PIANO TRANSCRIPTION USING PATTERN RECOGNITION: ASPECTS ON PARAMETER EXTRACTION

Luis I. Ortiz-Berenguer

Dept. Audiovisual and Comm. Eng.
Universidad Politécnica de Madrid, Spain
lortiz@diac.upm.es

F. Javier Casajús-Quirós, Marisol Torres-Guijarro, J. A. Beracoechea

Dept. Signals, Systems and Radiocommunications
Universidad Politécnica de Madrid, Spain

ABSTRACT

A method for chord recognition for piano transcription has been previously presented by the authors. The method presents some limitations due to errors in parameter extraction carried out during the training process. Parameter extraction of piano notes is not as straightforward as sometimes can be thought. Spectral components detection is necessary but not enough to obtain accurately some note parameters. The inharmonicity coefficient *B* is one of the parameters that are difficult to evaluate. The obtained value of *B* is different for every partial used to calculate it, and sometimes, these differences are high. Tuning with respect to tempered scale is another important note parameter. The problems arise when we try to measure the tuning of a note belonging to octaves 0 or 1, because the fundamental is radiated by the soundboard with a very low level and, therefore, it is not captured by the recording microphone and cannot be measured.

A method to avoid these drawbacks is presented in this paper, including an explanation of the basis.

1. INTRODUCTION

A method for chord recognition for piano transcription has been previously presented by the authors. The method is based on spectral pattern matching using a set of spectral patterns generated by a physical model of the piano. The identification of several notes belonging to a chord is performed using iterative note detection with spectral subtraction. Both the spectral patterns and the subtraction masks are generated using a physical model of the piano. The proposed model makes use of several parameters in order to calculate patterns and masks. These parameters are obtained by a training process, using only a few notes [1].

Some of the limitations of the chord recognition algorithm have to do with the fact that subtraction masks do not fit well the actual spectrum of the piano note. The reason for this drawback is related with some degree of imperfection during parameter extraction that is carried out during training process.

Piano notes have two main characteristics which are unusual in typical musical modeling. They are inharmonic (its partials are not harmonically related) and they are not exactly tuned to tempered scale.

Moreover, the lower octave notes have not significant radiation of its fundamental frequency, so it is not present on recorded signals using microphones.

This paper will present some aspects and solutions for extracting, the more precisely possible, the following parameters: inharmonicity coefficient *B* and fundamental tuning. Both parameters are essential for note identification and for spectral subtraction used in piano chord recognition [2].

2. INHARMONICITY COEFFICIENT

Inharmonicity appears due to string stiffness. As a result, every partial has a frequency that is higher than the corresponding harmonic value. Moreover, the higher the partial order, the higher the separation from the harmonic value.

The frequencies are calculated using [3]:

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

where f_0 is the fundamental frequency for a flexible string (without stiffness) with hinged ends and the "inharmonicity coefficient" *B* is defined as [3]:

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

where *E* is the Young's Module, *d* is the string diameter, *L* is the string length and *T* is the string tension.

It can be seen that the fundamental frequency of stiff strings differs from that of flexible strings. We can obtain:

$$f_n = n f_1 \frac{\sqrt{1+n^2 B}}{\sqrt{1+B}}\ , \tag{3}$$

which is very useful since f_0 cannot be directly measured in actual strings, because they always have stiffness and we do not know the a priori value of *B*.

Apparently, extracting the value of *B* during the training process is as simple as measuring a pair of partial (e.g. f_1 and f_n) and calculating:

$$B = \frac{\delta - 1}{n^2 - \delta} \tag{4}$$

where

$$\delta = \left(\frac{f_n}{n f_1}\right)^2 .$$

However, the results are different depending on the selected partials. This is due to the fact that frequency values of partials are not only affected by string stiffness, but are also affected by bridge impedance (i.e. soundboard impedance).

2.1. Soundboard induced inharmonicity: I_{SB}

The previous equation from Fletcher was obtained considering both ends of the string hinged. Hinged boundary conditions allow the end of the string to have slope but not to move, so the impedance of the string support is infinite. But actual bridges present finite impedance controlled by the soundboard impedance.

The soundboard impedance is a function varying with frequency, controlled by the joint effect of the modes. The effect of the soundboard impedance on the string vibration frequency is such that if the string tries to vibrate with a frequency above a resonant frequency of the soundboard, the resulting frequency is even higher. If string tries to vibrate with a frequency below a resonant frequency of the soundboard, the resulting frequency is even lower [4][5].

So, the effect of the soundboard seems to be to prevent the string to vibrate with a frequency equal to a resonance of the soundboard. The exception is when the string tries to vibrate with a frequency coincident with a soundboard resonance. In that case, due to the high resistive value of the soundboard impedance, the soundboard has no effect on the string frequency.

Actually, the partial frequencies are modified by the effect of the soundboard impedance. We have called this variation "Soundboard induced Inharmonicity" I_{SB} so the frequency of any partial must be modeled as:

$$f_n = \tilde{f}_n \cdot I_{SB}\left(\tilde{f}_n\right) \quad (5)$$

where

$$\tilde{f}_n = n\, f_0 \sqrt{1+n^2 B}$$

and then, substituting f_0:

$$f_n = n\, f_1 \frac{\sqrt{1+n^2 B}}{\sqrt{1+B}} \frac{I_{SB}(\tilde{f}_n)}{I_{SB}(\tilde{f}_1)} \quad (6)$$

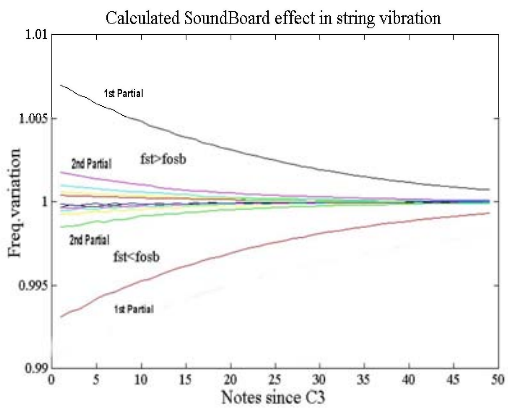

Figure 1: *I_{SB} values calculated for various notes and partials using certain values of impedance and resonances of the soundboard. Higher partials have less frequency variations than lower ones. This variation is lower for higher notes. fst>fosb indicates string frequency is above the nearest resonance of the soundboard and deviation is higher than unity.*

The difficult task of modeling soundboard impedance as well as the impossibility of knowing the soundboard characteristics from piano recordings, make it impossible to evaluate I_{SB} corresponding to any partial of any note.

Previous studies [6][7] show that the value of I_{SB} decreases with partial order and note(for the same kind of impedance behavior, i.e. above or below a soundboard resonance), and is not just a function of frequency (see Figure 1). So, the fundamental of lower notes are very much affected by I_{SB} whereas the higher partials are less affected.

2.2. Modified calculation of *B*

Taking into account equation (6), the value of B that is actually calculated from the measured partial's frequencies is:

$$B = \frac{\delta - \varepsilon}{\varepsilon n^2 - \delta}, \quad (7)$$

where

$$\delta = \left(\frac{f_n}{n\, f_1}\right)^2$$

and

$$\varepsilon = \left(\frac{I_{SB}(\tilde{f}_n)}{I_{SB}(\tilde{f}_1)}\right)^2.$$

As ε cannot be evaluated, the value of B has error, except if ε is near 1 in which case the equation (3) can be still used.

From Figure 1 it is evident that ε cannot have a value near 1 if we use the fundamental as a reference for the calculation of B. If we want to use as a reference one partial different from the first, we have to rewrite equation (6) as:

$$f_n = \frac{n}{m} f_m \frac{\sqrt{1+n^2 B}}{\sqrt{1+B}} \frac{I_{SB}(\tilde{f}_n)}{I_{SB}(\tilde{f}_m)} \quad (8)$$

where m is the order of the lower partial used in calculation (previously m was 1). Equation (7) becomes:

$$B = \frac{\delta - \varepsilon}{\varepsilon n^2 - \delta m^2}, \quad (9)$$

where

$$\delta = \left(\frac{m f_n}{n\, f_m}\right)^2$$

and

$$\varepsilon = \left(\frac{I_{SB}(\tilde{f}_n)}{I_{SB}(\tilde{f}_1)}\right)^2.$$

If partials m and n are correctly selected, ε can be considered nearly 1 and B can be calculated with a low error using the equation:

$$B = \frac{\delta - 1}{n^2 - m^2 \delta}, \quad (10)$$

where

$$\delta = \left(\frac{m f_n}{n\, f_m}\right)^2.$$

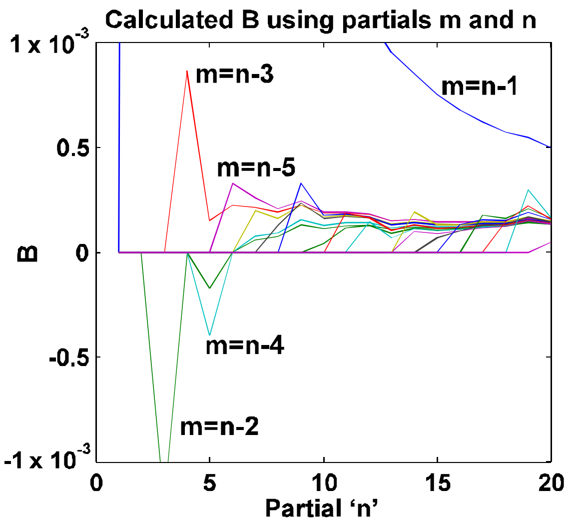

Figure 2: *Values of B calculated for A0 note. The higher partial n appears in the horizontal axis. The lines correspond to different values of reference partial m expressed as difference with respect to partial n. The zero values that appear have not been actually measured. They have been manually included in order to fit the needs of the plotting algorithm.*

Figure 2 shows the calculated values of *B*, using equation (10), for note A0 from a Steinway&sons grand piano (approx. 2.9 m. long). The actual value is about 2.10^{-4}.

It can be seen that some of the obtained values of *B* are negative, which is evidently erroneous, because it is physically impossible. It is evident that ε is not always near 1. It is important to note that the use of two consecutive partials (i.e. $m=n-1$) must be avoided. It also must be avoided the use of nearest partials except if they are very high order partials. After several tests, using different notes, we have concluded that for lower octaves, *n* must be selected between 14 and 20 and *m* must be selected between 1/2 and 2/3 of *n* (that is an intermediate partial order).

This selection of calculation partials is a bit more problematic in octaves 3 to 5 where the number of available partials decreases. And it is especially critical in the case of the higher octaves 6 to 8.

2.3. *B* calculation on octaves 6 to 8.

Notes belonging to octaves 6 to 8 have two problematic characteristics: only two partials have enough level to be accurately measured and the notes always present a high degree of non-linearity. The first issue is not very important due to the fact that I_{SB} tends to be very little noticeable at those frequencies above C6 with little difference between fundamental and second partial. The non-linearity is the main problem, because the second harmonic of the fundamental appears very close to the second partial, so they almost cannot be distinguished, except if values of *B* are high enough.

Figure 3 shows the second partial "zone" of note C7, for which the *B* value allows us to distinguish the second harmonic from the second partial. It can be seen that some of the spectral peaks are positioned at twice the fundamental frequency, so they are not the second partial but an IM (InterModulation) product (i.e. 2nd harmonic).

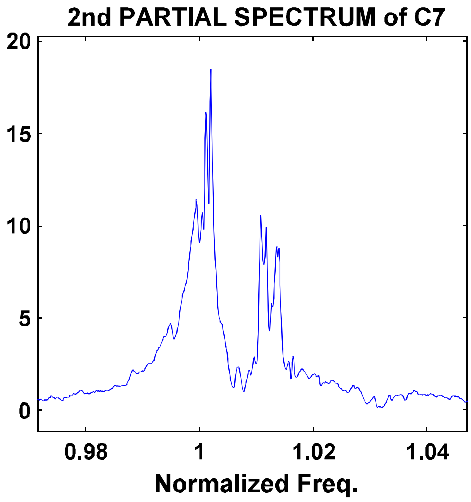

Figure 3: *Spectrum of C7 around second partial. It can be seen that more than 3 peaks are present. Frequency is normalized respect to two times the fundamental (2nd nonlinear harmonic).*

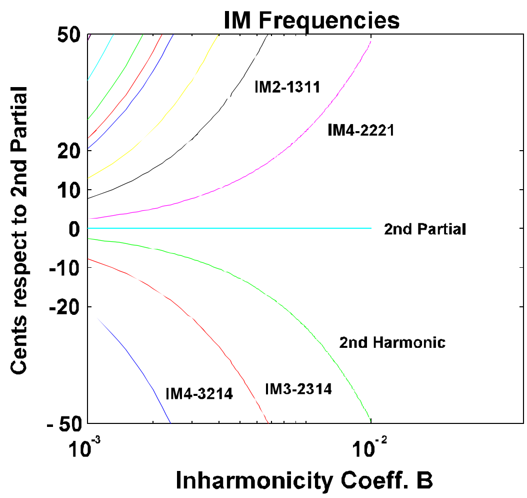

Figure 4: *Calculated position of IM products around second partial for several values of B. The shown values of B are typical for notes above C6 .*

It is also very interesting to notice that the second harmonic has higher level than second partial. This adds even more problems to parameter extraction.

The non-linearity effects have been modeled using InterModulation products (IM) [8]. Two IM products appear around the second partial. Those products are separated from the 2nd partial an amount that can be calculated depending on the value of *B* coefficient (Figure 4). As the *B* value is not known, but is being measured, a bounded value of the separation must be approximated in order to carry out the correct parameter extraction.

It is necessary to measure precisely the second partial in order to calculate B coefficient, but this is not an easy task. The several spectral peaks must me measured and the second harmonic values have to be discarded.

3. TUNING OF THE FUNDAMENTAL

Musical instruments tend to be tuned according to tempered scale. However, in the case of piano, inharmonicity establishes that the second partial of a note is slightly above the fundamental of the note one octave higher. If tempered tuning is performed, an audible beating that is not comfortable will be produced. This effect makes it necessary to tune the piano so that the fundamentals of higher notes are slightly above their tempered values, in order to be coincident with the second partial of the note one octave below [9]. In this case, beating is avoided. This can be expressed using:

$$f_{1,\,i\,+12} = 2 f_{1,\,i} \frac{\sqrt{1+4B_i}}{\sqrt{1+B_i}} \;, \tag{11}$$

where i is the number of the note and i+12 is the number of the note one octave above.

As tempered tuning is not used, the tuning of every note can be expressed as:

$$f_{1,\,i} = A \cdot f_{1T,\,i} \tag{12}$$

where $f_{1T,i}$ is the tempered value of the fundamental of note i, and A is the "Tuning Factor". This tuning factor used to be called simply "Tuning".

The actual tuning process begins with note A0, which is tuned to 440 Hz (sometimes is tuned to 442 or even 444 Hz). The remaining notes of octave 4 are tuned according to intervals, but for simplicity, we are going to consider that all the notes of octave 4 are tuned to their tempered value. With this assumption, Figure 5 shows the value of "tuning factor" calculated using equation (11).

It can be seen that the values calculated for octaves above octave 4 are very coincident with the values of the Railsback curve [10].

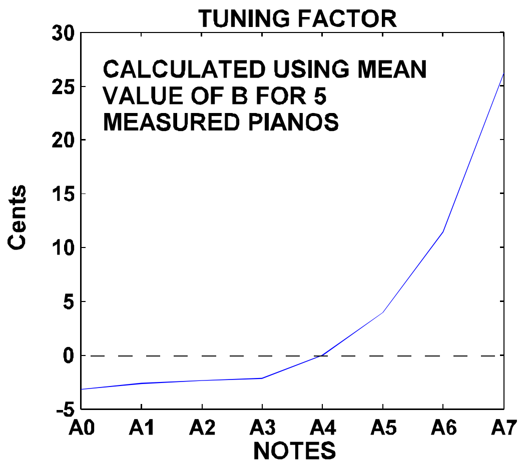

Figure 5: *Calculated tuning considering inharmonicity coefficient B.*

The Railsback curve is an average of several measures of actual piano tunings (Figure 6).

It can be noticed that the Railsback curve for octaves below octave 2 differs from the results obtained considering the B coefficient. In conclusion, inharmonicity does not justify the tuning factor for lower octaves. The effect of I_{SB} must be considered again to explain this difference.

3.1. Effect of I_{SB} on tuning

The condition for tuning can be rewritten as:

$$f_{1,i+12} = 2 f_{1,i} \frac{\sqrt{1+4B_i}}{\sqrt{1+B_i}} \frac{I_{SB}(\tilde{f}_{2,i})}{I_{SB}(\tilde{f}_{1,i})} \;, \tag{13}$$

where

$$\tilde{f}_{1,i} = f_{0,i}\sqrt{1+B_i}$$

$$\tilde{f}_{2,i} = 2 f_{0,i}\sqrt{1+4B_i} \;.$$

This introduces some changes in the tuning values for octaves above octave 4, but they are not very important. But for octaves below octave 4, the tuning process is the reverse, so note i+12 has been previously tuned and after, the note i has to be tuned. Then, the equation for tuning lower octaves is:

$$f_{1,i} = \frac{f_{1,i+12}}{2} \frac{\sqrt{1+B_i}}{\sqrt{1+4B_i}} \frac{I_{SB}(\tilde{f}_{1,i})}{I_{SB}(\tilde{f}_{2,i})} \tag{14}$$

For octaves 0 to 2, where the values of B range between 10^{-4} and 10^{-3}, the term depending on B only justifies a tuning factor of up to −3 cents. However, for those octaves, the soundboard presents only a few resonances that are separated and then, the values of the term depending on I_{SB} can be very high. This term can justify tuning factors of up to −40 cents.

These results are coincident with Railsback curve.

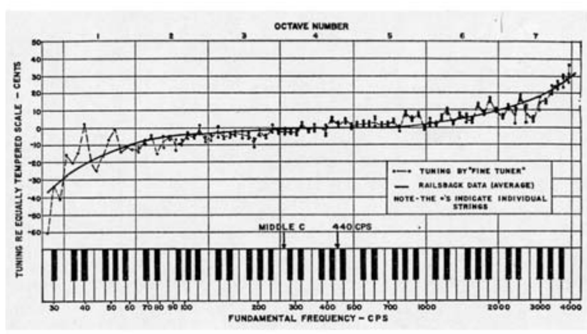

Figure 6: *Railsback curve and some piano tuning measured by Martin and Ward [10]. Note the great values of tuning factor (deviation respect to tempered tuning) at higher octaves (with high B values) and also at lower octaves (with low B values).*

4. CONCLUSIONS

Evaluation of the inharmonicity coefficient *B* requires to measure the frequency of the partials and to select the correct partials to be used in the calculation. Lower partials are very affected by the soundboard and their use must be avoided. Due to the fact that the frequency of the partials may be increased or decreased by the effect of soundboard, several values of *B* must be measured using several partials and the mean value of them can be considered to be the string *B* value.

Measuring *B* in higher octaves can be done using only the second and first partials. Soundboard effect is almost negligible but care must be taken in order to avoid the error of measure the second harmonic (non-linear product) instead of the second partial.

Tuning measure is almost straightforward for octaves above 3, but it can be difficult for lower octaves where the fundamental has very low level due to radiation limitations. In these cases, the second partial may be used as reference and the fundamental can be calculated using the *B* value previously determined and some approximation to the value of the ratio of I_{SB} in equation (6).

5. REFERENCES

[1] L. I. Ortiz-Berenguer, F. J. Casajús-Quirós, "Pattern Recognition of Piano Chords Based on Physical Model," *112th Convention of the Audio Engineering Soc.*, Munich, Germany May 2002. Paper 5525.

[2] L. I. Ortiz-Berenguer, F. J. Casajús-Quirós, "Polyphonic Transcription Using Piano Modeling for Spectral Pattern Recognition," *5th Int. Conf. on Digital Audio Effects (DAFX-02),* Hamburg. Germany, September 2002.

[3] H. Fletcher, "Normal vibration frequencies of a stiff piano string," *J. Acoust. Soc. Am.,* vol. 36, pp. 203–209, 1964.

[4] L. E. Kinsler, A. R. Frey, A. B. Coppens, J. V. Sanders, *Fundamentals of acoustics*. John Wiley & Sons, 4th edition, 1999.

[5] T. D. Rossing and N. H. Fletcher, *Principles of vibration and sound*. Springer-Verlag, New York, 1995.

[6] L. I. Ortiz-Berenguer, F. J. Casajús-Quirós, "Identification of Piano Chords Using Patterns from Acoustic Model," *9th Int. Congress on Sound and Vibration,* Florida, July 2002.

[7] L. I. Ortiz-Berenguer, *Transcripción automática de acordes musicales*. Tesis Doctoral. (*Automatic transcription of musical chords*. Ph.D. Thesis. Written in Spanish), Madrid 2002.

[8] L. I. Ortiz-Berenguer, F. J. Casajús-Quirós, "Non linearity modeling for spectral pattern recognition in piano chords," *116th Convention of the Audio Engineering Soc.,* Berlin, Germany, May 2004.

[9] J. Lattard, "Influence of inharmonicity on the tuning of a piano: Measurements and mathematical simulation," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 46–53, 1993.

[10] D. W. Martin, W. D. Ward, "Subjective evaluation of musical scale temperament in pianos," *J. Acoust. Soc. Am.,* vol. 33, no. 5, pp. 582–585, 1961.