# Measurement, Recognition, and Visualization of Piano Pedaling Gestures and Techniques

**BEICI LIANG,** *AES Student Member,* **GYÖRGY FAZEKAS,** *AES Associate Member,* **AND**
(beici.liang@qmul.ac.uk)                                    (g.fazekas@qmul.ac.uk)

**MARK SANDLER,** *AES Fellow*
(mark.sandler@qmul.ac.uk)

*Centre for Digital Music, Queen Mary University of London, London, UK*

This paper presents a study of piano pedaling gestures and techniques on the sustain pedal from the perspective of measurement, recognition, and visualization. Pedaling gestures can be captured by a dedicated measurement system where the sensor data can be simultaneously recorded alongside the piano sound under normal playing conditions. Using the sensor data collected from the system, the recognition is comprised of two separate tasks: pedal onset/offset detection and classification by technique. The onset and offset times of each pedaling technique were computed using signal processing algorithms. Based on features extracted from every segment when the pedal is pressed, the task of classifying the segments by pedaling technique was undertaken using machine learning methods. We compared Support Vector Machines (SVM) and hidden Markov models (HMM) for this task. Our system achieves high accuracies, over 0.7 F1 score for all techniques and over 0.9 on average. The recognition results can be represented using novel pedaling notations and visualized in an audio-based score following application.

## 1 INTRODUCTION

Pedaling is among the important playing techniques that lead to expressive piano performance. It is comprised of not only the onset and offset information that composers often indicate in the score but also gestures related to the musical interpretation by performers such as part-pedaling techniques.

Modern pianos usually have three pedals, among which the most frequently used is the sustain pedal. The use of the sustain pedal constitutes one of the main musical gestures to create different artistic expressions through subtly coloring resonance. The sustain pedal lifts all dampers and sets all strings free to vibrate sympathetically with the current note(s) being played. Given that detecting pedaling nuances from the audio signal alone is a rather challenging task [1], we propose to (a) sense pedaling gestures from piano performances using a non-intrusive measurement system; (b) devise a method for reliable recognition of pedaling techniques; and (c) visualize the results that indicate the onset and offset times of the sustain pedal and the use of part-pedaling, i.e., pedal depth or the extent to which the pedal is pressed. Pedaling is not consistently notated in sheet music and, as we noted, there are important subjective variations of pedal use related to expressive performance. Therefore, this study benefits many applications, including automatic music transcription and piano pedagogy.

This paper is organized as follows. We first introduce the background of piano pedaling techniques and related works in Sec. 2. We then present the measurement system for data acquisition in Sec. 3. The process of database construction is described in Sec. 4. The methods of pedaling recognition including onset/offset detection and part-pedaling classification are discussed in Sec. 5. A visualization system and other potential use cases are outlined in Sec. 6. We finally conclude and discuss our future works in Sec. 7.

## 2 BACKGROUND

### 2.1 Piano Pedaling Techniques

Pedals have existed in pianos since the 18th century when Cristofori introduced a forerunner of the modern soft pedal. It took many decades before piano designers settled on the standardization of three pedals in the late 19th century. Mirroring the development of the pedals themselves, the notations used for indicating pedaling techniques have likewise changed over the centuries. Composers like Chopin and Liszt actively indicated the use of pedals in their works [2], while Debussy rarely notated pedaling techniques despite the importance of pedal use for the intended or an elaborate interpretation of his music [3]. Experts agree that pedaling in the same piano passage can be executed in many different ways, even when pedal markings are provided [4]. This

is adjusted by the performer's sense of tempo, dynamics, textural balance as well as the settings or milieu in which the performance takes place [3].

Pedaling techniques can vary in two domains: timing and depth [2]. This is especially the case for the sustain pedal. There are three main pedaling techniques considering the timing of the pedal with respect to note onsets. *Rhythmic pedaling* is employed when the pedal is pressed at the same time as the keys. This technique supports metrical accentuation, which is an important aspect of Classical-era (extending roughly from the late 18th century to the mid 19th century) performance. Pressing the pedal immediately after the note attack is called *syncopated* or *legato pedaling*. This enables the performer to have "extra fingers" in situations where legato playing is not possible with any fingering. *Anticipatory pedaling*, first described in the 20th century, can only be applied after a silence and before the notes are played. This technique is used to produce greater resonance at the commencement of the sound. Besides variation in pedal timing, professional pianists apply part-pedaling techniques that change as a function of the depth of the sustain pedal. Apart from *full pedal*, Schnabel defined another three levels of part-pedaling in [5]. These are referred to as *1/4 pedal*, *1/2 pedal*, and *3/4 pedal*. It should be noted that these terms neither refer to specific positions of the pedal, nor to specific positions of the dampers, but only characterize the amount of sound that remains when the keys are released. The position of the pedal that produces the ideal sound effect of part-pedaling can vary from one piano to another and may even vary on the same piano under different conditions.

In summary, with the help of the pedals, pianists can add variations to the tones. Pedal onset and offset times may be annotated in music scores. However, no compositional markings exist to indicate the variety of part-pedaling techniques mentioned above [6]. Moreover, the role of pedaling as an instrumental gesture to convey different timbre nuances has not been adequately and quantitatively explored, despite the existence of some studies on the acoustic effect of the sustain pedal on isolated notes described in [7] and [8].

## 2.2 Related Works

There has been a significant amount of research on instrumental gestures in piano performance. The strongest focus so far has been placed on hand gestures, starting from an early study by Ortmann [9] who first approached the "mystery of touch and tone" on the piano through physical investigation, to an extensive review of the studies on piano touch in [10]. Meanwhile, arm gestures have been used in piano pedagogy application through sensing the arm movement and generating feedback to increase piano students' awareness of their gestures in [11]. However, no formal study on piano pedaling gestures can be found in the literature.

In terms of data acquisition, several measurement systems have been developed to be in place for multi-modal recordings, in order to capture comprehensive performance parameters. The Yamaha Disklavier and Bösendorfer CEUS pianos have the ability to record continuous pedal position for example, which was used by Bernays and Traube [12] as one of the performance features to investigate timbre nuances. However, these instruments are rather expensive and not easily moved, which remain a barrier to wider adaptation. To overcome these problems, the Moog Piano-Bar [13] was developed as a convenient and practical option for adding MIDI recording capability to any acoustic piano. Its pedal sensing capability however is limited to discrete positions that only provides on/off information. McPherson and Kim [14] modified the PianoBar in order to provide a continuous stream of position information, but thus far few detailed studies have made use of the pedal data. These problems have motivated us to develop a dedicated, portable, and non-intrusive system to measure continuous pedaling gestures for our further analysis. Our goal is to enable better understanding of the use of pedals in expressive piano performance, as well as to aid detailed capture and transcription of piano performances.

In the field of analysis of expressive gestural features in music performance, the use of machine learning has become a common approach. This is primarily because of the flexibility of statistical models to deal with inexact observations and their ability to adapt to individual differences between players or interpretations, owing to their probabilistic representations of underlying signal data, or the ability to learn and exploit dependencies between the techniques employed [15]. For instance, Van Zandt-Escobar et al. [16] developed *PiaF* to extract variations in pianists' performances based on a set of given gesture references. The estimated variations are used subsequently to manipulate audio effects and synthesis processes. Yet, the inclusion of pedaling techniques was not considered as part of gesture sensing in this or other related studies, let alone the provision of intuitive feedback to users.

The approach taken in this paper follows the aforementioned ideas but with a focus on the measurement and recognition of piano pedaling techniques. Our measurement system enables synchronously recording the pedaling gestures and the piano sound at a high sampling rate and resolution, with the ability to be deployed on common acoustic pianos. Using the sensor data collected from the system, we first detect onset and offset times of pedaling gestures on the sustain pedal using signal processing techniques. Relying on different assumptions discussed in Sec. 5, two machine learning methods (SVM and HMM) are proposed for classifying the segment between every pedal onset and offset. We focus on four known pedaling techniques: quarter, half, three-quarters, and full pedal. Good recognition results are obtained with the SVM-based method, which outperforms HMM in our case (see Sec. 5.3).

The developed algorithms are finally demonstrated in an audio-based score following system, extended with customized markings we devised to notate pedal use. These markings are visualized in the context of the music score in our application, which may be useful in musicology, performance studies or piano pedagogy. The possible use of the dataset created using our data acquisition system

Fig. 1. Schematic overview of the measurement system.

in the context of audio-based pedaling recognition is also discussed.

## 3 MEASUREMENT SYSTEM

This section describes a novel measurement system based on our previous work [17] to capture pedaling gestures on the sustain pedal. Fig. 1 illustrates the schematic overview of our system, consisting of a sensor and circuit system to collect pedal depth data, as well as an audio recorder and a portable single-board computer to capture both data sources simultaneously.

Near-field optical reflectance sensing was used to measure the continuous pedal position with the help of a reflective photomicrosensor (Omron EESY1200). This includes an LED and a phototransistor in a compact package. The sensor was mounted in the pedal bearing block, pointing down towards the sustain pedal. This configuration avoids interference with pianists. One of the major considerations in selecting this optical sensor is that its response curve is monotonic within the optimal sensing distance (0.7 mm to 5 mm). As the sustain pedal is pressed that the pedal-sensor distance is increased, the pedal reflects less of the optical beam projected by the sensor emitter, thus decreasing the amount of optical energy reaching the detector. However, when the sustain pedal is too close to the sensor, the current will drop off. We ensured that the measurement made use of the linear region of the sensor and remained in the optimal sensing range through a calibration procedure. Then the output voltage of the sensor was amplified and scaled to a suitable range through a custom-built Printed Circuit Board that employed a modified version of the circuit described in [18]. Another consideration is the reflectivity of the object being measured. A removable white sticker was affixed on the top of the sustain pedal in order to reflect enough light for the measurement to be robust. With this configuration,

the output voltage of the circuit is proportional to the incoming light and roughly follows the inverse square of the pedal-sensor distance.

The output of the circuit was then recorded at 22.05 kHz sampling rate using the analogue input of Bela[1], which is an open-source embedded system based on the BeagleBone Black single-board computer [19]. We opted for using this system because of the need to synchronously capture audio and sensor data using a high sampling rate and resolution. The Bela platform provides stereo audio input and output, plus several I/O channels with 16-bit analogue-to-digital converters (ADC) and 16-bit digital-to-analogue converters (DAC) for attaching sensors and/or actuators. It combines the resources and advantages of embedded Linux systems with the performance and timing guarantees typically available only in dedicated digital signal processing chips and microcontrollers. Consequently, Bela integrates audio processing and sensor connectivity in a single high-performance package for our use. These are the main reasons for choosing Bela, rather than other hybrid microcontroller-plus-computer systems, which typically impose limited sensor bandwidth and may introduce jitter between sensor and audio samples. Therefore, using our system shown in Fig. 1, the piano sound can be simultaneously recorded at 44.1 kHz on the recorder in a high quality and then fed through to the audio input of Bela. Finally both the sensor and audio data were captured with the same master clock and logged into the internal memory of Bela.

## 4 DATABASE CONSTRUCTION

The measurement system described in Sec. 3 was deployed on the sustain pedal of a Yamaha baby grand piano situated in the MAT studios at Queen Mary University of London. Ten well known excerpts of Chopin's piano music were selected to form our dataset. These pieces were chosen because of the expressive nature of Chopin's compositions, as well as because Chopin was among the first composers to consistently call for the use of pedals in piano pieces. A pianist was asked to perform the excerpts using music scores provided by the experimenter. Pedal onset and offset times were marked in several versions of Chopin's published scores. We adopted the version that most publishers accept. In these scores the pedal markings always coincide with the phrase markings. When the sustain pedal is pressed, the suggested pedal depth was also notated by the experimenter. This was roughly in accordance with the dynamics changes and metric accents, since more notes will remain sounding when the key is released in case the sustain pedal is pressed at a deeper level.

Since different techniques may not be used in equal proportion in real world performances, there was no intended coverage of the four different levels of pedal depth. Consequently the number of instances of each pedaling technique in the music excerpts we recorded remains unbalanced as can be observed in Table 1.

---

[1] http://bela.io

Table 1. Number of pedaling instances in the music excerpts from our database.

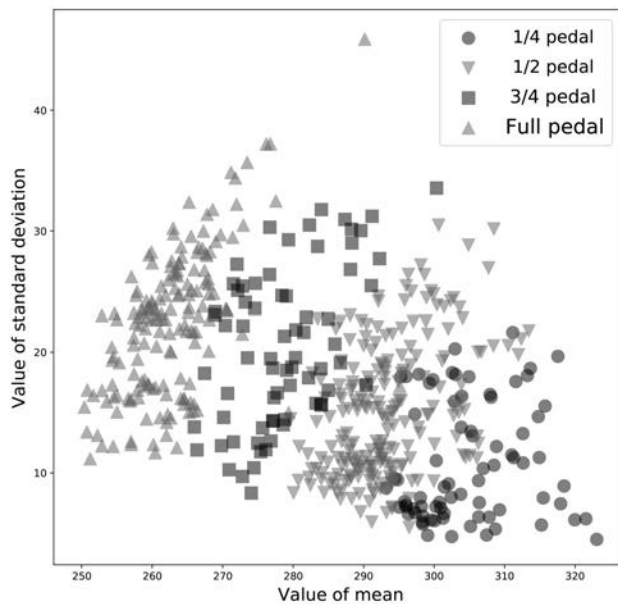| Music Excerpts | 1/4 | 1/2 | 3/4 | Full Pedal |
|---|---|---|---|---|
| Op.10 No.3 | 14 | 13 | 7 | 5 |
| Op.23 No.1 | 7 | 17 | 8 | 29 |
| Op.28 No.4 | 17 | 24 | 5 | 24 |
| Op.28 No.6 | 9 | 27 | 5 | 17 |
| Op.28 No.7 | 2 | 10 | 3 | 1 |
| Op.28 No.15 | 7 | 34 | 4 | 22 |
| Op.28 No.20 | 9 | 12 | 11 | 17 |
| Op.66 | 6 | 21 | 10 | 11 |
| Op.69 No.2 | 2 | 15 | 10 | 24 |
| B.49 | 3 | 51 | 8 | 17 |
| Sums | 76 | 224 | 71 | 167 |



Fig. 2. Scatter plot of the value of Gaussian parameters calculated from pedaling instances.

The gesture data were labelled frame by frame according to the notated scores to obtain a basic ground truth dataset. In order to evaluate to what extent the pianist followed the instructions provided in the scores, we computed descriptive statistics, visualized the data, and examined how well it matched the notation.

We first merged the frames that were consecutively labeled with the same pedaling technique into one segment. For the purpose of representing pedaling techniques we opted for using statistical aggregates of the sensor data in each segment. It was observed that the data in each segment fitted the normal distribution. Therefore Gaussian parameters were extracted to characterize the pedaling technique used within each segment. Fig. 2 presents the value of the parameters for each pedaling instance. We can observe fairly well defined clusters within the data with respect to pedal markings and also observe that the clusters are approximately linearly separable with the exception of half and quarter pedal. We also examined the consistency of pedal use with the markings and confirmed that the in-
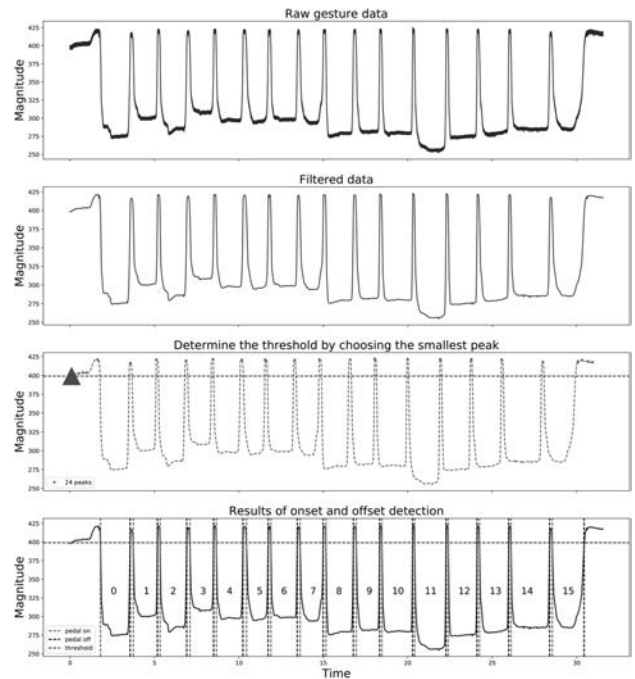


Fig. 3.   Onset and offset detection.

terpretation of the pianist was largely consistent with the pedaling notations provided by the experimenter.

## 5  PEDALING RECOGNITION

Given the dataset discussed in the previous section, our task is to recognize *when* and *which* pedaling technique were employed using the gesture data. "When" refers to the pedal onset and offset times, which can be detected using signal processing algorithms. "Which" refers to the level or class of pedal depth. We aim to classify this into quarter, half, three-quarters or full pedal technique. As we mentioned in Sec. 2, pianists vary their use of pedaling techniques with the music piece and/or the characteristics of the performance venue. This requires automatic adaptation to how a technique is used in a particular venue or by a particular musician. Manually setting the thresholds to classify the level of part-pedaling is therefore inefficient. We decided to use supervised learning methods to train SVM or HMM classifiers in a data-driven manner. To this end, we employed the scikit-learn [20] and hmmlearn[2] libraries to construct our SVM and HMM separately. In Sec. 5.2 we introduce SVM and HMM and discuss the rationale for choosing them as classifiers.

### 5.1  Onset and Offset Detection

Fig. 3 presents the process of segmenting the pedal data using onset and offset detection. The value of raw gesture data represents the position changes of the sustain pedal. The smaller the value the deeper the pedal was pressed. The Savitzky-Golay filter was used to smooth the raw data. It is a particular type of low-pass filter well-adapted for
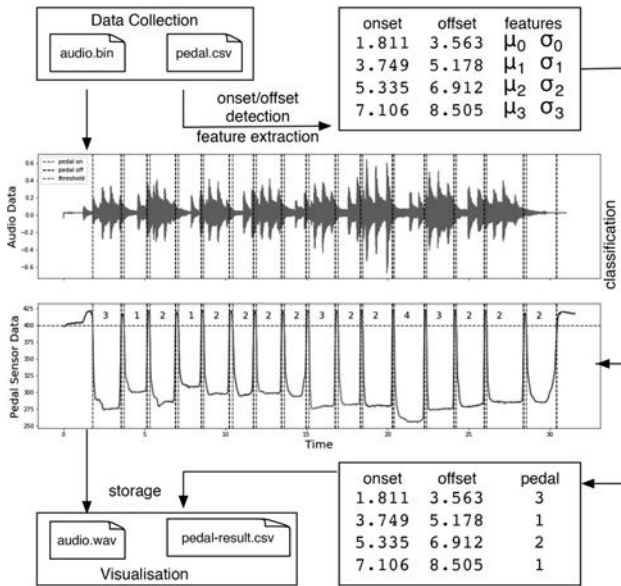
---

[2] http://hmmlearn.readthedocs.org

Fig. 4.   Classification process.

smoothing noisy time series data [21]. The Savitzky-Golay filter has the advantages of preserving the features of the distribution such as maxima and minima, which are often flattened by other smoothing techniques such as moving average or simple low-pass filtering. Thus it is often used to process time series data collected from sensors such as electrocardiogram processing [22]. Furthermore, filtering could avoid spurious peaks in the signal, which would lead to the false detection of pedaling onsets or offsets. Using the filtered data, pedaling onset and offset times were detected by comparing the data with a threshold (horizontal dashed line). This threshold is selected by choosing the minimum value from a peak detection algorithm, i.e., the smallest peak (represented by the triangle). The moment when the value of data crosses the threshold with a negative slope is considered as the onset time, while positive slope indicates the offset time. In this manner, each segment was defined by data between the onset time and its corresponding offset time. For example, there are 16 segments detected in Fig. 3.

### 5.2  Classification

Fig. 4 illustrates the overall classification procedure. After we defined the segments by the gesture data between the detected onset and offset times, Gaussian parameters were extracted from every segment to aid classification. This was motivated by the observation that the data in each segment largely fits the normal distribution as we discussed in Sec. 4. Using statistical aggregates as features can not only reduce the dataset size and improve computational efficiency, but also enable to focus on higher level information that represents each instance of pedal use. The statistical features used as input to the classifier were computed based on the Gaussian assumption and parametrised by Eq. (1), where $\mu$ is mean of the distribution and $\sigma$ is standard deviation.

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad (1)$$

We exploited SVM and HMM separately using the extracted features to classify the detected pedaling segments. A subset of our dataset was then used to train the classifiers in order to output the labels of pedaling techniques. Label number 1 to 4 correspond to the quarter, half, three-quarters, and full pedaling technique. Despite pedal position is measured in a continuous space, classification of pedaling as discrete events coincides with the interpretation by pianists and may benefit applications such as transcription and visualization, where discrete symbols corresponding to a recognized or intended technique are easier to read than a continuous pedal depth curve. The recognition results remained synchronized with the audio data. These were then used as the inputs of our visualization application discussed in Sec. 6.

The SVM algorithm was chosen because it was originally devised for classification problems that involve finding the maximum margin hyperplane that separates two classes of data [23]. If the data in the feature space are not linearly separable, they can be projected into a higher dimensional space and converted into a separable problem. For our SVM-based classification, we compared SVMs with different kernels and parameters in order to select one with the best discriminative capacity to categorize the extracted aggregate statistical features into pedaling techniques. SVM essentially learns an optimal threshold for classification from the features in training data, avoiding the use of heuristic threshold and may also account for possible non-linearities in the data.
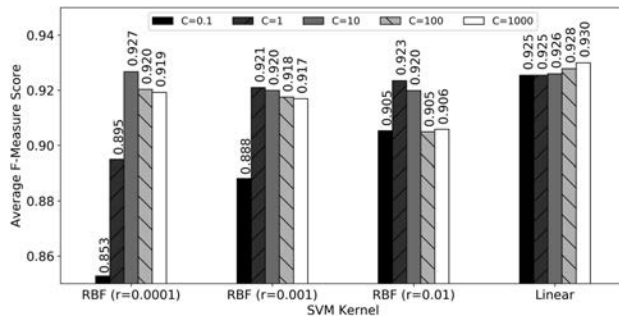
The second method we employed was HMM-based classification. HMM is a statistical model that can be used to describe the evolution of observable events that depend on hidden variables that are not directly observable [24]. In our framework the observations are the features from gesture data and the hidden states are the four pedaling techniques to be classified. In our dataset, which consists of Chopin's music, the levels of pedal depth among the segments were changed constantly. We assumed that learning the transition probability of the hidden states could reveal musicological meanings in terms of the extensive use of part-pedaling techniques for an expressive performance. The structure of our HMM was designed as a fully connected model with four states, where states may exhibit self transition or transition into any of the three other states. Gaussian emissions were used to train the probabilistic parameters. Our HMM-based classification was done by finding the optimal state sequence associated with the given observation sequence. The hidden state sequence that was most probable to have produced a given observation sequence can be computed using Viterbi decoding.

### 5.3  Results

Our ground truth dataset discussed in Sec. 4 contains labels for the pedal depth denoting the pedaling technique employed within each segment where the pedal is used. The performance of the classifiers were compared using this dataset by conducting *leave-one-group-out* cross-validation. This method is different from *leave-one-out*

Table 2. F-measure score of SVM and HMM.

| Music Excerpts | HMM F-score | SVM F-score |
|---|---|---|
| Op.10 No.3 | 0.744 | 0.969 |
| Op.23 No.1 | 0.902 | 0.924 |
| Op.28 No.4 | 0.914 | 0.976 |
| Op.28 No.6 | 0.759 | 0.959 |
| Op.28 No.7 | 0.688 | 0.893 |
| Op.28 No.15 | 0.627 | 0.943 |
| Op.28 No.20 | 0.816 | 0.906 |
| Op.66 | 0.938 | 0.970 |
| Op.69 No.2 | 0.804 | 0.881 |
| B.49 | 0.823 | 0.879 |
| Mean | 0.801 | 0.930 |

Table 3. Average F-measure scores of different machine learning techniques.

| | KNN | GNB | DT | RF | SVM |
|---|---|---|---|---|---|
| LTGO | 0.916 | 0.910 | 0.905 | 0.910 | 0.925 |
| SSS | 0.941 | 0.926 | 0.930 | 0.944 | 0.945 |



Fig. 5. Performance of SVM classifiers with different kernels (RBF and linear) and parameters ($\gamma$ and $C$).



Fig. 6. Normalized confusion matrix.

cross-validation, which is more commonly applied in the field of music information retrieval. In the leave-one-group-out scheme, samples were grouped in terms of music excerpts. Classifiers were validated in each music excerpt where the data need to be classified, while the rest of the excerpts constitute the training set. Fig. 5 presents the average F-measure scores for SVM classifiers with different kernels and parameters. The highest score was achieved by a linear-kernel SVM with the penalty parameter $C = 1000$. This largely confirms that the pedaling data for most pieces is linearly separable in the feature space we employed. We adopted this SVM model and compared it with HMM. Table 2 shows the F-measure scores of the evaluation. We can observe that SVM outperformed HMM in every music excerpt, while a mean F-measure score of 0.801 and 0.930 was obtained for the HMM and SVM respectively.

We hypothesize that the lower score of the HMM is resulting from the fact that it was trained in a non-discriminative manner. The HMM parameters were estimated by applying the maximum likelihood approach using the samples from the training set and disregarding the rival classes. Furthermore, a causality of one pedaling technique being followed by a certain other one may be unnecessary or adds very little value when the individual pedal events are separated from each other by long offset phases. For this reason the learning criterion was not related to factors that may yield an improvement of the recognition accuracy directly. While this does not allow us to dismiss potential dependencies between pedaling techniques, our simple
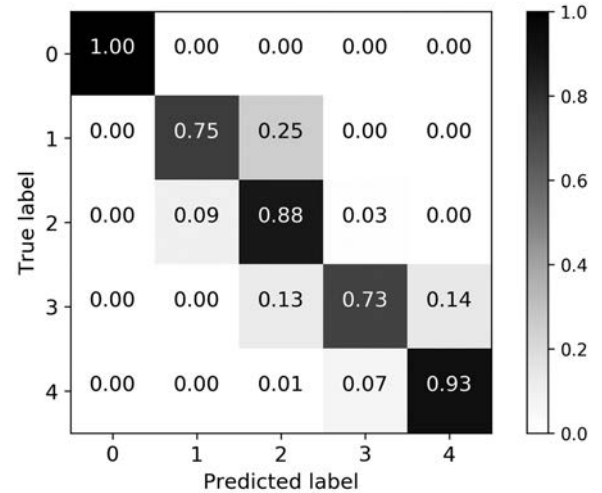
HMM model was not able to capture and exploit such dependencies. The reported results can possibly be improved using the hidden Markov SVM proposed in [25] as a discriminative learning technique for labeling sequences based on the combination of the two learning algorithms. Alternative or richer parametrization of the data instead of Gaussian parameters may also benefit the classification.

To take a detailed look at the SVM-based classification, we present a confusion matrix showing the cross-validation results with the highest average F-measure score in Fig. 6. It can be observed that the ambiguities between adjacent pedaling techniques can lead to misclassification. In most cases however, pedaling techniques can be discriminated from one another well. To avoid a potential over-fitting problem that the leave-one-group-out scheme may cause, we checked the results with two other cross-validation strategies, namely, leave-three-group-out (LTGO) and 10-iteration stratified shuffle split (SSS). For this, the test size was set to 0.3. The SVM model shows a mean F-measure score of 0.925 and 0.945 for these two strategies separately. The scores were also higher than the results using a range of common machine learning techniques we tested, including K-Nearest Neighbours (KNN), Gaussian naive Bayes (GNB), decision tree (DT), and random forest (RF). The average F-measure scores of these techniques obtained from the LTGO and SSS cross-validation are presented in Table 3.

## 6 USE CASES

### 6.1 Visualization

In order to demonstrate a practical application of our study, a piano pedaling visualization application was
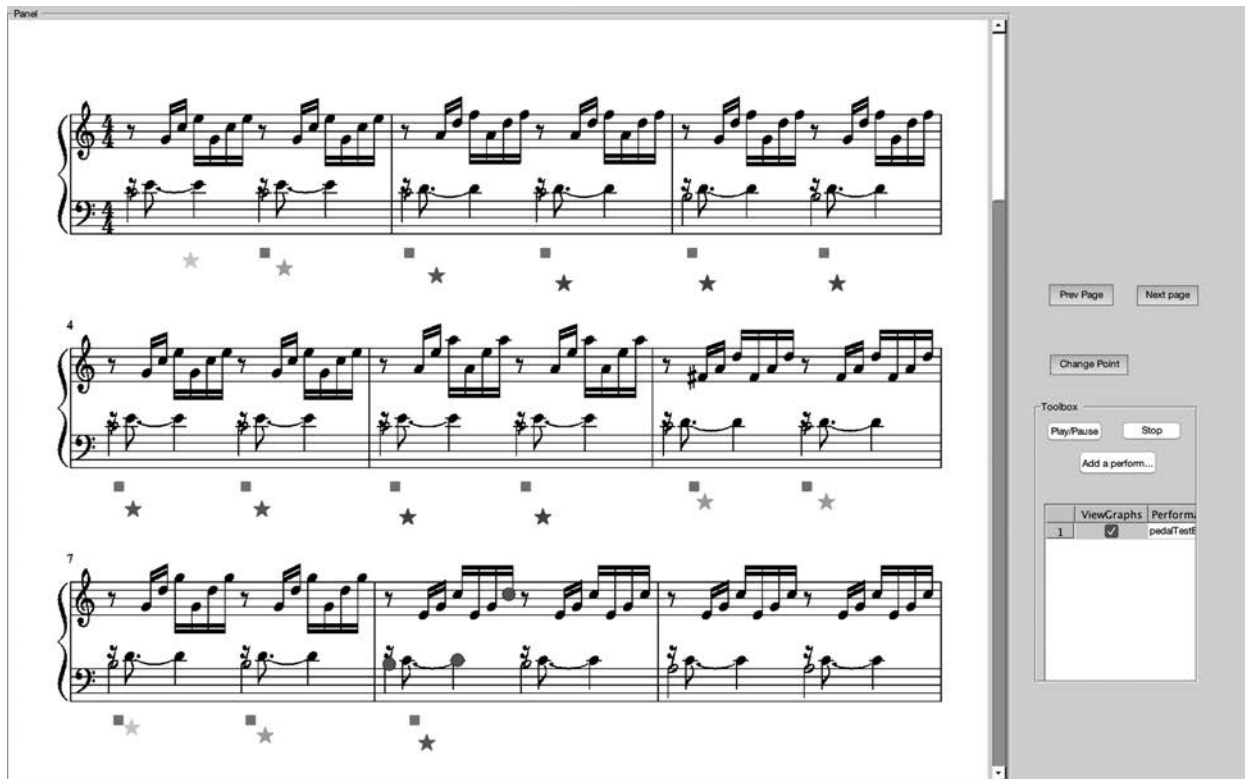
Fig. 7.   Screen shot of the visualization system.

developed that can present the recognition results in the context of the music score. This may be useful, for instance, in piano pedagogy or practice as well as musicological performance studies. We devised a simple notation system for pedaling that indicates pedal depth and timing. The application employed a score following implementation [26] implemented in Matlab, which aligns the music score with the audio recording of the same piece. Asynchronies between the piano melody and the accompaniment were handled by a multi-dimensional variant of the dynamic time warping (DTW) algorithm in order to obtain better alignments. We extended this implementation to align the pedaling recognition results of the same piece, given the detected onset and offset times and the classified pedaling technique. A screen shot of this system is shown in Fig. 7. The graphical user interface (GUI) allows the user to select a music score first. After importing the audio recording and the corresponding pedaling recognition results, they can be displayed by clicking the Play/Pause button. The GUI used the following markups for display purposes: circles show what notes in the score are sounding aligned with the audio; stars indicate pedal onsets while squares indicate pedal offset. Four different levels of color saturation plus the vertical location of the star delineate the four pedaling techniques. The levels are increased with the recognized pedal depth class.

The recognition and score alignment are completed offline so that our visualization application allows the player to review the pedaling techniques used in a recording. This could be used in music education, for instance, guiding students how to use the pedals in practice after class. We

obtained only informal feedback on the application so far. It was suggested that the visualization should be implemented as a real-time application to enable its use during live piano performance. This could also be used to trigger other visual effects in the performance, as pedaling is partly related to music phrasing. Because of the relatively high latency of the Matlab GUI, it was also recommended to implement our application using another platform.

## 6.2 Ground Truth Dataset for Audio-Based Pedaling Detection

Detection of pedaling techniques from audio recordings is necessary in the cases where installing sensors on the piano is not practical. Our measurement system is portable and easy to set up on any piano, therefore the techniques introduced in this study can be used to capture ground truth datasets for the development of pedaling recognition algorithms from audio alone. Thereafter, recognition can be done by learning a statistical model with the multi-modal data we collected from piano performances. No sensors should be required once the detection system is trained, i.e., onset and offset times plus pedal depth may be expected to be returned from audio only. This could help to analyze existing as well as historical recordings. We have exploited useful acoustic features and implemented the detection using isolated notes as a starting point in [27]. Our present work is dealing with pedaling detection in the context of polyphonic music. The measurement system presented here can also provide ground truth data for this work.

## 7 CONCLUSION

We presented a method for recognizing piano pedaling techniques on the sustain pedal using gesture data measured by a dedicated sensor system. The temporal locations of pedaling events were identified using onset and offset detection through signal processing methods. The employed pedaling technique was then recognized using supervised machine learning based classification. SVM- and HMM-based classifiers were trained and compared to assess how well we can separate the data into quarter, half, three-quarters or full pedal techniques. In our evaluation, SVM outperformed the HMM-based method and achieved an average F-measure score of 0.930. A practical use case was exemplified by our visualization application, where the recognition results are presented together with the corresponding piano recording in a score following system. A dataset was also created that can provide ground truth for related research. Our future work includes the development of audio-based pedaling detection algorithms. Techniques in this study can contribute to providing the ground truth dataset to test recognition algorithms designed to work from the audio alone. Evaluation of the visualization system has not yet been conducted with users. This also constitutes future work.

## 8 ACKNOWLEDGMENT

## 9 REFERENCES

[1] W. Goebl, S. Dixon, G. De Poli, A. Friberg, R. Bresin, and G. Widmer, "Sense in Expressive Music Performance: Data Acquisition, Computational Studies, and Models," Sound to Sense - Sense to Sound: A State of the Art in Sound and Music Computing, pp. 195–242 (2008).

[2] D. Rowland, *A History of Pianoforte Pedalling* (Cambridge University Press, 2004).

[3] S. P. Rosenblum, "Pedaling the Piano: A Brief Survey from the Eighteenth Century to Present," *Performance Practice Rev.*, vol. 6, no. 2, p. 8 (1993), https://doi.org/10.5642/perfpr.199306.02.08.

[4] E. Chew and A. R. François, "MuSA. RT and the Pedal: The Role of the Sustain Pedal in Clarifying Tonal Structure," *Proceedings of the 10th International Conference on Music Perception and Cognition* (2008).

[5] K. U. Schnabel, *Modern Technique of the Pedal: A Piano Pedal Study* (Mills Music, 1954).

[6] D. R. Sinn, *Playing Beyond the Notes: A Pianist's Guide to Musical Interpretation* (Oxford University Press, 2013).

[7] H.-M. Lehtonen, H. Penttinen, J. Rauhala, and V. Välimäki, "Analysis and Modeling of Piano Sustain-Pedal Effects," *J. Acoust. Soc. Amer.*, vol. 122, no. 3, pp. 1787–1797 (2007), https://doi.org/10.1121/1.2756172.

[8] H.-M. Lehtonen, A. Askenfelt, and V. Välimäki, "Analysis of the Part-Pedaling Effect in the Piano," *J. Acoust. Soc. Amer.*, vol. 126, no. 2, pp. EL49–EL54 (2009), https://doi.org/10.1121/1.3162438.

[9] O. Ortmann, *The Physical Basis of Piano Touch and Tone* (Kegan Paul, Trenc, Trubner & Co., London, 1925).

[10] J. MacRitchie, "The Art and Science behind Piano Touch: A Review Connecting Multi-Disciplinary Literature," *Musicae Scientiae*, vol. 19, no. 2, pp. 171–190 (2015), https://doi.org/10.1177/1029864915572813.

[11] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser, "The Elbow Piano: Sonification of Piano Playing Movements," *Proceedings of the 8th International Conference on New Interfaces for Musical Expression (NIME)*, pp. 285–288 (2008).

[12] M. Bernays and C. Traube, "Investigating Pianists' Individuality in the Performance of Five Timbral Nuances through Patterns of Articulation, Touch, Dynamics, and Pedaling," *Frontiers in Psychology*, vol. 5, no. 157 (2014), https://doi.org/10.3389/fpsyg.2014.00157.

[13] PianoBar, "Products of Interest," *Computer Music J.*, vol. 29, no. 1, pp. 104–114 (2005).

[14] A. McPherson and Y. Kim, "Piano Technique as a Case Study in Expressive Gestural Interaction," in *Music and Human-Computer Interaction*, pp. 123–138 (Springer, 2013), https://doi.org/10.1007/978-1-4471-2990-5_7.

[15] B. Caramiaux and A. Tanaka, "Machine Learning of Musical Gestures," *Proceedings of the 13th International Conference on New Interfaces for Musical Expression (NIME)*, pp. 513–518 (2013).

[16] V. Zandt-Escobar, B. Caramiaux, and A. Tanaka, "PiaF: A Tool for Augmented Piano Performance Using Gesture Variation Following," *Proceedings of the 14th International Conference on New Interfaces for Musical Expression (NIME)*, pp. 167–170 (2014).

[17] B. Liang, G. Fazekas, A. McPherson, and M. Sandler, "Piano Pedaller: A Measurement System for Classification and Visualisation of Piano Pedalling Techniques," *Proceedings of the 17th International Conference on New Interfaces for Musical Expression (NIME)*, pp. 325–329 (2017).

[18] A. McPherson, "Portable Measurement and Mapping of Continuous Piano Gesture," *Proceedings of the 13th International Conference on New Interfaces for Musical Expression (NIME)*, pp. 152–157 (2013).

[19] A. McPherson and V. Zappi, "An Environment for Submillisecond-Latency Audio and Sensor Processing on BeagleBone Black," presented at the *138th Convention of the Audio Engineering Society* (2015 May), convention paper 9331, http://www.aes.org/e-lib/browse.cfm?elib=17755.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and

E. Duchesnay, "Scikit-learn: Machine Learning in Python," *J. Machine Learning Res.*, vol. 12, pp. 2825–2830 (2011), https://dl.acm.org/citation.cfm?id=1953048.2078195.

[21] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, and P. B. Kramer, *Numerical Recipes: The Art of Scientific Computing* (AIP, 1987).

[22] K. Pandia, S. Ravindran, R. Cole, G. Kovacs, and L. Giovangrandi, "Motion Artifact Cancellation to Obtain Heart Sounds from a Single Chest-Worn Accelerometer," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 590–593 (2010), https://doi.org/10.1109/ICASSP.2010.5495553.

[23] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification* (Wiley-Interscience, 2000).

[24] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286 (1989), https://doi.org/10.1109/5.18626.

[25] Y. Altun, I. Tsochantaridis, and T. Hofmann, "Hidden Markov Support Vector Machines," *Proceedings of the 20th International Conference on Machine Learning (ICML)*, vol. 3, pp. 3–10 (2003), https://dl.acm.org/citation.cfm?id=3041838.3041839.

[26] S. Wang, S. Ewert, and S. Dixon, "Compensating for Asynchronies between Musical Voices in Score-Performance Alignment," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 589–593 (2015), https://doi.org/10.1109/ICASSP.2015.7178037.

[27] B. Liang, G. Fazekas, and M. Sandler, "Detection of Piano Pedalling Techniques on the Sustain Pedal," presented at the *143rd Convention of the Audio Engineering Society* (2017 Oct.), convention paper 9812, http://www.aes.org/e-lib/browse.cfm?elib=19209.

## THE AUTHORS

Beici Liang                Győrgy Fazekas                Mark Sandler

Beici Liang is a Ph.D. candidate at Media & Arts Technology Centre for Doctoral Training and Centre for Digital Music at Queen Mary University of London, UK. She received her Bachelor of Engineering in integrated circuit design and integrated system from Tianjin University, China, in 2014. Since then, she started her doctoral research on analysis and modelling of instrumental gestures. She is also interested in piano acoustics and music information retrieval.

•

Győrgy Fazekas is a lecturer at Queen Mary University of London (QMUL), UK, working at the Centre for Digital Music (C4DM), School of Electronic Engineering and Computer Science. He received his B.Sc. and M.Sc. in electrical engineering and subsequently completed his Ph.D. at QMUL in 2012 focussing on semantic audio analysis, se-

mantic web technologies and ontology based information management for audio applications. He is leading QMUL's research team on the EU funded Audio Commons project facilitating the use of open sound content in professional audio production. He has published over 100 papers, collaborates on several research projects, and he is a member of the IEEE, AES, and ACM.

•

Mark Sandler was born in 1955. He received the B.Sc. and Ph.D. degrees from the University of Essex, UK, in 1978 and 1984 respectively. He is a Professor of signal processing and the Founding Director of the Centre for Digital Music in the School of Electronic Engineering and Computer Science at Queen Mary University of London, UK. He has published over 400 papers in journals and conferences. He is a Fellow of the Royal Academy of Engineering, IEEE, AES, and IET.