CORRECTION SYSTEM FOR POLYPHONIC PIANO RECORDINGS

M. Isabel Martín Erdozain(1), Isabel Barbancho(1), Alejandro Rosa-Pujazon(1), Ana María Barbancho(1)
marisamartin87@gmail.com, ibp@ic.uma.es, alejandror@uma.es, abp@ic.uma.es
(1)Department of Ingeniería de Comunicaciones, Universidad de Málaga. Campus de Teatinos s/n, 29071 Málaga, Spain

Abstract— **In this paper, a support tool for piano rehearsal is presented. The system analyses a given piano polyphonic recording to find the times, pitch and duration of the notes and figures played, taking into account the possibility of playing more than one note simultaneously as well as covering the whole piano frequency range. In order to do so, the system uses an onset detection algorithm to segment the input signal into partitions which are then analysed in the time and frequency domains. Then, the system correlates the data extracted from the partitions with the score of the original piece, identifying the positions and type of the mistakes performed by the user, and providing her/him with the corresponding feedback. The experiments conducted showed that the application is capable of analysing a given recording and indicate the musician the mistakes made.**

I. Introduction

The advances in information and communication technologies in the recent years have spread massively to most aspects of our everyday-life, including also the educational field, as the use of Web 2.0 applications and emerging technologies has been proven to provide useful support tools in the school, helping students in their learning process and empowering their creativity [1], [2], [3], [4].

However, when it comes to the field of music studies, this array of tools might prove to be partially lacking. The use of web resources and applications can constitute a way to increase students' motivation towards learning, and in the particular case of music learning, it can be quite helpful in the early stages offer better accessibility to the abstract concepts of music theory. Learning music, however, relies heavily on practising, and therefore it is usually necessary to have a tutor or expert to show the student which aspects they need to improve, which mistakes they commonly perform, etc. In this sense, there is a need for a more specialised guide in music learning, which the conventional use of currently available applications fail to provide. Thus, to satisfy such needs, it is necessary to use more specialized tools and applications to provide a more specific interface for an adequate learning experience.

In this paper, we present a system that addresses this need, serving as a support tool for piano students, and allowing them to correct their mistakes when practising without the need of having another musician acting as an external reviewer. Concretely, the system allows for the analysis of a musical piano polyphonic recording to assess the correctness of the performance. The system is capable of segmenting and identifying the notes and figures played, and compares them with the score of the piece, finding where the practician has made any potential mistakes, as well as informing the student of the types of mistakes performed.

The next section will present the technical details of the algorithms implemented to analyse a piano recording, including the onset detector and the procedures followed to identify note length, pitch and time. The following section will cover how the piece analysed is corrected according to the data extracted from the original score. Next, the results of the tests performed with the system will be presented, and finally the article will end with a presentation of the conclusions extracted from this work.

II. Analysis of the recorded musical signal

In order to properly process a given musical piece, a time-frequency analysis is performed to find the notes played at each time. The system loads a piano recording from a wav file, with a sampling rate of 44100 Hz and a beats-per-minute value specified by the user. Then, the system divides the signal into temporal slots or "partitions", each of these corresponding to the time at which a given note is being played, and analyses the partition to find the pitch and duration of the note(s) played. The system designed takes into account the possibility of more than one note being played simultaneously and covers the whole frequency range of the piano. In the next following subsections, the most relevant features of the system's processing stage are presented.

A. *Temporal segmentation*

An onset detector [5], [6], [7] has been implemented in order to divide the musical signal into the aforementioned partitions, each onset corresponding to the time instant in which a note "attacks" or "appears" in the signal, i.e. when an energy peak corresponding to that note is introduced. The segmentation process consists of two major steps: the onset detection itself (to find the attack time of the note) and the delimitation of each attack slot (to find when the note "ends"). Before performing this segmentation, the signal is normalized to have an amplitude between 0 and 1.

General onsets and onsets masked by other onsets are located, and then, of all detected onsets, actual onsets are determined.

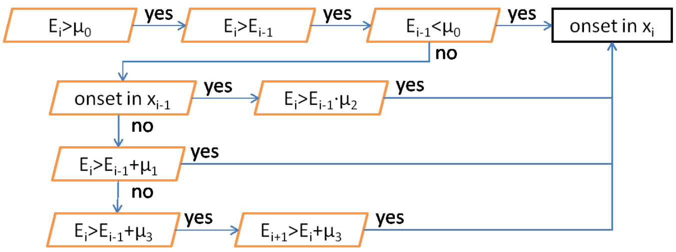

Fig. 1. Onset detection standard process
E_i represents the energy for the i-th window, while the different μ_j values represent the thresholds used to find whether an onset is present or not

In the onset detection step, a sliding window procedure akin to the one described in [8] is used to detect energy peaks. For each sliding window, the energy E_i is calculated as follows:

$$E_i = \sum_{j=x_i}^{x_i+L-1} \Big(y(j)\Big)^2, \qquad (1)$$

where x_i is the index for the initial sample of window i, L is the length of the window in samples and $y(j)$ represents the jth-sample of the piano piece.

We set a threshold value of 0.7 so that only those windows whose E_i exceeds this threshold can potentially contain an onset. According to the ADSR (attack-decay-sustain-release) model [8], it may be possible that part of one note's energy overlaps with the next one, or that a increase of energy is found because of the sustain-release of a previously detected note (false onset). To account for this, the energy of each window is compared with the energy had for the previous and next ones, as per the graph presented in Fig. 1, where μ_0 is the aforementioned 0.7 threshold value, and the parameters μ_1 to μ_3 define conditions to address the previously commented issue.

The different parameters were set to L=3000 samples, μ_1=6.05, μ_2=1.9 and μ_3=3 to find the onsets in the piece considered. It may be possible though that some onsets were masked by higher energy neighbouring notes for the window size considered. To prevent this issue, a second search for masked onsets is subsequently performed, this time with $L =$ 2000 samples and μ_1=5.5.

The piano signal is windowed without overlapping if no attacks are detected. However, if an attack is found in the ith-window, the next window is set to start 20 % windows samples before the location of the maximum amplitude sample found in the ith-window.

Finally, one single note might generate more than one onset peak in a short time period, but the note itself only has one real onset time. To detect and erase subsequent false onsets, a minimum separation distance is defined according to the shortest figure's duration considered in the piece used. In particular, following the ADSR model, the decay and release time is assumed to last 2/3 times the duration of the figure. The minimum separation is then defined as two thirds of the shortest figure duration in the score. Onsets that are separated less that this minimum are combined into a single onset attack, storing only the onset that has the biggest energy value associated.

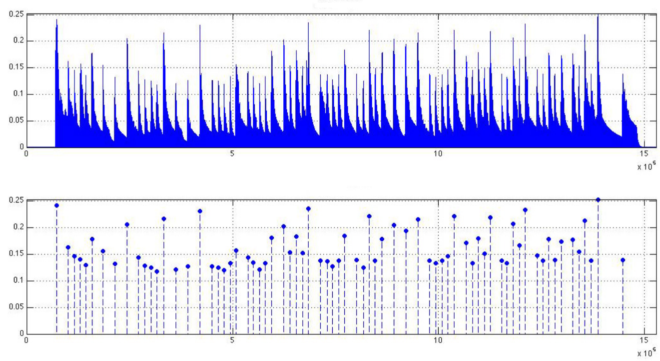

Fig. 2. Input signal (up) and its corresponding detected onsets (down)

After the different onsets have been found, the samples of the musical piece are divided into partitions according to these onsets. Concretely, following the ADSR model, there is a brief attack time before reaching the maximum amplitude (onset), which is modelled by the system by considering that a given partition starts 1000 samples before the onset time itself. In order to prevent overlapping between two consecutive notes near in time, the partition is set to finish 3200 samples before the next onset, as this number guarantees isolation between one note and the next one for the worst case scenario in the range of frequencies of the piano (which was found at 72.4 milliseconds, 3195 samples roughly at a sampling rate of 44100 Hz). An example of the output of the onset detector implemented can be seen in Fig. 2

B. Temporal analysis: finding the duration of the notes played

After finding the onsets and dividing the signal into partitions, the system has an effective segmentation of the notes played at each time. The next step needed is to find which is the duration of each of the notes found. This is easily achieved by simply dividing the number of samples in each partition by the sampling rate in the recording (by default, 44100 Hz). However, this measure of duration is dependent on the tempo at which the piece is being played. Thus, in order to objectively determine the duration of the notes independently of the velocity in the performance, this measure of duration is normalized by dividing it by the black figure's duration (which is actually the beats-per-minute value specified by the user).

As each partition normalized duration is calculated, the figure played in that partition is then classified as one of possible figures in the score (whole, half, black, half-time, quarter-time, etc.), assigning it the type of figure whose normalized duration is closest to. The system also takes into account the possibility of having dotted notes in the classification process.

C. Frequency analysis

The time of the notes as well as their duration has already been found using the previous modules. The last step to fully characterize the score of the piece recorded is to find the pitch of the notes actually played. This is performed by transforming each partition into the frequency domain by using a DFT (Discrete Fourier Transform). The resulting spectrum is normalized so that the frequency axis is scaled into a MIDI numbers axis.

From this spectrum, a peak detection process is performed to find the pattern of peaks associated with each partition (that is, which fundamental frequencies and partials are present in the partition), normalizing the amplitudes of the peaks found to the value of 1. This pattern of peaks is then correlated with the pattern of peaks that correspond to each of the notes that should be played in that instant according to the score of the piece assessed. If a given note was correctly played, then its pattern of peaks should be included in the pattern found for the corresponding partition.

III. Assessment of correction

A. *Parameters of a piano piece*

The previous section presented the most relevant aspect of the analysing blocks of the system that allow as to identify the notes being played, their position and duration. In order to check the correctness of the piece recorded, the system also requires to have additional parameters specified regarding the original score of the piece. Concretely, the system extracts the required information regarding the time signature, musical figures and notes from an auxiliary data base.

In this regard, the system stores information corresponding to the times of each note as well as their normalized duration. If two or more notes are played at the same time, the system will only consider the figure with shorter duration.

The system identifies which notes are being played at each instant according to the peaks found in the MIDI spectrum. Thus, for each note or group of notes present in the score, its corresponding pattern of fundamental and partial spectral peaks is stored as a reference. To account for potential variations in the frequency values of the peaks due to inharmonicity [9], the reference pattern stored is not actually a pattern of peaks, but rather a pattern of narrow filters of width 1 (in the MIDI scale) centered at the frequencies were the fundamental and partial peaks should lie in theory.

B. *Evaluation and correction*

The system uses the data stored as a reference and the information extracted from the analysis stage to judge the correctness of the piece recorded. Two correction processes have been implemented, depending on whether the number of note times identified is coincident with the number of partitions found.

1. Normal correction: If the number of partitions and note times is the same, the system evaluates if the duration of the notes in the performance is correct by simply comparing the duration estimated in each partition with the expected one in the reference.

To verify if the notes have been played at their correct frequencies, the pattern of peaks of each partitions is filtered by the corresponding reference pattern, and the resulting spectrum is summed along all frequencies. If both patterns match, the expected sum will be zero, otherwise the notes played were not correct (there are peaks in the partition spectrum that do not have their match in the reference pattern)

2. Special correction: The special correction is performed when the number of partitions is different from the number of note times. Given a piano performance, an error in the musical figure played is more probable than an error in the musical note played. Therefore, the latter will be given more importance than the former, as it is associated with a more meaningful mistake. When assessing each partition's correctness, there are three possible cases:

- The notes played are the expected ones: in this case, the system proceeds as per the normal case.
- The notes are different from the expected ones and there are more notes in the recording than in the score: this happens because new notes have been added or the same note has been played more than once. The notes played in the actual partition are then compared with the ones expected for the next one. If there is a coincidence, then it is assumed that the user played a new non-existent note in the current partition. Otherwise, it is assumed that the user simply played the note wrong.
- The notes are different from the expected ones and there are less notes in the recording than in the score: this will happen if the user skipped some notes when performing. Again, the next expected note in the score is compared with the note had in the actual partition. If they are the same, it is assumed that the user skipped one note, otherwise it is again assumed that the user played the note wrong.

IV. Results

A. *Onset detection*

It is crucial to corroborate that the onset detection algorithm proposed is indeed capable of adequately segmenting the signal and identifying the partitions which the score consists of. Thus, we conducted a set of tests in which a set of ten quavers or quarter-notes were played at different velocities, ranging from 40 to 230 beats-per-minute. For each tempo considered, three different indicators are used to assess the quality of the onset detector: the rate of detected notes over the total number of notes played (denoted by N), the rate of false notes detected (false negatives, FN), and a score indicator defined as follows:

$$Score = \frac{N}{N + FP + FN} \times 100, \tag{2}$$

where FP represents the false positives rate (non-existent onsets detected). The results yielded are summarized in Fig. 3. The onset detector used shows to be very effective at finding the notes being played as long as the tempo of the piece is not higher than 180 bpm. For faster pieces, the quality of the detection worsens gradually.

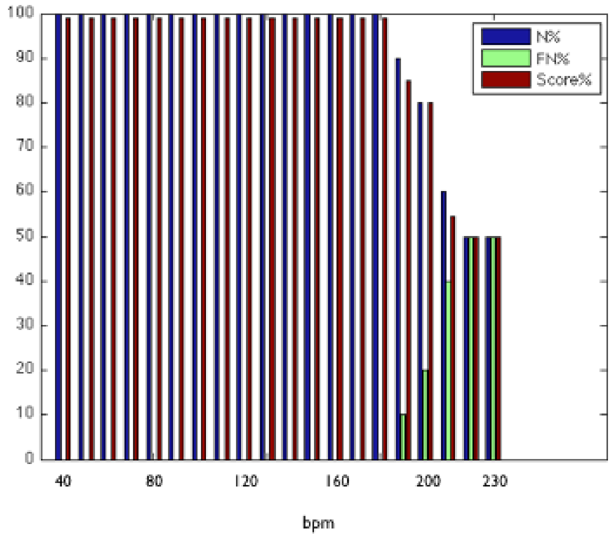

Fig. 3. Results for the evaluation of the onset detector
N = the rate of detected notes over the total number of notes, FN = the rate of false negatives), $Score = \frac{N}{N+FP+FN} \times 100$ with FP being the rate of false positives

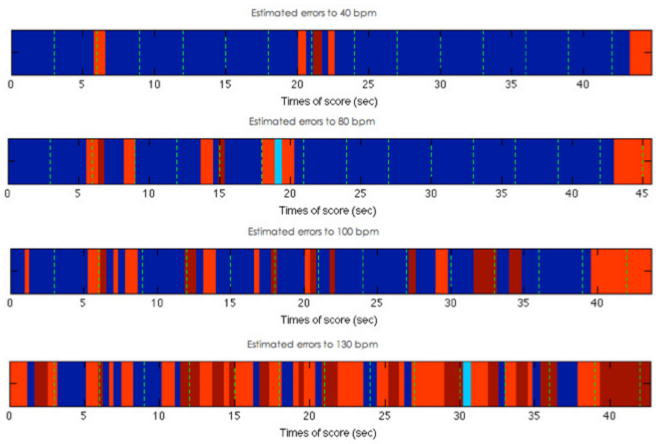

Fig. 4. Results for the evaluation of the onset detector:
each bar color is associated to one type of error, i.e. dark blue - no error, orange - wrong duration of note, red - wrong note, cyan - skipped note

B. *Correctness evaluation*

In order to assess the viability of the global system as a support tool for aided rehearsal, we conducted an experiment in which the song 'menuet 114' was played and recorded at 21 different speeds. The system presents the user a colored image that indicates him/her the types of errors that have been found in his/her performance. An example can be found in the figure Fig. 4. It was found that users tend to perform more errors when the tempo of the piece is higher, which is something to be expected; also, as can be noticed, the vast majorities of the mistakes made in the performance come from not keeping the duration of the figures as indicated in the score.

From the tests performed, it was found that the system detected correctly the notes played most of the time (as Fig 3 shows). However, it was found that the system failed to properly correct errors in some specific cases:

- The system cannot discriminate when the user is playing with one hand or another, but instead makes a global correction of both hands at the same time.
- It may be possible that a mistake when playing a figure is assumed as a a mistake in the note played. I.e., if the musician is playing a long note with his left hand and releases it slightly before time, the lack of its spectral contribution could affect the detection of the notes played with the right hand.

V. Conclusions

In this paper, we have presented a system that analyzes a polyphonic piano recording and assesses the correctness of the piece played. The system automatically segments the musical signal, identifying the notes played and their duration, as well as indicating the type of mistakes the musician performed in the piece recorded. The tests conducted showed that the system has a very good performance and does indeed fulfil its purpose as a support tool in the learning processes. The system designed works for any kind of piano and does not require prior training of any type.

Acknowledgements

This work has been funded by the Ministerio de Economia y Competitividad of the Spanish Government under Project No. TIN2010-21089-C03-02 and by the Ministerio de Industria, Turismo y Comercio under Project No. TSI-090100-2011-25.

References

[1] Ajjan, H. and Hartshorne, R. "Investigating faculty decisions to adopt Web 2.0 technologies: Theory and empirical testsc," *The Internet and Higher Education* vol. 11, no. 2, pp. 71-80, 2008.

[2] Condie, R. and Munro, B. "The impact of ICT in schools: Landscape review," *British Educational Communications and Technology Agency (BECTA)*, 2007.

[3] Harrison, C, Comber, C, Fisher, T, Haw, K, Lewin, C, Lunzer, E, McFarlane, A, Mavers, D, Scrimshaw, P, Somekh, B and others. "ImpaCT2: The impact of information and communication technologies on pupil learning and attainment," *British Educational Communications and Technology Agency (BECTA)*, 2002.

[4] Ajjan, H. and Hartshorne, R. "Sowing the seeds for a more creative society," *Learning and Leading with Technology* vol. 35, no. 4, pp. 18-22, 2007.

[5] Bello J.P, Daudet L, Abdallah S, Duxbury C, Davies M, Sandler M.B. "A tutorial on onset detection in music signals," *IEEE Transactions on Speech and Audio Processing,* vol.13 , no. 5, pp. 1035-45, IEEE. 2005.

[6] Boogaart, C.G and Lienhart, R. "Note onset detection for the transcription of polyphonic piano music, " *IEEE International Conference on Multimedia and Expo, ICME 2009*, pp. 446-449, 2009.

[7] Benetos, E and Dixon, S. "Polyphonic music transcription using note onset and offset detection, " *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 37-40, 2011.

[8] Barbancho I, Barbancho A.M, Jurado A, Tardón L.J. "Transcription of piano recordings," *Applied Acoustics,* vol.65 , no. 12, pp. 1261-66, 2005.

[9] Barbancho A.M, Barbancho I, Soto B, Tardón L.J. "SIC receiver for polyphonic piano music," *In Proc. ICASSP,* pp. 377-380, 2011.