IMPROVED METHOD FOR EXTRACTION OF PARTIAL'S PARAMETERS IN POLYPHONIC TRANSCRIPTION OF PIANO HIGHER OCTAVES

Luis Ortiz-Berenguer

Dept. Audiovisual and Comm. Engineering
EUIT Telecomunicacion-UPM,Madrid,Spain
lortiz@diac.upm.es

S.Torres-Guijarro,F.J.Casajus,J.Beracoechea

Dept.Signals, Systems and Radiocomm.
ETSI Telecomunicacion-UPM,Madrid,Spain

ABSTRACT

Polyphonic transcription is specially challenging in piano higher octaves due to the complexity of the spectrum of notes and therefore, chords. Besides the fundamental and second partial components, other spectral elements appears. The three peaks related to the unison as well as the second harmonic of the fundamental unison can be distinguished in most measures. Furthermore, intermodulation components are also present when non-linearity is high enough. This paper compares several methods to improve the training process that allows to synthesize the spectral patterns and masks used in transcription methods.

1. INTRODUCTION

One method for polyphonic transcription is based on matching a spectral patterns database. The spectral database consists on a set of spectra. These are called spectral patterns'. Mainly the whole set of note's spectrum has to be available. Obtaining the whole set by recording all the single notes would not be elegant, so it has to be obtained from a few training notes and an interpolation process. One possible interpolation process relies on using a physical modeling of the instrument, here piano [1].

When identifying single notes, only the spectral patterns are necessary. To identify polyphonic chords, a spectral subtractive approach is used. A second spectral database is necessary for this function. Those are the called spectral masks'. Chords are identified by iterative detection of the notes that compose them. Every time a note is detected, its spectral mask is applied to remove the spectral components of the note away from the analyzed chord.

Obtaining the spectral databases for a given piano requires a specific training stage. Pianos are enough different regarding their Inharmonicity and tuning. This make the use of a unique set of spectral patterns impossible, as previous works have shown [1]

Moreover, the intensity applied when playing a piano is related to the nonlinear behavior of the vibration. Some studies have shown that the nonlinear effect can be modeled using the intermodulation products (I.M.) [2], which can be easily calculated. In harmonic vibrations, IM products are coincident to partials so no effect rather than change of level appears to exist. When the vibration is inharmonic, the IM products deviate from the partials producing new spectral components. If spectral resolution is low (as it is the case of lower octaves) these components are not distinguished and it only appears to be a widening of the partial. Nevertheless, in higher octaves, the IM products are clearly distinguished from the partials. The non-linearity is too high and the IM products have high levels, sometimes higher than the ones of the partials.

It is a key factor for subtracting masks not to leave out those IM associated with its note. If not, those IM can be considered a different note in the following iterative stage.

Previous results obtained using this transcription algorithm on piano chords have shown that mistakes take place when attempting to identify higher octaves chords, specially when they are played with high force[2].

Therefore, spectral masks for higher octaves need to be synthesized highly accurately, and with even more detail than spectral patterns, which are used to detect the notes.

The position and width of the meaningful partial for each note have to be known to synthesize the spectral patterns and masks. Besides, masks require to know the expected IM products.

The position of partials is obtained from two parameters: the frequency of the fundamental (f_1) and the Inharmonicity coefficient (B). Any partial's frequency (f_n) is related to the fundamental by [3]:

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

The frequency of the fundamental may be specified using the tuning factor and the well-tempered scale.

$$f_1 = f_{wt} \cdot T \tag{2}$$

Besides, the effect of the soundboard impedance affects the definitive position of a partial [4]. This effect can be considered negligible in higher octaves [5].

The width of the partials is related to either its decay time or spectral estimation limits (window and data length).

Regarding non-linearity products, the meaningful of each of them are related to the level of the original partials involved. In higher octaves, the second harmonic of the fundamental is one of the more relevant. The products that depend on two partials have been studied in previous work [2] and their position depends a lot on the value of B (Inharmonicity coefficient). Specially important are those obtained from the fundamental and the second partial, which are the two higher level components.

Higher octaves (sixth and seventh) notes have values of B greater than 0.002, which lead to IM products clearly distinguishable from the partials. It also leads to the second harmonic of the fundamental to be clearly below the frequency of the second partial.

Moreover, every note is composed by the vibration of three strings that are slightly deviated in frequency (what is called unison') [6]. The above explained has to be applied to each string. It is widely assumed that the non-linearity can be considered to exist in each string, but not in the unison. This is explained by the fact that the main nonlinear process is the hammer excitation rather than the bridge movement [7].

2. ANALYZING HIGHER OCTAVES

Figure 1 shows the spectrum of note C6 played softly in a Steinway grand piano. Many other higher octave notes have been recorded and all of them show the same type of spectrum or with even higher second harmonic.

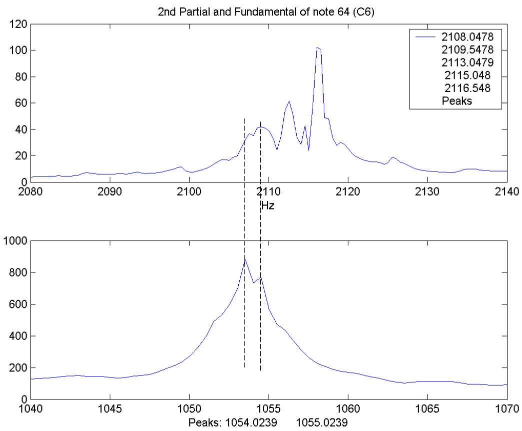

Figure 1: *Second partial and fundamental segments of the C6 spectrum. Note the existence of second harmonic near the second partial. Regarding tuning, the well-tempered frequency of C6 would be 1046.5 Hz. The effect of inharmonicity on tuning is clear. Take also into account that the maximum peaks in fundamental and second partial segments do not correspond to the same string of the unison*

How can we analyze this note during training to obtain tuning and Inharmonicity coefficients? How must be synthesized the spectral mask to remove all these components? Two approaches have been tested and are presented in this paper. The first approach is the simplest and consists in detecting fundamental and second partial as the peak with highest level on every segment of the spectrum. The second is to detect all the peaks and do a further analysis to decide which are the fundamentals and the second partials. The last is carried out taking into account the unison.

The first approach presents a drawback removed from the method for this paper. Some notes, specially if they are played loudly, present a second harmonic that has a higher level than the second partial, leading to a value of zero for the Inharmonicity coefficient. For this approach, the second harmonic zone has been masked.

The second approach has been carried out using three different methods to calculate B value.

Once the fundamental and the second partial have been detected, the values of tuning factor and Inharmonicity coefficient can be obtained, completing the parameter extraction process. The higher octaves present spectra that contain very little partials, so only the first two are used for this analysis.

The analyzing task is more challenging that it could seem. Figure 2 shows the spectrum of G#7 played in a Kawai grand piano. Effects of non-linearity are clear: higher second harmonic and more than three peaks around second partial (some IM product).

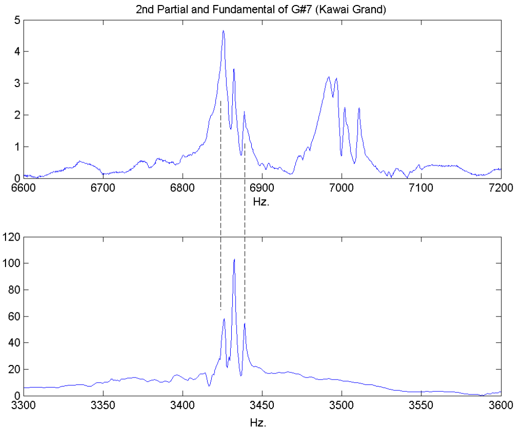

Figure 2: *Second partial and fundamental segments of the G#7 spectrum. Note the existence of second harmonic near the second partial. Its level is even higher than the second partial's one. Well-tempered frequency of G#7 would be 3322,5 Hz.*

2.1. First approach

The spectrum is segmented around both fundamental and second partial, and the maximum valued peak is detected in every segment. These are considered the fundamental and the second partial. The value of parameter B is calculated using the equation:

$$B = \frac{C_0^2 - 4}{16 - C_0^2} \quad (3)$$

where

$$C_0 = \frac{f_2}{f_1} \quad (4)$$

The effect of the soundboard impedance can be considered negligible. If the second harmonic of the fundamental (i.e., $2 \cdot f_1$) is detected instead of the second partial, the obtained value for B would be zero. The spectral segment around the second partial is corrected so it does not include the second harmonic segment. The second harmonic segment is calculated using the detected fundamental.

2.2. Second approach

The spectrum is segmented, but all the relevant peaks are detected in every segment, keeping the unison, existing IM products and the second harmonic. The second harmonic segment is removed using all the information in the fundamental segment (not only the maximum peak). This removal is carried out for the detection of the second partial, but the second harmonic segment is used for helping validation of the fundamental unison. The remaining peaks in the second partial segment include the second partial unison and some IM products. The three more likely peaks to be the unison are selected using mainly two considerations: Second partial uses to have higher level than the nearest IM products and the positions of the IM products can be approximated using an expected value of B [2]. Regarding the fundamental, up to four peaks can be detected sometimes (unison plus one IM product), but often only two or three peaks are detected. This is due to spectral resolution limits. All the unison is present but not all the peaks' frequency value are clearly known. Due to this, several methods start at this point.

The first method analyzes the Nx3 matrix that can be obtained by calculating the value of parameter B using each second detected partial and each detected fundamental. The theoretically expected matrixes present a clear behavior regarding their values.

Let us consider the three fundamentals of the unison. The first one is considered the main or tuning value. The second is deviated δ_1 times and the third is deviated δ_2 times from the first one. Both δ_1 and δ_2 are higher than unity, and $\delta_2 > \delta_1$ [6].

If the three fundamentals of the unison and the three second partials of the unison are measured, the following matrix could be obtained:

$$\begin{pmatrix} C_0 & \delta_1 C_0 & \delta_2 C_0 \\ \frac{1}{\delta_1} C_0 & C_0 & \frac{\delta_2}{\delta_1} C_0 \\ \frac{1}{\delta_2} C_0 & \frac{\delta_1}{\delta_2} C_0 & C_0 \end{pmatrix} \tag{5}$$

Every term in the matrix corresponds to the ratio between a second partial and a fundamental. Rows order is the fundamental's order and column's order is the second partial's order. The main diagonal has the value of C_0 repeated three times and will be called 'C_0-diagonal' in the following discussion. The Inharmonicity coefficient B is obtained from that ratio C_0 by using the equation (3). Real matrices are not so easy.

If δ_2 is δ_1^2, the unison is symmetrically tuned. This leads to more repeated values inside the matrix, making more difficult the identification of the C_0-diagonal in a real matrix.

$$\begin{pmatrix} C_0 & \delta_1 C_0 & \delta_2 C_0 \\ \frac{1}{\delta_1} C_0 & C_0 & \delta_1 C_0 \\ \frac{1}{\delta_2} C_0 & \frac{1}{\delta_1} C_0 & C_0 \end{pmatrix} \tag{6}$$

Nevertheless, C_0 is still the value repeated more times. Still, this can change if any of the rows are left, as happens when only two fundamentals are detected.

$$\begin{pmatrix} \frac{1}{\delta_1} C_0 & C_0 & \delta_1 C_0 \\ \frac{1}{\delta_2} C_0 & \frac{1}{\delta_1} C_0 & C_0 \end{pmatrix} \tag{7}$$

$$\begin{pmatrix} C_0 & \delta_1 C_0 & \delta_2 C_0 \\ \frac{1}{\delta_2} C_0 & \frac{1}{\delta_1} C_0 & C_0 \end{pmatrix} \tag{8}$$

$$\begin{pmatrix} C_0 & \delta_1 C_0 & \delta_2 C_0 \\ \frac{1}{\delta_1} C_0 & C_0 & \delta_1 C_0 \end{pmatrix} \tag{9}$$

Equation (8) shows three possible matrices when only two fundamentals are detected and unison are symmetrically tuned. Although in the theoretical matrix is possible to identify the C_0-diagonal, it is not the case if only calculated values are available. In some cases, we have two diagonals with two repeated values. In other case there is no repeated value.

The analysis of several cases has to be carried out by the algorithm and some of them lead to an unavoidable error. The algorithm tries to make that error the minimum possible when it detects one of those cases.

Knowing which is the lost fundamental is not possible. Therefore, the error cannot be completely corrected.

The second method simplifies the process taking the geometric mean of all the values in the matrix. Theoretically, calculated matrix values present a clear symmetrical distribution. The geometric mean of matrix in equation (5) is C_0.

The geometric mean is more accurate when all the values of the unison are present. In actual measures, the lack of fundamentals can lead to an error. For the three matrices of (8) the errors would be:

$$\begin{gathered} \sqrt[6]{1/(\delta_1\delta_2)} \\ \sqrt[6]{\delta_1^2/\delta_2} \\ \sqrt[6]{\delta_2^2/\delta_1}. \end{gathered} \tag{10}$$

These can be clearly lower than the obtained by using the first approach, which could be as high as δ_2.

Furthermore, in spite of the method, the measurement of the frequency values of fundamentals and second partials is also limited by the spectral resolution. This further affects to the calculated matrix values. In some cases, even the main diagonal has not three repeated values but three very similar ones.

The third method calculates the mean of all the values. In the case of the 3x3 matrix, this method produce error, but in some of the cases of 2x3 matrices, it can lead to a lesser error. This mean values are included in the study for comparison purposes mainly.

2.3. Expected values

To interpolate all the note's spectra from a few training notes, some kind of relationship between the values of the parameters for different notes has to exist. The strings for the higher octaves notes follow several designing rules. Values of B and tuning for a given piano depend on a few initial designing values. This permits to expect the values of B and tuning parameter to carry out the interpolation. Figure 3 shows the expected B values for several initial designing values (that is, several different pianos).

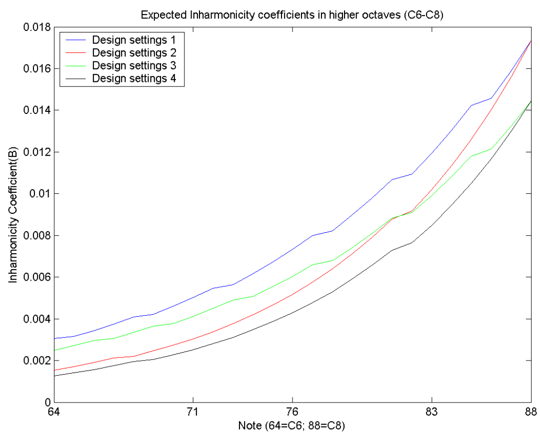

Figure 3: *Curves of B values calculated using the theoretical equations of the model. Only four expected curves are shown*

3. RESULTS

The presented results show a comparison of the obtained values of parameter B using all the three methods plus the expected curve.

Figure 4 shows the results for the three methods of the second (new) approach. For comparison, figure 5 shows the same plus the results for the first (maximum peaks) approach and one expected curve.

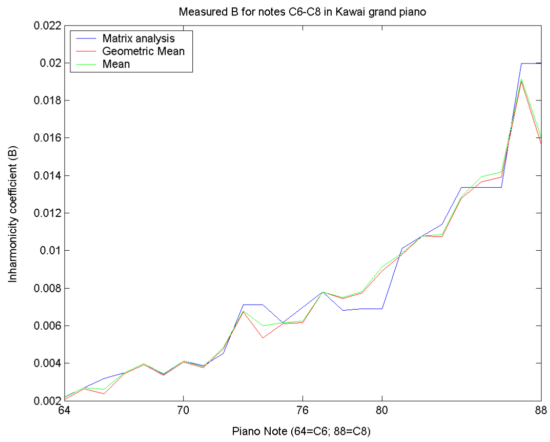

Figure 4: *Curves of B values calculated by using the second approach.*

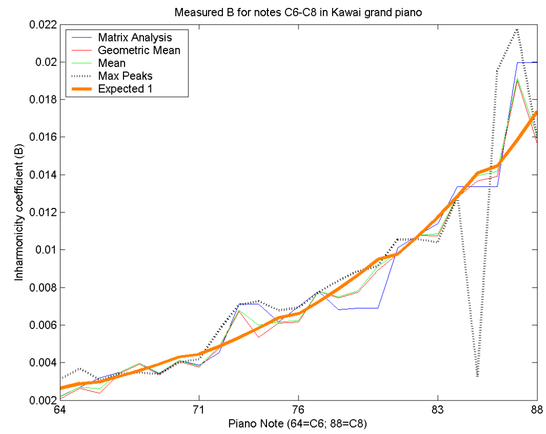

Figure 5: *Curves of B values calculated by using the second approach, plus the curve using the first (dotted line) and one possible expected curve (thick line)*

Results show that the new approach has lesser variance respect to the expected curve, especially in the higher notes of seventh octave.

4. DISCUSSION

The presented results show the values of the B parameter measured for all the notes. Then, the training set is the whole set. This is not actually the definitive training method. For real training, the complete curve has to be interpolated from a few trained notes. Different curves can be interpolated depending on the selected training notes.

Using the more accurate (lesser variance) analysis method for the training stage is very important. The more accurate might be the 'Geometric Mean' method corresponding to the second approach presented and proposed by the authors.

The first approach presents the highest variance respect to the expected curve. One reason is that maximum valued peaks are not related to one specific string in the unison. Moreover, spectral resolution limits makes possible to detect as a maximum the fundamental of one string and the second partial of other string, leading to an error in the calculated value of parameter B.

5. CONCLUSIONS

The spectral patterns and masks used in polyphonic transcription of piano chords can be improved by obtaining better estimation of the note's spectral parameters. The presented methods increase the accuracy of the measurement of training note's inharmonicity coefficients.

Results are specially good in the seventh octave, which is the one that produced more false detections in the transcription tests [8].

6. REFERENCES

[1] Ortiz-Berenguer L.I. Casajus-Quiros J. Torres-Guijarro M., "Multiple piano note identification using a spectral matching method with derived patterns," *Journal Audio Engineering Society*, vol. 53, no. 1-2, January/February 2005.

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

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

[4] Morse P.M., *Vibration and sound*, McGraw-Hill, New York, 1948.

[5] Ortiz-Berenguer L. Casajus J. Torres-Guijarro M. Bera coechea J., "Piano transcription using patter recognition: Aspects on parameter extraction," *7th Intl.Conference on Digital Audio Effects (DAFX-04)*, October 2004, Naples.Italy.

[6] Kirk R.E., "Tuning preferences for piano unison groups," *J.Acoust.Soc.Am.*, vol. 31, no. 12, pp. 1644–1648, 1959.

[7] Hall D., "Piano string excitation.vi:nonlinear modeling," *J.Acoust.Soc.Am.*, vol. 92, no. 1, pp. 95–105, 1992.

[8] Ortiz-Berenguer L.I. Casajus-Quiros J. Torres-Guijarro M., "Non-linear effects modeling for polyphonic piano transcription," *6th Intl.Conference on Digital Audio Effects (DAFX-03)*, September 2003, London.UK.