PIANO SOUND ANALYSIS USING NON-NEGATIVE MATRIX FACTORIZATION WITH INHARMONICITY CONSTRAINT

François Rigaud and Bertrand David

Institut Telecom; Telecom ParisTech;
CNRS LTCI; Paris, France
firstname.lastname@telecom-paristech.fr

Laurent Daudet

Institut Langevin; Paris Diderot Univ.;
ESPCI ParisTech; CNRS; Paris, France
and Institut Universitaire de France
firstname.lastname@espci.fr

ABSTRACT

This paper presents a method for estimating the tuning and the inharmonicity coefficient of piano tones, from single notes or chord recordings. It is based on the Non-negative Matrix Factorization (NMF) framework, with a parametric model for the dictionary atoms. The key point here is to include as a relaxed constraint the inharmonicity law modelling the frequencies of transverse vibrations for stiff strings. Applications show that this can be used to finely estimate the tuning and the inharmonicity coefficient of several notes, even in the case of high polyphony. The use of NMF makes this method relevant when tasks like music transcription or source/note separation are targeted.

Index Terms**—** non-negative matrix factorization, piano tuning, inharmonicity coefficient estimation

1. INTRODUCTION

The precise estimation of F_0, the fundamental frequency adjusted by the tuner, and of B, the inharmonicity coefficient of piano notes, has been dealt by several studies (e.g. [1, 2, 3]) and, to our knowledge, has always been achieved from single note recordings. In the polyphonic case, and for tasks such as transcription or source separation, these parameters are sometimes taken into account, but they are rarely jointly and finely estimated for all the played notes. For example, in [4], the (B, F_0) parameters are learned on some single note recordings and interpolated on the tessitura. In [5, 6], they are jointly, roughly estimated during a preprocessing step.

In this paper, we propose a Non-negative Matrix Factorization (NMF) framework to finely estimate (B, F_0) from the analysis of single notes or chord recordings, assuming that the played notes are known. Given a non-negative matrix V of dimension $K \times T$, the NMF [11] aims at finding an approximate factorization:

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

where W and H are non-negative matrices of dimensions $(K \times R)$ and $(R \times T)$, respectively. In the case of music analysis, V corresponds to the magnitude (or power) spectrogram of an audio excerpt, k corresponds to the frequency bin index and t to the frame index. Thus, W represents a dictionary containing the spectra (or atoms) of the R sources, and H represents their time-frame activations. Recently, harmonic structure [7, 8], temporal evolution of spectral envelopes [9] and vibrato [7] have been introduced as a parametrization of the matrices W and / or H, in order to take explicitly into account specific properties of different musical sounds.

The idea of this work is to introduce the parameters (B, F_0) of each note as constraints for the estimation of the partials frequencies in the matrix W. In [10], the ideal inharmonicity law was directly introduced in the parametrization of the spectra, however this led to a decrease of piano transcription performances when compared to a purely harmonic constraint. The key idea here is to include this parametrization as a relaxed constraint, in order to finely estimate every partial amplitude and frequency corresponding to a transverse vibration of the strings at the same time as (B, F_0). Note that in this paper, V is not strictly speaking a spectrogram but a set of magnitude spectra computed from single notes or chord recordings. Because the played notes are known, the elements of H are fixed to one when a note is played and zero when it is not. Thereby, only the magnitude spectra of the dictionary W are optimized on the data.

In order to quantify the quality of the approximation of (1), a distance (or divergence) is estimated. If the metric is separable, it can be expressed as:

$$D(V \mid WH) = \sum_{k=1}^{K} \sum_{t=1}^{T} d\left(V_{kt} \mid \widehat{V}_{kt}\right). \qquad (2)$$

In audio applications, the family of β-divergences is widely used [12], because it encompasses 3 common metrics: $\beta = 2$

This research was partially funded by the French Agence Nationale de la Recherche, PAFI project.

for the Euclidian distance, $\beta = 1$ for the Kullback-Leibler divergence and $\beta = 0$ for the Itakura-Saito divergence. These distances are used to define a cost function which is minimized with respect to W and H, respectively. The mathematical expressions which are given in this paper are derived within the general framework of β-divergences. The results presented in the application section are obtained for the Kullback-Leibler divergence:

$$d_{\beta=1}(x \mid y) = x(\log x - \log y) + (y - x). \tag{3}$$

The rest of this paper is constructed as follows: the general NMF model is extended to our framework in section 2, where a model for the magnitude spectra of the notes and a soft inclusion of the inharmonicity constraint are presented. A multiplicative algorithm is then proposed to solve the optimization problem. Section 3 presents the experimental validation, for the application to the estimation of (B, F_0) on single notes and chord recordings. Finally, section 4 discusses a few perspectives for future work.

2. MODEL AND OPTIMIZATION

This section first introduces a general model for a parametric atom composed of a sum of partials. The information of the inharmonicity of piano sounds is then included (section 2.2) as a relaxed constraint on the frequencies of the partials. Finally, the multiplicative update rules are given to compute the optimization of the model on the data.

2.1. General parametric atom

The general parametric atom used in this work is based on the additive model proposed in [7]. Each spectrum of a note, indexed by $r \in [1, R]$, is composed of the sum of N_r partials. The partial rank is 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 is denoted by $\theta = \{\theta_r \mid r \in [1, R]\}$. Finally, the expression of a parametric atom is given by:

$$W_{kr}^{\theta_r} = \sum_{n=1}^{N_r} a_{nr} \cdot g_\tau(f_k - f_{nr}), \tag{4}$$

where f_k is the frequency of the bin with index k and $g_\tau(f_k)$ the magnitude of the Fourier transform of the analysis window of size τ. Here, we limit the spectral support of $g_\tau(f_k)$ to its main lobe to obtain a simple expression of the update rules (*cf.* [7]) and a faster optimization. The results presented in this paper are obtained for a Hanning window. Its main lobe magnitude spectrum (normalized to a maximal magnitude of 1) is given by $g_\tau(f_k) = \frac{1}{\pi\tau} . \frac{\sin(\pi f_k \tau)}{f_k - \tau^2 f_k^3}$, for $f_k \in [-2/\tau, 2/\tau]$.

In order to learn the parameters of the model from the data, a cost function is defined using the β-divergence:

$$C_0(\theta, H) = \sum_{k=1}^{K} \sum_{t=1}^{T} d_\beta \left(V_{kt} \mid \sum_{r=1}^{R} W_{kr}^{\theta_r} \cdot H_{rt} \right). \tag{5}$$

2.2. Inclusion of the inharmonicity constraint

Piano tones are known to be inharmonic (see for instance [13]): the frequencies of the transverse vibration of an ideal plain stiff string with fixed endpoints are given by

$$f_n = nF_0 \sqrt{1 + Bn^2}, \quad n \in \mathbb{N}^*, \tag{6}$$

where F_0 is the fundamental frequency of a flexible string and B the inharmonicity coefficient. B depends on the piano string design and can be about $[10^{-5}, 10^{-2}]$ along the tessitura. In the spectrum, inharmonicity results in a slight frequency shift of every partial from the harmonic law nF_0, and the higher the rank of the partial, the largest the deviation. This ideal model does not take into account the bridge coupling between the strings and the soundboard, which modifies the partial frequencies, mainly in the low frequency domain [14].

The inharmonicity law (6) has already been introduced in a previous study on parametric NMF [10], constraining the partials frequencies to exactly follow the ideal inharmonic law. However, this study was not conclusive, as this inharmonic model had poorer performance in a task of piano transcription than the simpler harmonic constraint. In the model, given equation (4), it would correspond to a reduction of the N_r parameters f_{nr} of each note to only 2 parameters $\{B_r, F_{0r}\}$. In contrast with this previous study, inharmonicity is here included as a relaxed constraint, allowing for a local adaptation of the frequency of each partial, while constraining the entire set of partials to globally follow an inharmonic law. At the same time, for each partial it allows a slight frequency deviation from the inharmonicity law, as for instance due to the bridge coupling with the soundboard. The set of parameters related to the constraint is denoted by $\gamma = \{F_{0r}, B_r \mid r \in [1, R]\}$. Finally a new cost function is built by adding a regularization term:

$$C(\theta, \gamma, H) = C_0(\theta, H) + \lambda_1 C_1(f_{nr}, \gamma), \tag{7}$$

where $C_1(f_{nr}, \gamma)$ is defined as the sum on each note of the Euclidian distance between the estimated partial frequencies f_{nr} and those given by the inharmonicity law, normalized by the number of partials:

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

The empirical parameter λ_1 is fixed and sets the weight of the constraint in the global cost function.

2.3. Optimization algorithm

The optimization is performed using multiplicative update rules for a_{nr} and f_{nr} parameters[1]. B_r and F_{0r} parameters are updated by means of a simplex search method (as implemented in the *fminsearch* MATLAB$^{\text{TM}}$ function). The rules for a_{nr} and f_{nr} are obtained from the decomposition of the partial derivatives of the cost function given in equation (7), in a similar way to [7] :

$$a_{nr} \quad \leftarrow \quad a_{nr} \cdot \frac{Q_0(a_{nr})}{P_0(a_{nr})}, \tag{9}$$

$$f_{nr} \quad \leftarrow \quad f_{nr} \cdot \frac{Q_0(f_{nr}) + \lambda_1 \cdot Q_1(f_{nr})}{P_0(f_{nr}) + \lambda_1 \cdot P_1(f_{nr})}, \tag{10}$$

where

$$P_0(a_{nr}) = \sum_{k=1}^{K} \sum_{t=1}^{T} \left[\left(g_\tau(f_k - f_{nr}).h_{rt} \right).\widehat{V}_{kt}^{\beta-1} \right], \tag{11}$$

$$Q_0(a_{nr}) = \sum_{k=1}^{K} \sum_{t=1}^{T} \left[\left(g_\tau(f_k - f_{nr}).h_{rt} \right).\widehat{V}_{kt}^{\beta-2}.V_{kt} \right], \tag{12}$$

$$P_0(f_{nr}) = \sum_{k,t} \left[\left(a_{nr} \frac{-f_k.g'_\tau(f_k - f_{nr})}{f_k - f_{nr}}.h_{rt} \right).\widehat{V}_{kt}^{\beta-1} + \left(a_{nr} \frac{-f_{nr}.g'_\tau(f_k - f_{nr})}{f_k - f_{nr}}.h_{rt} \right).\widehat{V}_{kt}^{\beta-2}.V_{kt} \right], \tag{13}$$

$$Q_0(f_{nr}) = \sum_{k,t} \left[\left(a_{nr} \frac{-f_k.g'_\tau(f_k - f_{nr})}{f_k - f_{nr}}.h_{rt} \right).\widehat{V}_{kt}^{\beta-2}.V_{kt} + \left(a_{nr} \frac{-f_{nr}.g'_\tau(f - f_{nr})}{f_k - f_{nr}}.h_{rt} \right).\widehat{V}_{kt}^{\beta-1} \right], \tag{14}$$

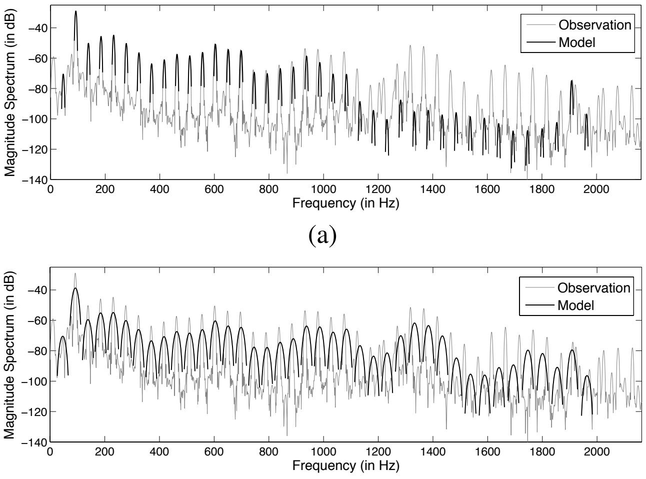
$$P_1(f_{nr}) = 2f_{nr}/N_r, \tag{15}$$

$$Q_1(f_{nr}) = 2nF_{0r}\sqrt{1 + B_r n^2}/N_r, \tag{16}$$

are all positive quantities. $g'_\tau(f_k)$ represents the derivative of $g_\tau(f_k)$ with respect to f_k on the spectral support of the main lobe.

3. APPLICATIONS

The proposed model is applied as an estimator of (B, F_0) on single notes and chord recordings taken from RWC [16] and MAPS [2] databases. Instead of processing spectrograms, the observation matrix V is built by concatenating magnitude spectra computed on each recording. A 500 ms Hanning window is used in order to get a sufficient spectral resolution for a good estimation of (B, F_0). For the estimation on single note recordings, V contains 88 columns corresponding to the 88 notes, from A0 to C8. Because we assume that the processed notes are known, H is set to the identity matrix of dimensions 88×88. For the estimation on chord recordings, the same protocol is applied for building V and H is filled with ones when a note is known to be present and zeros otherwise.



(a)

(b)

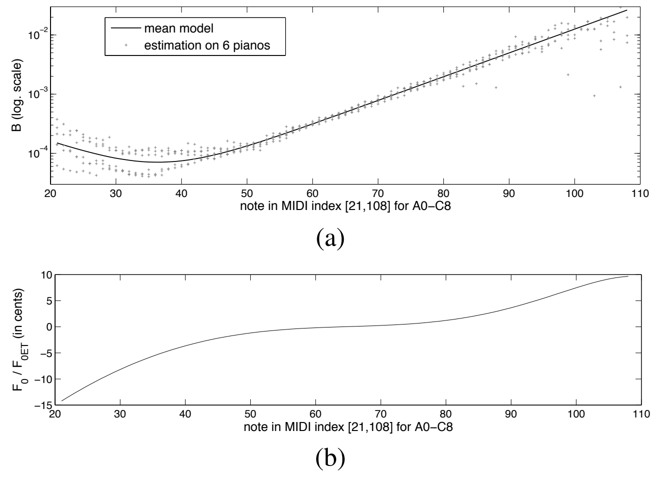
Fig. 1: Initialization for the analysis of a note $Gb1$. The observed magnitude spectrum is computed from a 500 ms Hanning window. (a) Despite a crude initialization of (B, F_0), the 23 first partials of the model and the data are overlapping. (b) The width of the partials of the model is increased to overlap a greater number of the partials of the data.

3.1. Initialization of W

A good initialization of W results in a majority of partials of the model overlapping the corresponding partials in the data. Because of the spectral width of the partials (linked to τ, the length of the analysis window), even a rough initialization of (B, F_0) leads to the overlap of the first partials (see figure 1(a)). In order to set the best possible initialization, we use the model of (B, F_0) along the tessitura proposed in [15]. From 6 different types of pianos (upright and grand pianos), mean curves of (B, F_0) are estimated by taking into account the invariances in the piano string set design and tuning rules. Results are depicted on figure 2. (B_r, F_{0r}) are initialized according to this mean model along the tessitura and then the f_{nr} according to the inharmonic law given equation (6). The a_{nr} are initialized to 1. Moreover, τ is initialized with a smaller value than the one used for the analysis window, in order to have partials with larger main lobe (see figure 1(b)). During the optimization, τ is gradually increased to its final value, i.e. the length of the analysis window.

[1] For a transcription task, H could be updated with standard NMF multiplicative rules [12].

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



(a)

(b)

Fig. 2: (a) B model along the tessitura estimated from 6 pianos estimates (see [15] for details). (b) F_0 model along the tessitura obtained from the tuning model averaged on 6 pianos. F_0 is depicted as the deviation from equal temperament in cents.

3.2. Dealing with partials initialized in noise

In practice, if too many partials are initialized in noisy frequency bands, they can get stuck and therefore lead to bad estimates of the inharmonicity law. For each iteration of the optimization algorithm, we cancel their influence in the estimation of the physical parameters γ by removing them from the regularization term given in equation (8), and by re-initializing them on the current inharmonic law.

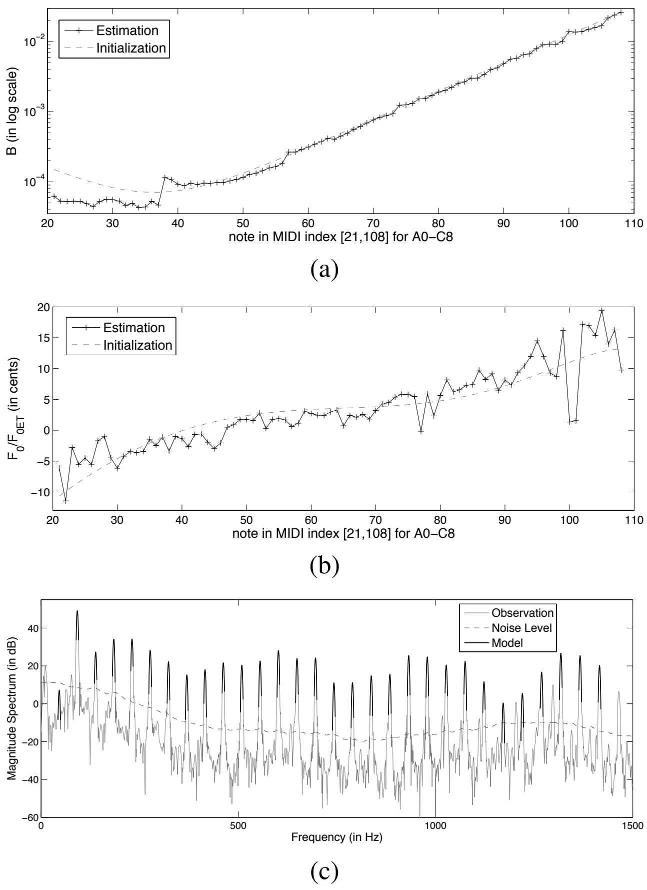
Then,

$$C_1(f_{nr}, \gamma) = \sum_{r=1}^{R} \frac{1}{\sharp\Delta_r} \sum_{n\in\Delta_r} \left(f_{nr} - nF_{0r}\sqrt{1+B_r n^2} \right)^2, \tag{17}$$

where Δ_r is the set of reliable partials (not located in noise) of the note indexed by r, and $\sharp\Delta_r$ its cardinal. For the proposed application, we first compute the noise level[3] $\mathrm{NL}(f_k)$ on each magnitude spectrum composing the matrix V, and at each iteration we look for the estimated partials that have a magnitude greater than the noise. Thus, we define the set of reliable partials of each note, being above the noise level, by $\Delta_r = \{n \mid a_{nr} > \mathrm{NL}(f_{nr}),\ n \in [1, N_r]\}$. This criterion is taken into account in the update rules (15) and (16) by replacing N_r by $\sharp\Delta_r$.

3.3. (B, F_0) estimation on the whole tessitura from single note recordings

N_r, the number of partials is computed as $\arg\min_{N_r}(30,$ $f_{N_r,r} < F_s/2)$, where F_s is the sampling frequency (22050 Hz in our experiments). We set $\beta = 1$ (Kullback-Leibler divergence) and $\lambda_1 = 10^{-1}$.



(a)

(b)

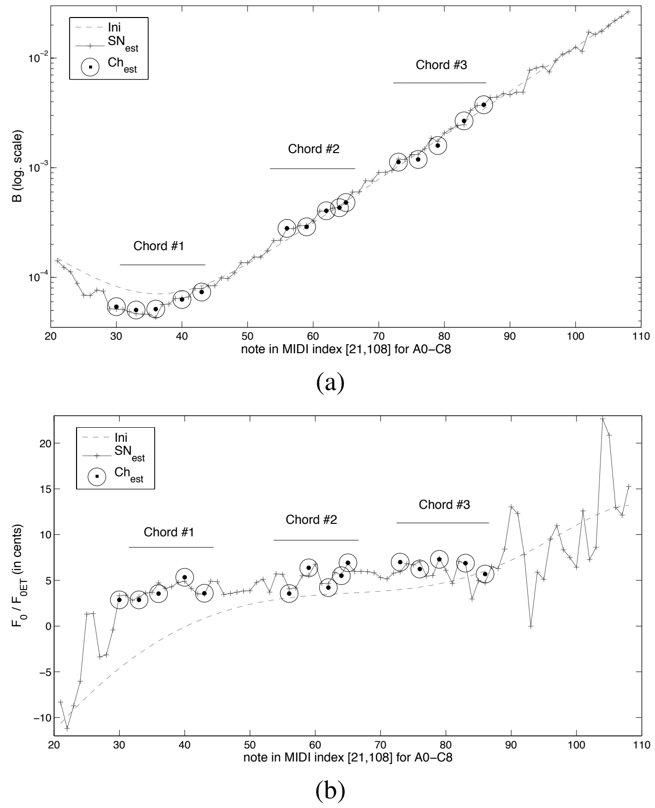
(c)

Fig. 3: (a) B along the tessitura estimated for the 2^{nd} grand piano of RWC database. (b) F_0 along the tessitura estimated (depicted as the deviation from equal temperament in cents). (c) Results of the partial estimation for the note Gb1 (index 30 in MIDI norm).

The results for the second grand piano of RWC are depicted on figure 3. The bass break between the bass and the treble bridges, which produces a discontinuity in B, is well estimated around the notes 37 and 38 (in MIDI index). The result of the algorithm for the estimation of the partial magnitudes and frequencies of the note Gb1 (MIDI note index 30) are depicted on figure 3(c). Each partial corresponding to a transverse vibration of the strings has been correctly estimated. Here, the inharmonic constraint avoids the selection of partials corresponding to longitudinal vibrations of the strings (visible around 1300 Hz).

In order to compare the estimation of (B, F_0) on single note and chord recordings, we apply the same method for the grand piano of the MAPS database (synthesized using high-quality samples). The curves are presented on figure 4.

[3] The method to compute the noise level is described in the appendix of [15].



(a)

(b)

Fig. 4: B (a) and F_0 (b) along the tessitura. In dashed line the initialization. '+' markers correspond to the single note estimation and circles to the 3 chord estimations.

3.4. Multiple (B, F_0) estimation on chord recordings

The same protocol as for the single note estimation is used, except that now the maximum number of partials N_r of the model is set to 15. The chords are taken in the treble, medium and bass range of the piano. The results for the grand piano of MAPS database are depicted with black circles and compared to the single note estimation values ('+' markers) on figure 4. We can see that even for a high degree of polyphony (here 5) the values of (B, F_0) of each note composing the chords are properly estimated when compared with the reference values obtained from single note estimation.

4. CONCLUSION

We introduced a model for estimating the inharmonicity coefficient and the tuning of piano tones. For now, the method has been applied to magnitude spectra computed from single note and chord recordings assuming that the played notes are known, and the preliminary results presented in this study validate our model. Because the method allows a simultaneous estimation of the amplitude and the frequency of each partial (corresponding to a transverse vibration) of each note, it will be interesting to extend it to piano note separation from a spectrogram, informed by the score (to initialize the matrix H). Another interesting application will be to learn the inharmonicity coefficients and the tuning of the piano from a whole musical piece : because the curves of B along the tessitura are related to the piano design, we may in ideal conditions be able to infer from these measurements the model of the piano and the tuning done by the tuner.

5. REFERENCES

[1] A. Galembo and A. Askenfelt, "Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano," *IEEE Transaction on Speech and Audio Processing*, vol. 7, pp. 197–203, march 1999.

[2] Simon Godsill and Manuel Davy, "Bayesian computational models for inharmonicity in musical instruments," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, oct. 2005.

[3] J. Rauhala, H.M. Lehtonen, and V. Välimäki, "Fast automatic inharmonicity estimation algorithm," *JASA Express Letters*, vol. 121, 2007.

[4] L.I. Ortiz-Berenguer and F.J. Casajús-Quirós, "Polyphonic transcription using piano modelling for spectral pattern recognition," in *Proc. of the* 5^{th} *Int. Conf. on Digital Audio Effects (DAFx-02)*, Sept. 2002.

[5] Valentin Emiya, Roland Badeau, and Bertrand David, "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 6, 2010.

[6] Emmanouil Benetos and Simon Dixon, "Joint multi-pitch detection using harmonic envelope estimation for polyphonic music transcription," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1111–1123, October 2011.

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

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

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

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

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

[12] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence. with application to music analysis.," *Neural Computation*, vol. 21, no. 3, March 2009.

[13] P.M. Morse, *Vibration and Sound*, American Institute of Physics for the Acoustical Society of America, 1948.

[14] G. Weinreich, "Coupled piano strings," *The Journal of the Acoustical Society of America*, vol. 62, no. 6, pp. 1474–1484, 1977.

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

[16] M. Goto, T. Nishimura, H. Hashiguchi, and R. Oka, "Rwc music database: Music genre database and musical instrument sound database," in *ISMIR*, 2003, pp. 229–230.