [9]:

$$a_{nr} \leftarrow a_{nr} \cdot \frac{\sum_{k,t} (g_\tau(f_k - f_{nr}).H_{rt}) .\widehat{V}_{kt}^{\beta-2}.V_{kt}}{\sum_{k,T} (g_\tau(f_k - f_{nr}).H_{rt}) .\widehat{V}_{kt}^{\beta-1}}. \tag{10}$$

Then, update rules of the remaining parameters are specific for each of the three different NMF models. In the following, $g'_\tau(f_k)$ represents the derivative of $g_\tau(f_k)$ with respect to f_k on the spectral support of the main lobe. For a Hann window (normalized to a maximal magnitude of 1) and $f_k \in [-2/\tau, 2/\tau]$:

$$g'_\tau(f_k) = \frac{1}{\pi\tau} \frac{(3\tau^2 f_k^2 - 1)\sin(\pi\tau f_k) + \pi\tau(f_k - \tau^2 f_k^3)\cos(\pi\tau f_k)}{(f_k - \tau^2 f_k^3)^2}.$$

• *Ha-NMF / Inh-NMF*

$$F_{0r} \overset{\text{Ha / Inh}}{\leftarrow} F_{0r} \cdot \frac{Q_0(F_{0r})}{P_0(F_{0r})}, \tag{11}$$

$$B_r \overset{\text{Inh}}{\leftarrow} B_r \cdot \frac{Q_0^{\text{Inh}}(B_r)}{P_0^{\text{Inh}}(B_r)}, \tag{12}$$

with

$$P_0(F_{0r}) = \sum_{k,t}\left[\left(\sum_{n=1}^{N_r} a_{nr}\frac{-C.f_k.g'_\tau(f_k-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-1} + \left(\sum_{n=1}^{N_r} a_{nr}\frac{-C.f_{nr}.g'_\tau(f_k-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-2}.V_{kt}\right] \quad (13)$$

$$Q_0(F_{0r}) = \sum_{k,t}\left[\left(\sum_{n=1}^{N_r} a_{nr}\frac{-C.f_k.g'_\tau(f_k-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-2}.V_{kt} + \left(\sum_{n=1}^{N_r} a_{nr}\frac{-C.f_{nr}.g'_\tau(f-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-1}\right] \quad (14)$$

$$P_0^{\text{Inh}}(B_r) = \sum_{k,t}\left[\left(\sum_{n=1}^{N_r} a_{nr}\frac{-D.f_k.g'_\tau(f_k-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-1} + \left(\sum_{n=1}^{N_r} a_{nr}\frac{-D.f_{nr}.g'_\tau(f_k-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-2}.V_{kt}\right] \quad (15)$$

$$Q_0^{\text{Inh}}(B_r) = \sum_{k,t}\left[\left(\sum_{n=1}^{N_r} a_{nr}\frac{-D.f_k.g'_\tau(f_k-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-2}.V_{kt} + \left(\sum_{n=1}^{N_r} a_{nr}\frac{-D.f_{nr}.g'_\tau(f-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-1}\right]. \quad (16)$$

For the strict harmonic model (*Ha-NMF*), $f_{nr} = nF_{0r}$ and $C = \partial f_{nr}/\partial F_{0r} = n$. For the strict inharmonic model (*Inh-NMF*), $f_{nr} = nF_{0r}\sqrt{1+B_r n^2}$, $C = \partial f_{nr}/\partial F_{0r} = n\sqrt{1+B_r n^2}$, and $D = \partial f_{nr}/\partial B_r = n^3 F_{0r}/2\sqrt{1+B_r n^2}$.

• *InhR-NMF:*

$$f_{nr} \overset{\text{InhR}}{\leftarrow} f_{nr}\cdot\frac{Q_0^{\text{InhR}}(f_{nr})+\lambda\cdot Q_1^{\text{InhR}}(f_{nr})}{P_0^{\text{InhR}}(f_{nr})+\lambda\cdot P_1^{\text{InhR}}(f_{nr})}, \quad (17)$$

$$B_r \overset{\text{InhR}}{\leftarrow} B_r\cdot\frac{Q_1^{\text{InhR}}(B_r)}{P_1^{\text{InhR}}(B_r)}, \quad (18)$$

$$F_{0r} \overset{\text{InhR}}{=} \frac{\sum_{n=1}^{N_r} f_{nr}n\sqrt{1+B_r n^2}}{\sum_{n=1}^{N_r} n^2(1+B_r n^2)}, \quad (19)$$

where

$$P_0^{\text{InhR}}(f_{nr}) = \sum_{k,t}\left[\left(a_{nr}\frac{-f_k.g'_\tau(f_k-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-1} + \left(a_{nr}\frac{-f_{nr}.g'_\tau(f_k-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-2}.V_{kt}\right], \quad (20)$$

$$Q_0^{\text{InhR}}(f_{nr}) = \sum_{k,t}\left[\left(a_{nr}\frac{-f_k.g'_\tau(f_k-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-2}.V_{kt} + \left(a_{nr}\frac{-f_{nr}.g'_\tau(f-f_{nr})}{f_k-f_{nr}}.H_{rt}\right).\widehat{V}_{kt}^{\beta-1}\right], \quad (21)$$

$$P_1^{\text{InhR}}(f_{nr}) = 2f_{nr}, \quad Q_1^{\text{InhR}}(f_{nr}) = 2nF_{0r}\sqrt{1+B_r n^2}, \quad (22)$$

$$P_1^{\text{InhR}}(B_r) = F_{0r}\sum_{n=1}^{N_r} n^4, \quad Q_1^{\text{InhR}}(B_r) = \sum_{n=1}^{N_r}\frac{n^3 f_{nr}}{\sqrt{1+B_r n^2}}. \quad (23)$$

Note that for F_{0r}, an exact analytic solution is obtained (eq. (19)) when cancelling the partial derivative of the cost function C_1.

3.1.2. Practical considerations

• *Amplitude normalization:* In order to obtain a unique decomposition, the a_{nr} are normalized to a maximal value of 1 for each atom indexed by r. Thus, after each update $\forall r \in [1, R]$ and $\forall n \in [1, N_r]$ of the a_{nr}, the following steps are applied:

$$\begin{aligned} A_r &= \max_{n\in[1,N_r]}(a_{nr}), \quad \forall r \in [1,R], \\ a_{nr} &= a_{nr}/A_r, \quad \forall r \in [1,R],\ \forall n \in [1,N_r], \\ H &= \text{diag}(A)\cdot H, \end{aligned} \quad (24)$$

where $\text{diag}(A)$ is a diagonal matrix of dimension $R \times R$ containing the $A_r, \forall r \in [1, R]$.

• *Algorithms:* The steps of *InhR-NMF* algorithm are summarized in table Algorithm 1. *Ha-NMF* and *Inh-NMF* algorithms are not given in this paper but are highly similar to *InhR-NMF*: instead of updating f_{nr}, then F_{0r} and B_r in lines 12-18, F_{0r} (*Ha-NMF*) or F_{0r} and B_r (*Inh-NMF*) are directly updated and W^θ recomputed.

Algorithm 1 InhR-NMF optimization

1: **Input:** V mag. spectrogram (normalized to a max. value of 1)
2: **Initialization:** $\forall\, r \in [1, R]$, $n \in [1, N_r]$,
3: $>(B_r, F_{0r})$, then $f_{nr} = nF_{0r}\sqrt{1+B_r n^2}$ / $a_{nr} = 1$
4: $> W^\theta$ (*cf.* eq (2)) / H ini. with rand. positive values
5: $> \beta$ / λ
6: **Optimization:**
7: **for** $it = 1$ to I **do**
8: $> H$ update (eq. (9))
9: $> a_{nr}$ update $\forall\, r \in [1, R]$, $n \in [1, N_r]$ (eq. (10))
10: a_{nr} normalization and H update (eq. (24))
11: W^θ update (eq. (2))
12: $> f_{nr}$ update $\forall\, r \in [1, R]$, $n \in [1, N_r]$ (eq (17))
13: W^θ update (eq. (2))
14: **for** $u = 1$ to 30 **do**
15: $\forall r, n \in N_r$
16: $> F_{0r}$ update (*cf.* eq (19))
17: $> B_r$ update (20 times) (*cf.* eq (18))
18: **end for**
19: **end for**
20: **Output:** H, a_{nr}, f_{nr}, B_r, F_{0r}

3.2. Post-processing

In order to obtain a list that contains the detected notes, their onset and offset time, a post-processing is applied to the activation matrix H. Each line is processed by a low-pass differentiator filter. The obtained matrix dH is then scaled so that its maximal element is 1. Finally, in each line, an onset is detected if dH is above a threshold $10^{T_{\text{on}}/20}$, and the corresponding offset found when dH crosses (from negative to positive values) a second threshold $-10^{T_{\text{off}}/20} < 0$. If the same note is found to be repeatedly played at less than 100 ms of interval it is then considered as a unique note. As discussed in the introduction, this very simple post-processing has been chosen in order to better highlight the differences in the model itself.

4. TRANSCRIPTION EVALUATION

4.1. Protocol

The three models have been tested on the MAPS database[1]. 45 pieces were randomly chosen (5 out of 30 for each of the 9 pianos),

[1] http://www.tsi.telecom-paristech.fr/aao/en/category/database/

re-sampled to 22050 Hz, and 30 seconds excerpts were taken starting from $t_0 = 5$ s. The mean polyphony level by time-frame is about 3.23. In order to estimate an appropriate value for the parameter λ of the *InhR-NMF* method, a learning set composed of 9 pieces was similarly built (1 piece for each piano, none of them in the test set). The spectrograms were computed with a Hann window of length $\tau = 90$ ms, with a hope–size of $\tau/8$ and a 2^{13}-point FFT. The number of spectra in the dictionary was fixed to $R = 64$, and initialized for notes having MIDI note number in $[33, 96]$. Usually the piano keyboard contains 88 notes, from A0 (21) to C8 (108). Our choice corresponds to a reduction of one octave in the extreme bass (where the spectral resolution is not sufficient to perform the analysis) and one octave in the high treble range (where the non-linear coupling between triplets of strings at the soundboard [18] produces complex spectra with multiple partials that cannot be fully explained by a simple harmonic or inharmonic model). However, these notes in the extreme parts of the keyboard are rarely played. Over the complete MAPS dataset (352710 notes for 159 different pieces), they only account for 1.66% of the notes.

For *Ha-NMF*, F_{0r} is initialized to exact equal temperament. For *Inh-NMF* and *InhR-NMF* two initializations are tested. The first one sets F_{0r} to equal temperament (no "octave stretching") and B_r to 5.10^{-3}, $\forall r \in [1, R]$. The second one is based on an average model of inharmonicity and tuning on the whole piano tessitura considering invariants in piano string design and tuner choices [16].

The experiments are run for $\beta = 1$ (KL divergence), 150 iterations and different values of N_r, the number of partials of the atoms (from 5 to 30). For the post-processing, the onset detection threshold T_{on} is varying from -80 to -1 dB in steps of 1 dB, and the offset detection threshold is fixed to -80 dB.

4.2. Results and discussion

For each excerpt, performances are evaluated in terms of Precision ($\mathcal{P}$), Recall ($\mathcal{R}$) and F-measure ($\mathcal{F}$) [19] (one note is assumed to be correctly detected if for a given pitch, the estimated onset time is contained in +/-50 ms interval around the ground truth). Mean and standard deviation are then computed on the whole dataset.

On the learning database, the influence of λ parameter of *InhR-NMF* with $N_r = 10$ has been studied on a grid covering the range $[10^{-6}, 100]$ and logarithmically distributed. The optimal F-measure was obtained for $\lambda = 1$, and it should be noted that the performance did not depend on a fine tuning of this parameter.

The performances obtained on the test database are presented in table 4.2 for all the methods and for T_{on}=-18 dB (the influence of T_{on} is depicted in figure 1 for *Ha/Inh2/InhR2-NMF* and $N_r = 20$). Standard deviations are not reported in the table but are around 10 to 14 %. For each method, increasing the number of partials N_r leads to higher performances. It seems that adding more partials avoids a situation where high notes explain high rank partials belonging to lower notes. *Ha-NMF* results for $N_r = 30$ are comparable to those obtained by the NMF under the harmonicity constraint presented in [20] (section V.B) for similar experimental setups.

For the first initialization, *Inh-NMF* and *InhR-NMF* perform less than *Ha-NMF* (this is consistent with the observation in [14]). Conversely, for the second initialization with the mean model of inharmonicity and piano tuning, these methods perform significantly better than *Ha-NMF* (ANOVA p-values lower than 0.05 for $N_r < 30$). Furthermore, both inharmonic models give comparable mean F-measures (p-values higher than 0.5).

This tends to demonstrate that such strategies are highly dependent on the initialization. Indeed, the reconstruction cost function C_0 is non-convex with respect to f_{nr}, F_{0r} or B_r parameters, and present a large amount of local minima (this has been checked by computing the cost function on a grid of these parameters). Hence, multiplicative update rules (as well as other optimization methods based on gradient descent) cannot ensure that these parameters will be correctly estimated. Regarding the results, taking into account the dispersion of the partial frequencies from a theoretical inharmonic relation in *InhR-NMF* does not seem valuable, when compared to *Inh-NMF*.

	N_r	5	10	20	30
Ha	$\mathcal{P}$	39.3	51.9	58.9	62.1
	$\mathcal{R}$	41.0	49.3	56.6	60.6
	$\mathcal{F}$	38.7	49.0	55.9	59.7
Inh 1	$\mathcal{P}$	32.6	44.6	54.9	55.6
	$\mathcal{R}$	34.5	46.0	56.7	57.7
	$\mathcal{F}$	32.4	43.8	54.0	54.8
InhR 1	$\mathcal{P}$	34.0	43.0	53.4	55.2
	$\mathcal{R}$	36.0	44.9	55.7	57.1
	$\mathcal{F}$	33.7	42.6	52.7	54.4
Inh 2	$\mathcal{P}$	44.3	59.6	66.4	66.9
	$\mathcal{R}$	45.1	55.9	60.9	62.5
	$\mathcal{F}$	**43.0**	**55.8**	**61.5**	**62.6**
InhR 2	$\mathcal{P}$	43.3	57.5	64.1	64.6
	$\mathcal{R}$	45.2	54.9	60.3	61.1
	$\mathcal{F}$	42.5	54.1	60.2	60.7

Table 1. Mean Precison, Recall and F-measure, in %, as a function of N_r for *Ha/Inh/InhR-NMF* algorithms, for T_{on}= -18 dB. Index 1 and 2 refer to the two different initializations of B_r and F_{0r}.

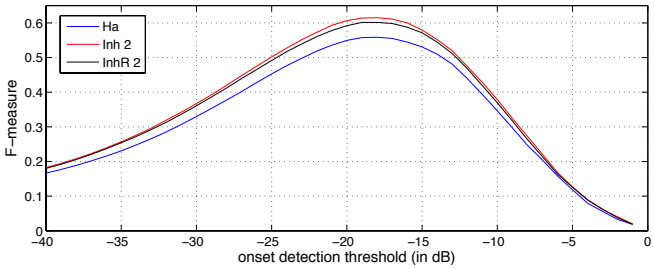

Fig. 1. F-measure as a function of the onset detection threshold $T_{on} \in [-40, -1]$ for *Ha/Inh2/InhR2-NMF* methods and $N_r = 20$.

5. CONCLUSION

Including inharmonicity in parametric NMF models has been shown to be relevant in a piano transcription task, provided that the inharmonicity and tuning parameters are sufficiently well initialized. More precisely, an initialization with the same average value for the inharmonicity of all notes, and equal temperament for the tuning, turns out to provide worse estimates than the simpler purely harmonic model. However, a note-dependent inharmonicity law, with fixed parameters, and the corresponding "stretched" tuning curves, provide a good initialization to our models, that lead to significant improvement in the transcription results. Further work will investigate how these models on partials frequencies can be combined with amplitude models (smooth spectral envelopes), or frame dependencies in time.

6. REFERENCES

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[2] Paris Smaragdis, Bhiksha Raj, and Madhusudana Shashanka, "Sparse and shift-invariant feature extraction from non-negative data," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 2069–2072.

[3] Paris Smaragdis and Judith C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proc. of the IEEE Workshop on Application of Signal Processing to Audio and Acoustics (WASPAA)*, 2003, pp. 177–180.

[4] Arnaud Dessein, Arshia Cont, and Guillaume Lemaitre, "Real-time polyphonic music transcription with non-negative matrix factorization and beta-divergence.," in *Proc. of the 11th Int. Soc. for Music Information Retrieval Conference (ISMIR)*, 2010, pp. 489–494.

[5] Jean-Louis Durrieu, Gaël Richard, Bertrand David, and Cédric Févotte, "Source/filter model for unsupervised main melody extraction from polyphonic audio signals," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 564–575, March 2010.

[6] Tuomas Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, March 2007.

[7] Bernhard Niedermayer, "Non-negative matrix division for the automatic transcription of polyphonic music," in *Proc. of the 9th Int. Soc. for Music Information Retrieval conference (ISMIR)*, 2008, pp. 544–549.

[8] Nancy Bertin, Roland Badeau, and Emmanuel Vincent, "Enforcing harmonicity and smoothness in bayesian non-negative matrix factorization applied to polyphonic music transcription," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 18, no. 3, pp. 538–549, March 2010.

[9] Romain Hennequin, Roland Badeau, and Bertrand David, "Time-dependent parametric and harmonic templates in non-negative matrix factorization," in *Proc. of the 13th Int. Conf. on Digital Audio Effects (DAFx-10)*, September 2010, pp. 246–253.

[10] Romain Hennequin, Roland Badeau, and Bertrand David, "Nmf with time-frequency activations to model nonstationary audio events," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 19, no. 4, pp. 744–753, May 2011.

[11] Patrik O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[12] Kazuki Ochiai, Hirokazu Kameoka, and Shigeki Sagayama, "Explicit beat structure modeling for non-negative matrix factorization-based multipitch analysis," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2012, pp. 133–136.

[13] N. H. Fletcher and T. D. Rossing, *THE PHYSICS OF MUSICAL INSTRUMENTS. 2nd Ed.*, pp. 64–69, 352–398, Springer, 1998.

[14] Emmanuel Vincent, Nancy Bertin, and Roland Badeau, "Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 109–112.

[15] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu, "Non-negative matrix factorization with the itakura-saito divergence. with application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, March 2009.

[16] François Rigaud, Bertrand David, and Laurent Daudet, "A parametric model of piano tuning," in *Proc. of the 14th Int. Conf. on Digital Audio Effects (DAFx-11)*, September 2011, pp. 393–399.

[17] François Rigaud, Bertrand David, and Laurent Daudet, "Piano sound analysis using non-negative matrix factorization with inharmonicity constraint," in *Proc. of the 20th European Signal Processing Conference (EUSIPCO)*, August 2012, pp. 393–399.

[18] G. Weinreich, "Coupled piano strings," *J. Acoust. Soc. Am.*, vol. 62, no. 6, pp. 1474–1484, 1977.

[19] C.J. van Rijsbergen, *INFORMATION RETRIEVAL*, Butterworths, London, UK, 2nd edition, 1979.

[20] Emmanuel Vincent, Nancy Bertin, and Roland Badeau, "Adaptive harmonic spectral decomposition for multiple pitch estimation," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 18, no. 3, pp. 528–537, March 2010.