Available online at www.sciencedirect.com
SCIENCE DIRECT®
Applied Acoustics 65 (2004) 1261–1287
applied acoustics
www.elsevier.com/locate/apacoust

Transcription of piano recordings

I. Barbancho *, A.M. Barbancho, A. Jurado, L.J. Tardón

Departamento de Ingeniería de Comunicaciones, E.T.S. Ingeniería de Telecomunicación, Universidad de Málaga, Campus Universitario de Teatinos s/n, 29071 Málaga, Spain

Received 22 January 2003; received in revised form 14 November 2003; accepted 4 May 2004
Available online 2 September 2004

Abstract

A system for automatic identification of polyphonic piano recordings based on a heuristic approach is presented. The input to the transcription system is any piano piece sampled at 44.1 kHz. The system does not need any kind of training and there are no restrictions on the number of notes played simultaneously or on the notes' frequency range. The transcription system initially divides the piano piece into attack slots. This temporal segmentation is based on onset detection and it is made by means of a sliding window procedure. Afterwards, a frequency analysis is performed on each attack slot. Finally, musical parameters such as beat, time signature, key signature and tonality are determined. It is shown that the proposed system achieves very good performance using simple signal processing techniques.

Keywords: Music transcription; Polyphonic piano recording

1. Introduction

The aim of a transcription system is to convert an acoustic music signal into its symbolic representation. This transformation involves several tasks designed to extract from the music signal all the necessary elements to write down the score: pitches

* Corresponding author. Tel.: +34 95 213 2587; fax: +34 95 213 2027.
E-mail address: ibp@ic.uma.es (I. Barbancho).

0003-682X/$ - see front matter
doi:10.1016/j.apacoust.2004.05.007

and durations of the notes, starting times, time signature and key signature. These tasks become particularly complex when polyphonic music is considered. The word 'polyphonic' is here intended to mean the opposite of monophonic, while in music theory it usually means the opposite of homophonic. Since 1975, when Moorer presented the first transcription system [1,2], several transcription systems have been proposed, however they still suffer substantial limitations. Historical reviews of the evolution of the transcription systems can be found in [3–6].

In this paper, we focus on a transcription system for piano music [7–11]. The transcription system proposed is based on a heuristic approach, which has been found to achieve very good performance using simple signal processing techniques. Some particular features of our system are:

- Sampling rate: 44.1 kHz. This is the standard CD sampling rate; therefore, the user can analyze, directly, any CD piano recording.
- The system does not need any kind of training provided that the quality of the recording and the piano tuning comply with some minimum requirements.
- There are no restrictions on the number of notes played simultaneously or on the notes' frequency range.
- The time resolution is 56 ms. This is a trade-off between frequency resolution and time resolution to characterize the rhythmic structure and to quantify the timings of each note.

A general overview of the proposed system is presented in Fig. 1. This paper is organized as follows. Section 2 describes the onset detection process employed to perform the temporal segmentation of the signal. Section 3 is devoted to the frequency analysis, i.e. the estimation of the pitch of the notes played in each temporal slot of the piano piece. Section 4 deals with the calculation of musical parameters such as time signature, key signature, tonality, etc. In Section 5, the results obtained are shown and, finally, some conclusions, limits of the system and future improvements are presented in Section 6.

2. Onset detection and temporal segmentation

In order to analyze any musical piece, a time-frequency analysis is needed. Short-time Fourier transforms (STFT), constant-Q and wavelet transforms are some of the most commonly employed solutions in several transcription systems [12–14]. Our system divides the piano piece into temporal slots and, afterwards, a frequency analysis of each slot is done. This temporal segmentation is based on the detection of onsets.

Each temporal slot starts when a new onset is detected, i.e., when a new note is played. These temporal slots are referred to as attack slots. This type of temporal segmentation is very well suited for the frequency analysis because new frequencies appear in the music signal when a new note is played.

Several onset detection methods have been proposed in the literature, which use either time domain or frequency domain signal processing algorithms or a combina-

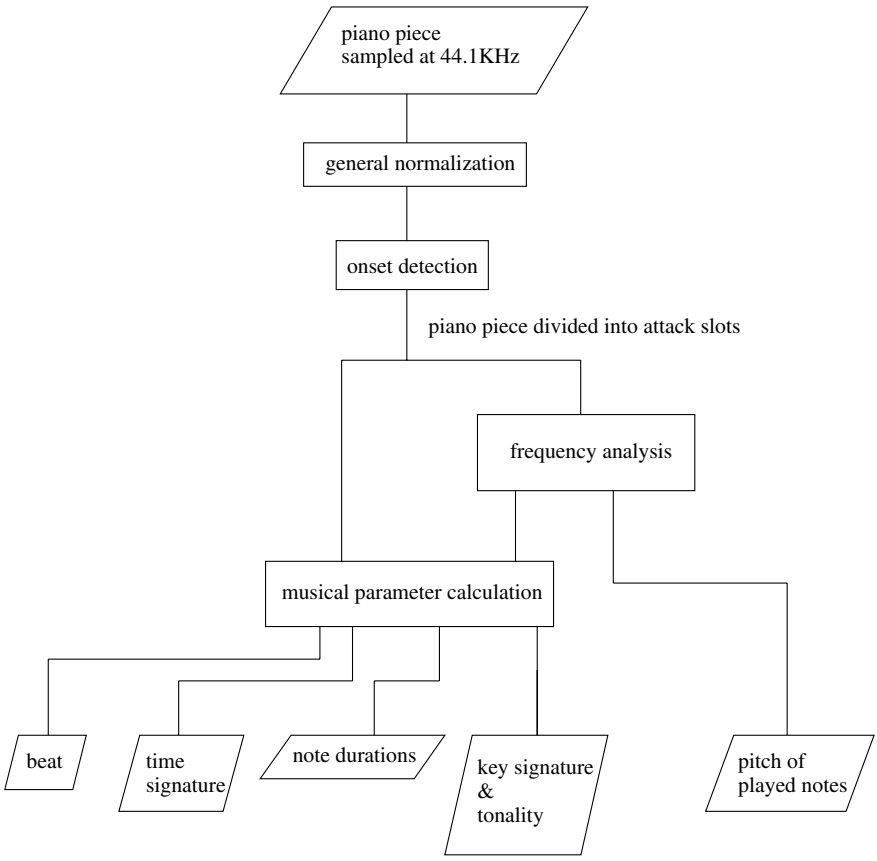

Fig. 1. Overview of the proposed transcription system.

tion of both [15]. In our case, the onset detection method is implemented in the time domain. Consider the attack-decay-sustain-release (ADSR) envelope for the piano: its attack is fast, i.e., the time in which the sound builds up to its maximum intensity is quite short, and then the sound amplitude shows an exponential decrease. Therefore, when a note is pressed, there is a sudden, large increase in the energy of the signal. The implemented method detects local maxima in the amplitude of the signal due to this energy increase, using a sliding windows procedure.

The whole segmentation process consists of three major steps: inner normalization, onset detection and attack slots delimitation. Before these stages, the entire piano piece is normalized to its maximum sample.

2.1. Inner normalization

After the first global normalization, a second normalization process is carried out on the piano composition. This second normalization is referred to as inner normalization and it is aimed at detecting onsets in passages played with low intensity.

All samples with an amplitude larger than 0.5 are located. If two successive samples of this type (i.e. of amplitude greater than 0.5) are more than 4.5 s apart, then we say that the intervening passage has been played with low intensity. In this case, a new local normalization is made with respect to the local maximum of the passage. This process is repeated along the whole piano composition; afterwards, the sliding window procedure for the onset detection is started.

This inner normalization is used only for onset detection. For all the subsequent processes, the original normalized signal is used.

2.2. Onset detection: sliding windows procedure

A sliding window procedure is employed to detect any increases in energy that exceed a certain threshold. This threshold has been selected to characterize the appearance of an onset.

Rectangular windows that contain 3000 samples (68 ms) of the signal to analyze are employed. The piano signal is windowed without overlapping if no attacks are detected. However, if an attack is found in window i, the location of the maximum amplitude sample in that window defines the instant of the onset (although this is not entirely accurate) and the next window (window $i + 1$) is specified to start 600 samples (13.6 ms) before that sample.

A detailed flowchart of this procedure is shown in Fig. 2. The energy E_i of the samples in each window i is calculated as follows:

$$E_i = \sum_{j=x_i}^{x_i+\text{window\_length}-1} (y(j))^2, \tag{1}$$

where x_i is the initial sample of window i, window_length = 3000 and $y(j)$ represents the sample j of the piano piece. If $E_i < \varepsilon$ then it is decided that there is a rest and, therefore, the frequency analysis should not be performed on this fragment. The threshold ε has been set to 0.75.

A detailed analysis must be made if $E_i > \varepsilon$ and $E_i > E_{i-1}$, since it must be checked whether there has been an onset in window i or not. The tests and the thresholds used to find onsets must be selected from among the following cases:

1. Attack after a rest, i.e., $E_{i-1} < \varepsilon$ and $E_i > \varepsilon$. In this case, it is decided that an attack has occurred, regardless of the difference of energy between windows $i-1$ and i.
2. An attack has gone through the four stages of the ADSR model before a new attack takes place. In this case, no attack would have been detected in window $i-1$. Then, if $E_i > E_{i-1} + \mu_1$ ($\mu_1 = 6.05$), it is decided that a new attack has taken place in window i.
3. Attacks played very close to one another, i.e., a note is played before the previous one has reached the release state. In this case, an attack would have been detected in window $i-1$. Taking into account the ADSR model, the energy of window i will be bigger, but we must discover whether this increase of energy is due to either the previously played note or a new attack. If $E_i > \mu_2 E_{i-1}$ ($\mu_2 = 1.9$), then it is decided that there is a new attack in window i.

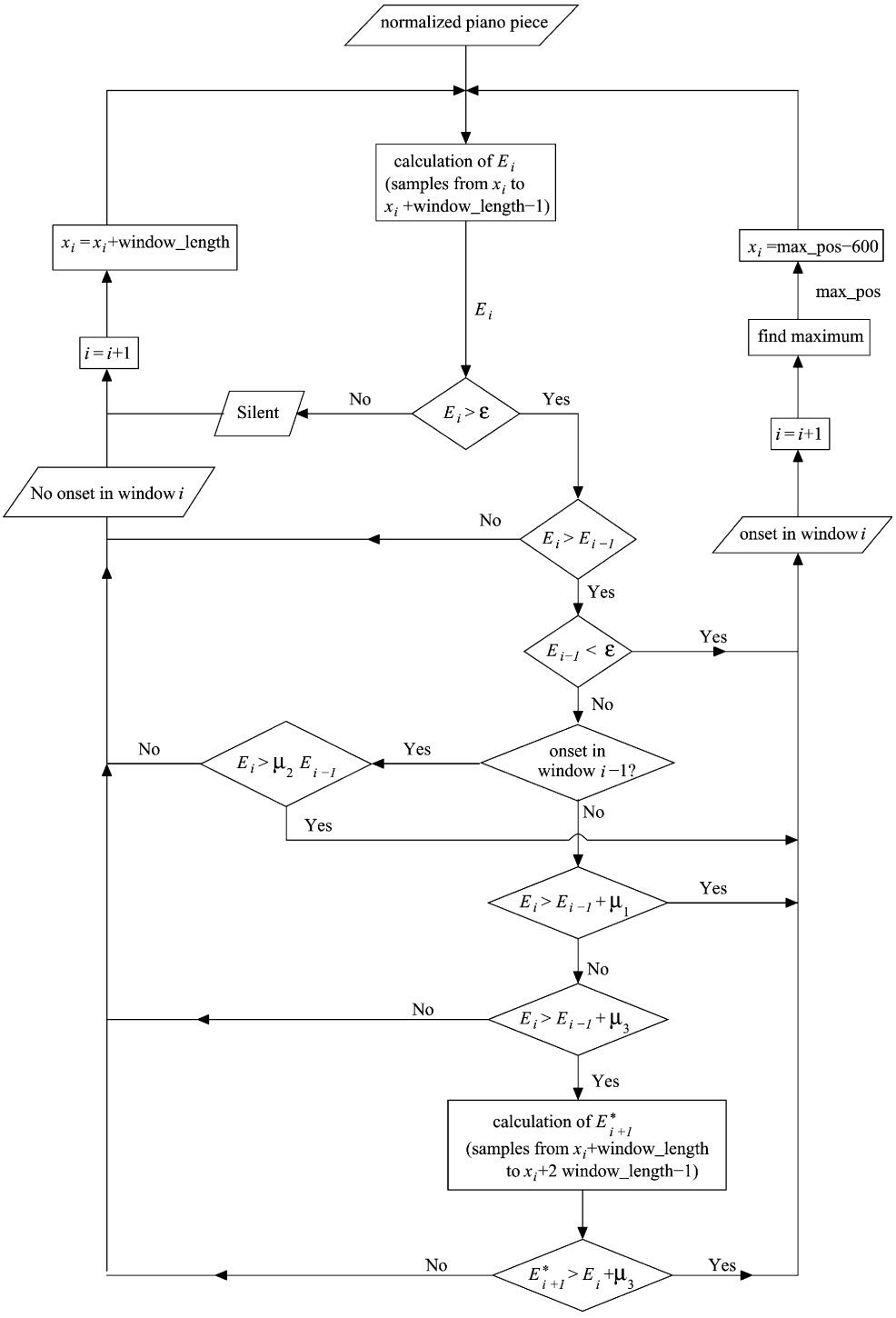

Fig. 2. Flowchart of the sliding window procedure.

4. Weak attacks and attacks whose samples with the maximum increase of energy lie between the end of the window under analysis and the beginning of the following window. In these cases, the energy of consecutive windows is decreasing but, suddenly, an increase of energy, which is not enough to reach the threshold μ_1, is detected. In order to decide whether there is a new attack or not, it is checked whether $E_i > E_{i-1} + \mu_3$ ($\mu_3 = 3$). If this condition is fulfilled, it is checked whether $E^*_{i+1} > E_i + \mu_3$, where E^*_{i+1} is the energy of the 3000 samples placed immediately after window i. If this second condition is also fulfilled, then there is a new attack in window i.

To locate any onsets masked by other onsets, the sliding window procedure is repeated over each pair of consecutive onsets previously detected. In this case, windows containing 2000 samples (45.35 ms) are used and the threshold μ_1 is set to 5.5.

Fig. 3 shows an example of the attack detection procedure. Window 1 starts at $x_1 = 400$, 600 samples before the attack detected in sample 1000 ($t = 22.67$ ms). In window 2, no attack is found since $E_2 < E_1$. In window 3 a new attack is detected at position 7390 ($t = 167.57$ ms). Therefore, window 4 starts at $x_4 = 6790$, 600 samples before the sample of maximum amplitude in window 3.

2.3. Attack slots delimitation

Until this point, we have found when onsets take place. Now, it is necessary to determine the samples that belong to each attack slot. According to the ADSR model, each attack starts before the sample with maximum amplitude. Therefore an attack slot n will last from the instant when the onset n took place to 1000 samples (22.67 ms) before onset $n + 1$ takes place. With this definition, we do not analyze the samples in which the end of an attack and the beginning of the next attack are mixed. This is highly convenient, because when this occurs there is large harmonic distortion that can cause the detection of false notes.

Finally, short attack slots defined by onsets separated by less than 2500 samples (56.69 ms) are not considered, because the system has insufficient frequency resolution. With this criterion, we avoid the detection of false notes, although some notes may happen to be missed in some arpeggios and trills.

3. Frequency analysis

The flowchart of the frequency analysis procedure is shown in Fig. 4. The aim of this analysis is to estimate the notes present in each attack slot and to determine which of those notes were played in the current attack and which are due to previous attacks.

The first stage of the frequency analysis is the selection of the frequency band that comprises 90% of the energy of the attack. To this end, a full binary tree filter bank with five stages is employed [16]. The subband selected for the frequency analysis is defined as the smallest frequency band that contains at least 90% of the total energy

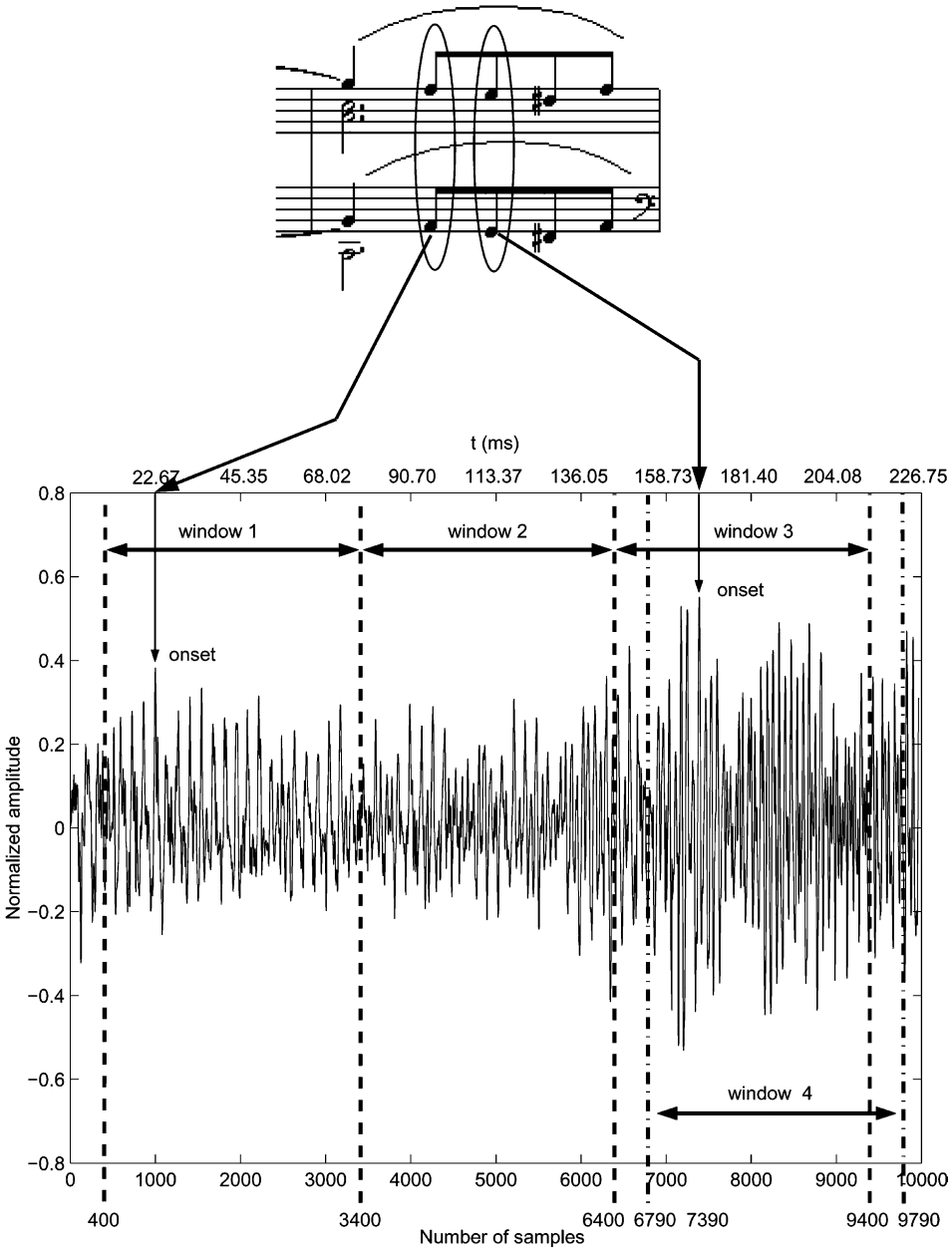

Fig. 3. Example of the sliding windows procedure. Window 1 starts at $x_1 = 400$, 600 samples before the attack detected in sample 1000. In window 2, no attack is found ($E_2 < E_1$). In window 3 a new attack is detected located at position 7390. Window 4 starts at $x_4 = 6790$, 600 samples before the sample of maximum amplitude in window 3.

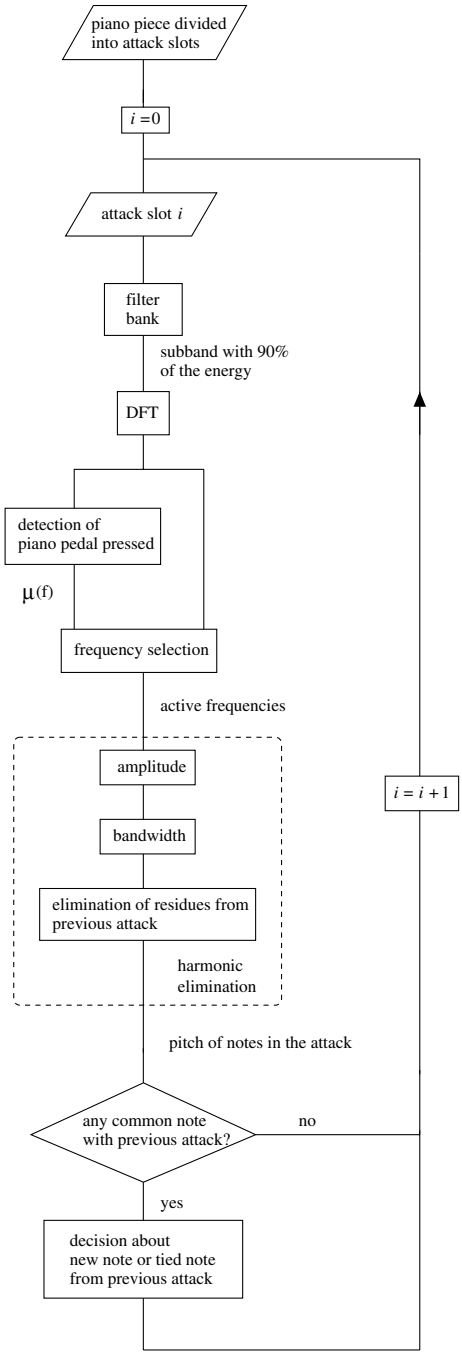

Fig. 4. Flowchart of the frequency analysis procedure.

of the attack. The objective is to reduce the number of samples used in the frequency analysis to make it more efficient.

The output signal of the selected subband is transformed into the frequency domain using a DFT and, afterwards, it is normalized so that its maximum amplitude in the frequency domain is 1. Then, all frequencies above a certain threshold are selected. We call these frequencies active frequencies. Afterwards, the system checks which ones, among these active frequencies, correspond to the pitch of the played notes. The threshold used to determine the active frequencies ($\mu(f)$) depends on two factors: the frequency range and the damper pedal of the piano:

$$\mu(f) = \begin{cases} 0.2 + \Delta & 0 < f < 225 \text{ Hz}, \\ 0.25 + \Delta & 225 \text{ Hz} < f < 800 \text{ Hz}, \\ 0.35 + \Delta & f > 800 \text{ Hz}, \end{cases} \tag{2}$$

where Δ depends on how long the damper pedal is pressed. If, in the attack slot under study, the pedal is not pressed, $\Delta = 0$, otherwise:

$$\Delta = \begin{cases} 0 & p = 0, \\ 0.75 & p = 1, \\ 1 & p = 2, \\ 1.25 & p > 3, \end{cases} \tag{3}$$

where p represents the number of consecutive attack slots before the present attack slot in which the pedal has been pressed.

The threshold in (2) can be used for any piano composition, since it has been previously normalized.

Some comment must be made regarding how the threshold used to determine the active frequencies has been chosen. It has been experimentally confirmed that the threshold must increase as the frequency increases. Also, whether the rightmost pedal of the piano is pressed or not and how long it has been pressed must be taken into account. Pressing the damper pedal on the piano causes all the dampers to lift from the strings, thus allowing the strings to oscillate and produce sound. This causes the noise floor to increase; therefore, it is necessary to raise the threshold so as not to confuse spurious frequencies with played notes. The longer the damper pedal is pressed, the larger this effect will be.

To determine if the pedal is pressed, the noise floor is estimated as the mean value of the magnitude of the DFT transform of the attack slot under analysis. If the noise floor is larger than 0.066, then it is decided that the pedal is pressed.

Now, we must distinguish between the active frequencies that correspond to played notes and those that correspond to the harmonics of played notes. Several tests are applied to the active frequencies to find the pitches of the played notes. It is known that errors in the frequency analysis are mainly due to the higher harmonics of the played notes. Therefore, the tests are specially designed to deal with these frequencies.

Fig. 5(a) shows the notes played in a certain attack slot and Fig. 5(b) its spectrum. The frequencies corresponding to the pitches D3, D4, A4 and A5 all exceed the threshold and, therefore, they are active frequencies. However only the notes D3 and A4 have actually been played. In the following subsection, we will see how the frequencies of the harmonics of played notes are eliminated.

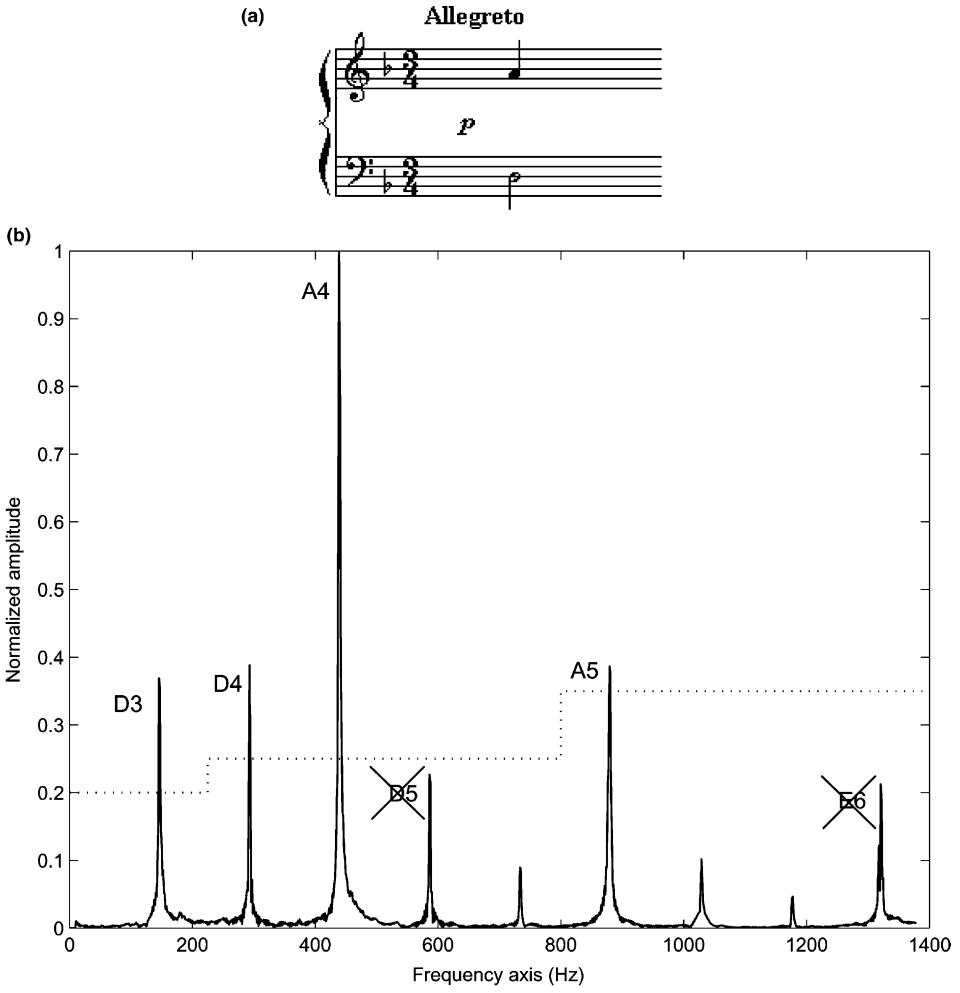

Fig. 5. (a) Score of the played notes in a certain attack slot. (b) Spectrum of the attack slot. The frequencies corresponding to the pitches D3, D4, A4 and A5 are all over the threshold and, therefore, they are active frequencies. However, only the notes D3 and A4 have actually been played.

3.1. Harmonic elimination

3.1.1. Amplitude

The amplitude of a harmonic frequency depends on the strength with which the note was played and on the number of the harmonic with respect to the played note. Therefore, when a frequency is marked active, new thresholds for its harmonic frequencies are set. These thresholds are chosen as follows: if the normalized amplitude of the fundamental frequency is A, then the threshold at the position of its second harmonic will be $0.67A$ and at the position of the third harmonic it will be $0.26A$.

In Fig. 6, the same attacks as in Fig. 5 are shown. Using the thresholds chosen for the harmonic frequency set, $A5$ is eliminated.

This test does not eliminate all the potential harmonics. There may be harmonics with larger amplitudes for several reasons, e.g., the overlap of harmonics of several notes, or one of the notes in the attack may have a fundamental which is weaker than its harmonics.

3.1.2. Bandwidth

The inharmonicity of a string is the amount by which the actual mode frequencies differ from a harmonic series [17]. Due to the inharmonicity of the string vibration, the second harmonics of one octave are slightly sharper than the fundamental of the next octave up and so on for the higher harmonics. So, when an octave interval is played the lower string's second harmonic is out of tune with the higher string's first harmonic. As a result, when a note and the fundamental of its next octave up are played simultaneously, intermodulation components appear in the spectrum. These intermodulation components may appear like a single wide spectral component if the frequency resolution is low.

In Fig. 7, the spectra of two attacks of different lengths (6 and 1 s) in which only C3 has been played are compared with the spectra of two attacks (again 6 and 1 s), in which the chord C3C4 has been played. For each attack slot length, when only C3 is played, the bandwidth around C4 (Figs. 7(c) and (d), dash-dot lines) is similar to the bandwidth around C3 (Fig. 7(a) and (b), dash-dot lines). If chord C3C4 is played, when the frequency resolution is high, some components due to the intermodulation are observed (Fig. 7(c), solid line). However, when the resolution is lower, only a widening of the spectral component around the fundamental frequency of C4 is observed (Fig. 7(d), solid line).

The magnitude of this effect depends on several factors: how the piano is tuned (stretch tuned or tuned to equal temperament), the size of the piano (spinet piano, grand piano...), how the chord is played (legato, staccato) [18], which notes have been played (the inharmonicity is especially noticeable in the case of the large bass strings), etc.

The important point is that this effect allows us to define another test for potential harmonics. Taking into account the resolution of our system, if the 3 dB bandwidth of a harmonic frequency is equal to the bandwidth of the corresponding fundamental frequency $\pm 10\%$, then the harmonic frequency is removed. Otherwise it is considered a played note.

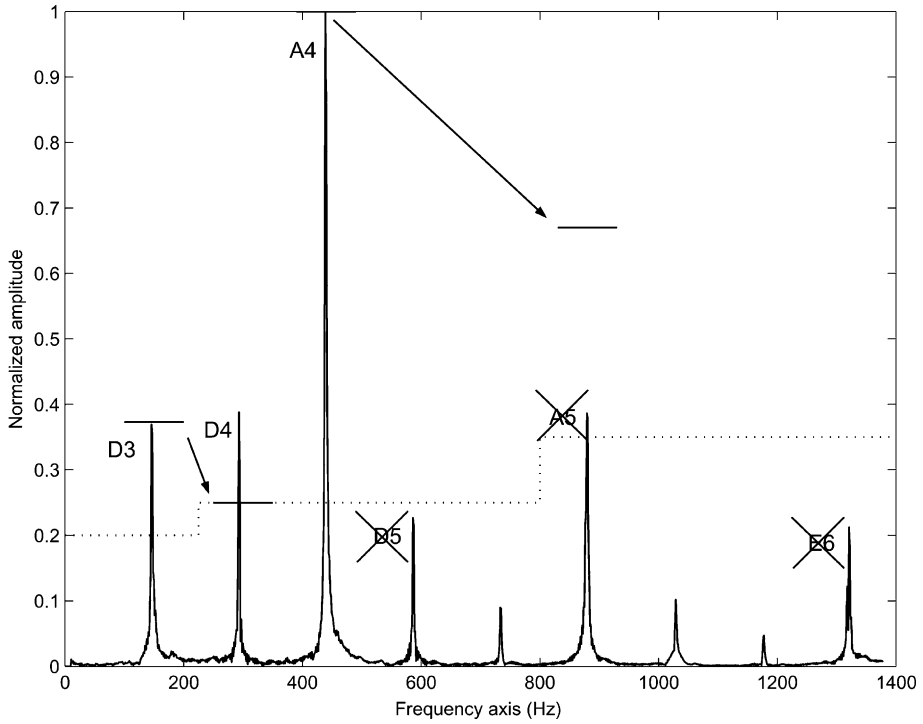

Fig. 6. Amplitude harmonic elimination. The frequency corresponding to pitch A5 is eliminated, but the frequency corresponding to D4 remains under consideration.

In Fig. 8, it can be observed that the bandwidth of the frequency component at pitch D3 is 3.5 Hz and the bandwidth of the frequency component at pitch D4 is 3.2 Hz, so the note D4 is finally eliminated using this test.

3.1.3. Harmonic from previous attack

Occasionally, a residual harmonic frequency of a note played in the previous attack, appears to be an active frequency. To avoid this problem, it should be checked whether any of the active frequencies, which passed the two previous tests, is due to such phenomenon.

In Fig. 9, two consecutive attacks are shown. For convenience, these attack slots have not been normalized in the frequency domain, so the thresholds have been scaled. In both cases, the active frequencies that did not pass the two previously defined tests have been crossed out. In the second attack, the active frequency corresponding to the pitch D4 has not been crossed out since it passed the tests. However, a note containing a D4 frequency component has not been played in the present attack.

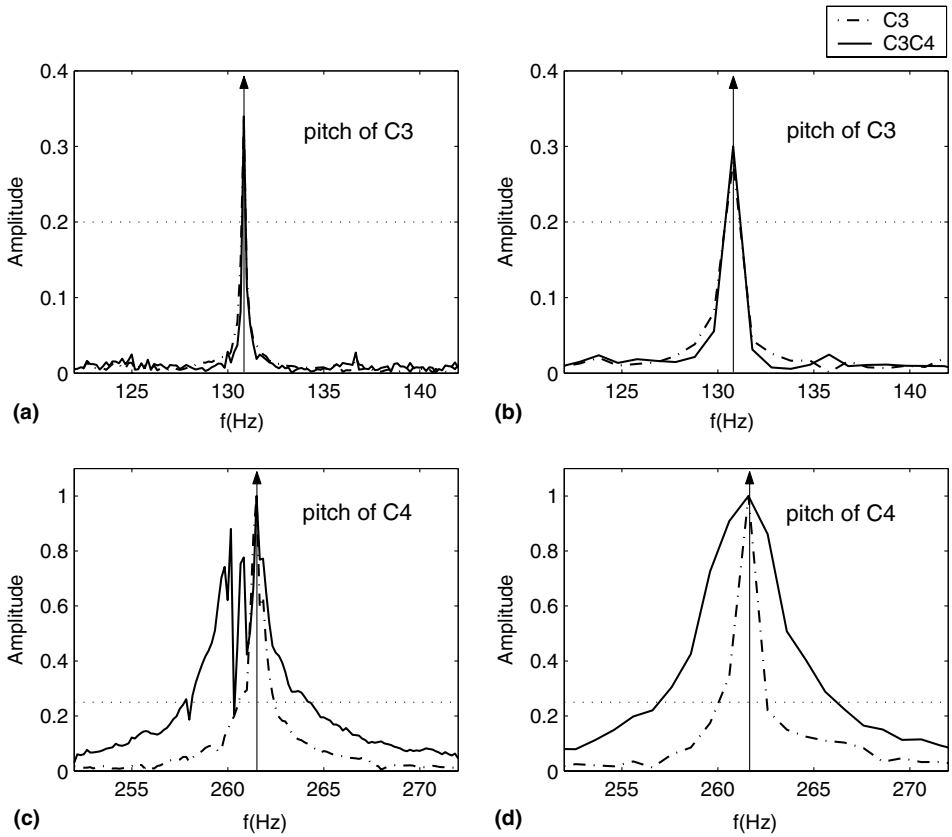

Fig. 7. The spectra of two attack slots of different length in which only C3 has been played (dash-dot lines) are compared with the spectra of two other similar attacks in which the chord C3C4 has been played (solid lines). (a) Spectrum around the fundamental frequency of C3 (6 s attack slot). (b) Spectrum around the fundamental frequency of C3 (1 s attack slot). (c) Spectrum around the fundamental frequency of C4 (6 s attack slot). (d) Spectrum around the fundamental frequency of C4 (1 s attack slot). For each attack slot length, when only C3 is played, the bandwidth around the frequency component at pitch C4 is similar to the bandwidth around the frequency component at pitch C3. If chord C3C4 is played, when the frequency resolution is high, some components due to the intermodulation are observed. However, when the resolution is lower, we can only appreciate a widening of the spectral component around the fundamental frequency of C4.

The solution adopted to identify such situations is to check if any active frequency coincides with any harmonic eliminated in the previous attack. In that case, the magnitude of that harmonic in the previous attack is compared with its magnitude in the present attack. If its magnitude in the present attack is bigger than in the previous one, then the system decides that it is the pitch of a played note. In Fig. 9, the frequency corresponding to D4 has a lower amplitude than in the previous attack, so D4 is not considered to be a played note.

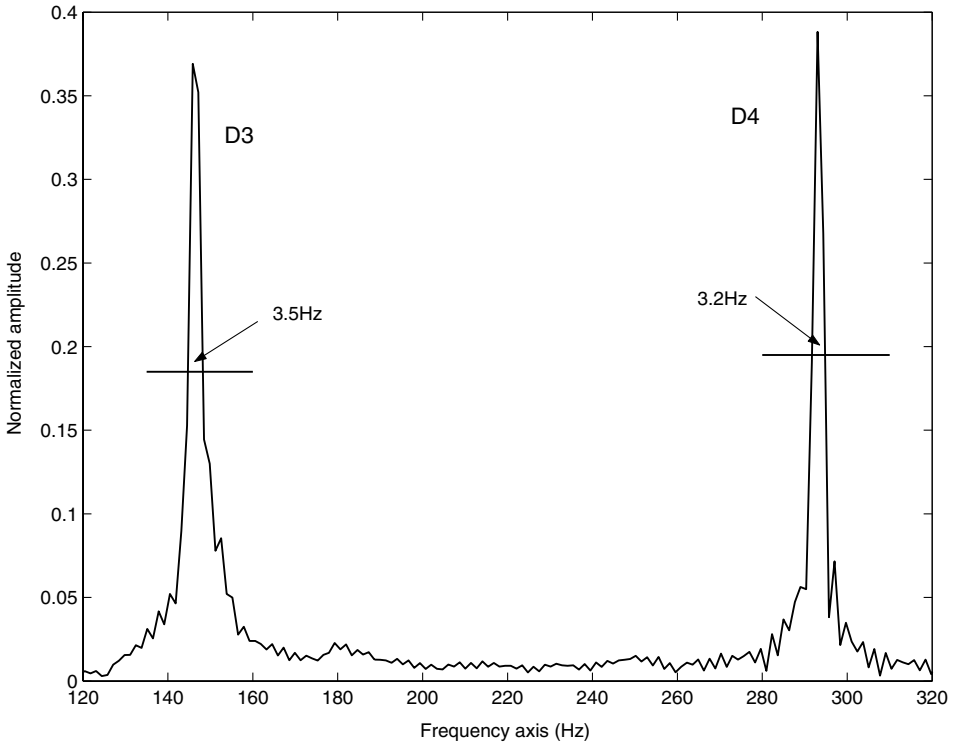

Fig. 8. Bandwidth harmonic elimination. The 3 dB bandwidth of the frequency component at pitch D3 is 3.5 Hz and the 3 dB bandwidth of the frequency component at D4 is 3.2 Hz, so the note D4 is removed.

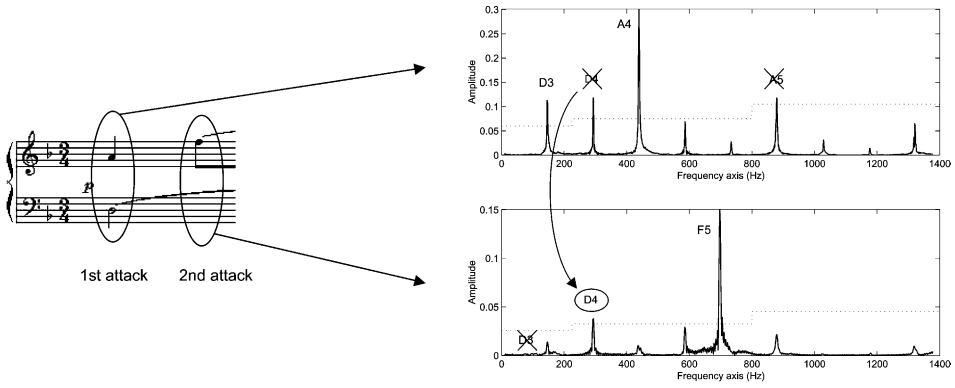

Fig. 9. Elimination of residual harmonics from previous attack. In the figure, two consecutive attacks are shown. The spectra of these attack slots have not been normalized (the thresholds have been scaled). In the spectrum of the second attack, D4 is marked as an active frequency. However, the frequency corresponding to pitch D4 coincides with a harmonic eliminated in the previous attack, whose amplitude is bigger than in the present attack, so the note D4 is eliminated.

3.2. Decision about new played note or tied note from previous attack

To complete the frequency analysis, it must be checked whether any of the remaining active frequencies that coincide with active frequencies from the previous attack, are due to a new played note or to a tied note.

This information is extracted from the correlation between the piece of signal that comprises the present and the previous attacks and a sinusoidal reference signal of frequency f_{note}, where f_{note} is the fundamental frequency of the note under study. In the frequency domain, this can be described as a filter matched to the target harmonic. The output of this matched filter provides the necessary information.

Fig. 10 shows an example of this test. In Fig. 10(a), G5 has been played twice, and in Fig. 10(b), F3 is tied. Subplots (c) and (d) represent the time-domain signals that comprise the two attacks under study. Subplots (e) and (f) show the outputs of the filters matched to G5 and F3 respectively. It can be noted that in (e) the first derivative of the envelope of the correlation modulus has a discontinuity at the point where G5 is played again, whereas in (f) the envelope does not show such a discontinuity.

4. Calculation of musical parameter

By this stage, the system has determined the pitches of the notes, the instant when they were played and their duration, but some musical parameters are still needed to write down the score: tempo, time signature, note durations, tonality and key signature. In this section, how these parameters are estimated is described.

4.1. Estimation of beat and time signature

To estimate the time signature, first the beat is determined, then the denominator of the time signature is obtained and, finally, the numerator is found.

The time signature denominator indicates what kind of note defines the beat (half note, quarter note...). The duration of the beat, τ, is estimated on the basis of the average duration of the attack slots. First, the average value is calculated using all the attack slots, then, to obtain more accurate results, a new average value is calculated using only the attack slots whose duration is ±25% apart from the duration previously calculated. Once τ has been determined, the kind of note that will represent the beat is decided. If $\tau > 0.75$ s, the beat is defined as being a half note, if $0.75\ \text{s} > \tau > 0.4$ s the beat is a quarter note and if $\tau < 0.4$ s it is an eighth note. The relation between τ and the kind of note that represents the beat has been set empirically after studying several musical pieces. The beat is presented in the usual way in the scores, in which the number of beats per minute (bpm) is also shown. All these relations are summarized in Table 1.

Next, the numerator of the time signature is obtained. To this end, the first beat of the measure, i.e. the beat immediately after the bar, is assumed to have more energy than the rest of beats in the measure. Taking into account that each measure lasts

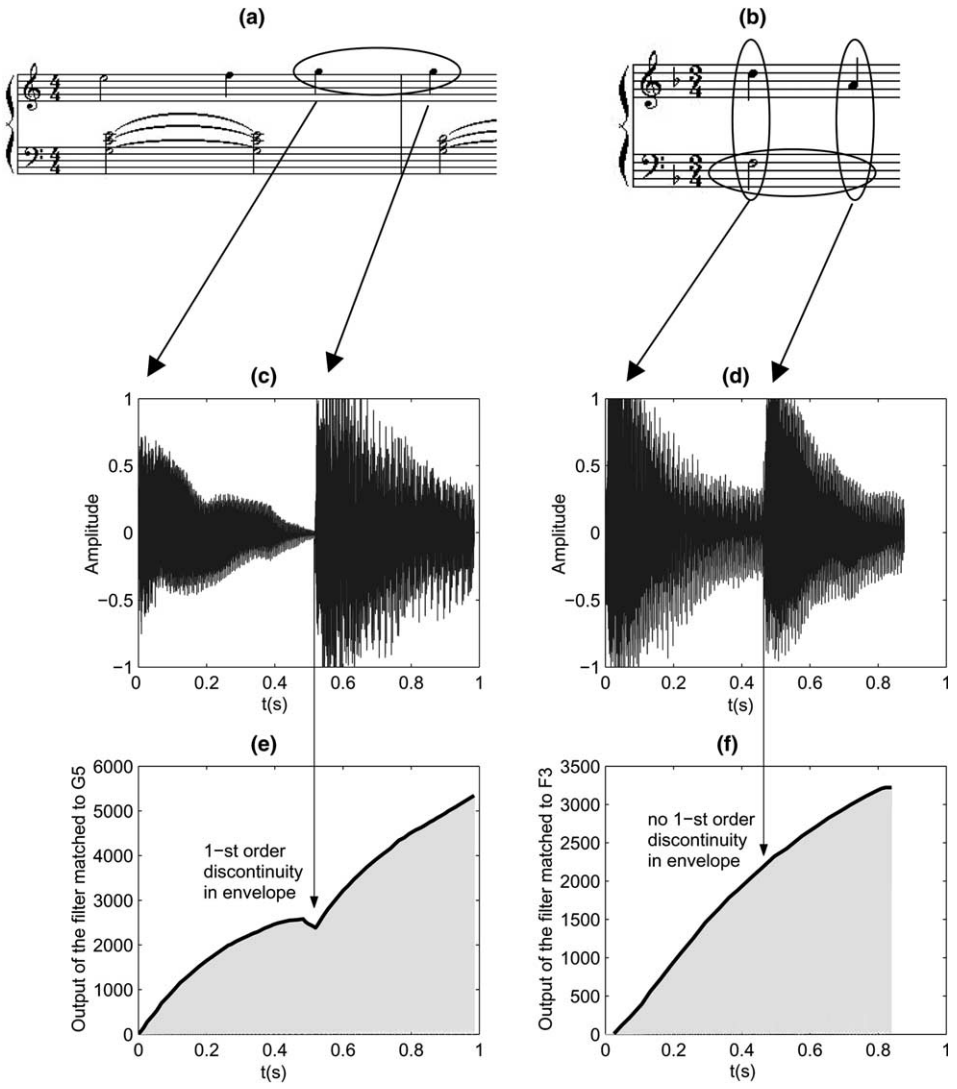

Fig. 10. Procedure to decide new/tied note. (a) Score in which G5 has been played twice. (b) Score in which F3 is tied. (c) Waveform of the attacks in subplot (a). (d) Waveform of the attacks in subplot (b). (d) Output of the filter matched to G5. (f) Output of the filter matched to F3. In (e), the envelope of the matched filter output has a first-order discontinuity at the point where G5 is played again, whereas in (f) the matched filter output does not shown any first order discontinuity. So, we conclude that in (a) G5 was played twice and in (b) F3 is tied.

(numerator$\times\tau$) s, a bar division of the piano piece is carried out for each possible value of the numerator. Then, the average intensity of the beginning of each measure is calculated. The bar division procedure also takes into account that the piano piece

Table 1
Relation between the estimated duration of beat (τ), the number of beats per minute (as it should appear in the score) and the time signature denominator

Estimated duration of the beat	$\tau > 0.75$ s	0.75 s > τ > 0.4 s	$\tau < 0.4$ s
# Beats per minute (bpm)	$\mathbf{\mathit{o}} = 60/\tau$	$\mathbf{\mathit{♩}} = 60/\tau$	$\mathbf{\mathit{♪}} = 60/\tau$
Time signature denominator	2	4	8

can start with an upbeat. The program chooses the numerator with which the maximum average intensity at the beginning of measures is achieved.

The calculation of the time signature does not have a unique solution. For example, consider a piano piece with signature 4/4 and ♩ = 120. If the most repeated note duration was the eighth-note, then $\tau \approx 0.25$ s and the time signature decided would be 8/8. However, the performance of the original and the transcribed scores would be the same.

In spite of this uncertainty, the system solves the ambiguity between 3/4 and 6/8 time signatures as follows: the system counts the number of measures with three attacks of similar duration (#3a) and the number of measures with two attacks of similar duration (#2a). If #3a > #2a, the system decides a 3/4 time signature, otherwise 6/8 is decided.

4.2. Estimation of note durations

The first step of this process assigns a normalized duration, among the set of normalized durations, to each note x. The normalized duration selected minimizes δ_x

$$\delta_x = \left| \left(\frac{\text{length of note } x(s)}{\tau(s)} \times \frac{4}{\text{time signature denominator}} \right) - \text{normalized duration of note } x \right|. \tag{4}$$

The normalized durations of each kind of note are summarized in Table 2. Note that 4/(time signature denominator) is the normalized duration of the note that represents the beat.

Finally, it should be checked if the normalized duration of each measure is (time signature numerator) × 4/(time signature denominator). If a measure is not properly filled, the relative error between the estimated duration of each note x in the measure under study and its normalized duration is defined as:

$$\Delta_x(\%) = \frac{\left(\frac{\text{length of note } x(s)}{\tau(s)} \times \frac{4}{\text{time signature denominator}} \right) - \text{normalized duration of note } x}{\text{normalized duration of note } x} \times 100. \tag{5}$$

Starting with the notes with largest value of $|\Delta_x|$, and checking the difference between the normalized duration of the measure and its actual duration, the normalized durations of the notes are changed until the measure is properly filled.

Table 2
Normalized duration of the notes and their symbolic representation

Normalized duration	Symbolic representation	Normalized duration	Symbolic representation
4		0.75	
3		0.66 (×3)	
2		0.5	
1.5		0.33 (×3)	
1		0.25	

4.3. Determination of tonality and key signature

The last elements needed to write down the score are the key signature and the tonality. First, the key signature is determined and then, the tonality of the piano piece is chosen between a major key and its relative minor key.

The determination of the key signature is based on the minimization of the number of accidental notes in the composition. Initially, the system proposes a key signature that uses sharps. Then, the system counts the number of times that the note that corresponds to the first possible sharp in the key signature appears in the piano piece. If that note is found to be sharp on over half the time that it appears in the piece, then it is decided to include that sharp in the key signature. The process is then repeated for the next possible sharp in the key signature. When one note does not fulfill this criterion, the procedure is finished leaving the key signature composed of the sharps thus far decided. If the first note does not fulfill this criterion, the same procedure is employed to test the key signatures that use flats. If, again, the first note does not fulfill the criterion, then no key signature is put in the score.

Now, the tonality is selected between the major and minor tonalities that share the same key signature. To do this, the note that corresponds to the seventh degree of the minor tonality is considered. The number of times that this note is used as the leading tone, i.e., halftone separated from the tonic of the minor tonality (#leading_tone) is counted and the number of times this note is used as the subtonic, i.e., separated a tone from the tonic of the minor tonality (#subtonic) is counted. If #leading_tone>#subtonic, then the system selects the minor key, otherwise, the system decides on the major key.

Finally, if the tonality of a piano piece is minor, it could be natural, harmonic or melodic minor [19]. The harmonic and melodic minor scales have the seventh note raised, unlike the natural minor scale. Therefore, the proposed method to decide between the major and minor scales works properly if either the major scale or the harmonic or melodic minor scales have been used. If the natural minor is employed, the proposed method cannot distinguish between the major key and the natural minor

Table 3
HM piano pieces analyzed, characteristics and results

HM piano pieces	Original time signature	Estimated time signature	Original key signature and tonality	Estimated key signature and tonality	Estimated bpm	Maximum # of simultaneous notes	Average # of simultaneous notes
1 Beethoven 2nd movement, Sonata no 9	3 4	3 4	1 sharp E minor	1 sharp E minor	♩ = 116	6	4.4
2 Chopin Nocturne Op. 55 No. 1	4 4	4 4	4 flats F minor	4 flats A b major	♩ = 90	7	3.1
3 Albéniz Spanish Suite Granada	3 8	3 4	1, 4 & 5 flats F major	1 flat F major	♪ = 170	8	4.2
4 Sergei Prokofiev Drehorgel	2 4	8 8	No C major	No C major	♪ = 211	3	2.3
5 Hal Leonard adaptation Himmlische, dein Heiligtum (bars 1–3)	4 4	4 4	No C major	No C major	♩ = 98	4	4
6 Ana Magdalena Bach Notebook 20 pieces Minuet 20	3 4	2 8	1 flat D minor	1 flat D minor	♪ = 231	2	2
7 Ana Magdalena Bach Notebook 20 pieces Minuet 2	3 4	3 4	1 sharp G major	1 sharp G major	♩ = 169	4	2.1
8 Schumman The Happy Farmer	4 4	4 8	1 flat F major	1 flat F major	♪ = 205	4	2.8

key. This final case could be addressed by analyzing the final chord of the piano piece.

5. Results

To test the system, some polyphonic piano pieces, played on different pianos, by different performers and in very different recording conditions, have been analyzed.

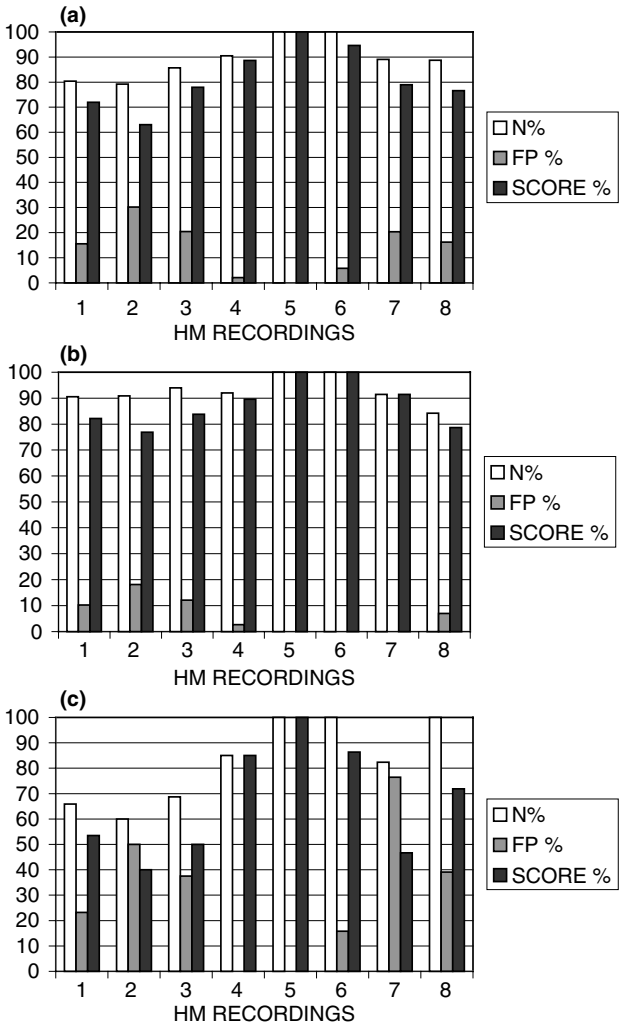

Fig. 11. Results for eight home made piano pieces: Percentage of detected notes (N%), percentage of false notes (FP%) and score ($\mathrm{SCORE}\% = (N)/(N + \mathrm{FN} + \mathrm{FP}) \times 100$). (a) Overall results. (b) Partial results for notes above C4. (c) Partial results for notes below C4.

Table 4
CD piano pieces analyzed, characteristics and results

	CD piano pieces	Original time signature		Estimated time signature		Original key signature and tonality	Estimated key signature and tonality	Estimated bpm	Maximum # of simultaneous notes	Average # of simultaneous notes
1	Chopin Nocturne Op. 9 No. 2	12 8		3 4		3 flats Eb major	3 flats Eb major	♩ = 126	6	3.1
2	Beethoven Minuet in G	3 4		8 8		1 sharp G major	2 sharps D major	♪ = 152	3	2.6
3	Chopin Polonaise Op. 40 No. 1	3 4		8 8		3 sharps A major	3 sharps A major	♪ = 226	9	5
4	Bartok Ballad	4 4		4 4		1 flat D minor	1 flat D minor	♩ = 100	5	3
5	Beethoven Für Elise	3 8		6 8		No A minor	No C major	♪ = 274	6	2
6	Beethoven Sonatine 5	4 4	6 8	8 8	6 8	1 sharp G major	1 sharp G major	♪ = 269|303	5	2
7	Bach March	4 4		3 8		2 sharps D major	2 sharps D major	♪ = 288	2	2
8	Bach Polonaise	3 4		4 8		2 flats G minor	2 flats Bb major	♪ = 289	4	2.6

(*continued on next page*)

Table 4 (*continued*)

CD piano pieces		Original time signature	Estimated time signature	Original key signature and tonality	Estimated key signature and tonality	Estimated bpm	Maximum # of simultaneous notes	Average # of simultaneous notes
9	Schumman Marcha militar	2 4	2 4	1 sharp G major	1 sharp G major	♩ = 128	4	3
10	Bartok Children at Play	2 4	4 8	No C major	No C major	♪ = 248	2	2
11	Bartok Quasi Adagio	2 4	2 4	No A minor	1 sharp G major	♩ = 132	3	3
12	Chopin Nocturne Op. 55 No. 1	4 4	4 4	4 flats F minor	4 flats A b major	♩ = 90	7	3.1

Table 3 and Fig. 11 summarize the results obtained for eight home made (HM) recorded piano pieces, played by three different players on different pianos (Offberg, Yamaha and Steinway). In these HM pieces, the signal level is low and there is a

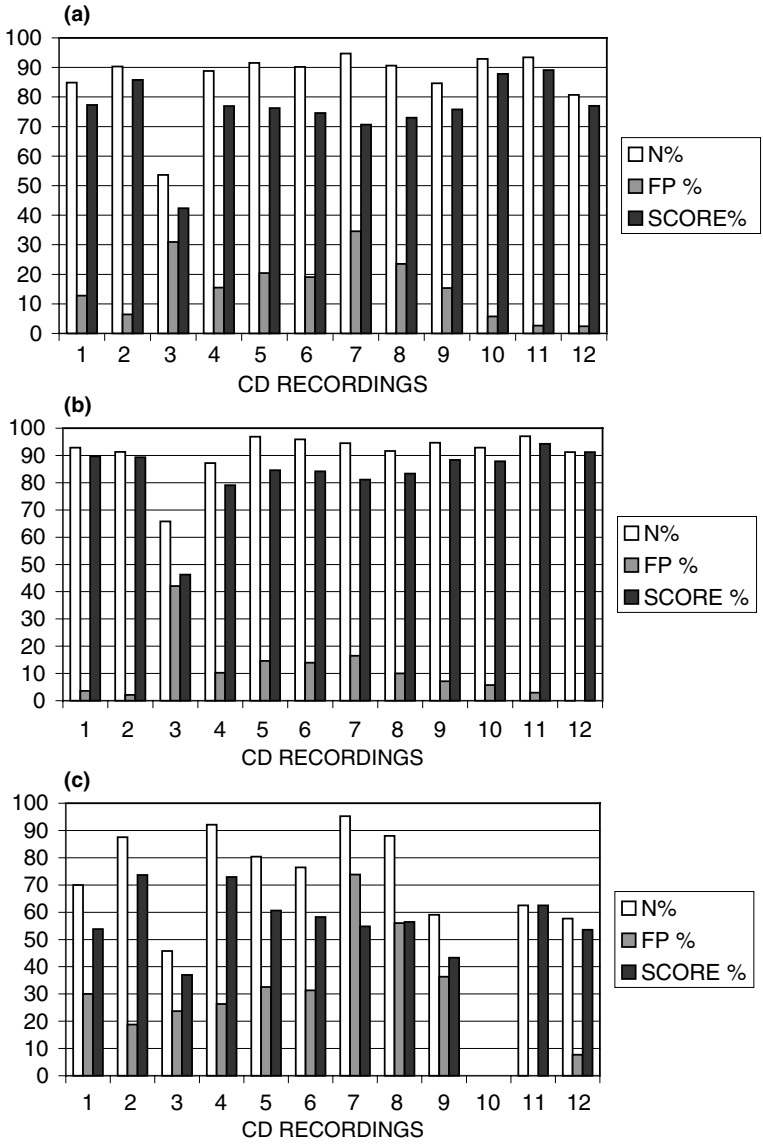

Fig. 12. Results for 12 high quality CD piano pieces: Percentage of detected notes ($N\%$), percentage of false notes (FP%) and score ($\mathrm{SCORE}\% = (N)/(N + \mathrm{FN} + \mathrm{FP}) \times 100$). (a) Overall results. (b) Partial results for notes above C4. (c) Partial results for notes below C4.

lot of noise. Table 4 and Fig. 12 summarize the results obtained for 12 piano pieces extracted from different high quality CD recordings.

In Tables 3 and 4, the title of the analyzed piece is shown, together with the average number of notes played simultaneously and the results for the time signature, the tonality analysis and the estimated tempo. In Figs. 11 and 12, the percentage of detected notes over the total played notes, the percentage of false notes over the total played notes and the score are shown. The score is defined as in [8]:

$$\text{Score} = \frac{N}{N + \text{FP} + \text{FN}} \times 100, \tag{6}$$

where N = Number of correctly identified notes, FN (false negative) = Number of notes played that were not reported by the system and FP (false positive) = Number of notes reported by the system that were not played. The total number of played notes is $N + \text{FN}$.

Figs. 11 and 12 show the overall results obtained (case (a)) and also the detailed partial results obtained analyzing only the notes above C4 (case (b)), and partial results analyzing only the notes below C4 (case (c)). In the partial analyses, the FP refers to the FP produced by the notes contained in the set of notes under study, i.e. if the played note is a C3 and there is a FP in C4, it will be included as a FP in the case of analysis below C4. The specific analyses for the notes above and below middle C have been made because for all piano sounds below middle C the fundamental is weaker than its harmonics [17]. Consequently, the amplitude test to find the harmonics will not work properly, though the bandwidth criterion will solve part of the problem. It can be observed that, although the results for all the played notes are quite good, better results are found if we consider only the notes above C4.

There are no large differences between the results obtained with the CD recordings and the HM recordings. The proposed system is not severely affected by the noise present in the HM recordings. With regard to specific pieces, the worst results were obtained for Chopin's Polonaise Op. 40, No.1 (the third column of Fig. 12(a)–(c)). In most of the attack slots of this piece, there are seven notes played simultaneously and, also, most of the time the damper pedal is pressed.

In general terms, the score and the false positive ratio are better in the system outlined in this paper than in the tool proposed in [8], while the percentage of detected notes is approximately the same. Also, if the present system is compared with the one using neural networks proposed in [7,20], the current system is simpler, it achieves approximately the same score with a lower percentage of false notes and only the percentage of detected notes over the total is slightly lower. If the present system is compared with the system proposed in [3], which is specifically designed for transcribing piano pieces of four-voice Bach chorales, it is clear that the current system is much more versatile, because there are no restrictions on the type of classical piano music that can be analyzed.

If the key signature and tonality results are examined, it can be seen that the system has identified the correct key signature except in two cases (pieces 2 and 11 of the CD recordings) and, in most cases, also the right tonality. In CD recordings 2 and 11, the error is only one sharp more in the estimated key signature than in the orig-

inal score. In fact, in piece 2, the real tonality is actually D major for the majority of the piece, although the key signature is only written with one sharp. The case of piano piece 11 is different since it is in doric mode on A [19]. The corresponding key signature has neither flats nor sharps, but the sixth note (F in this case) is raised by a semitone. This causes the system to decide upon a key signature with one sharp. Therefore, this fault is due to the particular mode of the piano piece (one which is not considered in the algorithm). Also, the case of HM piece three should receive a special mention because the key signature changes along the piano piece. This possibility is not considered by the system described in this paper, which decides the key signature that lasts longest in the piece.

Most of the time signatures estimated by the system can be considered correct, even in some pieces in which the estimated time signature and the original time signature are not the same (i.e. HM piece 8). This difference cannot be considered an error because of the inherent ambiguity of the time signature decision. In other pieces, the system estimates a time signature, which is a subdivision of the original time signature, like HM piece 6, or a multiple of the original time signature, like CD piece 5. The only important failure is found with the time signature of CD piece 7 in which the estimated time signature has no relation with the original one.

6. Conclusions and future improvements

In this paper, a system has been presented that determines all the necessary elements to write down the score of a polyphonic piano composition from a recording. The system identifies the notes, their duration and the time instants when the notes were played. Additionally, the system finds the key signature, the tonality, the beat and the time signature of the piano piece under analysis. The designed system has good performance. It does not need any kind of training and it is very versatile in the type of piano pieces that can be properly transcribed. Also, the system developed is not computationally cumbersome and, so, a real time implementation would be affordable.

Although the system performance is good when compared with other systems, a number of conclusions can be extracted that could help to improve the system. First of all, the time resolution selected (56 ms) can cause the system difficulties when it comes to analysing thrills or some arpeggios. Also, the score of the system depends on the number of notes per second, the number of simultaneous notes and other factors like how much and how long the damper pedal is pressed, the type of played chords, etc. Roughly speaking, if the average number of simultaneous notes is around 2, the score will be above 80% for up to 300 bpm. If the average number of simultaneous notes is around 3, the score will be around 80% for up to 150 bpm. When 3.5 notes are played simultaneously on average, the maximum speed which will enable a score of around 80% to be achieved is 130 bpm. Finally, when the mean of the number of simultaneous notes is between 4 and 5, the number of bpm can only be as high as 110 to allow the same score to be achieved.

After a detailed analysis of the system, it has been found that it could be improved in a number of ways. Specifically, the following topics could be addressed:

- The low performance for notes below C4. This problem can be recast in terms of false positives and false negatives. To avoid false positives, the amplitude criterion should be changed so that the thresholds for the harmonic frequency set depend on the fundamental frequency of the note under consideration. These thresholds would be designed to cope with harmonics whose amplitude is larger than the amplitude of the fundamental frequency. With respect to the false negatives, a possible solution would be to take into account that, for the lowest sounds, the harmonic frequencies are much clearer than the fundamental frequency. Then, the system could perform a backward search to discover a specific harmonic frequency configuration and find local peaks at the location of the fundamental frequency.
- When the number of simultaneous notes played is high, the performance of the system decreases. In this paper, the entire detection problem has been addressed in an iterative way, i.e. searching the notes one by one. Potentially, a method to search chords, instead of individual notes, would perform better in these situations.
- The algorithm to decide between major and minor key fails if the natural minor scale is employed. This algorithm could be improved by, for example, analyzing the final chords of the piano piece.

References

[1] Moorer JA. On the segmentation and analysis of continuous musical sound. PhD Dissertation, CCRMA report STAN-M-3, July 1975.
[2] Moorer JA. On the transcription of musical sound by digital computer. In: Second USA-JAPAN Computer Conference, August 1975. Reprinted in the Computer Music Journal, November 1977;1(4):32–38.
[3] Martin, KD. A blackboard system for automatic transcription of simple polyphonic music. MIT Media Laboratory Perceptual Computing Section Technical Report No. 399, 1996.
[4] Martin, KD. Automatic transcription of simple polyphonic music: robust front end processing. MIT Media Laboratory Perceptual Computing Section Technical Report No. 385, 1996.
[5] Sterian, A, Wakefield, GH. Music transcription systems: from sound to symbol. In: Proceedings of the AAAI-2000 Workshop on Artificial Intelligence and Music, Austin, Texas, July 2000.
[6] Klapuri, A. Automatic transcription of music. Master of Science Thesis, Department of Information Technology, Tampere University of Technology, 1997.
[7] Marolt, M. SONIC: transcription of polyphonic piano music with neural networks. In: Proceedings of Workshop on Current Research Directions in Computer Music, Barcelona, 15–17 November 2001.
[8] Dixon S. On the computer recognition of solo piano music. Mikropolyphonie 2000;6.
[9] Barbancho AM, Jurado A, Barbancho I. Identification of rhythm and sound in polyphonic piano recordings. Forum Acousticum, Sevilla, September 2002.
[10] Rossi L, Girolami G, Leca M. Identification of polyphonic piano signals. Acta/Acta Acustica 1998;83(6):1077–84.
[11] Rossi L, Girolami G. Instantaneous frequency and short term Fourier transform: application to piano sounds. J Acoust Soc Am 2001;110(5):2412–20.

[12] Pielemeier WJ, Wakefield GH, Simoni MH. Time-frequency analysis of musical signals. Proc IEEE 1996;84(9):1216–30.
[13] Brown JC. Calculation of a constant Q spectral transform. J Acoust Soc Am 1991;89(1):425–34.
[14] Bobrek M, Koch DB. Music signal segmentation using tree-structured filter banks. J Audio Eng Soc 1998;46(5):412–27.
[15] Dixon S. Learning to detect onsets of acoustic piano tones. In: MOSART Workshop on Current Research Directions in Computer Music, Barcelona, Spain, November 2001.
[16] Burrus CS, Gopinath RA, Guo H. Introduction to wavelets and wavelet transforms. A primer. Englewood Cliffs, NJ: Prentice Hall; 1998.
[17] Rossing TD, Moore FR, Wheeler PA. The science of sound. 3rd ed. Reading, MA: Pearson Addison Wesley; 2001.
[18] Ortiz-Berenguer LI, Casajús-Quirós FJ, Torres-Guijarro M. Non-linear effects modeling for polyphonic piano transcription. In: 6th International Conference on Digital Audio Effects (DAFX-03), London, UK, September 2003.
[19] Michels U. Atlas de música I. Alianza Editorial, 2001.
[20] Marolt M. Transcription of polyphonic piano music with neural networks. In: 10th Mediterranean Electrotechnical Conference, MELECON 2000, vol. 2. p. 512–5.