POLYPHONIC TRANSCRIPTION USING PIANO MODELING FOR SPECTRAL PATTERN RECOGNITION

Luis I. Ortiz-Berenguer

Dept. Audiovisual and Communication Engineering

Polytechnic University of Madrid. Spain

lortiz@diac.upm.es

F.Javier Casajús-Quirós

Dept. Signals, Systems and Radiocommunications

Polytechnic University of Madrid. Spain

javier@gasp.ssr.upm.es

ABSTRACT

Polyphonic transcription needs a correct identification of notes and chords. We have centered the efforts in piano chords identification. Pattern recognition using spectral patterns has been used as the identification method. The spectrum of the signal is compared with a set of spectra (patterns). The patterns are generated by a piano model that takes into account acoustic parameters and typical manufacturer criteria, that are adjusted by training the model with a few notes. The algorithm identifies notes and, iteratively, chords. Chords identification requires spectral substraction that is performed using masks. The analyzing algorithm used for training, avoids false partials detection due to nonlinear components and takes into account inharmonicity for spectrum segmentation. The method has been tested with live piano sounds recorded from two different grand pianos. Successful identification of up to four-notes chords has been carried out.

1. INTRODUCTION

Automatic transcription is a difficult task that is being tried to solve. Identification of single notes has been solved in several ways [1][2][3], but identifying several simultaneous notes (chords) is a much more difficult issue [4][5]. Two main problems are: partial separation and octave ambiguity. The different behavior of the musical instruments makes more difficult to solve the problem in a generalized way. We have concentrated our efforts in piano notes and chords identification.

Identification of piano notes using pattern recognition of simple spectral parameters (i.e. frequency of partials) has been used previously with some degree of success [6], but with the need for a complete spectral database of the piano used to play the notes and chords. This database is obtained previously by recording all the notes.

We present a solution based in spectral pattern recognition by comparing not only a set of parameters but the whole spectrum note or chord with full spectrum patterns. These full spectrum patterns are generated by a simplified acoustical model that also has been developed. This model only needs a little set of notes (about 20), previously recorded from the piano, to be trained in order to generate the 88 notes spectra.

2. PIANO MODEL

2.1. Overview

The model generates the spectral patterns taking into account: inharmonicity, soundboard impedance effect on frequency vibration, tuning of fundamentals, unison tuning and error margin due to the training process.

2.2. Basic Description

Each note is generated by the vibration of a set of one,two or three strings. These string are struck simultaneously by a hammer that is felt covered. Every string is under tension and the two extremes are fixed to an iron frame. The hammer strikes near one of the extremes. Near the other extreme, the string is in contact with a wooden piece, called "bridge", that vibrates and transmit the string vibration to the "soundboard". The soundboard acts as a diaphragm becoming the real sound radiator of the piano. The vibrating length of the string is considered to be the length between the extreme nearest the hammer and the bridge.

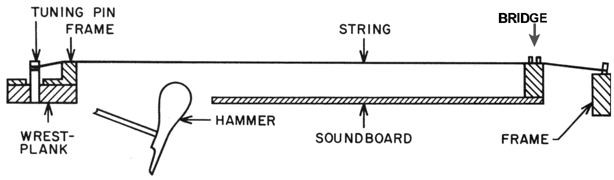

Fig. 1. *Basic description of the vibrational system*

2.3. Basis of the Model

The frequency of a piano string is mainly determined by the vibrating modes of an elastic string clamped in both extremes:

$$f_n = n\frac{c}{2L} = n\frac{1}{2L}\sqrt{\frac{T}{r_L}} \quad \textbf{(1)}$$

where L is the length (m), T is the tension (N), ρ_L is the linear density (kg/m) and n is the partial order [7].

The actual vibration frequency has to be calculated taken into account inharmonicity due to the stiffness of the string [8]

$$f_n = nf_0(1+n^2B)^{1/2} = n\frac{1}{2L}\sqrt{\frac{T}{r_L}}\sqrt{1+n^2\frac{Ep^3d^4}{64L^2T}} \quad \textbf{(2)}$$

and also the soundboard impedance effect [7]:

$$\tan(k_nL) = j\frac{Tk_n}{\omega_n Z_{SB}} \quad \textbf{(3)}$$

The piano strings are made of steel. The diameter of the string can be selected between 0.031" and 0.063" in steps of 0.001". The length of the string ‘n’ is given by the formula: $L_n=L_{88}\, s^{(88-n)}$, where L88 is the length of the 88th note string (the highest of the piano) and s is the scaling factor. L88 is usually selected to be between 5.0 and 5.4 cm, and ‘s’ has a value between 1 and 1.06 [9]. Each piano can be designed with a different value of L88 and s. The tension of strings is almost constant for octaves 3 to 7 [9] and has a value between 650N and 870N depending on L88 and d88 (diameter of string 88, that has a value between 0.031" and 0.033").

Strings of notes below C3 use to be mounted on a separated bridge and the manufacturers decisions about length and diameter are quite different, to avoid using strings several meters long. This bridge is called “Bass-bridge” and the other is called ”Trebble-bridge”

The term B in equation 2 is called “Inharmonicity factor” and it is characteristic of each string, varying along the piano, with higher values for higher notes.

2.4. Application of the Model: Pattern Generation

2.4.1. Partials frequencies

The model has to predict the value of the frequencies of several partials corresponding to all the 88 notes of a piano. Three parameters are important: fundamental frequency, inharmonicity factor B and frequency variation due to the effect of soundboard. The model takes into account the designing rules, described above, for some of its calculations and carries out previous training using a few notes to adjust some of the parameters.

2.4.2. Width of spectral pulses.

The width of the spectral pattern pulses has been selected taking into account the unison tuning (separation between the three spectral components of a note) and the soundboard effect. The pattern pulse must be wide enough to contain the spectral pulse of the partial but not too wide to avoid containing partials of adjacent notes.

2.4.3. Spectral amplitudes.

We have assigned different levels to fundamentals and other partials, in order to improve the octave ambiguity. This is a spectral weighting and does not resemble spectral distribution of the signal

3. MATERIAL AND METHODS

The model has been tested and validated using two different grand pianos (one by Yamaha and the other by Steinway) that have been recorded at the rehearsal hall of the Opera Theater of Madrid.

All the 88 notes have been recorded, to verify single note identification. For polyphonic identification purposes, two three-note chords and four four-note chords belonging to octaves 1 to 7, were also recorded.

3.1. Training the Model

Several notes (15 to 23) along the piano are used for training.

3.1.1. Training Algorithm: Parameter Extraction

This algorithm calculates and segments the spectrum looking for the fundamental of the known analyzed note and its partials. Values of fundamental frequencies, inharmonicity coefficient B and unison tuning are measured using an algorithm developed specific for this step. Approximation of tuning curve and coefficient B curve are calculated. The approximation errors are also evaluated.

This algorithm calculates the expected inter-modulation products due to non-linearity [10] and avoid to detect them as valid partials. It also takes care of rejecting the non-linear second harmonic of notes that use to have higher level than the second partial, specially in octaves 6 and 7. If these non linear products are not taken into account, false partial frequencies values can be obtained.

3.2. Identification

Identification is carried out by calculation of the inner product of the spectrum to be identified and the set of spectral patterns generated by the model. This process is called "metric calculation". The components that fit the pattern "scores" increasing the inner product. Previously, the signal spectrum undergoes a little thresholding to reject low level spectral components (mainly noise). The identification is improved using octave-predetection. The octave ambiguity has been fully removed.

In the case of chord detection, the identification is performed identifying each of the notes in an iterative process. Figure 2 shows the diagram of the identification process.

Although the basic algorithm is able to identify every single note with no error, in the case of chords, the spectral components of the other notes may produce false detection on some notes of the chords. For instance, the algorithm applied to chord C4,E4,G4 may detect C3 instead of C4 due to the existence of G4. To avoid this, a post-detection validation process is carried out on every detected note. The validated notes are grouped identifying a chord ("mapping to chord" process)

After a note is validated, the corresponding mask is applied to the spectrum of the signal in order to substract the spectral components of that note and perform the next iterative step of note detection. The mask are generated by the model in a way similar to patterns but with two differences: i)all the partials of the mask have level unity and ii)the shape of the mask pulses may be the same than patterns shape or may be rectangular, the first are named "progressive masks" and the second "hard masks".

Masks are very important because if they do not fit the signal, may result a residual spectral component after substraction. This residual component may cause detection errors in the following step.

3.3. Octave predetection

The octave predetection is not made in a octave by octave basis as we tried to do previously [11]. Instead of it, we have segmented the spectrum in 4 slots. Octaves 0,1 and 2 belongs to slot 1. Octaves 3 and 4 to slot 2, octave 5 to slot 3 and octaves 6, 7 and 8 to slot 4.

Predetecting to what slot belongs a note or a chord is performed by evaluating two parameters: spectral zone and band predetection coefficients.

We have define two spectral zones (below or above 700 Hz). We have verified from typical spectral distributions [12] and from measurements of both pianos, that notes which

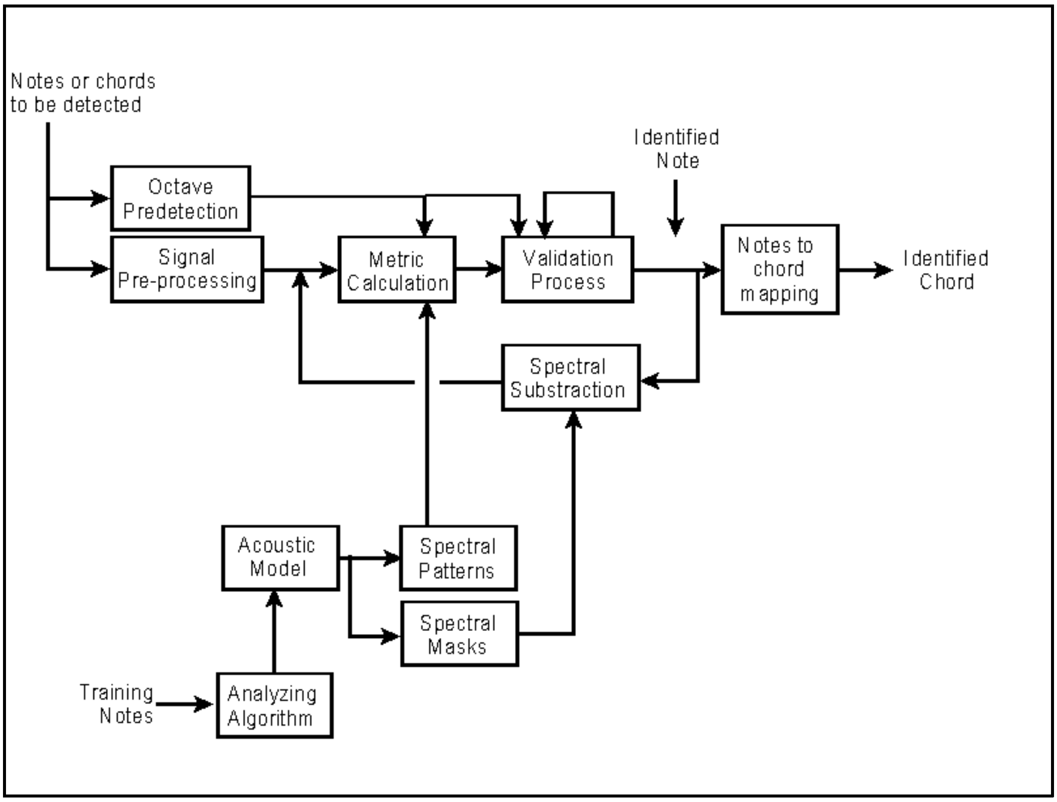

Fig. 2. *Diagram of chord identification process*

maximum level partial (easy to detect) is above 700Hz are always notes belonging to octaves 5 or higher.

The bands are three and are defined so its boundaries are the limits of octave 2,3 and 5. The bands are used to calculate the three predetection coefficients that are defined as follows:

$$CPni = \frac{1}{N^2} \sum_n x_n^2 \qquad \textbf{(4)}$$

where 'n' is the index of every spectral value of the signal (x_n) in the band "i" and N is the total number of spectral values in that band. The figure 3 shows the three coefficients calculated for every note of the recorded piano (1 to 88).

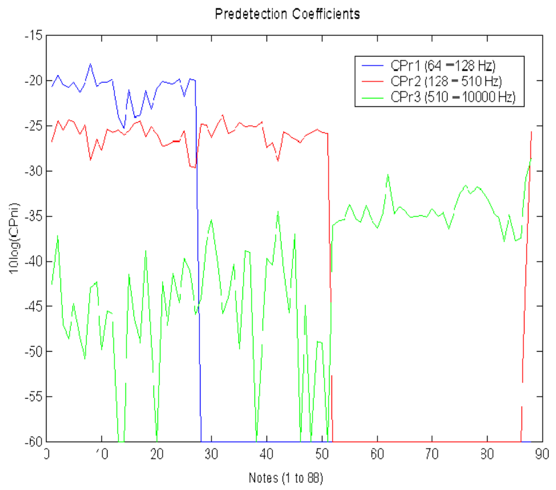

Fig. 3. *Predetection coefficients for every note*

Can be seen that CPn1 falls to very low levels for notes above 27 (B2:limit between bands 1 and 2) and CPn2 also falls for notes above 51 (B4:limit between bands 2 and 3). There is a possible ambiguity for notes 87 and 88 but is solved using the zone, because a note of zone "high" must belong to band 3 (octave 5 or higher).

It is possible to define a threshold, that can be considered independent of the piano, so each time a signal is predetected, the three coefficients are calculated and depending on the value of the calculated coefficients respect to the threshold, the band of the signal is predetected. If also the zone is calculated, the spectral slot to which the signal belongs is determined.

4. RESULTS: IDENTIFICATION OF NOTES AND CHORDS

4.1. Identification of notes

All the 88 notes of the piano, analyzed one by one, have been identified without error, using the called "basic algorithm".

Basic algorithm is the whole algorithm but without validation process and without substraction.

4.2. Identification of chords

Identification of chords is made by iterative identification of notes followed by a spectral subtraction step. The subtraction is made using the mask corresponding to previously identified pattern. The table shows the detected notes for every chord of every octave analyzed.

The validated notes are shown following the order they are validated by the system. The failed detections are bold typed.

3 note- chords identification

Analyzed Chords		Identified Chords		
C,E,G	oct1	'G1'	'E1'	'C1'
	oct2	'G2'	'E2'	'C2'
	oct3	'C3'	'G3'	'E3'
	oct4	'C4'	'G4'	'E4'
	oct5	'C5'	'E5'	'G5'
	oct6	'E6'	'G6'	'C6'
	oct7	'E7'	'G7'	'C7'
C,D#,G	oct1	'G1'	'D#1'	'C1'
	oct2	'G2'	'D#2'	'C2'
	oct3	'C3'	'G3'	'D#3'
	oct4	'C4'	'D#4'	'G4'
	oct5	'C5'	'G5'	'D#5'
	oct6	'G6'	'D#6'	'C6'
	oct7	'D#7'	'G7'	'C7'

4 note-chords identification

Analyzed Chords		Identified Chords			
C,E,G,A#	oct1	'A#1'	'G1'	'E1'	'C1'
	oct2	'A#2'	'G2'	'E2'	'C2'
	oct3	'C3'	'A#3'	'G3'	'E3'
	oct4	'A#4'	'C4'	'G4'	'E4'
	oct5	'C5'	'E5'	'G5'	'A#5'
	oct6	'E6'	'A#6'	'C6'	'G6'
	oct7	'E7'	'A#7'	'C7'	**'E7'**

C,D#,F#,A	oct1	'A1'	'F#1'	'D#1'	'C1'
	oct2	'A2'	'D#2'	'F#2'	'C2'
	oct3	'C3'	'D#3'	'F#3'	'A3'
	oct4	'A4'	'C4'	'F#4'	'D#4'
	oct5	'C5'	'A5'	'D#5'	'F#5'
	oct6	'F#6'	'D#6'	'C6'	'A6'
	oct7	'D#7'	'F#7'	'C7'	'A7'
C,D#,G,A#	oct1	'G1'	'A#1'	'D#1'	'C1'
	oct2	'A#2'	'G2'	'D#2'	'C2'
	oct3	'C3'	'D#3'	'A#3'	'G3'
	oct4	'C4'	'A#4'	'G4'	'D#4'
	oct5	'C5'	'D#5'	'G5'	'A#5'
	oct6	'D#6'	'G6'	'C6'	'A#6'
	oct7	'D#7'	'G7'	'A#7'	'C7'
C,E,G,B	oct1	'B1'	'G1'	'E1'	'C1'
	oct2	'G2'	'E2'	'C2'	**'B1'**
	oct3	'C3'	'G3'	'E3'	'B3'
	oct4	'C4'	'G4'	'B4'	'E4'
	oct5	'C5'	'E5'	'G5'	'B5'
	oct6	'E6'	'B6'	'G6'	'C6'
	oct7	'E7'	'G7'	'C7'	**'E7'**

Octaves 1 and 2 are more prone to errors because the postdetection validation process is different for those octaves than for octaves 3 to 7. The latest validates the notes verifying that the fundamental of the detected note exists in the analyzed signal . If that fundamental does not exist, the algorithm search for a new detected note.

The fundamental of notes belonging to octaves 1 and 2 have very low level and they almost do not exist, so this validation strategy can not be applied because it would yield false negatives.

The validation technique that is performed in octaves 1 and 2 is based on selecting the candidate of higher frequency. Can be seen that identified notes of octaves 1 and 2 are always in decreasing order. There is no additional verification of the existence of the validated note in the original signal and errors can occur.

Both validation strategies are based on the fact that when the basic algorithm fail, it always detects a note of lower octave than the real note. This behavior is due to the selection of the number of partials and its weighting, made by the model when the patters are generated.

The only three errors that can be seen in the tables are:

-B1 instead of B2. If the detection does not include B2 between the candidates, the validation process can not choose it. Can be seen that the fail occurs in the last detected note so some degree of spectral residual can be the cause.

-E7 detected twice instead of G7 or B7. In these cases, have been noted that the level of the failed notes, in the original analyzed signal, is too low , so the metric calculation “scores” higher even a little residual of a note previously substracted..

5. CONCLUSIONS

A polyphonic identification method for pianos has been presented. Detection of four-notes chords has been carried out. The errors (bold text in table)occur only in octaves 2 and 7, and are due to well known drawbacks.

The correct detection of the notes alone indicates that the errors in chord detection are due to the existence of other notes which spectral components scores, erroneously, for other patterns. This has been corrected with octave predetection and postdetection validation criteria.

6. REFERENCES

[1] Brown JC, “Musical frequency tracking using the methods of conventional and narrowed autocorrelation”. J Acoust Soc Am 1991; 89(5): 2346-2354.

[2] Klapuri A, “Pitch estimation using multiple independent time-frequency windows”. Proc.1999 IEEE Workshop on App.of Signal Processing to Audio and Acoustics. New York. Oct.1999. 115-118.

[3] Monti G, Sandler M. “Monophonic transcription with autocorrelation” Proceedings of the COST G-6 Conference on digital audio effects (DAFX-00), Verona, Italy, 2000:257-260

[4] Klapuri A, Virtanen T, Holm JM. “Robust multi pitch estimation for the analysis and manipulation of polyphonic musical signals”. Proceedings of the COST G-6 Conference on digital audio effects (DAFX-00), Verona, Italy, 2000:141-146

[5] Verfaille V, Duhamel P, Charbit M. “LIFT:Likehood frequency time analysis for partial tracking and automatic transcription of music”. Proceedings of the COST G-6 Conference on digital audio effects (DAFX-01), Limerik, Ireland, 2001

[6] Rossi L and Girolami G, “Identification of polyphonic piano signals”. Acústica-acta acustica 1997; 83;10777-1084.

[7] Kinsler LE, Frey AR, Coppens AB, Sanders JV. *Fundamentals of acoustics.*John Wiley & sons 1972. New York

[8] Fletcher,H. “Normal vibration frequencies of a stiff piano string”. J.Acoust. Soc. Am. 1964; 36:203-209.

[9] Conklin HA. “Design and tone in the mechanoacoustic piano. Part III. Piano strings and scale design”. J Acoust Soc Am 1996; 100(3): 1286-1298.

[10] Conklin HA. “Generation of partials due to nonlinear mixing in a stringed instrument”. J Acoust Soc Am 1999; 105(1): 536-545.

[11] Ortiz-Berenguer L, Casajús-Quirós F.J. “Identification of piano notes using wavelet transform”. Proc. 8th Intern. congress on sound and vibration, Hong-Kong, China, 2001.

[12] Rossing T.D. *The science of sound.* 2ªed. Addison Wesley. Massachusetts. 1990