A parametric model and estimation techniques for the inharmonicity and tuning of the piano[a)]

François Rigaud[b)] and Bertrand David
TSI Department, Institut Télécom, Télécom ParisTech, CNRS LTCI, 37-39 rue Dareau, 75014 Paris, France

Laurent Daudet[c)]
Institut Langevin, Paris Diderot University, ESPCI ParisTech, CNRS UMR 7587, 1 rue Jussieu, 75005 Paris, France

(Received 29 October 2012; revised 26 February 2013; accepted 20 March 2013)

Inharmonicity of piano tones is an essential property of their timbre that strongly influences the tuning, leading to the so-called octave stretching. It is proposed in this paper to jointly model the inharmonicity and tuning of pianos on the whole compass. While using a small number of parameters, these models are able to reflect both the specificities of instrument design and tuner's practice. An estimation algorithm is derived that can run either on a set of isolated note recordings, but also on chord recordings, assuming that the played notes are known. It is applied to extract parameters highlighting some tuner's choices on different piano types and to propose tuning curves for out-of-tune pianos or piano synthesizers.
 [http://dx.doi.org/10.1121/1.4799806]

PACS number(s): 43.75.Zz, 43.75.Mn, 43.60.Uv [TRM] Pages: 3107–3118

I. INTRODUCTION

Modeling timbre variations of a specific musical instrument across its whole compass is an issue of particular importance to musical acoustics, for instance, in synthesis or instrument recognition. It could also be useful for many other tasks related to Music Information Retrieval (MIR) such as automatic music transcription or source separation. The case of the piano is particularly relevant, as it has been central to Western music in the last two centuries, with an extremely wide solo or orchestral repertoire. In this paper, a parametric model which accounts for specifics of both the piano type and tuning is proposed. More specifically, the variations of the string inharmonicity and tuning are modeled along the whole pitch range and estimated from monophonic or polyphonic recordings.

Despite considerable differences in shape, size, and design, all pianos share construction elements: Keyboard, hammers, steel strings, bridges, soundboard—all these contribute to its distinct timbre. In return, these physical characteristics lead to strong constraints on the tuning technique, which targets equal temperament (ET).[1,2] Whereas the transverse vibrations of an ideal string produce spectra with harmonically related partials, the stiffness of actual piano strings leads to a slight inharmonicity.[3] For instance, the frequency ratio between the second and first partials is slightly higher than 2, between the third and second it is higher than 3:2, and so on. This effect depends on many physical parameters of the strings (material, length, diameter, etc.), and then differs not only from a piano to another, but also from one note to another. As a consequence, simply adjusting the first partial of each note on ET would produce unwanted beatings, in particular for octave intervals. Aural tuning consists of controlling these beatings.[4] The final result then depends not only on the specific design of each piano, but also on the tuner practice—he usually focuses on particular beatings, which may not necessarily be the same for different tuners.[5,6] Thus, according to the model of the piano and the choices/abilities of the tuner, the resulting tuning is unique, but within some physically-based constraints. From a musical acoustics perspective its modeling is hence an interesting challenge that has been tackled by different viewpoints. A simulation of aural piano tuning has been proposed[2] to help pianists in tuning their own pianos, replicating the tuner's work by iteratively tuning different intervals. The method is based on a mathematical computation of the beat rates, and requires the frequencies of the first five partials of each note. More recently, an approach based on psycho-acoustic considerations has been introduced.[7] This algorithm adjusts the 88 notes at the same time, by an optimization procedure on modified spectra of the notes according to psycho-acoustic laws and tuning updates. Besides these works, a number of authors have proposed algorithms to estimate inharmonicity from isolated note recordings (cf. state-of-the-art in Sec. III A).

This paper takes a different global approach by jointly modeling tuning and inharmonicity laws for the whole compass. This global estimate is made possible thanks to recent advances in optimization techniques, here based on a non-negative decomposition scheme. The model can be run, with no hand tuning of the parameters, either on isolated notes or chord recordings, assuming that we know which notes are being played. On sets of isolated notes for the whole 88-notes compass, this model compares favorably with some algorithms of the state of the art. However, to the best of our knowledge, it is the only approach that can still build a

---

[a)]Portions of this work were presented in two conference proceedings: Rigaud *et al.*, in *Proceedings of the 14th International Conference on Digital Audio Effects (DAFx-11)*, pp. 393–399 (2011) and Rigaud *et al.*, in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO* 2012), pp. 393–399 (2012).

[b)]Author to whom correspondence should be addressed. Electronic mail: francois.rigaud@telecom-paristech.fr

[c)]Also at: Institut Universitaire de France, 103 boulevard Saint-Michel, 75005 Paris, France.

global model from a small subset of the notes, or even from chord recordings. Although such an interpolated model can only capture the main trends in the inharmonicity and tuning of a given piano, it should be reminded that one of the objectives of piano manufacturing and tuning is to have a timbre that is as homogeneous as possible, smoothing out as much as possible the discontinuities of physical origins: Bass break, change in the number of strings per note, change of strings diameter, and winding. Therefore, it is not only realistic, but also relevant, to try to globally parametrize the inharmonicity and tuning with only a few parameters—at least as a first-order approximation.

The obtained synthetic description of a particular instrument, in terms of its tuning/inharmonicity pattern, can be useful to assess its state and also provides clues on some of the tuner's choices. In the field of musical acoustics, the use of such a model could be helpful, for instance, for the tuning of physically-based piano synthesizers, where we are otherwise faced with the problem of having to adjust a large number of parameters, all of them being inter-dependent. Here a higher-level control can be obtained, with few physically meaningful parameters. In the fields of audio signal processing and MIR, including *a priori* knowledge is often done when trying to enhance the performance of the algorithms.[8–12] The herein proposed method is a first step for further use in tasks such as piano model identification or automatic transcription of polyphonic piano recordings.

The joint model of inharmonicity and tuning on the whole compass is introduced in Sec. II. The estimation of the parameters is then presented in Sec. III. Section IV describes the results obtained from experimental data and discusses possible applications. Finally, conclusions and perspectives are drawn in Sec. V.

II. PARAMETRIC MODEL OF INHARMONICITY AND PIANO TUNING

A. Inharmonicity and aural tuning principles

First consider the transverse vibration of a plain stiff string fixed at end-points. Because of the bending stiffness, the resulting partial frequencies are given by an inharmonic relation[13]

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

where n is the partial rank, B is the inharmonicity coefficient, and F_0 is the fundamental frequency of vibration of an ideal flexible string. F_0 is related to the speaking length of the string L, the tension T, and the linear mass μ according to

$$F_0 = \frac{1}{2L}\sqrt{\frac{T}{\mu}}. \tag{2}$$

The stiffness is taken into account in B

$$B = \frac{\pi^3 E d^4}{64\,TL^2}, \tag{3}$$

where E is the Young's modulus and d is the diameter of the plain string. Since the mechanical characteristics of the strings differ from one note to another, obviously F_0 but also B are varying along the compass (typical values for B are in the range $[10^{-5}, 10^{-2}]$). Hereafter, these quantities will be indexed by the MIDI note number, $m \in [21, 108]$ (from $A0$ to $C8$) as $(B(m), F_0(m))$.

It is worth noting that Eq. (1) assumes a string fixed at both ends and then neglects the bridge motion. The actual partials deviate[14] upwards or downwards from the frequencies given in Eq. (1), mostly in the low frequency domain. Moreover, the coupling between doublets or triplets of strings can lead to multiple partials and produce double decays and beatings in piano tones.[15–17] These phenomena are not considered in the model presented in this paper although they could slightly affect the estimation results (this is discussed in Secs. IV B 1 and IV C).

Aural tuning is based on the perception and the control of beatings between partials of two different tones played simultaneously,[4] and is then affected by inharmonicity. It always begins by the tuning of a reference note, in most cases the $A4$ at 440 Hz (sometimes 442 Hz). To do so, the tuner adjusts the tension of the strings to cancel the beatings produced by the difference of frequency of the tuning fork and the first partial of the note. Thus, $f_1(m=69) = 440$ Hz. Even if there are different methods, skilled tuners usually begin by the scale tuning sequence: the $F3$ to $F4$ octave is set by approximate ET.[4,6] The rest of the keyboard is tuned by adjusting beatings between the partials of two different notes, typically octave-related.

When tuning an octave interval by canceling the beatings produced by the second partial of a note indexed by m and the first partial of a note indexed by $m+12$, the resulting frequency ratio $f_1(m+12)/f_1(m)$ is higher than 2 because $f_2(m) > 2f_1(m)$. This phenomenon is called octave stretching. Depending on where the notes are in the range of the compass, the amount of stretching can be different. This fact is linked to the underlying choice of the octave type (related to perceptual effects and tuner's personal choices) during the tuning.[5] For instance, in a 4:2 type octave, the fourth partial of the reference note is matched to the second partial of its octave. Depending on the position in the compass, the piano can be tuned according to different octave types: 2:1, 4:2, 6:3, 8:4, etc., or a trade-off between two. This means that the tuner may not focus only on canceling beatings between a pair of partials, but that he controls an average beating generated by a few partials of the two notes.

In order to highlight this stretching, the tuning along the compass is usually depicted as the deviation, in cents (¢), of the first partial frequency of each note from ET,

$$d(m) = 1200 \cdot \log_2 \frac{f_1(m)}{F_{0,\mathrm{ET}}(m)}, \tag{4}$$

where $F_{0,\mathrm{ET}}(m)$ is the theoretical fundamental frequency given by the ET,

$$F_{0,\mathrm{ET}}(m) = 440 \cdot 2^{(m-69)/12}. \tag{5}$$

Usually[1,13] the stretching increases gradually from the mid-range (deviation about ± 5¢) to the extreme parts of the keyboard, producing deviations down to -30¢ in the low bass and up to $+30$¢ in the high treble. The goal of the proposed model is to explain the main variations of $d(m)$ along the compass (also known as the Railsback curve) by taking into account the piano string set design characteristics [model of $B(m)$ along the compass] and the tuner's choices (model related to the octave type).

B. A parametric model for inharmonicity and tuning

The proposed model which simulates aural tuning on the whole compass is based on octave interval tunings. Its successive steps are a simplified version of those actually performed by a tuner, but the most important global considerations (stretching inherent in the inharmonicity and the octave type choice) are taken into account. The model starts by tuning all the octave intervals relative to a reference note (for example, the $A4$ at 440 Hz). From these notes, the tuning is then interpolated on the whole compass. Finally, the possibility of a global deviation is added in order to allow for different tuning frequencies for the reference note.

1. Octave interval tuning

When tuning an "upper" octave interval (for instance $A5$ from $A4$), the cancellation of the beatings produced by the 2ρth partial ($\rho \in \mathbb{N}^*$) of a reference note, indexed by m ($A4$), and the ρth partial of its octave, indexed by $m+12$ ($A5$), can be done by tuning $F_0(m+12)$ such as

$$F_0(m+12) = 2\,F_0(m)\sqrt{\frac{1+B(m)\cdot 4\rho^2}{1+B(m+12)\cdot \rho^2}}. \tag{6}$$

Equation (6) clearly shows the influence of the note-dependent inharmonicity coefficient (B) and of the octave type (related to ρ) in the stretching of the octave. In the case of "lower" octave tuning (for instance $A3$ from $A4$), the same relation can be inverted and applied by considering $m+12$ ($A4$) as the reference note and m ($A3$) as the note to tune. Sections II B 2 and II B 3, respectively, describe parametric models for B and ρ along the whole compass.

2. Whole compass model for B

a. String set design influence on B. In order to keep a homogeneous timbre along the compass, the strings are designed in such a way that discontinuities due to physical parameter variations are smoothed.[18–20] Three main design considerations might produce such discontinuities in B along the keyboard: The bass break between the bass and treble bridges (jump[18] in L), the transitions between adjacent keys having a different number of strings (jump[13,19] in T), and the transition between plain strings to wrapped strings (jump[19] in d).

On the treble bridge, from $C8$ note downwards, B is decreasing because of the increase of L. Down to middle C ($C4$ note, $m=60$), the values of B are roughly the same for all the pianos and B follows a straight line in logarithmic scale.[3] This result is mainly due to the fact that string design in this range is standardized, since it is not constrained by the limitation of the piano size.[18]

In the low pitch range, the strings use a different bridge (the bass bridge) to keep a reasonable size of the instrument. Then, the linear mass of the strings is increased in order to adjust the value of F_0 according to Eq. (2). Instead of increasing only the diameter d, which increases B, the strings are wound with a copper string wire, which increases the linear mass. Thus, on the bass bridge, B is increasing from the sharpest notes downwards. Note that the number of keys associated with the bass bridge and the design of their strings are specific to each piano.

b. Parametric model. According to the string design considerations, B could be modeled by two distinct functions corresponding to the two bridges, and could present discontinuities at the bass break or at the changes single-doublets and doublets-triplets of strings. The difficulty when modeling B on the whole compass is to know the position of these possible discontinuities, because it is specific to each piano model. Therefore, we propose a "continuous" additive model on the whole compass, discretized for $m \in [21, 108]$. We denote it by $B_\xi(m)$, ξ being the set of modeling parameters.

Usually the evolution of B along the keyboard is depicted in logarithmic scale and presents two linear asymptotes. We denote by $b_T(m)$ [respectively, $b_B(m)$] the treble bridge (respectively, the bass bridge) asymptote of $\log B_\xi(m)$. Each asymptote is parametrized by its slope and its Y-intercept,

$$\begin{cases} b_T(m) = s_T \cdot m + y_T, \\ b_B(m) = s_B \cdot m + y_B. \end{cases} \tag{7}$$

According to Young,[3] $b_T(m)$ is similar for all the pianos so s_T and y_T are fixed parameters. Then, the set of free (piano dependent) parameters reduces to $\xi = \{s_B, y_B\}$. $B_\xi(m)$ is set as the sum of the contributions of these two curves [Eq. (7)] in the linear scale

$$B_\xi(m) = e^{b_B(m)} + e^{b_T(m)}. \tag{8}$$

It should be emphasized that this additivity does not arise from physical considerations, but it is the simplest model that smoothes discontinuities between the bridges. Experimental data will show that it actually describes well the variations of B in the transition region around the two bridges.

The model is presented in Fig. 1(a) for three different typical values of the set of parameters: ξ_1, ξ_2, and ξ_3, corresponding to low, medium, and high inharmonic pianos, respectively. The asymptotes corresponding to the bass and treble bridges are also drawn for $B_{\xi_2}(m)$.

3. Whole compass model for ρ

The octave tuning relation, given in Eq. (6), considers the cancellation of the beatings produced by a single pair of partials. In practice, the deviation $F_0(m+12)/2F_0(m)$ could be a weighted sum of the contribution of two pairs of partials, because the amount of stretching may result from a compromise between two octave types.[5] An alternative model to take into account this weighting is to allow non-integer

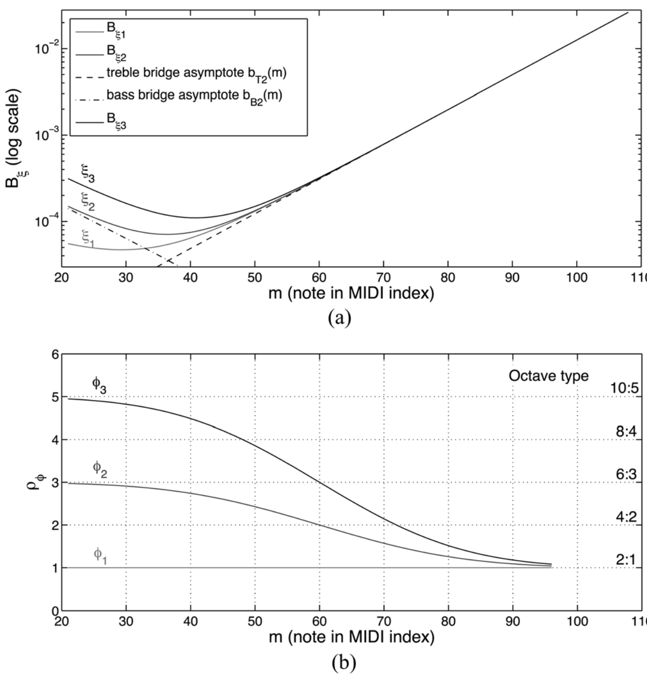

FIG. 1. Model for (a) inharmonicity coefficient $B_\xi(m)$ and (b) octave type parameter $\rho_\phi(m)$ along the compass for different values of the sets of parameters.

values for $\rho \in [1, +\infty[$. For example, if the octave tuning is a compromise between a 2:1 and 4:2 type octave, ρ will be in the interval $[1, 2]$. This model loses the physical meaning because ρ is not anymore related to a partial rank; it will however be shown in Sec. III B 2 that it allows the inversion of Eq. (6), in order to estimate ρ from the data.

We choose arbitrarily to model the evolution of ρ along the compass as follows:

$$\rho_\phi(m) = \frac{\kappa}{2} \cdot \left(1 - \operatorname{erf}\left(\frac{m - m_0}{\alpha}\right)\right) + 1, \tag{9}$$

with erf the error function, and $\phi = \{\kappa, m_0, \alpha\}$ the set of parameters. Note that ρ_ϕ is indexed by the note m, and not by the note $m + 12$ [cf. Eq. (6)]. It is then defined for $m \in [21, 96]$. κ is related to the value of the asymptote in the low bass range. m_0 is a parameter of translation along m, and α rules the slope of the decrease.

This model expresses the fact that the amount of stretching inherent in the octave type choice is decreasing from the low bass to the high treble range and that it is limited by horizontal asymptotes at each extremity. It may be justified by the fact that the perception of the pitch of complex tones is not only based on the first partial of the notes, but on a set of partials contained in the "dominant region" of the human hearing.[21–23] For bass tones (with fundamental frequencies around 100 to 400 Hz, i.e., in the range $G2$ to $G4$, $m \in [43, 67]$), this dominant region covers the third to fifth partials.[22] While going up to the treble part of the compass, the dominant region tends to be localized on the partials with a lower rank. For tones having a first partial frequency above 1400 Hz (i.e., for a higher note than $F6$, $m = 89$) the perception of the pitch is mainly linked to the first partial.[23] Then, in the model high treble asymptote is set to 1. It corresponds to the minimal octave type (2:1), and means that the tuner focuses on the first partial of the highest note. In the low bass range, the asymptote is set by the value of $\kappa + 1$.

The model is represented in Fig. 1(b) for three different values of the set of parameters: ϕ_1, ϕ_2, and ϕ_3, respectively, corresponding to a low, mid, and high octave type choice in the low bass range.

4. Interpolation on the whole compass

From the estimation of the sets of parameters, ξ related to the design of the strings, and ϕ related to the choices of the tuner, it is possible to tune all the octaves of a reference note. If $A4$ is tuned such as $f_1(m = 69) = 440$ Hz, all the A notes of the keyboard can be iteratively tuned by using Eq. (6). To complete the tuning on the whole compass, a Lagrange polynomial interpolation is performed on the deviation from ET of the tuned notes of the model [computed by using Eq. (4)]. The interest of this method is that the interpolated curve is constrained to coincide with the initial data. The interpolated model of deviation from ET is denoted by $d_{\xi,\phi}(m)$.

5. Global deviation

Finally, in order to take into account the fact that the reference note is not necessarily an $A4$ at 440 Hz (other tuning forks exist, for instance $A4$ at 442 Hz or $C5$ at 523.3 Hz) we add in the model the possibility of a global "detuning." In the representation of the deviation from ET in cents, it corresponds to a vertical translation of the curve. Then, the deviation from ET of the model is set to $d_{\xi,\phi}(m) + d_g$, where d_g is an extra parameter of the model, corresponding to the global deviation.

The whole compass tuning model is depicted in Fig. 2 for different values of the sets of parameters ξ and ϕ (corresponding to those used in Fig. 1), and for $d_g = 0$. The tuning of the A notes from an $A4$ at 440 Hz is indicated with black dots on the middle curves. Figure 2(a) corresponds to the influence on the tuning of B_ξ (for ξ_1, ξ_2, and ξ_3), for ϕ_2 fixed. Since the string design is standardized in the range $C4$ to $C8$, the tuning changes significantly only in the bass range. Figure 2(b) represents the influence on the tuning of ρ_ϕ (for ϕ_1, ϕ_2, and ϕ_3), for ξ_2 fixed. Its influence is visible on the whole compass but it is mainly important in the bass range, where it can produce a deviation up to -20¢.

III. ESTIMATION OF THE PARAMETERS FROM ISOLATED NOTE AND CHORD RECORDINGS

A. Automatic estimation of (B, F_0)

The parameter estimation of the proposed tuning model requires a prior precise estimation of (B, F_0) of several notes along the compass. This task has been dealt with by several authors, and often achieved from isolated note recordings. For instance, Galembo and Askenfelt[24] carry it out by means of an inharmonic comb filtering of the magnitude spectrum. The output of the comb filter is computed on a grid of (B, F_0) and the maximal value is selected (after a local

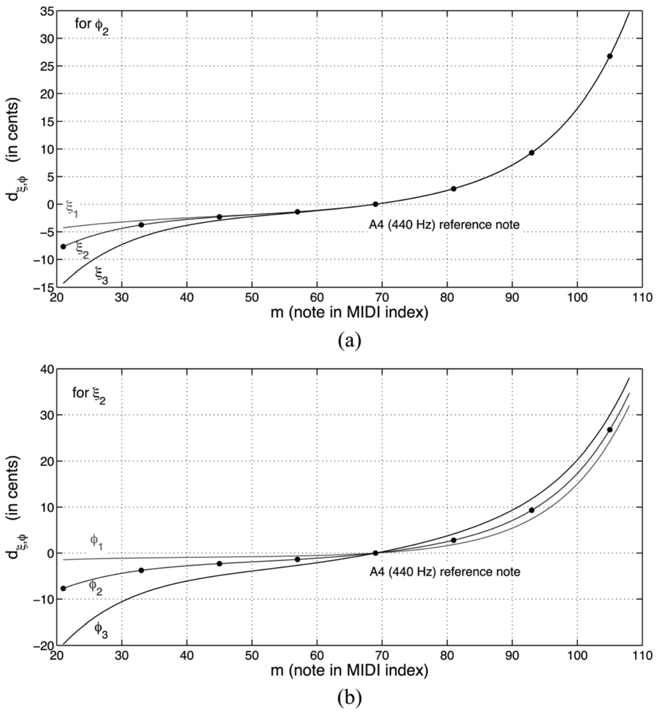

FIG. 2. Model for the deviation of tuning from ET along the compass. (a) Influence of ξ in the tuning for ϕ fixed. (b) Influence of ϕ in the tuning for ξ fixed. The different values for ξ and ϕ correspond to those used to generate the curves of Fig. 1.

interpolation) as the best estimate. Rauhala *et al.*[25,26] propose the Partial Frequencies Deviation (PFD) algorithm that minimizes, with respect to (B, F_0), the deviation between the theoretical partial frequencies of the model and the frequencies of high amplitude peaks previously selected in the magnitude spectra. Godsill and Davy[9] introduce a Bayesian framework to model piano sounds in time domain. (B, F_0) are parameters of the model and are estimated by maximizing the *a posteriori* probability density function. Besides, in the case of polyphonic harpsichord music—where the inharmonicity effect ($B < 10^{-4}$) is less important—an iterative method has been recently proposed by Dixon *et al.*[27] to estimate inharmonicity and temperament, together with a transcription task.

Here, a robust new algorithm based on the Non-negative Matrix Factorization (NMF) frameworks is proposed in order to finely estimate (B, F_0) from isolated notes, but also chord recordings.

1. NMF framework

Given a non-negative matrix V of dimension $K \times T$, the NMF consists of finding an approximate factorization[28]

$$V \approx WH \quad \Longleftrightarrow \quad V_{kt} \approx \hat{V}_{kt} = \sum_{r=1}^{R} W_{kr} H_{rt}, \tag{10}$$

where W and H are non-negative matrices of dimensions $(K \times R)$ and $(R \times T)$, respectively. In the case of music transcription,[29] V corresponds to the magnitude (or power) spectrogram of an audio excerpt, k corresponds to the frequency bin index, and t corresponds to the frame index. Thus, W represents a dictionary containing the spectra (or atoms) of the R sources, and H their time-frame activations. Recently, harmonic structure,[30,31] temporal evolution of spectral envelopes,[32] vibrato,[30] beat structure,[33] etc., 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 purpose of this section is to introduce the information of the inharmonicity of piano tones explicitly into the dictionary of spectra W. The idea is to take into account the parameters (B, F_0) as constraints on the partial frequencies of each note, so as to perform a joint estimation. In order to limit the number of parameters that we need to retrieve, besides the amplitude and frequency of each partial, we make the assumption that for every recording we know which notes are being played, and the corresponding time activations. Then, short-time spectra are extracted from the recordings and concatenated to build the observation matrix V (it is therefore not strictly speaking a spectrogram). Because for each column of V the played notes are known, the elements of H are fixed to one whenever a note is played, and zero when it is not. Thereby, only the dictionary W is optimized on the data. In that case, we should notice that the proposed model is not a factorization. However, the model has been developed in the NMF framework for further inclusion in transcription or source separation algorithms, where NMF is a very competitive approach.

In order to quantify the quality of the approximation of Eq. (10), a distance (or divergence) is introduced. For a separable metric, it can be expressed as

$$D(V|WH) = \sum_{k=1}^{K} \sum_{t=1}^{T} d(V_{kt}|\hat{V}_{kt}). \tag{11}$$

In audio applications, the family of β-divergences is widely used,[34] because it encompasses three common metrics: $\beta = 2$ for the Euclidian distance, $\beta = 1$ for the Kullback-Leibler divergence, and $\beta = 0$ for the Itakura-Saito divergence. These divergences 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 Sec. IV B are obtained for the Kullback-Leibler divergence

$$d_{\beta=1}(x \mid y) = x(\log x - \log y) + (y - x). \tag{12}$$

2. Modeling piano sounds in W

The model for the spectra/atoms of the dictionary W is based on an additive model: The spectrum of a note is composed of a sum of partials, in which the frequencies are constrained to follow an inharmonicity relation.

a. General additive model for the spectrum of a note. The general parametric atom used in this work is based on the additive model proposed by Hennequin *et al.*[30] 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 $\theta_r = \{a_{nr}, f_{nr} | n \in [1, N_r]\}$ and the set of parameters for the dictionary is denoted by $\theta = \{\theta_r | r \in [1, R]\}$. Finally, the expression of a parametric atom is given by

$$W_{kr}^{\theta_r} = \sum_{n=1}^{N_r} a_{nr} \cdot g_\tau(f_k - f_{nr}), \tag{13}$$

where f_k is the frequency of the bin with index k and $g_\tau(f_k)$ is 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 updated rules[30] 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 $g_\tau(f_k) = (1/\pi\tau).\sin(\pi f_k \tau)/(f_k - \tau^2 f_k^3)$, for $f_k \in [-2/\tau, 2/\tau]$.

Finally, the cost function is defined by using the β-divergence

$$C_0(\theta, H) = \sum_{k=1}^{K} \sum_{t=1}^{T} d_\beta\left(V_{kt} | \sum_{r=1}^{R} W_{kr}^{\theta_r} \cdot H_{rt}\right). \tag{14}$$

b. Inclusion of the inharmonicity constraint. A previous study on parametric NMF (Ref. 10) has already tested an inharmonicity relation, as given in Eq. (1), for piano music transcription. This study that constrained the partial frequencies to exactly follow this ideal relation was not conclusive: Adding inharmonicity in their model did not increase the transcription performance, compared to a simpler harmonic constraint. Similarly, in our model, given Eq. (13), this constraint would be equivalent to a reduction of the N_r parameters f_{nr} of every note to only 2 parameters $\{B_r, F_{0r}\}$ by setting $f_{nr} = nF_{0r}\sqrt{1 + B_r n^2}$. This constraint turns out to be too stringent, and numerically it leads to ill-convergence of the B parameter update.

In contrast with these studies, inharmonicity can be included as a *relaxed* constraint,[35] allowing for a local adaptation of the frequency of every partial, while constraining the entire set of partials to globally follow an inharmonic relation. At the same time, for each partial it allows a slight frequency deviation from the inharmonicity relation, as for instance due to the bridge coupling with the soundboard. The set of parameters related to the constraint is denoted by $\gamma = \{F_{0r}, B_r | r \in [1, R]\}$. Finally, a new cost function is built by adding a regularization term

$$C(\theta, \gamma, H) = C_0(\theta, H) + \lambda \cdot C_1(f_{nr}, \gamma), \tag{15}$$

where $C_1(f_{nr}, \gamma)$ is defined as a sum on each note of the mean square error between the estimated partial frequencies f_{nr} and these given by the inharmonicity relation

$$C_1(f_{nr}, \gamma) = \sum_{r=1}^{R} \sum_{n=1}^{N_r} \left(f_{nr} - nF_{0r}\sqrt{1 + B_r n^2}\right)^2. \tag{16}$$

λ is a scalar parameter, empirically tuned, which sets the weight of the inharmonicity constraint in the global cost function.

3. Optimization algorithm

a. Update of the parameters. As commonly proposed in NMF modeling, the optimization is performed iteratively, using multiplicative update rules for a_{nr}, f_{nr}, and B_r parameters. These update rules are obtained from the decomposition of the partial derivatives of the cost function given in Eq. (15), in a similar way to Hennequin *et al.*[30] (for the interested reader, the derivation is detailed in the associated supplementary material[36]). For F_{0r}, an exact analytic solution is obtained when canceling the partial derivative of the cost function C_1. Then, at each iteration the following update rules are applied:

$$a_{nr} \leftarrow a_{nr} \cdot \frac{Q_0(a_{nr})}{P_0(a_{nr})}, \tag{17}$$

$$f_{nr} \leftarrow f_{nr} \cdot \frac{Q_0(f_{nr}) + \lambda \cdot Q_1(f_{nr})}{P_0(f_{nr}) + \lambda \cdot P_1(f_{nr})}, \tag{18}$$

$$B_r \leftarrow B_r \cdot \frac{Q_1(B_r)}{P_1(B_r)}, \tag{19}$$

$$F_{0r} = \frac{\sum_{n=1}^{N_r} f_{nr} n \sqrt{1 + B_r n^2}}{\sum_{n=1}^{N_r} n^2 (1 + B_r n^2)}, \tag{20}$$

where

$$P_0(a_{nr}) = \sum_{k=1}^{K} \sum_{t=1}^{T} [(g_\tau(f_k - f_{nr}) \cdot H_{rt}) \cdot \hat{V}_{kt}^{\beta-1}], \tag{21}$$

$$Q_0(a_{nr}) = \sum_{k=1}^{K} \sum_{t=1}^{T} [(g_\tau(f_k - f_{nr}) \cdot H_{rt}) \cdot \hat{V}_{kt}^{\beta-2} \cdot V_{kt}], \tag{22}$$

$$P_0(f_{nr}) = \sum_{k,t} \left[\left(a_{nr} \frac{-f_k . g'_\tau(f_k - f_{nr})}{f_k - f_{nr}} \cdot H_{rt}\right) \cdot \hat{V}_{kt}^{\beta-1} + \left(a_{nr} \frac{-f_{nr} \cdot g'_\tau(f_k - f_{nr})}{f_k - f_{nr}} \cdot H_{rt}\right) \cdot \hat{V}_{kt}^{\beta-2} \cdot V_{kt}\right], \tag{23}$$

$$Q_0(f_{nr}) = \sum_{k,t} \left[\left(a_{nr} \frac{-f_k \cdot g'_\tau(f_k - f_{nr})}{f_k - f_{nr}} \cdot H_{rt}\right) \cdot \hat{V}_{kt}^{\beta-2} \cdot V_{kt} + \left(a_{nr} \frac{-f_{nr} \cdot g'_\tau(f - f_{nr})}{f_k - f_{nr}} \cdot H_{rt}\right) \cdot \hat{V}_{kt}^{\beta-1}\right], \tag{24}$$

$$P_1(f_{nr}) = 2f_{nr}, \tag{25}$$

$$Q_1(f_{nr}) = 2nF_{0r}\sqrt{1 + B_r n^2}, \tag{26}$$

$$P_1(B_r) = F_{0r} \sum_{n=1}^{N_r} n^4, \tag{27}$$

$$Q_1(B_r) = \sum_{n=1}^{N_r} \frac{n^3 f_{nr}}{\sqrt{1 + B_r n^2}}, \tag{28}$$

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. We remind that $\hat{V} = W^\theta H$. Note that for a transcription task, H could be updated with standard NMF multiplicative rules.[34]

b. Initialization. The initialization of (B_r, F_{0r}) is done using the inharmonicity and tuning model along the whole compass (cf. Sec. II), with typical values of the parameters. For the applications presented in Sec. IV, we set for the model of B: $s_B = 8.9 \cdot 10^{-2}$, $y_B = -7$; for the model of octave type: $\kappa = 3.5$, $\alpha = 25$, $m_0 = 60$ and for the global deviation $d_g = 5$. Then, the f_{nr}'s are initialized according to the inharmonic relation given in Eq. (1). The a_{nr}'s are initialized to 1.

c. Dealing with noise and partials related to longitudinal vibrations. 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 relation parameters. For each iteration of the optimization algorithm, we cancel their influence in the estimation of γ by removing them from the regularization term given in Eq. (16). For the proposed application, we compute, during a preprocessing step, the noise level[37] (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 NL, by $\Delta_r = \{n | a_{nr} > \mathrm{NL}(f_{nr}),\ n \in [1, N_r]\}$. This information is taken into account in the updated rules of B_r [Eqs. (27) and (28)] and F_{0r} [Eq. (18)] by replacing the sums over the entire set of partials $\sum_{n=1}^{N_r}$ by sums over the reliable set of partials $\sum_{n \in \Delta_r}$.

Moreover, in order to avoid the partials of the model to match with the wrong partials of the observed spectra (as for instance corresponding to longitudinal vibrations of the strings) the parameter λ is set to $1.25 \cdot 10^{-1}$. This value corresponds to a large weight for the inharmonicity constraint in the global cost function. In the last iterations the weight of the constraint is relaxed to a smaller value ($\lambda = 5 \cdot 10^{-3}$), so that the partials of the model can exactly match the measured partials, which are expected to correspond to transverse vibrations of the strings, and can slightly deviate from the theoretical inharmonicity relation. Note that these values of the parameter λ, given for the analysis of spectra normalized to a maximal magnitude of 1, do not have to be finely tuned on each piano (the same values have been used throughout our analysis).

Finally, the steps of the algorithm are summarized in Algorithm 1.

B. Estimation of the whole compass tuning model

1. B_ξ estimation

We first estimate the fixed parameters $\{s_T, y_T\}$, corresponding to the string set design on the treble bridge and being almost equal for all the models of pianos, by using $B(m)$ estimates of six different pianos (the databases are presented in Sec. IV) in the range $C4$ to $C8$. These are obtained by an $L1$ regression (in order to reduce the influence of potential outliers), i.e., by minimizing the least absolute deviation, between the model and the average of the estimated inharmonicity curves over the six different pianos. We find $s_T \simeq 9.26 \cdot 10^{-2}$, $y_T \simeq -13.64$. These results are in accordance with estimates based on physical considerations[3] $s_{T[\mathrm{Yo52}]} \simeq 9.44 \cdot 10^{-2}$, $y_{T[\mathrm{Yo52}]} \simeq -13.68$.

Each piano is then studied independently to estimate the particular parameters $\xi = \{s_B, y_B\}$ on a set of notes M. ξ is estimated minimizing the least absolute deviation between $\log B(m)$ and $\log B_\xi(m)$,

$$\hat{\xi} = \arg\min_{\xi} \sum_{m \in M} |\log B(m) - \log B_\xi(m)|. \tag{29}$$

2. ρ_ϕ estimation

For each piano, the data $\rho(m)$ is estimated for $m \in [21, 96]$ from $(B(m), F_0(m))$ values by inverting Eq. (6),

$$\rho(m) = \sqrt{\frac{4F_0(m)^2 - F_0(m+12)^2}{F_0(m+12)^2 B(m+12) - 16F_0(m)^2 B(m)}}. \tag{30}$$

Then, the set of parameters ϕ is estimated by minimizing the least absolute deviation distance between $\rho_\phi(m)$ and $\rho(m)$ on a set M of notes

$$\hat{\phi} = \arg\min_{\phi} \sum_{m \in M} |\rho(m) - \rho_\phi(m)|. \tag{31}$$

Algorithm 1: NMF with inharmo. constraint

Input:
V set of magnitude (normalized to a max. of 1) spectra
H filled with 0 and 1

Preprocessing:
for each column of V compute $\mathrm{NL}(f_k)$ the NL
see the Appendix of Rigaud *et al.*[37]

Initialization: $\forall\, r \in [1, R], n \in [1, N_r]$,
(B_r, F_{0r}) according to the model of Sec. II
$f_{nr} = nF_{0r}\sqrt{1 + B_r n^2}$, $a_{nr} = 1$
W^θ computation [cf. Eq. (13)]
$\beta = 1 / \lambda = 0.125$

Optimization:
for $it = 1$ to 150 **do**
(1) a_{nr} update $\forall\, r \in [1, R], n \in [1, N_r]$ [Eq. (17)]
  W^θ update [Eq. (13)]
  deduce Δ_r by comparing a_{nr} with $\mathrm{NL}(f_{nr})$
(2) f_{nr} update $\forall\, r \in [1, R], n \in [1, N_r]$ [Eq. (18)]
  W^θ update [Eq. (13)]
for $u = 1$ to 30 **do**
  $\forall r, n \in \Delta_r$
  F_{0r} update [cf. Eq. (20)]
  B_r update (20 times) [cf. Eq. (19)]
end for
if $it \geq 100$ **then**

$\lambda = 5 \cdot 10^{-3}$
end if
end for

Output: $a_{nr}, f_{nr}, B_r, F_{0r}$

---

3. d_g estimation

Once the ξ and ϕ sets of parameter have been estimated, the octaves of the reference note are tuned according to Eq. (6). Then, the deviation from ET of the model $d_{\xi,\phi}$ is obtained on the whole compass after the Lagrange interpolation stage. Finally, d_g is estimated by minimizing the least absolute deviation, on the reference octave $F3$ to $F4$ ($m \in [53, 65]$) between $d(m)$, the deviation from ET estimated on the data [see Eq. (4)], and $d_{\xi,\phi}(m) + d_g$,

$$\hat{d}_g = \arg\min_{d_g} \sum_{m=53}^{65} |d(m) - (d_{\xi,\phi}(m) + d_g)|. \tag{32}$$

IV. EXPERIMENTAL RESULTS

A. Experimental data

The results presented in this section are obtained from three separate databases, covering a total of six pianos: Iowa (Ref. 38) (one grand piano), RWC (Ref. 39) (three grand pianos), and MAPS (Ref. 40) (one upright piano and one grand piano synthesizer using high quality samples).

In the paper, only selected examples are shown. For a more extensive set of results, one can refer to the associated supplementary material.[36]

B. (B, F_0) estimation results

1. Isolated note analysis

The proposed algorithm has been applied to the estimation of (B, F_0) from isolated note recordings played with *mezzo-forte* dynamics and re-sampled to $F_s = 22\,050$ Hz. In order to obtain a sufficient spectral resolution, the observation spectra have been extracted from 500 ms Hanning windows, applied to the decay part of the sounds. Then, the matrix V has been built by concatenating the 88 spectra (each column corresponding to the magnitude spectrum of a note, from $A0$ to $C8$) and here H is fixed to the identity matrix. For each note, N_r has been set to $\arg min_{N_r}(30, f_{N_r,r} < F_s/2)$. Figure 3 shows (a) the initialization and (b) the result of the optimization for the analysis of the F♯1 note of RWC grand piano #1 database. At the initialization 14 partials of the model are overlapping the partials of the data corresponding to transverse vibrations of the strings. After the optimization procedure, the amplitudes and frequencies of the first 30 partials have been correctly estimated and it can be seen that the selection of longitudinal vibration partials (visible from approximately 700 to 1500 Hz) has been avoided, although the initialization was close to some of the corresponding peaks. The result is also displayed for the G♭6 note in Fig. 4. In this case, the validity of the estimation cannot be assessed

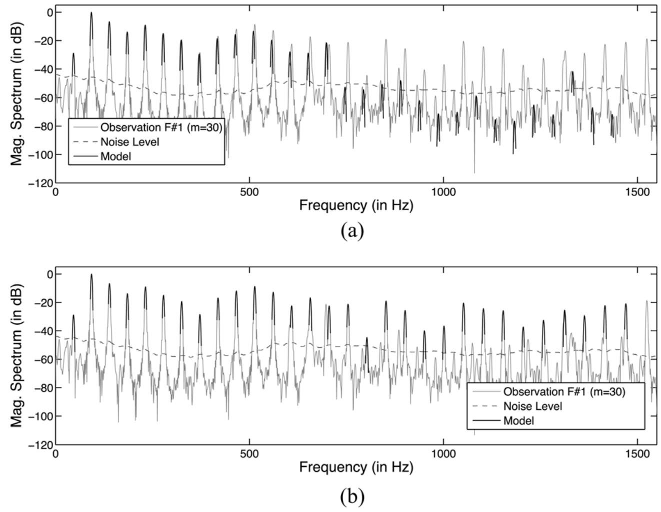

FIG. 3. (a) Initialization and (b) result of the algorithm for the analysis of the F♯1 note from the RWC grand piano #1 database.

so easily for some partials (as for instance around 5000, 6500, and 8000 Hz), mainly because these partials aggregate multiple peaks when the model only assumes a single component. This issue typically happens in the treble range, where the notes are associated with triplets of slightly detuned strings. Then, the algorithm selects one peak per group that has the best balance between peak strength and model fitting, and might return slightly biased estimates for (B, F_0).

The results for the estimation of the inharmonicity coefficient have been compared to the PFD algorithm[25,26] on both synthetic and real piano tones (corresponding to these used by Rauhala *et al.*[25]). As suggested,[25] the evaluation is performed on the $A0$ to $G3$ range ($m \in [21, 55]$). NMF achieves an average relative deviation from ground truth of 0.33% on synthetic samples and 0.76% on real samples, whereas PFD returns 0.78% and 3.3%, respectively. The interested reader can find a more in depth description of the evaluation and results in the supplementary material.[36] A benefit of using the NMF algorithm is that it still performs well in the high pitch range, while PFD (not optimized there) has some robustness issues.

2. Chord analysis

The same protocol has been applied to the analysis of four chords (from MAPS grand piano synthesizer), respectively taken in the extreme bass, bass, middle, and treble range of the compass. Each chord is composed of five notes. In order to have a sufficient spectral resolution, the analysis window length was set to 1 s for the chords played in the

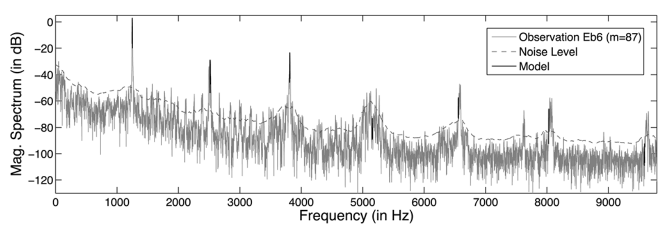

FIG. 4. Result of the algorithm for the analysis of the G♭6 note from the RWC grand piano #1 database.

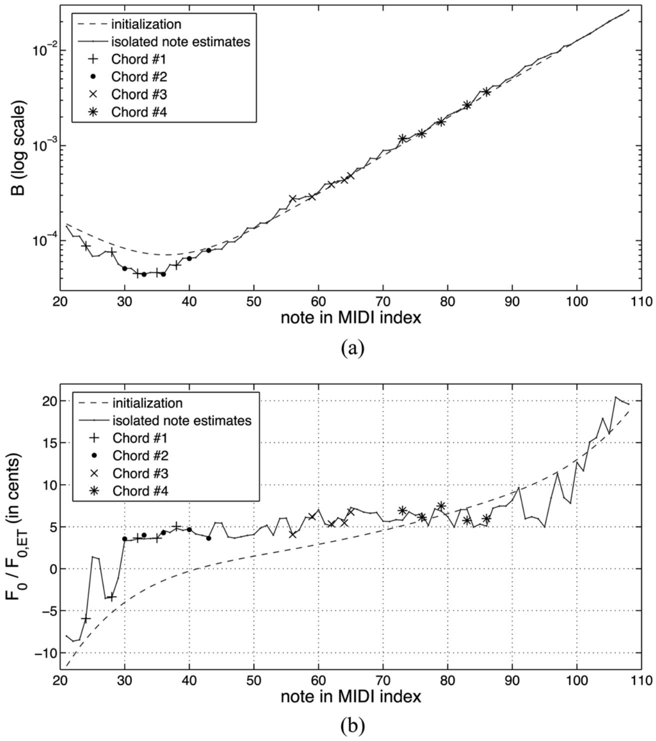

FIG. 5. Isolated note vs chord analysis for the grand piano synthesizer of MAPS database. (a) Inharmonicity coefficient and (b) deviation from ET along the compass.

extreme bass/bass ranges and 500 ms in the medium/treble ranges. In Fig. 5, the results of (B, F_0) estimates obtained from isolated notes (in thin gray lines) are compared with the ones obtained from chords (one type of marker for each chord). The initialization is drawn as a dashed line. It can be observed that both types of estimations lead to remarkably similar results. The slight deviations in the estimation from chord recordings could be explained by the overlapping of the partials belonging to different notes that could corrupt the estimation of the frequencies. Moreover, it has been shown in Sec. IV B 1 on isolated note analysis that, in the treble range, the precise estimation of (B, F_0) cannot be always guaranteed since the model of inharmonicity with one frequency peak per partial, as given by Eq. (1), is not sufficient to explain the spectrum of the notes. The estimated spectrum corresponding to chord #1 is given in Fig. 6, where one can see that, despite a considerable spectral overlap between the notes, the partials are well identified for every note up to a high order around 30.

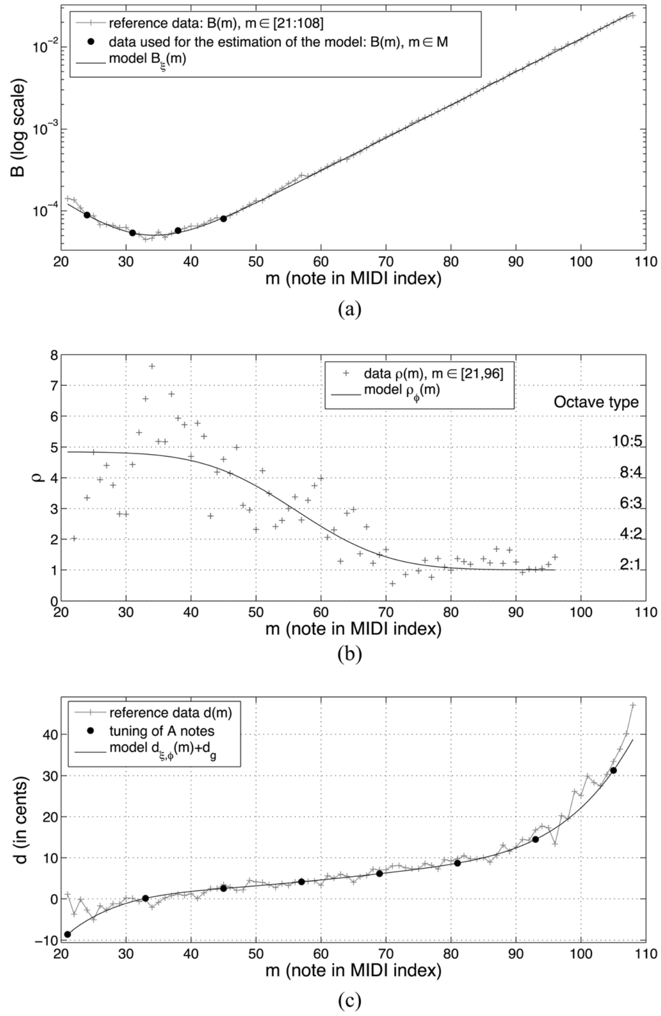

FIG. 7. RWC grand piano #3. (a) Inharmonicity coefficient B, (b) octave type parameter ρ, and (c) deviation from ET along the whole compass. The data are depicted as gray + markers and the model as black lines.

C. Whole compass tuning model results

1. Modeling the tuning of well-tuned pianos

The results of the estimation of the whole compass tuning model for two different pianos are presented in Figs. 7 and 8. Figures 7(a)–7(c) and 8(a)–8(c) correspond to the inharmonicity coefficient B, the octave type parameter ρ, and the deviation from ET curves along the whole compass, respectively. The data corresponding to the estimation of

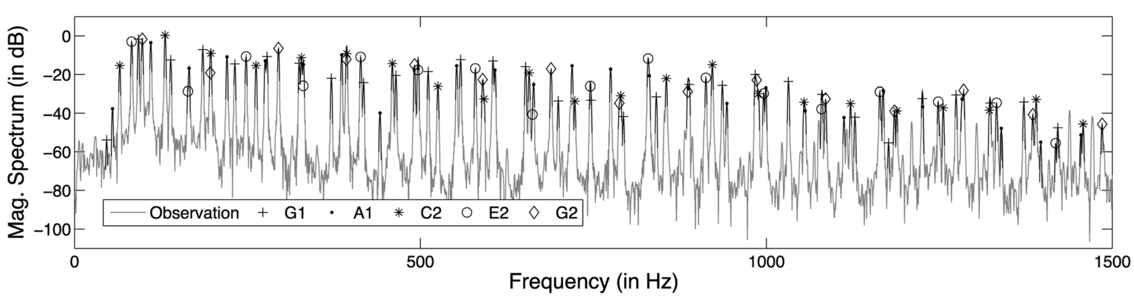

FIG. 6. Result of the algorithm for the analysis of chord #1 ($G1$-$A1$-$C2$-$E2$-$G2$) of MAPS grand piano synthesizer.

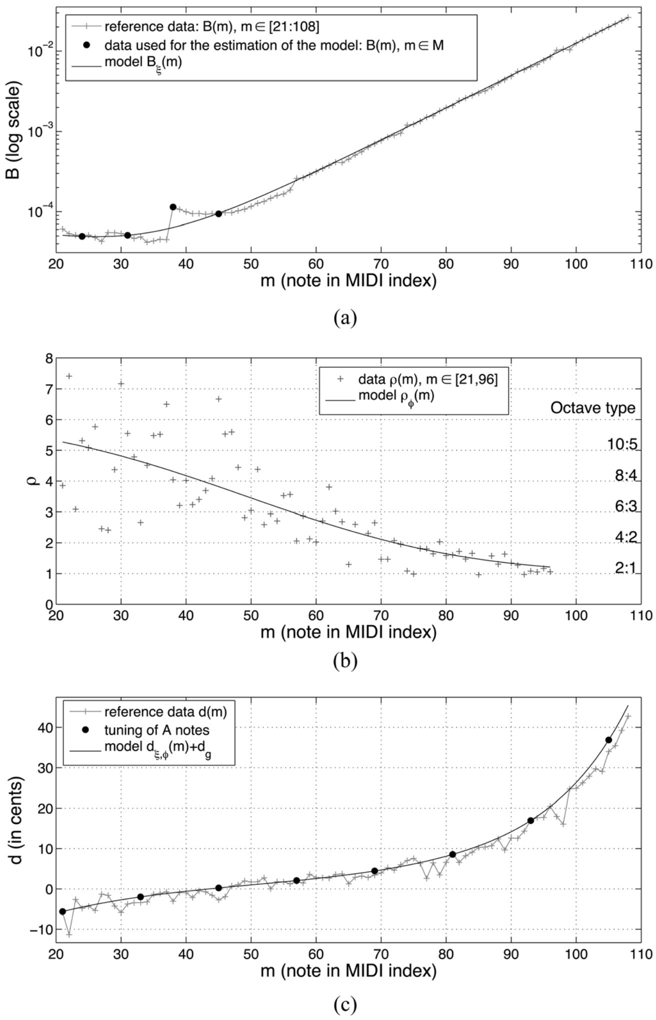

FIG. 8. RWC grand piano #2. (a) Inharmonicity coefficient B, (b) octave type parameter ρ, and (c) deviation from ET along the whole compass. The data are depicted as gray + markers and the model as black lines.

(B, F_0) from isolated note recordings is depicted as "+" markers, and the model as black lines.

B along the compass [Figs. 7(a) and 8(a)]. The estimation of the parameters has been performed from a limited set of four notes (black dots markers), taken in the bass range and equally spaced by fifth intervals. As the string set design on each bridge is quite regular, a few notes can be used to correctly estimate the model. In the case where an important discontinuity is present in the variations of $B(m)$ [for instance, between $C2$, $m = 37$, and $D2$, $m = 38$, notes in Fig. 8(a)] the 2-bridge additive model produces a smoothed curve. It is worth noting from RWC grand piano #2 design characteristics that the slight jump between $m = 27$ (D♯1) and $m = 28$ ($E1$) might be explained by the single to doublet of strings transition, and the important jump between $m = 37$ ($C2$) and $m = 38$ ($D2$) by the bridge change joint to the doublet to triplet of strings transition.

ρ along the compass [Figs. 7(b) and 8(b)]. The curves of $\rho(m)$ can present an important dispersion around the mean model $\rho_\phi(m)$, but the global variations are well reproduced. In the medium range, the estimated octave types are a trade-off between 6:3 and 4:2, which is common in piano tuning.[5] The variations, more important in the bass range, could be explained by the fact that the model of the partial frequencies [cf. Eq. (1)] does not take into account the frequency shifts caused by the bridge coupling, mainly appearing in the low frequency domain. Moreover, the proposed tuning model is a simplification of a real tuning procedure, it is based on octave interval tuning, while an expert tuner would jointly control different intervals along the keyboard and can do local readjustments after a global tuning. Note that some values of $\rho(m)$ can be missing when the quantity under the square root of Eq. (30) is negative. This happens if the corresponding octave interval is compressed instead of being stretched.

Deviation from ET along the compass [Figs. 7(c) and 8(c)]. The curves demonstrate that the model reproduces the main variations of the tuning in a satisfactory manner. This confirms that, besides the well-known influence of the inharmonicity on the tuning, perceptual effects (taken into account through the octave type consideration) can take part in the stretching, mainly in the bass range. Note that the tuning of A notes is marked with black dot markers.

2. Tuning pianos

Because the model of octave type choice $\rho_\phi(m)$ is defined for well-tuned pianos (the stretching of the octaves is implicitly assumed to be higher than 2), it cannot be used to study the tuning of strongly out-of-tune pianos. In this case, we propose to generate tuning curves deduced from a mean model of octave type choice. The model is obtained by averaging the curves $\rho(m)$ over three pianos (RWC #2, #3, and Iowa grand pianos), that were assumed to be well-tuned, by looking at the shape of their deviation from ET curves. From this averaged data, a mean model $\bar{\rho}_\phi(m)$ is estimated. In order to give a range of fundamental frequencies in which the pianos could be reasonably re-tuned, we arbitrarily define a high (respectively, low) octave type choice as $\bar{\rho}_{\phi,H}(m) = \bar{\rho}_\phi(m) + 1$ (respectively, $\bar{\rho}_{\phi,L}(m) = \min(\bar{\rho}_\phi(m) - 1, 1)$). These curves are shown in Fig. 9.

Tuning curves are then computed from the estimation of ξ and $\bar{\rho}_\phi(m)$. The global deviation parameter d_g is set to 0. The results are presented in Fig. 10 for RWC grand piano #1. The current tuning is depicted as + gray markers and clearly shows that the piano is not well-tuned, mainly in the bass range where the tuning is "compressed." The space

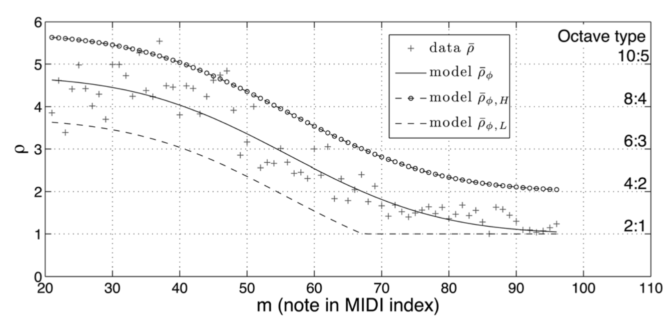

FIG. 9. Mean octave type choice for tuning application. + gray markers correspond to an average of $\rho(m)$ over three different pianos. The black line corresponds to the estimated model. Circle markers (respectively, in dashed line) represent the high (respectively, low) octave type choice model.

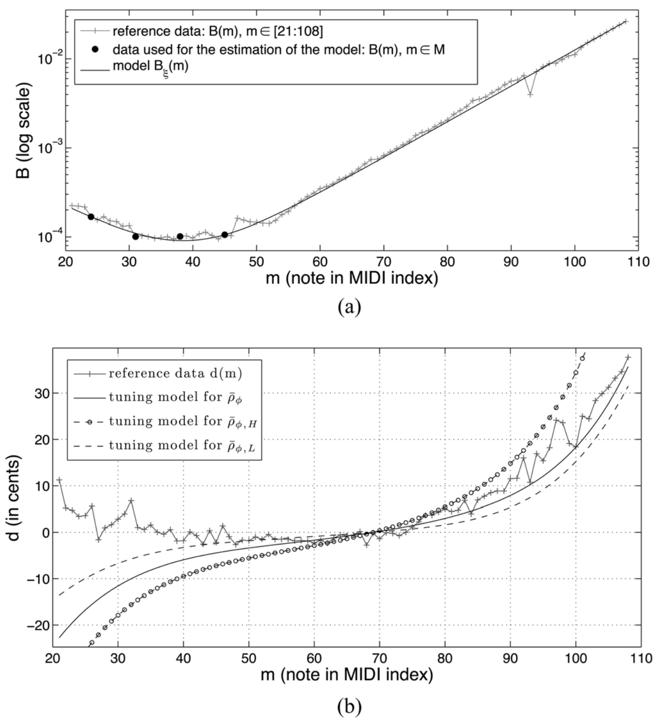

FIG. 10. RWC grand piano #1. (a) Inharmonicity curves along the compass. (b) Actual tuning and proposed tuning. Plus gray markers correspond to the data. The model corresponding to the octave type choice $\bar{\rho}_\phi(m)$ [respectively, $\bar{\rho}_{\phi,L}(m)$, $\bar{\rho}_{\phi,H}(m)$] is depicted as a black line (respectively, black dashed line, black dashed line with circle markers).

between the tuning curves obtained from $\bar{\rho}_{\phi,H}(m)$ and $\bar{\rho}_{\phi,L}(m)$ corresponds to a range in which we assume the piano could be well-tuned. For a quantitative interpretation, it will be interesting to compare our curves with those obtained after a re-tuning done by a professional tuner.

V. CONCLUSION

A parametric model of piano tuning has been proposed in this paper. It takes into account a model on the whole compass for the inharmonicity coefficient, and the octave type choices of the tuner. The complete algorithm takes as input either isolated notes, or chords recordings. From this, (B, F_0) is finely estimated and used as data to estimate the parameters of the tuning model. It has been successfully applied to model the main variations of the tuning along the compass of different types of pianos, and it provides re-tuning curves for out-of-tuned pianos.

The current algorithm assumes a prior knowledge of which notes are being played, and when, in the input recordings. A long-term goal of this study is to be able to parametrize the piano inharmonicity and tuning, from any arbitrary piano recording, in a musical context. Toward this achievement, future work will focus on the interplay between a transcription algorithm, and the estimation of the physical parameters—the topic of this article—possibly in an iterative way.

ACKNOWLEDGMENT

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

[1] D. W. Martin and W. D. Ward, "Subjective evaluation of musical scale temperament in pianos," J. Acoust. Soc. Am. **33**, 582–585 (1961).

[2] J. Lattard, "Influence of inharmonicity on the tuning of a piano," J. Acoust. Soc. Am. **94**, 46–53 (1993).

[3] R. W. Young, "Inharmonicity of plain wire piano strings," J. Acoust. Soc. Am. **24**, 267–273 (1952).

[4] B. Bremmer, "Midrange piano tuning," Technical Report, Piano Technicians Guild (2007), www.billbremmer.com/articles/ (Last viewed 11/06/12).

[5] B. Bremmer, "Aural tuning tests for 2:1, 4:2 and 6:3 type octaves," Technical Report, Piano Technicians Guild (2007), www.billbremmer.com/articles/ (Last viewed 11/06/12).

[6] B. Capleton, *Theory and Practice of Piano Tuning. A Manual on the Art, Techniques and Theory* (Amarilli Books, Malvern, UK, 2007), pp. 1–626.

[7] H. Hinrichsen, "Entropy-based tuning of musical instruments," Rev. Bras. Ensino de Física **34**, 2301–2314 (2012).

[8] L. I. Ortiz-Berenguer, F. J. Casajús-Quirós, M. Torres-Guijarro, and J. A. Beracoechea, "Piano transcription using pattern recognition: aspects on parameter extraction," in *Proceedings of the 7th International Conference on Digital Audio Effects (DAFx-04)*, pp. 212–216 (2004).

[9] S. Godsill and M. Davy, "Bayesian computational models for inharmonicity in musical instruments," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 283–286 (2005).

[10] E. Vincent, N. Bertin, and R. Badeau, "Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 109–112 (2008).

[11] V. Emiya, R. Badeau, and B. David, "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle," IEEE Trans. Audio, Speech, Lang. Process. **18**, 1643–1654 (2010).

[12] E. Benetos and S. Dixon, "Joint multi-pitch detection using harmonic envelope estimation for polyphonic music transcription," IEEE J. Sel. Top. Signal Process. **5**, 1111–1123 (2011).

[13] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer-Verlag, New York, 1998), pp. 64–69, 352–398.

[14] C. E. Gough, "The theory of string resonances on musical instruments," Acta. Acust. Acust. **49**, 124–141 (1981).

[15] G. Weinreich, "Coupled piano strings," J. Acoust. Soc. Am. **62**, 1474–1484 (1977).

[16] M. Aramaki, J. Bensa, L. Daudet, P. Guillemain, and R. Kroland-Martinet, "Resynthesis of coupled piano string vibrations based on physical modeling," J. New Music Res. **30**, 213–226 (2001).

[17] A. Chaigne and J. Kergomard, *Acoustique des Instruments de Musique (Acoustics of Musical Instruments)* (Belin, Paris, France, 2008), pp. 235–253.

[18] H. A. Conklin, Jr., "Design and tone in the mechanoacoustic piano. Part III. Piano strings and scale design," J. Acoust. Soc. Am. **100**, 1286–1298 (1996).

[19] J. Engelbrecht, A. Mägi, and A. Stulov, "Grand piano manufacturing in Estonia: The problem of piano scaling," Proc. Est. Acad. Sci., Phys., Math. **5**, 155–167 (1999).

[20] A. Stulov, "Physical modeling of the piano string scale," Appl. Acoust. **69**, 977–984 (2008).

[21] B. C. J. Moore, B. R. Glasberg, and R. W. Peters, "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853–1860 (1984).

[22] R. J. Ritsma, "Frequencies dominant in the perception of the pitch of complex sounds," J. Acoust. Soc. Am. **42**, 191–198 (1967).

[23] R. Plomp, "Pitch of complex tones," J. Acoust. Soc. Am. **41**, 1526–1533 (1967).

[24] A. Galembo and A. Askenfelt, "Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano," IEEE Trans. Speech Audio Process. **7**, 197–203 (1999).

[25] J. Rauhala, H. M. Lehtonen, and V. Välimäki, "Fast automatic inharmonicity estimation algorithm," J. Acoust. Soc. Am. **121**, 184–189 (2007).

[26] J. Rauhala and V. Välimäki, "F0 estimation of inharmonic piano tones using partial frequencies deviation method," in *Proceedings of the International Computer Music Conference (ICMC'07)*, pp. 453–456 (2007).

[27] S. Dixon, M. Mauch, and D. Tidhar, "Estimation of harpsichord inharmonicity and temperament from musical recordings," J. Acoust. Soc. Am. **131**, 878–887 (2012).

[28]D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature **401**, 788–791 (1999).
[29]P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proceedings of the IEEE Workshop on Application of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 177–180 (2003).
[30]R. Hennequin, R. Badeau, and B. David, "Time-dependent parametric and harmonic templates in non-negative matrix factorization," in *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx-10)*, pp. 246–253 (2010).
[31]N. Bertin, R. Badeau, and E. Vincent, "Enforcing harmonicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription," IEEE Trans. Audio, Speech, Lang. Process. **18**, 538–549 (2010).
[32]R. Hennequin, R. Badeau, and B. David, "Nmf with time-frequency activations to model non stationary audio events," IEEE Trans. Audio, Speech, Lang. Process. **19**, 744–753 (2011).
[33]K. Ochiai, H. Kameoka, and S. Sagayama, "Explicit beat structure modeling for non-negative matrix factorization-based multipitch analysis," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 133–136 (2012).
[34]C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence with application to music analysis," Neural Comput. **21**, 793–830 (2009).
[35]F. Rigaud, B. David, and L. Daudet, "Piano sound analysis using non-negative matrix factorization with inharmonicity constraint," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO 2012)*, pp. 393–399 (2012).
[36]See supplementary material at http://dx.doi.org/10.1121/1.4799806 for the results on the complete data set, mathematical derivations of the NMF update rules and performance evaluation of the inharmonicity coefficient estimation algorithm.
[37]F. Rigaud, B. David, and L. Daudet, "A parametric model of piano tuning," in *Proceedings of the 14th International Conference on Digital Audio Effects (DAFx-11)*, pp. 393–399 (2011).
[38]http://theremin.music.uiowa.edu/MISpiano.html (Last viewed 11/06/12).
[39]M. Goto, T. Nishimura, H. Hashiguchi, and R. Oka, "RWC music database: Music genre database and musical instrument sound database," in *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR)*, pp. 229–230 (2003).
[40]http://www.tsi.telecom-paristech.fr/aao/en/category/database/ (Last viewed 11/06/12).