Machine Learning-Driven Detection and Visualization of Sustain Pedal Technique for Piano Pedagogy

Vincent Cao

Cedar Park High School,
Cedar Park-78613, United States of America

Abstract. The sustain pedal is a vital component of piano performance, adding warmth and resonance to the instrument's sound. As a result, a pianist's pedaling choices can significantly alter the perceived quality of their music. Studying the choices that professionals make when approaching the sustain pedal is a valuable method for students seeking to improve their playing, as mastering the technique is one of the most challenging components of a performance. However, pedal alterations are often difficult to distinguish by ear, since human perception of sustainment can be inconsistent and vary based on recording quality and piano tone. Thus, this project presents a novel methodology for analyzing piano performances through machine learning-driven pedal detection and visualization. The system accepts audio recordings of a student and reference as inputs, processes them using signal analysis techniques to identify instances of sustain pedal activation and release, and outputs these events onto time-aligned graphs. By visualizing pedal activity alongside performance timing, the project facilitates quantitative comparison of pedaling technique between the student and reference, offering applications in music education, performance evaluation, and expressive analysis. This approach aims to provide pianists, teachers, and researchers with an objective, data-driven perspective on an often-subjective element of piano education.

Keywords: Sustain pedal, digital signal processing, music analysis, music information retrieval, piano pedagogy.

1 Introduction

In both modern and classical piano performance, the sustain pedal is one of the most expressive components available to pianists, shaping tone, resonance, and phrasing in ways that are essential to musical interpretation. In his treatise on piano technique, renowned pianist Heinrich Neuhaus remarks that "one cannot speak of the pedal separately from tone," [10] arguing that the two qualities are inextricably linked together. Gieseking and Leimer [2] similarly recognize this idea, emphasizing the dramatic tonal effects created by slight variations in timing and depth of pedal.

Additionally, pedaling has also received widespread coverage in scientific studies about piano performance. Thompson et al. [12] presented an analysis of how different reviewers rate piano performances using the Personal Construct Theory [4]. The adjudicators were observed to rate piano performances based on various characteristics, including pedaling, as well as overall expression, phrasing, dynamics, rubato, and tempo. Russell [11], on the other hand, conducted a study evaluating piano recordings using the Piano Performance Evaluation Rating Scale, which included tone/pedaling, tempo, rhythm, articulation, technique, interpretation, dynamics, and memory.

While listeners and performers often evaluate pedaling subjectively, there is a growing interest in computational methods that enable objective, quantitative analysis of this technique. Pedal timing, duration, and coordination with note events influence both clarity of harmonic textures and emotional impact, making them measurable dimensions of music. At the same time, mastering the pedal remains one of the most essential yet elusive

aspects of performance. Unlike notes on the page, which can be learned with repetition and technical precision, the use of the pedal resists strict codification. It depends on depth, timing, and coordination with the hands to shape the resonance and color of the sound. Even subtle differences in timing—lifting a fraction of a second earlier or later—can mean the difference between clarity and muddiness. The right balance of depth must also be practiced tirelessly, since too much can blur harmonies while too little leaves the sound dry.

Thus, this paper presents a machine learning-based system for comparing pedal usage between students and professionals by visualizing their data in a time-aligned format. This paper details the project's methodology, features, and pedagogical value for music education.

1.1 Computational Analyses of Pedal

In the field of music information retrieval, studies have often pursued sustain pedal detection as a binary classification task, simply determining whether the pedal is on or off [5–7, 13, 14]. However, this assumption overlooks the nuances of actual pedaling, as purposeful usage of techniques such as half-pedals and flutter pedaling can be integral to piano performance yet are often neglected by simple binary algorithms.

To remedy this, Fang et al. [1] presented a machine learning-based solution for pedal analysis. Their high-resolution sustain pedal depth detection model is capable of performing both binary classification and determining the depth with which the pedal is pressed. The model employed in this study adopts a three-stage design. First, a convolutional block comprising three two-dimensional convolutional layers with batch normalization, rectified linear unit (ReLU) activation, and max pooling along the frequency axis is applied to log-mel spectrogram inputs. This step compresses spectral information and yields high-level feature representations. Second, the resulting sequence is passed to an eight-layer Transformer encoder with a hidden size of 256, eight attention heads, and a feed-forward dimension of 1024. The encoder captures temporal dependencies across frames, thereby extending the network's ability to model context beyond local convolutional features. Finally, prediction layers in the form of four distinct multilayer perceptrons (MLPs) are used to generate task-specific outputs, enabling 4-class classification of frame-wise pedal depth, onset, offset, and a global pedal depth estimate. Frame-level predictions retain the original sequence length (500 frames), while global prediction is achieved via mean pooling over the encoded representation. To enhance generalization, training incorporates a dropout rate of 0.15.

This model was trained on the MAESTRO dataset [3], which comprises hundreds of piano recordings from various genres. Though its binary accuracy is comparable to previous pedal detection models, such as those from [5] and [14], its F1 score was found to be substantially higher than other models for 4-class classification. Thus, this project primarily focuses on the model by Fang to provide a better analysis of pedal depths across performances.

In addition to architectural advances, recent studies have explored multi-task learning frameworks that integrate regression and classification objectives to address diverse prediction targets in temporal audio modeling. Such approaches are particularly effective when tasks involve both continuous control signals and discrete event boundaries. In [1], the model adopts a composite loss function that combines frame-wise mean squared error (MSE) for pedal depth estimation, MSE for global pedal depth averaging, and binary cross-entropy (BCE) for detecting pedal onsets and offsets. This formulation enables the network to simultaneously capture fine-grained continuous variations in pedal control as

well as higher-level event boundaries. By manually weighting the individual loss components, the framework balances the contributions of low-level regression and high-level classification objectives, ultimately enabling robust performance across complementary prediction tasks.

This paper applies the processes created by Fang to enable a practical, machine learning-driven approach to extracting and visualizing pedal usage within performances. Specifically, its novelty lies in enabling direct comparisons between expert and student pedaling, allowing students to learn and model from professionals with years of experience performing the piano.

2 Pedal Analysis

2.1 Simplifying Assumptions

The program requires input from two audio files: one of the student (X_1) and one from the reference pianist (X_2). For this paper, both recordings are of Beethoven's Piano Sonata No. 30 in E major, Op. 109, I. Vivace ma non troppo—Adagio espressivo. The student recording was taken during practice, while the reference is a performance by Japanese pianist Mitsuko Uchida [9]. All audio files are assumed to contain the complete sonata movement without major, unexpected pauses. While clean audio is preferred, the program remains compatible with recordings from a smartphone or a basic recording setup.

2.2 Pre-Preprocessing

To process the recordings, each audio file is first resampled to 16 kHz, sufficiently compressing the audio signal while still being able to capture the piano's highest pitch—C8, 4186 Hz—as described by Kong et al. [5]. Each file is then converted to log-mel spectrograms with 229 frequency bands, utilizing a short-time Fourier transform with a 2048-sample Hann window size and a 160-sample hop size, yielding a frame rate of 100 frames per second (FPS). In addition, 20 MFCC coefficients were extracted and vertically concatenated onto the spectrograms. Finally, each performance file is split into segments of 500 frames, yielding approximately 5 seconds of audio per clip, assuming a frame rate of 100 FPS. Each extracted segment has a shape of $[T, F]$, where $T = 500$ and refers to the number of frames per segment, while $F = 249$ and is the feature dimension of the segment obtained from concatenating the spectrogram and the MFCC coefficients. All pre-processing steps were conducted using Librosa [8], a prominent library for audio analysis. Additionally, the finished feature arrays were stored as HDF5 files in accordance with the input required by the pedal detection model.

2.3 Computation of the Pedal Graph

Following the pre-processing steps, the model was used to infer the pedal data of each input recording. The model's output consists of four parts: global pedal depth (a single numerical value indicating the average pedal depth of an audio segment), pedal onset (a binary sequence where a value of 1 indicates the pedal is pressed down), pedal offset (a binary sequence where a value of 1 indicates the pedal is released), and pedal depth (a continuous sequence indicating the frame-wise depth at which the pedal is pressed, normalized to a range of $[0, 1]$).

For each recording, the model's inference results were overlaid on top of each other and graphed together to provide a visual representation of pedaling. To provide a clearer frame

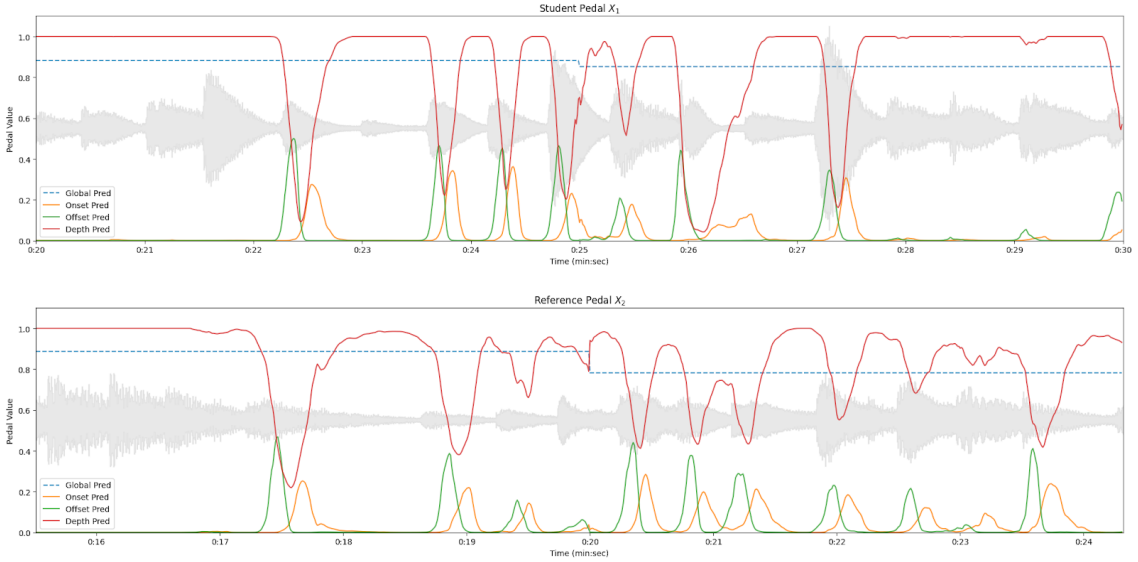


Fig. 1: Pedal graphs indicating pedal depth over time for the student recording X_1 and the reference recording X_2 within a 10-second sample clip from X_1. The graphs are aligned with DTW such that their domains correspond to the same musical passage within the piece. The global pedal depth average for each segment of time is also graphed to provide a baseline value.

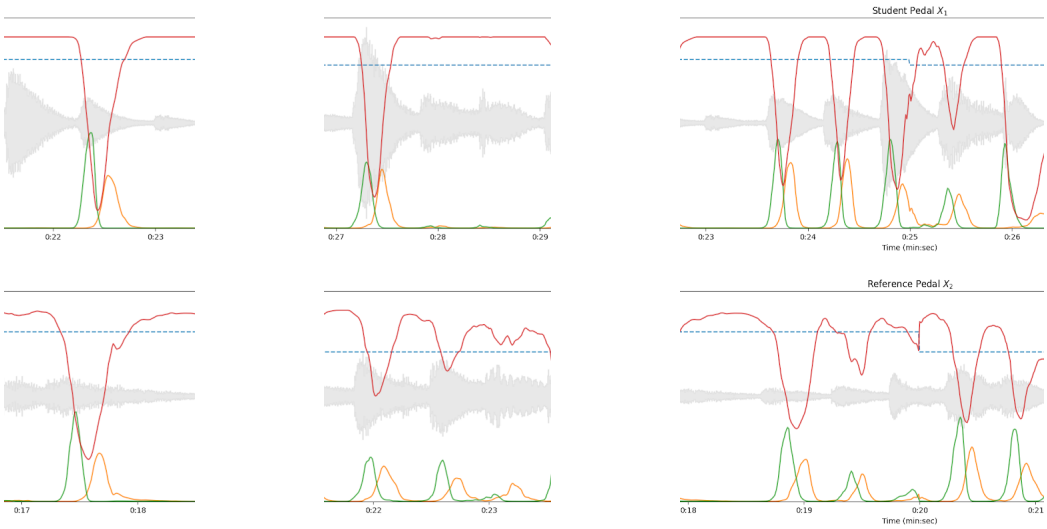
of reference, however, the two recordings were aligned with dynamic time warping (DTW) to account for mismatches in speed and rhythm between the student and reference.

To do this, Constant-Q Transform was used to extract each recording's chromagram, which represents represent the intensity of each pitch class (C, C#, D, etc.) over time, providing a robust descriptor of harmonic content. Dynamic Time Warping (DTW) was used on the chromagrams to find the best alignment path, minimizing the cumulative cosine distance between the feature matrices with Librosa's `librosa.sequence.dtw`. This process returned the cost matrix and an optimal warping path, which mapped the corresponding frames of the two recordings together. Specifically, the domains of the graph were limited such that $[t_{\min,X_1}, t_{\max,X_1}]$ and $[t_{\min,X_2}, t_{\max,X_2}]$, where each t represents the time domain of the pedal graphs. These bounds ensure that both the X_1 and X_2 graphs display the same section within the piano piece, despite potential timing differences due to speed, rhythm, etc.

3 Results and Discussion

The pedal detection model successfully identified pedal onsets, offsets, and depth values across both recordings, producing time-series curves that were subsequently aligned using DTW. The key contribution of this work is the extraction and visualization of the pedal curves along with the alignment procedure, which allows for a direct comparison of pedaling behavior between performances, despite differences in overall tempo and local timing. Fig. 1 presents the aligned pedal depth curves for recordings X_1 and X_2.

In general, the locations of pedal lifts and depressions were consistent across the two performances. For example, the first detected pedal event in X_1 at $t_1 = 22.4$ aligned with the corresponding event in X_2 at $t_2 = 17.6$, as shown in fig. 2a. This pattern of consistent alignment extended through much of the excerpt. However, discrepancies were



(a) $t_1 \in [22, 23]$, $t_2 \in [17, 18]$

(b) $t_1 \in [27, 29]$, $t_2 \in [22, 23]$

(c) $t_1 \in [23, 26]$, $t_2 \in [18, 21]$

Fig. 2: Excerpts of the original pedal curves showing time domains t_1 and t_2, where $[a, b]$ represents the bounds of each excerpt.

also observed, indicating differences between the pedal choices of the two recordings. For instance, fig. 2b indicates a pedal event was detected in X_2 at $t_2 = 22.7$ without a corresponding event in X_1.

In addition to timing, differences in pedal depth were evident. Fig. 2c indicates that X_1 exhibited substantial pedal lifts on the interval $t_1 \in [23, 26]$, reaching depths as low as 0.2. The corresponding passage in X_2 ($t_2 \in [18, 21]$) showed shallower lifts, with depth values rarely exceeding 0.4. Such differences illustrate how performers may achieve similar pedaling locations while differing in the degree of pedal motion. Deeper pedals might generate a warmer and more expressive music, while shallower pedals still provide warmth while evoking a lighter and less-cluttered sound.

Overall, the results indicate that DTW-based alignment provides a robust means of comparing pedaling strategies across performances. The method captures both consistencies in pedal timing and divergences in depth, enabling a detailed analysis of interpretive choices

4 Practical Applications

The methods presented in this work have several potential applications across music performance, pedagogy, and music information retrieval. First, the ability to detect and compare pedal usage across multiple performances provides valuable assistance for performance analysis and piano study. Pedaling is a subtle but critical expressive device in piano playing, and visualizing differences in pedal depth and timing allows performers and teachers to study interpretive choices in a systematic way. For instance, instructors can demonstrate how variations in pedal depth or release timing affect phrasing and resonance, offering students a clearer understanding of interpretive nuance.

Second, the alignment of performances via DTW enables direct comparisons even when tempo fluctuations are present. This functionality could be extended to automated feedback systems in music education, where a student's performance is aligned with a reference

recording and differences in pedaling or other expressive controls are highlighted. Such a system would provide students with actionable insights into how their pedaling differs from that of an expert, complementing traditional instruction.

Finally, this framework also contributes to broader research in music information retrieval. By combining automatic pedal detection with time-series alignment, the approach supports quantitative studies of performance practice at scale. Large corpora of recordings could be analyzed to reveal trends in pedaling across eras, styles, or performers, offering empirical evidence for questions traditionally studied only qualitatively in musicology. Moreover, the visualization methods outlined here may serve as a bridge between computational analysis and traditional score-based annotation, facilitating the integration of performance data into scholarly editions of musical works.

5 Conclusion

This work demonstrates a machine learning-driven framework for analyzing pedaling strategies in piano performance through the combination of automatic pedal detection, dynamic time warping, and visualization. By aligning recordings of the same piece, it was able to directly compare pedal timing, duration, and depth between performances despite differences in tempo. The results highlight both consistencies, such as shared pedal placements across aligned passages, and divergences, including variations in pedal depth and the presence or absence of certain pedal events. The machine learning aspect greatly facilitates analysis, as it automates the otherwise labor-intensive process of manually identifying and annotating pedal events.

The analysis shows that DTW-based alignment is effective for mitigating tempo differences and enabling meaningful comparison of expressive parameters. Moreover, the visualization of pedal depth curves provides interpretable evidence of stylistic choices, which can be further mapped onto the musical score for use in pedagogical or analytical contexts.

While the current study focused on two recordings of a single work, the methodology generalizes to larger corpora and other expressive parameters, such as dynamics or articulation. Future research may involve synchronizing and annotating the detected pedals from this project with sheet music, as well as integrating them with automated feedback systems to provide students and teachers with new tools for performance evaluation.

Overall, this project illustrates how computational methods integrating artificial intelligence can complement traditional approaches to performance analysis, offering quantitative and visual insights into one of the most nuanced aspects of piano education.

References

1. Fang, K., Zhang, H., Wang, Z., Fujinaga, I.: High-resolution sustain pedal depth estimation from piano audio across room acoustics (2025), https://arxiv.org/abs/2507.04230, _eprint: 2507.04230
2. Gieseking, W., Leimer, K.: Piano technique. Courier Corporation (Apr 2013), google-Books-ID: xwTDAgAAQBAJ
3. Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C.Z.A., Dieleman, S., Elsen, E., Engel, J.H., Eck, D.: Enabling factorized piano music modeling and generation with the MAESTRO dataset. In: 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019. OpenReview.net (2019), https://openreview.net/forum?id=r1lYRjC9F7
4. Kelly, G.: The psychology of personal constructs. Routledge (1955)
5. Kong, Q., Li, B., Song, X., Wan, Y., Wang, Y.: High-resolution piano transcription with pedals by regressing onset and offset times. IEEE/ACM Transactions on Audio, Speech, and Language Processing **29**, 3707–3717 (2021). https://doi.org/10.1109/TASLP.2021.3121991, https://ieeexplore.ieee.org/document/9585550

6. Liang, B., Fazekas, G., Sandler, M.: Transfer learning for piano sustain-pedal detection. In: 2019 International Joint Conference on Neural Networks (IJCNN). pp. 1–6. IEEE, Budapest, Hungary (Jul 2019). https://doi.org/10.1109/IJCNN.2019.8851724, https://ieeexplore.ieee.org/document/8851724/
7. Liang, B., Fazekas, G., Sandler, M.: Piano sustain-pedal detection using convolutional neural networks. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 241–245 (May 2019). https://doi.org/10.1109/ICASSP.2019.8683505, https://ieeexplore.ieee.org/document/8683505, iSSN: 2379-190X
8. McFee, B., McVicar, M., Faronbi, D., Roman, I., Gover, M., Balke, S., Seyfarth, S., Malek, A., Raffel, C., Lostanlen, V., Niekirk, B.v., Lee, D., Cwitkowitz, F., Zalkow, F., Nieto, O., Ellis, D., Mason, J., Lee, K., Steers, B., Halvachs, E., Thomé, C., Robert-Stöter, F., Bittner, R., Wei, Z., Weiss, A., Battenberg, E., Choi, K., Yamamoto, R., Carr, C.J., Metsai, A., Sullivan, S., Friesch, P., Krishnakumar, A., Hidaka, S., Kowalik, S., Keller, F., Mazur, D., Chabot-Leclerc, A., Hawthorne, C., Ramaprasad, C., Keum, M., Gomez, J., Monroe, W., Morozov, V.A., Eliasi, K., nullmightybofo, Biberstein, P., Sergin, N.D., Hennequin, R., Naktinis, R., beantowel, Kim, T., Åsen, J.P., Lim, J., Malins, A., Hereñú, D., Struijk, S.v.d., Nickel, L., Wu, J., Wang, Z., Gates, T., Vollrath, M., Sarroff, A., Xiao-Ming, Porter, A., Kranzler, S., Voodoohop, Gangi, M.D., Jinoz, H., Guerrero, C., Mazhar, A., toddrme2178, Baratz, Z., Kostin, A., Zhuang, X., Lo, C.T., Campr, P., Semeniuc, E., Biswal, M., Moura, S., Brossier, P., Lee, H., Pimenta, W.: Librosa (May 2024). https://doi.org/10.5281/zenodo.11192913, https://zenodo.org/records/11192913
9. Mitsuko Uchida: Beethoven Piano Sonata No. 30 in E Major, Op. 109: I. Vivace ma non troppo – Adagio (Jan 2006), https://www.youtube.com/watch?v=6vaKY5ODIRE
10. Neuhaus, H.: The art of piano playing (1958), oCLC: 1087008917
11. Russell, B.E.: The empirical testing of musical performance assessment paradigm. Ph.D. thesis, University of Miami (May 2010), https://scholarship.miami.edu/esploro/outputs/doctoral/The-Empirical-Testing-of-Musical-Performance/991031447324002976
12. Thompson, W.F., Diamond, C.T.P., Balkwill, L.L.: The adjudication of six performances of a Chopin etude: a study of expert knowledge. Psychology of Music **26**(2), 154–174 (Oct 1998). https://doi.org/10.1177/0305735698262004, https://doi.org/10.1177/0305735698262004
13. Yan, Y., Cwitkowitz, F., Duan, Z.: Skipping the frame-level: event-based piano transcription with neural semi-CRFs. In: Advances in Neural Information Processing Systems. vol. 34, pp. 20583–20595. Curran Associates, Inc. (2021), https://proceedings.neurips.cc/paper_files/paper/2021/hash/ac53fab47b547a0d47b77e424cf119ba-Abstract.html
14. Yan, Y., Duan, Z.: Scoring time intervals using non-hierarchical transformer for automatic piano transcription. In: Proceedings of the 25th International Society for Music Information Retrieval Conference. pp. 973–980. ISMIR, San Francisco, California, USA and Online (Nov 2024). https://doi.org/https://doi.org/10.5281/zenodo.14877493, https://zenodo.org/records/14877493

Authors

Vincent Cao is a current high school senior from Cedar Park High School. His research interests include signal processing, piano pedagogy, and integrating AI systems to advance music education.