Joint Estimation of Multiple Notes and Inharmonicity Coefficient Based on f_0-Triplet for Automatic Piano Transcription

Changhyun Kim, Wonil Chang, Sang-Hoon Oh, and Soo-Young Lee

***Abstract*—With the increasing importance of smart gadgets in our daily lives, there is a need for an automatic piano transcription system in various multimedia services. For automatic piano transcription, the string inharmonicity coefficient (B) and fundamental frequency (f_0) should be detected robustly and accurately. The proposed triplet-sequentially additive partial (SAP) algorithm improves the current B estimation algorithm in terms of both performance and speed with less prior knowledge. Additionally, this joint (B, f_0) estimation algorithm is applied directly to the transcription of real piano recordings, and the 4.41% improvement of accuracy was achieved over another transcription system that had both similar processing steps and feature extraction method.**

***Index Terms*—Automatic Transcription System, Inharmonicity Coefficient, Multiple-Fundamental Frequency Estimation, Piano**

I. Introduction

MANY music application services are currently employed, including music query by humming, music recommendation by genre classification, and automatic transcription of polyphonic music. Of these services, the latter is considered to be a milestone representing the beginning of the digital library era. Because the piano is the most representative instrument for polyphonic music, this study attempts to improve the performance of automatic piano transcription.

Inharmonicity is the phenomenon of nonlinear partial peaks in string instruments such as the piano, violin, and guitar. In these instruments, high-rank partial peaks deviate more from integer multiples of the fundamental frequency. To improve the accuracy of inharmonicity coefficient (B) values obtained from these partial peaks, Rauhala et al. [1], [2] developed a B estimation algorithm with iterative minimization of the partial frequency deviation (PFD) using given pitch numbers. More recently, Rigaud et al. [3], [4] developed a B estimation method with nonnegative matrix factorization (NMF). Furthermore, most multiple note detection algorithms have also tried to model the nonlinear characteristics of string instruments using the B concept [5]-[7]. Klapuri [5] applied harmonicity and spectral smoothness to the iterative note subtraction model. Emiya et al. [6] employed a probabilistic model to describe the spectrum amplitude as the sum of the signal and noise. Benetos et al. [7] included a B concept with the salience function, which determines the pitch by scoring. Then, it erases the halving or multiplying false positive note candidates using the flatness and spectral irregularity. However, these methods are limited in terms of their ability to discriminate between true notes and false positives in particular doubling and halving errors in real piano recordings.

To overcome these problems, we propose the triplet-sequentially additive partial (SAP) algorithm with four false positive (FP) error eraser blocks. The triplet-SAP algorithm simultaneously returns the pitch candidates (f_0), the inharmonicity coefficient (B), and the partial sequence ($\hat{f}_n$). With this information, we can erase three different types of false positive notes: trivial FP notes originating from the small noise peaks, harmonically related (multiplying or halving) FP notes by sharing the partial peaks, and decaying sound FP notes, which should be differentiated from the repeat note. The first trivial FP notes are erased based on the number of harmonics and amplitude score block. The second harmonically related FP notes are erased by the (B,f_0) thresholding block. Finally, the third decaying sound FP notes are erased by the repeat note detection block.

II. Proposed Methods

The proposed piano note onset transcription system comprises two steps: a preprocessing step to extract accurate spectrum peaks with high-frequency resolution, and a multiple note detection step to extract pitch candidates. In the second step, two main processes are used to perform multiple note detection: a triplet multiple-fundamental frequency estimation (MF0E) process to detect polyphonic notes, and a SAP process to estimate a more accurate value of B. The remaining four FP error eraser blocks filter out different types of false positive error notes.

A. Preprocessing Step

As shown in Fig. 1, five processes are included in the preprocessing step. The piano recording is given to the system as an input, and we used the corresponding MIDI file to extract the onset information. To handle the asynchronous onsets of simultaneous notes that deviate from 20 to 70 ms in real piano play [8], we merged onsets within a 30 ms time gap. In the segmentation block, the recording signal is split into segments by the merged onsets. A segment is a time unit for multiple note detection. To discriminate f_0 of the lowest notes, the frequency resolution should be lower than 1 Hz. Therefore, 2^{16} is chosen as the number of FFT points with a 44.1 kHz sampling

This work was supported in part by the Qualcomm Technologies, Inc.
C. Kim is with the Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon, 305-701 Korea. (e-mail : flipflop98@gmail.com).
W. Chang is with the Electronics and Telecommunications Research Institute, Daejeon 305-700 Korea. (e-mail : chang.wonil@gmail.com).
S.-H. Oh is with the Division of Information and Communication Convergence Engineering, Mokwon University, Daejeon, 302-318 Korea. (e-mail : shoh@mokwon.ac.kr).
S.-Y. Lee is with the Department of Electrical Engineering and Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon, 305-701 Korea. (e-mail : sylee@kaist.ac.kr).

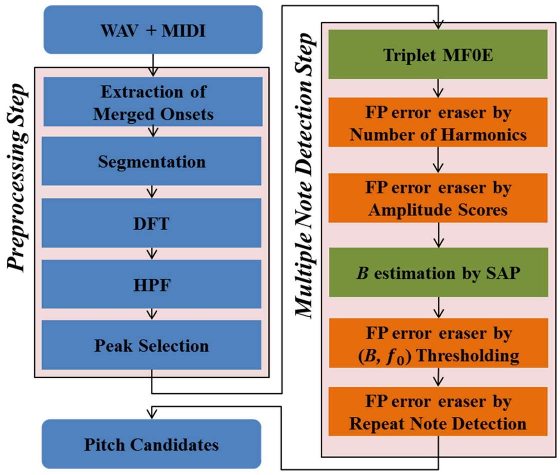

Fig. 1. Flow chart of the overall piano note onset transcription system.

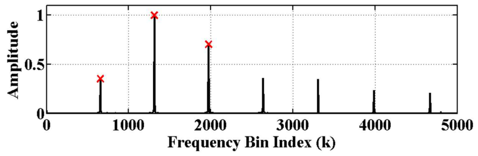

Fig. 2. f_0-triplet of Steinway & Sons piano 440 Hz (A4) note.

frequency. To enhance the higher-frequency partial amplitudes of the middle-range notes (notes over A4 (69) = 440 Hz), we use a finely designed high-pass filter (HPF) with a stop band gain = 0 dB, pass band gain = 12 dB, stop band edge frequency = 680 Hz, and pass band edge frequency = 2,850 Hz. Finally, we select 50 peaks from the spectrum. Experimentally, 50 peaks are sufficient for the detection of 6 to 7 multiple notes. First, we select the 20 lowest frequency peaks starting from 0 Hz. Then, the 30 highest amplitude peaks are added. In this way, we can avoid missing the low fundamental notes with small energy.

B. *Multiple Note Detection Step*

Among the 50 selected peaks, to reduce running time, our algorithm intelligently selects any three peaks that satisfy the following two conditions:

$$i)\ 4(f_j - f_i) > f_k - f_i, \qquad i\ )\ 4(f_k - f_j) > f_k - f_i,$$

where f_i, f_j, f_k are three different peaks from the spectrum. The factor 4 is an optimally selected value from the experiments (III-A, and III-B). Then, we calculate the possible values of the harmonic number (n), f_0, and B.

1) Triplet Multiple-Fundamental Frequency Estimation (MF0E) process

Here, B is a physical quantity of a string instrument, which expresses the tension of strings. The n^{th} partial frequency f_n is given by

$$f_n = n f_0 \sqrt{1 + Bn^2}, \tag{1}$$

where f_0 corresponds to the piano tuning frequency.

The f_0-triplet concept refers to three consecutive or one apart partial peaks that are used to simultaneously extract n, f_0, and B. As shown in Fig. 2, the first three partial peaks are selected to calculate B using two pairings. By pairing f_n/f_{n+1} and substituting (1), we can derive

$$B = \frac{n^2 f_{n+1}^2 - (n+1)^2 f_n^2}{(n+1)^4 f_n^2 - n^4 f_{n+1}^2}. \tag{2}$$

In addition, B can also be derived by pairing f_{n+1}/f_{n+2}. Because B in (2) must be equal to B obtained from pairing f_{n+1}/f_{n+2}, we can derive

$$(n+1)(an^4 + b'n^3 + c'n^2 + d'n + e') = 0, \tag{3}$$

where

$$a = 4f_{n+1}^4 - 2f_n^2 f_{n+1}^2 - 2f_{n+1}^2 f_{n+2}^2, \tag{4}$$

$$b = 20f_{n+1}^4 - 15f_n^2 f_{n+1}^2 - 5f_{n+1}^2 f_{n+2}^2, \tag{5}$$

$$c = 32f_{n+1}^4 - 44f_n^2 f_{n+1}^2 - 4f_{n+1}^2 f_{n+2}^2, \tag{6}$$

$$d = 16f_{n+1}^4 - 63f_n^2 f_{n+1}^2 - f_{n+1}^2 f_{n+2}^2, \tag{7}$$

and

$$e = -44f_n^2 f_{n+1}^2. \tag{8}$$

Furthermore, $b' = b - a, c' = c - b + a, d' = d - c + b - a$, and $e' = e - d + c - b + a$.

By applying the companion matrix eigenvalue method [9], the solution of (3) is easily calculated. Among the many solutions, real positive decimals should be selected. In addition to the above $(n, n+1, n+2)$ f_0-triplet set formulation, two other f_0-triplet sets of $(n, n+2, n+3)$ and $(n, n+1, n+3)$ are considered simultaneously for the robust multiple-f_0 estimation algorithm.

The tolerance intervals of the following n, f_0, and partial peak frequencies are sufficiently large to cover any type of piano considering the semitone intervals, which have about 6% higher frequency than the neighboring lower note. The harmonic numbers having a fractional-part greater than 6% of the integer-part are discarded because they are higher than the partial peak frequencies of the neighbor semitone notes. By inserting the selected n into (2), valid B values can be calculated inside the generous tolerance interval of [-0.0029, 0.02], which covers the whole-note scale of any piano. Finally, by inserting f_n, n, and B into (1), f_0 can be derived. The $\pm 2\%$ tolerance range of f_0 is a non-overlapped interval of half semitone notes. The other is the allowable partial peak frequency, which also has a half semitone tolerance interval shifted from [-3%, 3%] to [-1%, 4.5%] (a bit shortened) for high rank partials. This process lasts until the end of possible f_0-triplet peak selection.

2) False Positive (FP) error eraser by Number of Harmonics and Amplitude Scores

Trivial false positive error notes are discarded by the two FP error eraser blocks. The threshold values in both blocks are optimized with the training dataset. As shown in Table I, there are minimum numbers of valid harmonics, which are examined in the true positive samples on each octave. These threshold values are piano model free. Also, the amplitude score values of each note are set by summing the amplitudes of the limited numbers, which are generated after three process steps: spectrum normalization (max = 1), linear interpolation considering two neighbors, and spectral smoothing using Dresslers' method [10]. To maximize the overall system performance, the threshold score values for each note are

TABLE I
CRITERIA FOR FP ERROR ERASER BY NUMBER OF HARMONICS

Midi note number	Criteria for acceptance (# of detected partials/first # of harmonics)
21-38	5/7
39-48	4/8
49-67	3/6
68-83	3/4
84-108	3/3

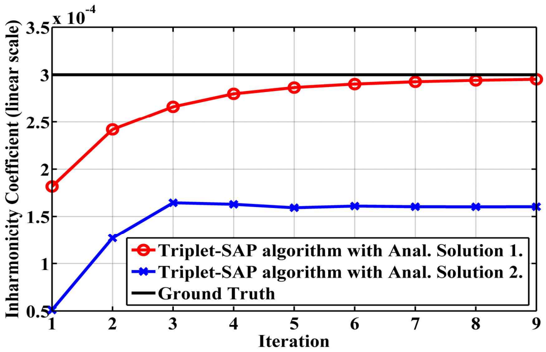

Fig. 3. B value update by SAP process.

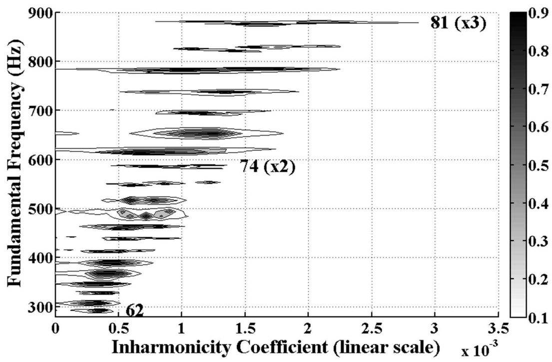

Fig. 4. Normalized probability density of true positive piano note on the (B, f_0) plane drawn by gray-scale contour map.

empirically set to be the 20th percentile of the true positive pitch candidates.

3) B Estimation by Sequentially Additive Partial (SAP) process

From the previous triplet MF0E process, we obtain the number of pitch candidates (f_0), along with an n sequence and n^{th} partial frequency ($\hat{f}_n$) sequence per candidate. Using equation (1), we can obtain the local estimation of inharmonicity coefficient $\hat{B}_n$ that is directly derived from the n^{th} observed partial frequency $\hat{f}_n$ as

$$\hat{B}_n = \frac{\left(\frac{\hat{f}_n}{n f_0}\right)^2 - 1}{n^2} = \frac{\left(\frac{\hat{f}_n}{n}\right)^2 \tau_0 - 1}{n^2}, \tag{9}$$

where we substitute $\tau_0 = 1/{f_0}^2$ for simplicity. The joint estimation of f_0 and B is performed by minimizing the following error function:

$$\mathrm{E}_1(\tau_0, B) = \textstyle\sum_{n=s}^{e} n^4 \{B - \hat{B}_n\}^2. \tag{10}$$

To give more weight to high-rank partials, the n^4 term is multiplied in the error function. Here, s and e represent the start and the end harmonic numbers, respectively, and these numbers are composed of non-redundant integer numbers by selecting the maximum amplitude peak from the three different f_0-triplet sets. After inserting (9) to (10), the minimum point of (10) with respect to τ_0 and B can be derived by

$$\frac{\partial \mathrm{E}_1(\tau_0, B)}{\partial \tau_0} = 0 \longrightarrow \tau_0 = \frac{\sum_{n=s}^{e}\left\{\left(\frac{\hat{f}_n}{n}\right)^2 + B{\hat{f}_n}^2\right\}}{\sum_{n=s}^{e}\left(\frac{\hat{f}_n}{n}\right)^4}, \tag{11}$$

$$\frac{\partial \mathrm{E}_1(\tau_0, B)}{\partial B} = 0 \longrightarrow B = \frac{\sum_{n=s}^{e}\left({\hat{f}_n}^2 \tau_0 - n^2\right)}{\sum_{n=s}^{e} n^4}. \tag{12}$$

By inserting (11) into (12) and substituting $\check{f}_n = \hat{f}_n / n = f_0\sqrt{1 + n^2 B}$, we can obtain analytic solution 1 as

$$B = \frac{\left\{\sum_{n=s}^{e}(\check{f}_n)^4 \sum_{n=s}^{e} n^2 - \sum_{n=s}^{e}(\check{f}_n n)^2 \sum_{n=s}^{e}(\check{f}_n)^2\right\}}{\left\{\left(\sum_{n=s}^{e}(n\check{f}_n)^2\right)^2 - \sum_{n=s}^{e}(\check{f}_n)^4 \sum_{n=s}^{e} n^4\right\}}. \tag{13}$$

Next, by inserting (13) into (12), we obtain f_0.

However, the maximum harmonic number in the partial sequence from the triplet MF0E process is limited. Because the partial frequency error is larger for the high-rank partial than the low-rank partial originating from the estimation error of B, higher rank partials should be added to estimate a more accurate value of B. To accomplish this, high-rank partials are added to the given f_0-triplet partial sequence by the max amplitude peak selection in the tolerance range of $[f_n \pm 0.4 f_0]$. An interval of $\pm 0.4 f_0$ is selected to give both non-overlapped and piano independent tolerance. In the isolated note case, five additional partial peaks are selected in a single iteration. In polyphonic music, only one additional partial peak is added to the existing partial peak sequence per iteration because of the overlapped partial problem. To verify the effectiveness of partial weights, the error function without the n^4 term is also examined as analytic solution 2. As shown in Fig. 3, as the number of iterations increases, the estimated value of B obtained from analytic solution 1 gradually approaches the ground truth. However, analytic solution 2 stops increasing after iteration 3, with a large gap between the estimated value and the ground truth value of B.

4) FP error eraser by $(B,\ f_0)$ Thresholding

Fig. 4 presents the normalized probability density of the true positive piano note on the (B, f_0) plane. The 0.1-level-contour of the normalized probability density, whose level is experimentally tuned from the training dataset, is used as the threshold of the (B, f_0) pair for FP erasing. In the example of note D4 (62) in Fig. 4, its multiplying notes (D5 (74), and A5 (81)) exhibit the distinguished distribution of B. This shows the possibility of filtering out false positive pitch candidates using the (B, f_0) pair.

5) FP error eraser by Repeat Note Detection

Finally, in this block, we discriminate a decaying sound note from a repeat note by checking the following two conditions: $i)\ A(m,t) > \alpha * A(m, t-1), i\ )\ B(m,t) > \beta * B(m, t-1)$, where $A(m,t) = \sum_n \hat{f}_n(m,t)$ and $B(m,t) = \hat{f}_1(m,t)$, m : note number, t : segmentation time index, and α, β : a constant derived from the training dataset.

III. Simulations

The performance of the proposed algorithm was tested using different databases and was measured in two experiments. Table II summarizes the tolerance intervals of Triplet-SAP variables, which are applicable to different pianos because they were selected based on the theoretically calculated semitone frequency interval ratio, generous whole-note scale min-max B values, and general integer multiple harmonics locations.

TABLE II
Tolerance Interval for Triplet-SAP Variables

Process	Variables	Tolerance Interval
	fractional part of n	$\pm$6% of round(n)
Triplet MF0E	f_0	$\pm$2% of Equal Temperament (ET) note frequency.
	B	-0.0029~0.02
	$\hat{f}_n$	-1%~4.5% of n multiple ET note freq.
SAP	$\hat{f}_n$	$f_n \pm 0.4 f_0$

TABLE III
AVERAGE RMS ERRORS AND AVERAGE RMS DEVIATIONS WITH TIME CONSUMPTION

Method	Synthetic tones		Real tones	
	RMS error	Running time	RMS deviation[a]	Running time
ICF[17]	1.19×10^{-6}	11.4s	1.73×10^{-5}	313.9s
ICF+[1]	1.16×10^{-6}	24.1s	1.88×10^{-5}	946.5s
PFD[1]	1.19×10^{-6}	7.0s	1.66×10^{-5}	58.4s
NMF[3]	5.85×10^{-7}	420.0s	4.54×10^{-6}	420.0s
Proposed	1.14×10^{-6}	2.0s	7.54×10^{-6}[b]	11.8s

[a]Difference obtained from manually estimated values. The exact inharmonicity coefficient is unknown.
[b]Ground Truth B values were derived from the manually selected peaks.

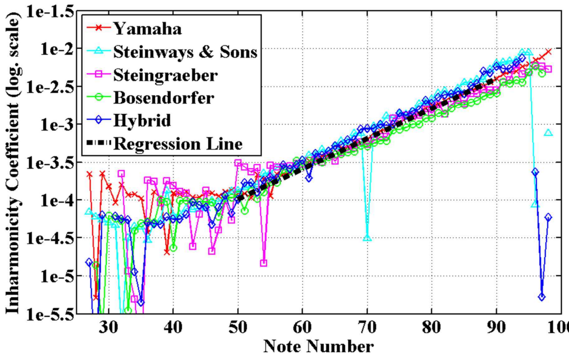

Fig. 5. B values estimated by analytic solution 1 on different pianos.

A. Experiment 1 - B Estimation on Isolated Notes

For this test, recordings from 35 synthetic isolated notes (A0 (21)-G3 (55)) [11] and 34 real isolated notes (Bb0 (22)-G3 (55)) [12] were used with ground truth B values of synthetic tones from Rauhala et al. [1]. We reused the test results obtained from their paper, as shown in Table III. In this isolated notes' experiment, four FP error eraser blocks are not needed because of the exact pitch from the triplet-SAP algorithm. The triplet-SAP algorithm with analytic solution 1 outperforms the inharmonic comb filter (ICF) and PFD algorithms in terms of the RMS error and the running time because of the high-rank partial emphasis term and analytic solution, respectively. Additionally, the PFD algorithm has a fluctuating B curve over the A#4 (70) note [3]. This makes it impossible to apply this algorithm directly for the transcription system. Even the most recent B estimation algorithm that is based on the NMF method performs more accurately in terms of RMS error than the proposed method, the speed is very slow. Moreover, this NMF algorithm requires finely tuned fundamental frequencies for the *a priori* initialization of NMF learning. When this prior knowledge is not provided, the NMF algorithm exhibits poor performance [13]. Thus, this NMF algorithm cannot be used as a standalone B estimation system. Because of the unknown ground truth B values in real-tone simulations, it is impossible to make a direct comparison of the RMS deviation obtained by the proposed algorithm and other systems.

B. Experiment 2 - Real Piano Transcription

This experiment was performed on the MIDI Aligned Piano Sounds (MAPS) [14] five instrument sets (Yamaha, Steinway & Sons, Steingraeber, Bosendorfer, and Hybrid). We randomly selected 12 files and extracted polyphonic music from the first interval of [30sec~2min] to form a total of about 14 minutes per piano. Therefore, the total music length is about 70 minutes.

TABLE IV
MAPS TEST SET NOTE ONSET TRANSCRIPTION ACCURACY (%)

	Y[b]	S&S[b]	S[b]	B[b]	H[b]	Avg.	Std.
Benetos[a][7]	28.37	27.83	26.91	22.73	10.85	23.34	7.32
Proposed[a]	27.04	32.19	29.39	22.82	27.32	27.75	3.44

[a]The accuracy was calculated in the limited key range 36 (C2)-95 (B6), since the Benetos[3] system is designed for these notes only.
[b]Y: Yamaha, S&S: Steinway & Sons, S: Steingraeber, B: Bosendorfer, H: Hybrid

We used five-fold cross-validation for the experiment: to test on the dataset of a single piano model, we tuned the threshold values for four FP error eraser blocks (II-B-2, II-B-4, and II-B-5) using the dataset from the remaining four piano models. This process was repeated for all five piano models. Figure 5 shows that the B value curves of five piano models are similar. Therefore, it is possible to extract generalized model of piano inharmonicity.

Some of previous works on automatic piano transcription used spectro-temporal feature extraction with moderate frequency resolution and time dynamics concerning temporal partial envelope transition of attack–decay–sustain–release (ADSR) [15], [16]. On the contrary, we are interested in multiple note detection based on delicate modeling of piano inharmonicity using high frequency-resolution spectrum. Therefore, we selected the Benetos' method [7] for performance comparison among multiple note estimation algorithms [15], [16], [5]-[7]. Both the Benetos' system and our system have similar processing steps, and they use similar type of feature extraction.

For performance-evaluation, we used both f-measure and accuracy [15]. These criteria values were also used in the MF0E task 2 (e.g., piano note onset only transcription) of the Music Information Retrieval Exchange (MIREX) competition.

Before experimenting on the test set described previously, the performance was checked with the MAPS Disklavier isolated random chord set with note level 5. The f-measure values were 0.72 and 0.70 with the proposed system and the Benetos' system [7], respectively. In this real music note onset test, the proposed system with analytic solution 1 has outperformed his method [7] by 4.41% on average (Table IV). In addition, our system shows smaller standard deviation of accuracy among five piano models, which supports that our system is less sensitive to the change of the piano model.

IV. CONCLUSIONS

In this letter, we presented a novel method to estimate the B value of piano sounds. Compared to previous methods, our algorithm show superior accuracy and speed in an isolated note test with less prior knowledge. In addition, it was successfully combined into an automatic piano transcription system, which exhibits both high accuracy and fairly good generalization abilities. The proposed automatic transcription system might be extended to other string instruments in the future.

REFERENCES

[1] J. Rauhala, H. M. Lehtonen, and V. Valimaki, "Fast automatic inharmonicity estimation algorithm," *J. Acoust. Soc. Am*., vol. 121 no.5, pp. 184–189, May. 2007.

[2] J. Rauhala, and V. Valimaki, "F0 estimation of inharmonic piano tones using partial frequencies deviation method," in *Int. Computer Music Conf.(ICMC),* Copenhagen, Denmark, Aug. 2007, pp. 453-456.

[3] F. Rigaud, B. David, and. L. Daudet, "A parametric model and estimation techniques for the inharmonicity and tuning of the piano," *J. Acoust. Soc. Am.,* vol. 133, no.5, pp. 3107–3118, Mar. 2013.

[4] F. Rigaud, B. David, and L. Daudet, "Piano sound analysis using nonnegative matrix factorization with inharmonicity constraint," in *Signal Process. Conf. (EUSIPCO)*, 2012 Proc. of the 20th Eur., Bucharest, Romania, 2012, pp. 2462-2466.

[5] A. P. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *IEEE Trans. Speech Audio Process.,* vol. 11, no. 6, pp. 804–816, Nov. 2003.

[6] V. Emiya, R. Badeau, and B. David, "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle," *IEEE Trans. Speech Language Process.,* vol. 18, no. 6, pp. 1643–1654, Aug. 2010.

[7] E. Benetos and S. Dixon, "Multiple-F0 estimation of piano sounds exploiting spectral structure and temporal evolution," in *ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Process.,* Makuhari, Japan, Sep. 2010, pp. 13–18.

[8] Large, W. Edward, and C. Palmer, "Perceiving temporal regularity in music," *Cognitive Science.*, vol. 26, no. 1, pp. 1-37, 2002.

[9] A. Edelman and H. Murakami, "Polynomial roots from companion matrix eigenvalues," *Mathematics of Computation.*, vol. 64, no. 210, pp. 763–776, Jan. 1994.

[10] K. Dressler, "Pitch estimation by the pair-wise evaluation of spectral peaks." *Audio Eng. Soc. Conf.: 42nd Int. Conf.: Semantic Audio*. Audio Eng. Soc., Jul. 2011.

[11] (2007. Feb.) Test data for inharmonicity estimation. [Online]. Available: http://www.acoustics.hut.fi/publications/papers/pfd/

[12] (1997) University of Iowa Electronic Music Studios. [Online]. Available: http://theremin.music.uiowa.edu/index.html

[13] F. Rigaud, "Models of music signals informed by physics. Application to piano music analysis by non-negative matrix factorization," PhD thesis, Telecom Paris Tech, Paris, France, 2013.

[14] V. Emiya, N. Bertin, B. Bertrand, and R. Badeau, "A piano database for multipitch estimation and automatic transcription of music," Telecom ParisTech, Paris, France, 2010. Available: http://www.tsi.telecom-paristech.fr/aao/en/2010/07/08/maps-database-a-piano-database-for-multipitch-estimation-and-automatic-transcription-of-music/

[15] F. Weninger, C. Kirst, B. Schuller and H. J. Bungartz, "A discriminative approach to polyphonic piano transcription using supervised non-negative matrix factorization," in *Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP),* Vancouver, Canada, May 2013, pp. 6–10.

[16] S. Bock, and M. Schedl, "Polyphonic piano note transcription with recurrent neural network," in *Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP),* Kyoto, Japan, Mar. 2012, pp. 121–124.

[17] A. Galembo, and A. Askenfelt, "Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano," *IEEE Trans. Speech and Audio Process.,* vol. 7, no. 2, pp.197-203, Mar. 1999.

[18] B. Lehman, "Bach's extraordinary temperament: our Rosetta Stone," *Early Music.,* vol. 33, no. 1, pp.3-23, Feb. 2005.