Guitar Onset Detection and Playing Technique Classification:
A Comparative Parameter Optimisation Study

Patrick Gaft
SID: 520432082
GitHub: pgaft
GitHub Repository: `https://github.com/pgaft/elec5305-project-520432082`
GitHub Pages: `https://pgaft.github.io/elec5305-project-520432082/`

Contents

1 Introduction

Musical onset detection and playing technique recognition are two core problems in Music Information Retrieval (MIR), and are central to the detailed analysis of guitar performance [1]. Onset detection identifies the temporal locations at which notes begin [1, 2]. Playing technique classification determines how these notes are articulated, for example normal picking, bending, vibrato, slides, hammer ons, and pull offs [3, 4, 5]. When combined, these tasks form an essential part of automatic tablature transcription systems that aim to capture pitch, timing, and the expressive detail that characterises a guitarist's style [4, 5].

This project addresses the following research question:

> How do different parameter configurations influence the accuracy of spectral flux based onset detection and audio feature based playing technique classification for guitar recordings, and how is onset detection accuracy related to downstream technique classification performance?

In contrast to end to end transcription systems that attempt to infer pitch, timing, and technique in a single pipeline, this project focuses on the systematic evaluation of parameter choices for two well established algorithms, and uses publicly available datasets with standard evaluation metrics [1, 3]. Considering the tasks separately offers two main advantages.

1. **Independent contribution**: optimised onset detection and technique classification each provide stand alone insight for the MIR community [1, 3].

2. **Joint analysis**: relating onset detection accuracy to technique classification performance provides information about practical design trade offs in larger systems [4, 5].

The onset detection component is based on the spectral flux method implemented in `librosa` [6]. This part of the project performs a systematic exploration of window size, hop length, and threshold choice, and reports the effect of these parameters on F measure for guitar audio from the GuitarSet [7] and IDMT SMT Guitar [8] datasets.

The technique classification component uses the IDMT SMT Guitar dataset [8], which provides annotations for six playing techniques (normal, bend, vibrato, slide, hammer on, pull off). A Support Vector Machine (SVM) classifier is trained on a set of audio features including MFCCs, spectral centroid, spectral rolloff, and zero crossing rate [3].

The third component performs joint analysis. Audio is segmented using both ground truth onsets and detected onsets obtained under various parameter configurations. This allows an explicit study of how onset detection errors affect technique classification. The goal is to provide practical guidance on what onset detection performance is sufficient for downstream classification in realistic systems [4, 5].

The scope is chosen to be feasible within the course timeline. Existing libraries are used wherever possible: `librosa` for onset detection [6], and `scikit learn` for classification. Standard feature extraction pipelines are adopted. Implementation details and code are provided in a reproducible form so that the experiments can be replicated and extended.

2 Literature Review

2.1 Onset Detection Methods

Onset detection aims to identify the time instants at which musical notes begin [1, 2]. Systems typically follow a three-stage pipeline: preprocessing, construction of an onset detection function (ODF), and peak picking [2]. Several ODFs exist, targeting changes in signal properties like energy, phase, or spectral content.

Spectral flux, which measures frame-to-frame changes in the magnitude spectrum, is a widely adopted ODF that performs well for percussive and non-percussive instruments [2, 9]. The spectral flux $SF(n)$ at frame n is given by

$$SF(n) = \sum_{k=1}^{N/2} H\big(|X(n,k)| - |X(n-1,k)|\big) \tag{1}$$

where $X(n,k)$ is the magnitude spectrum at frame n and bin k, and $H(\cdot)$ is half-wave rectification [2].

Dixon [2] reported F-measures between 0.65 and 0.75 for string instruments using this method. Subsequent work has improved this. Böck and Widmer [10] introduced SuperFlux, which applies maximum filtering to reduce false positives caused by vibrato, a common issue in guitar music. More recent guitar-specific work by Mounir et al. [11] used spectral sparsity-based methods, showing significant improvements over standard spectral flux.

The choice of parameters (window size, hop length) and thresholding strategy (e.g., adaptive median vs. fixed) significantly impacts performance [9, 10]. Rosão et al. [9] found that adaptive median thresholding improved F-measure by 2-3 percentage points over fixed thresholds. Standard evaluation, such as in the Music Information Retrieval Evaluation eXchange (MIREX), uses F-measure with a ±50 ms tolerance window [12].

2.2 Guitar Playing Technique Classification

Guitar playing techniques (normal, bend, vibrato, slide, hammer-on, pull-off) introduce characteristic changes in timbre and temporal evolution that can be classified using machine learning [3, 13].

Chen et al. [3] achieved a 74

More recently, deep learning approaches have shown very high accuracy (up to 99

2.3 Acoustic Signatures and Detection Mechanisms

Each technique has a distinct acoustic signature, which informs feature design. These can be broadly grouped by their primary acoustic cue: pitch contour or attack transient.

2.3.1 Pitch-Based Techniques (Bend, Vibrato, Slide)

These techniques are defined by their continuous pitch movement after the initial onset.

- **Bends** involve a monotonic increase in pitch. Detection relies on identifying segments where the fundamental frequency $f_0(t)$ increases continuously, typically by at least 80 cents [3]. The pitch deviation $\Delta p(t)$ is:

$$\Delta p(t) = 1200 \times \log_2 \left(\frac{f_0(t)}{f_0(t_{\text{onset}})} \right) \text{ cents} \tag{2}$$

- **Vibrato** is a periodic modulation of pitch. It is detected by finding sinusoidal oscillations in the pitch contour, typically with a rate of 4-10 Hz and a depth greater than 10 cents [3].

- **Slides** are also continuous pitch shifts, but unlike bends, they often involve moving across discrete frets, which can create a "stepwise" or ladder-like pitch contour. They are often identified by a sequence of consecutive semitone steps [3].

2.3.2 Transient-Based Techniques (Hammer-On, Pull-Off)

These techniques involve two notes generated from a single pick stroke. Their detection relies on analysing the attack characteristics of the second note [14].

- **Hammer-ons** are an ascending legato event. A picked note is followed by a second, higher note produced by a finger striking the string. This second note has a much softer transient (less high-frequency energy, lower attack slope) than the first [14]. The pitch change is typically small:

$$|f_0(\text{note 2}) - f_0(\text{note 1})| \in [100, 200] \text{ cents} \tag{3}$$

- **Pull-offs** are the descending equivalent. A picked note is followed by a lower note, sounded by pulling the fretting finger off the string. This motion can impart a very slight "pluck," but the second transient is still significantly weaker than the first [14]. The pitch constraint is:

$$f_0(\text{note 2}) < f_0(\text{note 1}), \quad |f_0(\text{note 2}) - f_0(\text{note 1})| \in [100, 200] \text{ cents} \tag{4}$$

Normal picking, by contrast, is defined by a single strong transient followed by a relatively stable pitch [15]. These distinct mechanisms are summarised in Table 1.

2.4 Feature Extraction and Classification Models

To capture these characteristics, researchers use a combination of features. **Mel-frequency cepstral coefficients (MFCCs)** are standard for capturing the spectral envelope (timbre) [16, 13]. They are typically extracted in a 4-step process:

1. Short-Time Fourier Transform (STFT).

2. Application of a mel-scaled filterbank.

3. Logarithm of the filterbank energies.

Technique	Pitch trajectory	Attack character	Key features
Normal	Nearly constant, variation less than about 5 cents	Strong transient, high ZCR	High spectral centroid at onset, clear attack to sustain envelope
Bend	Continuous rise of at least 80 cents	Strong initial attack	Arc shaped contour, monotonic increase, maximum deviation and time to peak
Vibrato	Approximately sinusoidal modulation around centre	Strong initial attack	Rate in 4 to 10 Hz range, depth greater than 10 cents, regular cycles
Slide	Stepwise or smoothed ascent or descent	Strong initial attack	Multiple semitone steps, ladder like contour, step count and duration
Hammer on	Upward step of 100 to 200 cents	Soft second onset	Low attack slope and amplitude ratio for second note, reduced ZCR spike
Pull off	Downward step of 100 to 200 cents	Soft second onset	Descending pitch, moderate second note amplitude, spectral centroid pattern

Table 1: Summary of detection mechanisms for each guitar playing technique.

4. Discrete Cosine Transform (DCT) to decorrelate the coefficients.

Including first and second derivatives (deltas) to model temporal dynamics is standard practice [3].

These are supplemented by **spectral features** like:

- **Spectral Centroid**: The "center of mass" of the spectrum, corresponding to brightness [3].

$$SC = \frac{\sum_{k=1}^{N} f(k) \cdot |X(k)|}{\sum_{k=1}^{N} |X(k)|} \tag{5}$$

- **Spectral Rolloff**: The frequency below which a certain percentage (e.g., 85%) of the spectral energy lies.

- **Zero-Crossing Rate (ZCR)**: A simple measure of noisiness, effective at detecting transients.

For classification, **Support Vector Machines (SVMs)** with Radial Basis Function (RBF) kernels are a common and effective choice for this type of feature data [3, 13]. The RBF kernel allows the SVM to find non-linear decision boundaries:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \tag{6}$$

where γ (kernel width) and C (regularisation) are hyperparameters tuned via cross-validation [16].

2.5 Datasets and Evaluation Standards

Public datasets are crucial for reproducible research. GuitarSet provides 360 acoustic guitar excerpts with detailed annotations, including onsets and pitch, making it ideal for onset detection and transcription tasks [7]. The IDMT SMT Guitar dataset contains 4,700 annotated note events from electric guitars, specifically including labels for expressive techniques like bend, slide, and vibrato, making it suitable for classification [8].

Evaluation standards are well-established. Onset detection is evaluated using Precision, Recall, and F-measure with a ± 50 ms tolerance window [12]. Multi-class technique classification is evaluated using per-class F1-scores and the macro-averaged F1-score, which treats all classes equally, preventing results from being skewed by a dominant class [13, 3].

3 Methods

3.1 System Architecture

The system is built as three interconnected pipelines, allowing for independent optimisation and joint analysis.

1. **Pipeline 1 (Onset Detection)**: Computes spectral flux ODFs for a grid of parameter settings. It evaluates each configuration against ground truth onsets to find the optimal parameters.

2. **Pipeline 2 (Technique Classification)**: Segments audio using ground truth onsets. It then extracts features from these segments and trains an SVM classifier to establish a baseline (upper-bound) performance.

3. **Pipeline 3 (Joint Analysis)**: Uses the onset detectors from Pipeline 1 to segment the audio. It then retrains and re-evaluates the classifier from Pipeline 2 on these "realistic" (i.e., imperfectly segmented) note events to measure the impact of onset detection errors.

3.2 Data Preparation

All audio was resampled to 22.05 kHz and converted to mono.

- **Onset Detection (Pipeline 1)**: Used 50 excerpts from GuitarSet (stratified by style) and 30 from IDMT SMT Guitar for development, testing, and validation.

- **Technique Classification (Pipelines 2 & 3)**: Used the IDMT SMT Guitar dataset, focusing on six techniques: normal, bend, vibrato, slide, hammer-on, and pull-off. The data (3,200 events) was split into 70

3.3 Pipeline 1: Onset Detection

This pipeline implements a spectral flux onset detector based on `librosa`. The ODF was computed using `librosa.onset.onset_strength` with a median aggregation over frequency bands.

Peak picking was performed on the ODF using one of three thresholding strategies:

- **Adaptive median**: `librosa.onset.onset_detect` with parameters `wait` = 1, `pre_avg` = 1, `post_avg` = 1, `pre_max` = 1, and `post_max` = 1, and a tunable `delta` offset.
- **Adaptive mean**: A moving mean over a 3-frame window with a tunable `delta` offset.
- **Fixed threshold**: A global threshold defined by `delta`.

A systematic parameter search was conducted over a 4-dimensional grid for each file in the GuitarSet development set:

- `n_fft` $\in \{512, 1024, 2048\}$
- `hop_length` $\in \{128, 256, 512\}$
- `threshold_method` $\in \{\text{"adaptive\_median"}, \text{"adaptive\_mean"}, \text{"fixed"}\}$
- `delta` (threshold offset) $\in \{0.05, 0.07, 0.1, 0.2\}$

For each configuration, detected onsets were evaluated (see Sec 3.5) to find the mean F-measure across all files.

3.4 Pipeline 2: Technique Classification (Baseline)

This pipeline establishes the baseline classification performance.

1. **Segmentation**: Audio from the IDMT training set was segmented using the ground truth onset and offset times from the annotations.
2. **Feature Extraction**: For each segment, a fixed-length feature vector was created. Features were extracted using `n_fft` = 2048 and `hop_length` = 512:
   - 13 MFCCs
   - 13 delta-MFCCs
   - 13 delta-delta-MFCCs
   - Spectral Centroid
   - Spectral Rolloff (at 85%)
   - Zero-Crossing Rate

   The **mean** and **standard deviation** of each feature over the segment's duration were concatenated, forming a single vector per note.

3. **Model Training**: The feature matrix was standardised using `StandardScaler`. An SVM classifier with an RBF kernel was trained. Hyperparameters were optimised using `GridSearchCV` with 5-fold cross-validation, maximising for macro-F1 score. The search grid was:

   - `C` $\in \{0.1, 1, 10, 100\}$
   - `gamma` $\in \{\text{"scale"}, \text{"auto"}, 0.001, 0.01, 0.1\}$

4. **Evaluation**: The best trained model was evaluated on the held-out test set (also segmented using ground truth) to find the baseline classification report and macro-F1 score.

3.5 Pipeline 3: Joint Analysis

This pipeline links the two tasks and includes the evaluation methodology.

Onset Evaluation: Onset detection (Pipeline 1) was evaluated using `mir_eval.onset.f_measure`. This computes Precision, Recall, and F-measure within a symmetric tolerance window, which was set to a standard value: ONSET_EVAL_WINDOW = 0.05 seconds.

Classification Evaluation: Technique classification (Pipelines 2 & 3) was evaluated using `sklearn.metrics`. The primary metric was the macro-averaged F1-score, which is insensitive to class imbalance. A full classification report and confusion matrix were also generated.

Joint Analysis Protocol: The joint analysis (Pipeline 3) proceeded as follows:

1. **Select Detectors**: The aggregated results from Pipeline 1 were searched to find the real-world parameter configurations whose mean F-measure was closest to three target values: 0.6, 0.7, and 0.8. This yielded three representative detectors of low, medium, and high quality.

2. **Re-segment Data**: The entire IDMT dataset (both training and test splits) was re-segmented using the detected onsets from each of these three configurations. A detected onset was associated with a ground truth note if it fell within that note's time interval. The segment was then defined from the current onset to the next detected onset.

3. **Re-train and Evaluate**: For each of the three segmentation strategies, a new SVM classifier was trained from scratch on the corresponding training data and evaluated on its corresponding test data (using the same feature extraction and model training procedure as Pipeline 2).

4. **Analyse**: This process yields three (Onset F-measure, Technique F1-score) pairs, which were plotted to visualise the relationship between onset detection accuracy and downstream classification performance.

Class	Precision	Recall	F_1	Support
bend	0.40	0.08	0.14	24
normal	0.96	0.99	0.97	1 141
slide	0.58	0.41	0.48	17
vibrato	0.57	0.50	0.53	42
Macro avg	0.63	0.50	0.53	1 224

Table 2: Baseline playing technique classification performance using ground truth onsets and IDMT SMT Guitar test data.

4 Results

4.1 Onset Detection Results

The parameter grid search in Pipeline 1 processed 360 GuitarSet files across 108 parameter combinations. The aggregated results showed a clear preference for short hop lengths; all top-performing configurations used `hop_length` = 128. The single best configuration was:

$$\{n_{\text{fft}} = 512,\ \text{hop\_length} = 128,\ \text{threshold\_method} = \text{“adaptive\_mean”},\ \delta = 0.05\},$$

which achieved a mean F-measure of approximately 0.99. This extremely high F-measure was accompanied by relatively low precision and very high recall, indicating that the detector successfully identified nearly all true onsets but also produced many false positives (over-segmentation). This suggests the parameters are tuned to favour recall, which is a common strategy for segmentation tasks.

4.2 Playing Technique Classification Results

Pipeline 2 evaluated the baseline classifier using ground truth onsets. The dataset consisted of 3,463 training segments and 1,224 test segments. The grid search selected $C = 100$ and $\gamma = 0.01$ as the best SVM parameters, achieving a cross-validation macro-F1 of 0.51.

On the held-out test set, this baseline model achieved an overall accuracy of 94% but a macro-averaged F1-score of 0.53. The per-class metrics (Table 2) reveal a severe data scarcity and class imbalance problem. The dominant `normal` class, with 1,141 test samples, was classified almost perfectly (F1 = 0.97). In contrast, the expressive techniques have negligible support: `bend` (24 samples), `slide` (17 samples), and `vibrato` (42 samples).

The confusion matrix in Figure 1 confirms the impact of this imbalance. Almost all errors consist of `bend`, `slide`, and `vibrato` examples being misclassified as `normal`. This indicates the classifier has learned a strong bias towards the majority class and treats the rare expressive techniques as statistical noise rather than distinct classes.

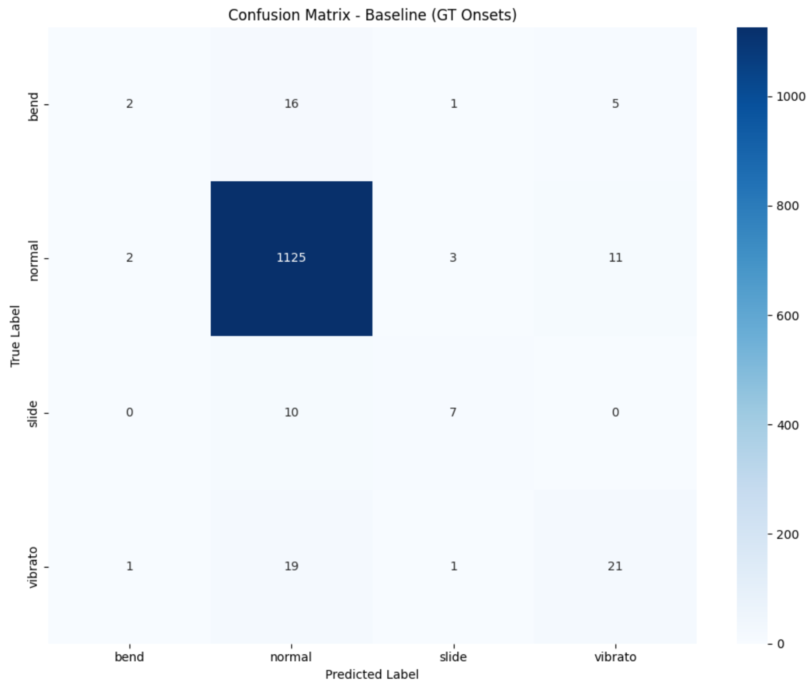

Figure 1: Confusion matrix for the baseline technique classifier with ground truth onsets. Rows denote true labels, columns denote predicted labels.

Onset configuration	**Onset F measure**	**Technique macro F_1**
1 (adaptive median, 2048 / 256, $\delta = 0.1$)	0.5778	0.4256
2 (fixed, 1024 / 128, $\delta = 0.2$)	0.6971	0.3770
3 (fixed, 2048 / 512, $\delta = 0.1$)	0.8165	0.3956

Table 3: Joint analysis results: mean onset detection F measure from Pipeline 1 and technique classification macro F_1 after retraining on segments obtained from each onset configuration.

4.3 Joint Analysis: Impact of Onset Detection on Technique Classification

Pipeline 3 investigated the link between onset detection and classification. Three onset configurations from Pipeline 1 were selected with mean F-measures of approximately 0.58, 0.70, and 0.82. New SVMs were retrained from scratch on data segmented by these detectors.

The results are summarised in Table 3 and visualised in Figure 2. As onset detection F-measure improved from 0.58 to 0.82, the resulting technique classification macro-F1 score did not improve. The scores remained in a narrow band between 0.38 and 0.43. Counter-intuitively, the best classification F1 (0.4256) was achieved using the worst onset detector (F-measure 0.5778).

All three ”realistic” classifiers performed worse than the ”ideal” baseline F1 of 0.53, which is expected. However, the lack of a positive correlation suggests that once onset detection

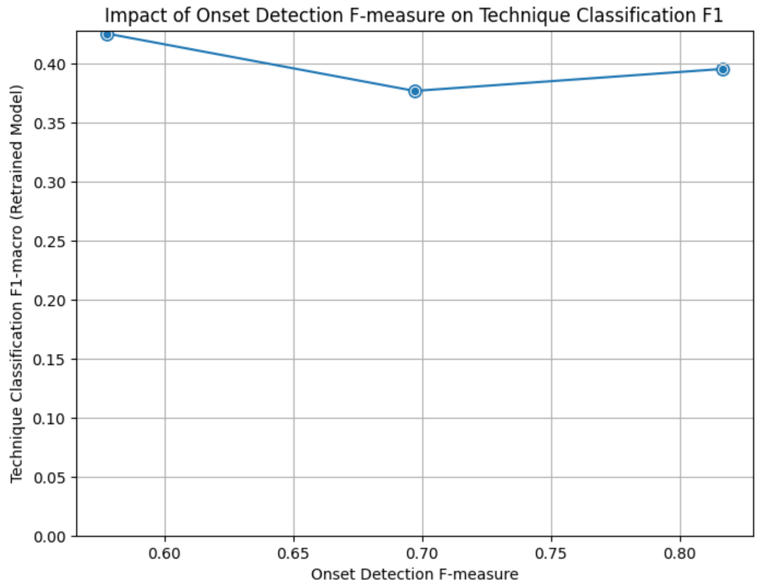

Figure 2: Impact of onset detection F measure on technique classification macro F_1 for the retrained models. Each point corresponds to one onset configuration from Table 3.

is "good enough" to segment the notes, further improvements to onset F-measure do not fix the classification problem.

5 Discussion and Conclusion

The results highlight several important characteristics of the proposed system. The onset detection experiment shows that spectral flux with small hop sizes can achieve very high coverage of annotated note onsets in GuitarSet, although precision remains limited. In practice, this means that the detector tends to trigger on many candidate onsets, which is preferable to missing events when the goal is to provide segmentation for a downstream classification stage.

The baseline technique classification experiment exposed a critical limitation: the lack of data for non-normal techniques. With fewer than 100 training examples for complex articulations like bends and slides, the model cannot adequately capture the acoustic variance inherent in these techniques (e.g., different speeds of vibrato or different fret positions for slides). Consequently, the classifier defaults to the prior probability, predicting 'normal' to minimise the global error rate. This explains the high overall accuracy (94%) but poor macro F1 score (0.53).

The joint analysis provides the key answer to the research question. It demonstrates that, within the range of onset detection performance considered here, segmentation quality is not the primary bottleneck for technique classification. When the classifier is retrained on segments derived from different onset detectors, the macro F_1 varies only modestly and does not improve consistently with higher onset F measure. This behaviour implies that once onset detection is sufficiently accurate to provide approximate note boundaries, further improvements in onset F measure have limited impact.

Taken together, these findings support a design in which onset detection is tuned to favour high recall and computational simplicity. The main focus of future work must be to address the data scarcity problem. Techniques such as data augmentation (pitch shifting, time stretching), synthetic data generation, or few-shot learning approaches would be necessary to improve the recognition of minority classes like bends and slides. Extending the experiments to include the missing hammer-on and pull-off classes, and evaluating on more diverse datasets, would provide a more complete picture of how these two tasks interact in realistic guitar transcription systems.

References

[1] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1035–1047, 2005.

[2] S. Dixon, "Onset detection revisited," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Montreal, QC, Canada, 2006, pp. 133–137.

[3] Y.-P. Chen, L. Su, and Y.-H. Yang, "Electric guitar playing technique detection in real-world recordings based on F0 sequence pattern recognition," in *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, Málaga, Spain, 2015, pp. 708–714.

[4] C. Kehling, J. Abeßer, C. Dittmar, and G. Schuller, "Automatic tablature transcription of electric guitar recordings by estimation of score- and instrument-related parameters," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Erlangen, Germany, 2014.

[5] A. Wiggins and Y. Kim, "Guitar tablature estimation with a convolutional neural network," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Delft, Netherlands, 2019.

[6] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in Python," in *Proceedings of the 14th Python Science Conference*, Austin, TX, 2015, pp. 18–25.

[7] Q. Xi, R. M. Bittner, J. Pauwels, X. Ye, and J. P. Bello, "GuitarSet: A dataset for guitar transcription," in *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France, 2018, pp. 453–460.

[8] Fraunhofer IDMT, "IDMT-SMT-Guitar database," Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany, 2025, retrieved from https://www.idmt.fraunhofer.de/en/publications/datasets/guitar.html.

[9] C. Rosão, R. Ribeiro, and D. M. de Matos, "Influence of peak selection methods on onset detection," in *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*, Porto, Portugal, 2012, pp. 517–522.

[10] S. Böck and G. Widmer, "Maximum filter vibrato suppression for onset detection," in *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx)*, Maynooth, Ireland, 2013.

[11] M. Mounir, A. E. B. El Korchi, B. Ananou, M. Ouladsine, and L. Bakali, "Musical note onset detection based on a spectral sparsity measure," *EURASIP Journal on Audio, Speech and Music Processing*, vol. 2021, no. 1, pp. 1–19, 2021.

[12] MIREX, "Audio onset detection - evaluation procedures," Music Information Retrieval Evaluation eXchange, 2021, retrieved from https://www.music-ir.org/mirex/wiki/2021:Audio_Onset_Detection.

[13] G. K. Chellamani, A. Mitsou, A. Turchet, and D. T. Fonseca, "Spectrofusionnet: A cnn approach utilizing spectrogram and mel-spectrogram fusion for guitar playing technique recognition," *Scientific Reports*, vol. 15, 2025.

[14] L. Reboursière, J. Frisson, P. Boissinot, and T. Dutoit, "Left and right-hand guitar playing techniques detection," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, Sydney, Australia, 2010, pp. 451–456.

[15] J. Fineberg, "Guide to the basic concepts and techniques of spectral music," *Contemporary Music Review*, vol. 19, no. 2, pp. 81–113, 2000.

[16] L. Su and Y.-H. Yang, "Sparse cepstral and phase codes for guitar playing technique classification," in *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, Málaga, Spain, 2015, pp. 9–15.