Piano Legato-Pedal Onset Detection Based on a Sympathetic Resonance Measure

Beici Liang, György Fazekas and Mark Sandler
Centre for Digital Music, Queen Mary University of London
London, United Kingdom
Email: \{beici.liang,g.fazekas,mark.sandler\}@qmul.ac.uk

Abstract—In this paper, the problem of legato pedalling technique detection in polyphonic piano music is addressed. We propose a novel detection method exploiting the effect of sympathetic resonance which can be enhanced by a legato-pedal onset. To measure the effect, a specific piano transcription was performed using the templates of pre-recorded isolated notes, from which partial frequencies were estimated. This promotes the acquisition of residual components associated to the weak co-excitation of damped notes due to the legato pedalling technique. Features that represent the sympathetic resonance measure were extracted from residuals. We finally used a logistic regression classifier to determine the existence of legato-pedal onsets.

Index Terms—onset detection, sympathetic resonance, piano acoustics, piano pedalling techniques

I. INTRODUCTION

Automatic music transcription (AMT) focuses on transforming an audio signal into symbolic data using features that correspond to acoustic and musical properties. Many of these features however are designed to capture the basic units of music such as pitch and note onset. Few features are related to playing techniques, a main component of expressive performance. Identifying playing techniques has become a particular new direction in AMT.

In piano music, pedalling techniques subtly colour the timbre to create different artistic expressions by adjusting the timing and depth of pedal press and release. Of the three standard pedals, the most frequently used is the sustain pedal which prolongs the sound by lifting all dampers off the strings. Besides sustaining the sounding notes, the sustain pedal also allows strings associated to other keys to vibrate due to coupling via the bridge. This phenomenon is known as sympathetic resonance and is defined in the dictionary of acoustics as “resonant or near-resonant response of a mechanical or acoustical system excited by energy from an adjoining system in steady-state vibration” [1]. Pianists embrace the phenomenon to produce seamless legato through a technique called legato pedalling [2]. Idealised spectrograms of two chords played with and without the sustain pedal are presented in Figure 1 to illustrate the effect of legato pedalling.

Prolong the first Cmaj chord, the sustain pedal is pressed while the fingers are still holding down the keys. When the fingers are lifted to reach for the Fmaj chord, the Cmaj chord is sustained because the pedal prevents the dampers from falling onto the strings. Immediately after the Fmaj chord onset, the pedal is released to avoid blurring effect caused by the overlap of the two sonorities. The dampers stop the vibrations in all strings whose keys are not currently pressed. Then, the pedal is pressed again to sustain the Fmaj chord, lifting dampers off the strings. This can slightly co-excite the damped strings associated to the previous Cmaj chord with the playing Fmaj chord. A detailed study of such an indirect excitation on piano tones is introduced in [3]. In this paper, a novel audio-based method is proposed for the detection of legato pedalling, i.e., legato-pedal onset times. Pedalling techniques can be adjusted by the performer’s sense of tempo, dynamics, textural balance as well as the settings or milieu in which the performance takes place [4]. Our legato-pedal onset detection could contribute to applications of piano pedagogy and the growing field of music information retrieval.

Existing methods for playing technique detection generally employ a two-stage framework: feature extraction followed by a decision-making mechanism. Features are extracted based on the different characteristics of the techniques. The decision-making stage can use machine learning methods such as...
Support Vector Machines (SVM) [5]. For instance, a system for drum playing technique detection in polyphonic mixtures of music was designed in [6] by using features extracted from activation functions of a Non-Negative Matrix Factorisation (NMF) algorithm [7], then classifying those features into strike, buzz roll, flam and drag using a trained SVM. To undertake the detection of techniques such as “bend” and “slide” in guitar playing, low-level spectral features were extracted and fed into an SVM with sparse coding in [8]. Another approach for detecting playing techniques, commonly used across different kinds of instruments, is modelling the observable patterns which span a certain duration in the time-frequency plot. Frame-wise vibrato detection can be achieved using the Filter Diagonalisation Method (FDM). A method proposed in [9] allows frame sizes to be set at values small enough for identifying local vibrato characteristics.

In the field of piano pedalling techniques, detection methods in the current literature are all based on isolated notes. The main features of the sustain-pedal effect outlined in [10] are the energy of residuals, decay time and amplitude beating. The values of these features are increased when the sustain pedal is fully engaged. The energy of residuals was used in [11] to separate notes played with or without the sustain pedal through autoregressive modelling of the estimated residuals and then selecting a threshold to define the two classes. We proposed audio features based on the analysis of both harmonics and residuals of recorded piano notes in [12]. In contrast to binary classification, we trained a decision-tree-based SVM model to identify isolated notes played with or without pedalling techniques of different pedal timing and depth.

In our prior work, there are strong assumptions about the ability to extract clean features from isolated notes that may not apply in more realistic situations. Despite the importance of pedalling technique in piano performance, its detection in polyphonic piano music has not been previously investigated. In this paper, we propose such an approach to detect legato-pedal onset using a sympathetic resonance measure which deals with feature extraction in the presence of overlapping partials when different notes are sounding. Our method is designed and evaluated on a novel dataset dedicated for research on piano pedalling techniques. Modelling the specific instrument being transcribed can efficiently improve AMT performance [13]. For this reason, our dataset was built using a specific piano to employ knowledge about the physics and acoustics of the instrument. We believe having access to model parameters of a specific piano is a reasonable assumption for many performance scenarios.

II. METHOD

The proposed method is motivated by piano acoustics introduced in the previous section. A method of measuring sympathetic resonance is developed using the weak co-excitation of damped notes, which is due to the legato pedalling technique. Figure 2 shows a flow chart describing our system. We first obtain the note transcription through a current state-of-the-art specific piano transcription method proposed in [14]. Isolated notes from our dataset are used to form the templates for the transcription. Their partial frequencies are estimated using the method proposed in [15]. Sinusoidal components are determined by the partial frequencies of the notes from their onset to offset times based on the transcription results. We then obtain the residuals by subtracting the sinusoidal components from the original sound. Features are extracted from residuals using a sympathetic resonance measure, which we consider the main contribution of this paper. Finally, existence of legato-pedal onset is determined via a classification mechanism. Each step is explained in the following subsections. All code used in our study is made publicly available1.

A. Transcription for Specific Piano

Transcription converts piano audio into a set of note events, each consisting of pitch, onset and offset times. Piano transcription can be implemented using NMF, which factorises a spectrogram of a piano recording into 88 spectral bases and corresponding activations. The sound to be transcribed can be reconstructed using:

$$V_{ft} = \sum_{k=1}^{K} W_{fk} H_{kt},$$  (1)

where $V$ is the reconstructed spectrogram, $W$ is the note template, and $H$ is the note activation. $f \in [1, F]$ is the frequency bin, $t \in [1, T]$ indicates the time frame, and $k \in [1, 88]$ is the pitch index. We employed the method proposed in [14] for specific piano transcription, where the 88 isolated notes played with mezzo-forte dynamics were pre-recorded using the same piano we wish to transcribe. Considering the different spectral and temporal characteristics at the attack and decay phases of a piano note, these two phases were reconstructed individually. The 88 isolated notes were used to obtain the attack and decay templates. For transcription, only the activations were updated. Thereafter onsets were detected from attack activations by peak-picking, and offsets were detected by dynamic programming proposed in [16].

1https://github.com/beiciliang/eusipco2018-legatopedal
B. Partial Estimation

Due to piano string stiffness, note partials occur at frequencies slightly higher than perfect harmonics. The theoretical partial frequencies of a note can be computed using Eq.2:

\[ f_n = n f_0 \sqrt{1 + B n^2}, \]

where \( n \) is the partial index, \( f_0 \) is the corresponding frequency, and \( f_0 \) is the fundamental frequency. \( B \) is the inharmonicity coefficient, which varies from note to note [17]. The method proposed in [15] was applied to the estimation of \( f_0 \) and \( B \) from the 88 pre-recorded notes played with *mezzo-forte* dynamics. We set the maximum \( f_0 \) less than \( f_s/3 \), where \( f_s \) is the sampling frequency in Hz. Thereafter partial frequencies for piano notes were obtained. The influence of dynamics is beyond the scope of this paper.

C. Sympathetic Resonance Measure Based on Residuals

A well-established spectral modelling of musical sound is to represent it as two separate components: stable sinusoids (partials) and noise (residuals) [18]. With pianos, the partial frequencies are stable so the frequencies can be fixed across frames between the note onset and offset [19]. Based on the results from transcription and partial estimation, sinusoidal components can be obtained by tracking the amplitude and phase of the transcribed notes’ partial frequencies from their detected onsets until offsets. For instance, a note with estimated partials \( f_n \) is detected playing from the input signal \( x \) at time \( t \), which can be modelled by:

\[ x(t) = \sum_{n=1}^{N} A_n(t) \cos(\theta_n(t)) + r(t), \]

where \( A_n(t) \) and \( \theta_n(t) \) are the instantaneous amplitude and phase of the \( n \)-th sinusoid respectively. Then \( \theta_n(t) \) is taken to be the integral of the instantaneous frequency \( f_n \). \( r(t) \) is the residual component at time \( t \). From Eq. 3, residual components were thus obtained by subtracting the sinusoidal components from the original sound.

At this stage, the residual consists of background noise, the sound of hammer-string strikes from note attacks, piano tones whose note events were not correctly transcribed and the effect of sympathetic resonance. As discussed in Section I, when a legato pedalling is played, the effect of sympathetic resonance is enhanced by string coupling via the bridge. This transfers energy from excited string vibrations of a played tone to unstruck strings in the time domain in order to detect legato-pedal onset.

The nature of legato pedalling implies that the pedal onset is between two note onsets. According to the note onsets detected in Section II-A, we defined segments, from which legato-pedal onset was detected. To determine the segments, we fused note onsets that are within a fixed temporal tolerance window of an estimated 16th note duration. This is because different onsets may be detected for the notes played as a single chord, and it is not possible to change the pedal with every note [20]. Therefore frames between every two successive note onsets were defined as a segment if the time range they covered was above the tolerance window. The extent of sympathetic resonance in each segment was then measured by the root-mean-square (RMS) energy using Eq.4:

\[ RMS[S[m]] = \sqrt{\frac{\sum_{m} S[m]^2}{M}}, \]

where \( S \) is the residual spectrogram within the current segment, \( m \) is the index of the frequency bins corresponding to the selected partials, and \( M \) is the number of selected partials. The whole procedure of sympathetic resonance measure based on residuals is illustrated by Algorithm 1, where \( P \), \( ON \) and \( OFF \) refer to the lists of transcribed pitch index, onset and offset times in seconds. \( PF \) is the estimated partial frequencies of the 88 pre-recorded notes. \( R \) is the residual spectrogram.

**Algorithm 1** Sympathetic resonance measure

```
Require: \( \theta \): estimated duration of a 16th note in seconds; \( \alpha, \beta \): frame and frequency resolution of the spectrogram
procedure MEASURE(P, ON, OFF, PF, R)
    I, J ← 0: len(R) − 1, 0: len(P) − 1
    SRM ← zeros(I)
    for all \( k \in J[0 : −1] \) do
        if \( ON[k+1] − ON[k] > \theta \) then
            \( t ← ON[k] \)
            \( SFreq ← empty list \)
            for all \( j \in P \) do
                if \( ON[j] < t < OFF[j] \) then
                    \( SFreq ← append(SFreq, PF[P[j]]) \)
                    \( frmS, frmE ← ON[k]/\alpha, ON[k+1]/\alpha \)
                    \( freqBin ← SFreq/\beta \)
                    \( SRM[freqBin] ← RMS[R[freqBin]] \)
            return SRM
```

D. Feature Extraction

Our intuitive observation relates the changes in the value of RMS energy to pedalling onsets. An energy peak appears when sympathetic resonance is enhanced by the legato-pedal onset. Maximum RMS energy was extracted as a feature per segment and recorded on both linear and decibel scale (denoted as \( Max_{linear} \) and \( Max_{dB} \) respectively). Because legato pedalling follows a note attack, the peak location with respect to the number of frames away from the note onset, i.e., starting point of the segment, was extracted from every segment (denoted as \( Peak_{loc} \)). A new 2-dimensional feature was designed by combining \( Peak_{loc} \) with either \( Max_{linear} \)
or $Max_{dB}$, depending on which is more representative of the maximum RMS energy. This was assessed through the evaluation presented in Section III. This aims at detecting legato-pedal onset with a higher accuracy. Finally, the existence of legato-pedal onset in each segment was determined via logistic regression as a binary outcome.

### III. Experiments

#### A. Dataset

Most public annotated piano datasets are either for research on multi-pitch estimation [21] or lack isolated-note recordings from the same piano [22]. We therefore built our own dataset for piano pedalling techniques detection. MIDI files of four classical pieces by different composers were selected from the SMD dataset [22]. These pieces were performed by professional pianists on a Yamaha Disklavier which allows key and pedal movements to be captured and stored as MIDI files. They were rendered into audio using a reproducing piano. This yielded an audio dataset with fully-automatic and reliable annotation. The recording was carried out at the Yamaha recording studio in Milton Keynes, United Kingdom, in March 2017. The instrument was a Yamaha Disklavier grand piano which was tuned directly prior to the recording session. The audio of the four pieces and isolated notes were recorded using the spaced-pair stereo microphone technique using Earthworks QTC40 omidirectional condenser microphones positioned about 50 cm above the strings. The positions were kept constant during the recording. The signals were recorded with a sampling rate of 44.1 kHz and a resolution of 24 bits.

Each audio signal was segmented as discussed in Section II-C based on note onsets from the transcription. The pedal movement from MIDI-annotation is represented using integers (0-127). To annotate each segment and designate the existence of legato-pedal onset, we first quantised pedal movement and represented it using a binary state variable. Values below 64 are considered the “off” state, while values above are considered “on”. Each segment was then labelled “1” to denote the presence of legato-pedal onset, when the pedal state within the segment had transitioned from off to on, otherwise segments were labelled “0”. Table I lists the number of segments annotated using 0 or 1 in each piece representing our ground truth data. The current dataset is limited in the number of music pieces, however there are almost 3000 segments in total to be classified.

The input signals were divided into frames of 2048-sample Hann window (hop size = 512) to compute the spectrogram. To further minimize the influence of percussive components from the residuals, based on which the sympathetic resonance was measured in Section II-C, harmonic percussive source separation using a median-filtering technique [23] was applied to the original spectrogram as a pre-processing step. Residuals were obtained by removing the sinusoids from the harmonic components.

#### B. Evaluation

Our goal is to evaluate the efficiency of the proposed signal processing method and features representing a measure of sympathetic resonance, instead of classifier selection. Logistic regression was chosen as a binary classifier because the model should discriminate the presence or absence of sympathetic resonance in note segments.

As we introduced in Section II-D, $Max_{linear}$ and $Max_{dB}$ in every segment were extracted as features. To determine which one represents the segments better, we evaluated logistic regression models with the two features separately, using the Akaike information criterion (AIC) [24]. A logistic regression model with the more representative feature should yield a smaller AIC value. $Max_{dB}$ was selected because it returns a smaller AIC value of 2507.39 than 2555.36 by $Max_{linear}$. We also evaluated the logistic regression model with 2-dimension features consisting of $Max_{dB}$ and $Peak_{loc}$. This was chosen as the final model because it gives the smallest AIC value of 2361.35.

To evaluate the model, data in each piece was separated into two halves, one for training and the other one for testing. This piece-level evaluation was selected because the overall tempo and dynamics in a piece effect the attributes of the trained model. Moreover, models were trained with weighted classes in the Beethoven and Ravel pieces, which exhibit unbalanced data containing less than 100 segments with legato-pedal onset. Given the test results for every piece, we then calculated precision ($P_1$), recall ($R_1$) and F-measure ($F_1$) with respect to label “1”. In addition, we show the overall performance using micro-averaged F-measure ($F_{micro}$). We used the scikit-learn\(^2\) library to construct the model and compute the performance metrics.

#### C. Results and Discussion

Table II presents the model performances for each piece. The overall results indicate that our method extracts relevant features to represent the effect of sympathetic resonance and helps to detect legato-pedal onsets from audio. In terms of performance metric for label 1 (indicating legato-pedal onset in a segment), results of the Chopin piece achieve the best $P_1$ and $F_1$, while those of the Beethoven piece are the worst.

---

\(^2\)http://scikit-learn.org
We assume this is partially due to the different nature of their music. Our method works better on the pieces where legato pedalling is an essential ingredient to create contrast between pedalled and unpedalled notes through the effect of sympathetic resonance. If this effect is also correlated with other musical attributes, our features are less representative as an indicator of legato-pedal onset which leads to poor classification performance. This adverse effect can be observed particularly in terms of precision due to the increased number of false positives.

IV. CONCLUSIONS AND FUTURE WORK

This paper presented a method for detecting legato-pedal onsets of the piano based on a measure of sympathetic resonance. The energy of unstruck strings can represent the extent of sympathetic resonance which changes with the legato pedalling technique. In our method, residual note segments were obtained using piano transcription and partial estimation. Based on the energy in each segment, maximum value in decibel scale and peak location were extracted as features. The existence of legato-pedal onset per segment was determined using a logistic regression classifier trained on the features. The overall performance shows that this technique can be used as an indicator of legato pedalling. Our method, although still preliminary, is the first to detect pedalling technique in polyphonic piano music.

Our evaluation results are well grounded in musicology and may be explained by differences in pedal use between composers and musical eras. It is well acknowledged that Chopin took particular care of pedal notations. This could yield cleaner features and consequently the best results for the Chopin piece. If the piece itself has cross rhythms and dense harmonic structure, which Brahms’ music is firmly rooted in, the extent of sympathetic resonance may not be significantly changed by legato-pedal onset. In this case, our features are less discriminative. Similarly, other playing techniques that are correlated with the effect of sympathetic resonance may degrade classification performance. This is observed in the Ravel piece, which puts emphasis on timbral nuances, expanding the keyboard and pedalling techniques more than the use of legato pedalling. Finally the results in the Beethoven piece can also be explained by highly unbalanced training data, due to the fact that legato pedalling was rarely used in Beethoven’s time.

Our method used piano transcription technique as an intermediate step for the residual acquisition and segmentation. Future work will investigate how to reduce the effects of note transcription errors on the sympathetic resonance measure. In addition, the current logistic regression model learned a fixed threshold for each piece to perform classification. We will assess adaptive thresholding techniques on the calculated residual energy. This could facilitate more accurate estimation of pedal onset times without a training process.

REFERENCES